# Gene recruitments and dismissals in argonaut octopus genome provide insights to pelagic lifestyle adaptation and shell-like eggcase reacquisition

Masa-aki Yoshida[1,#], Kazuki Hirota[2,3], Junichi Imoto[4,*], Miki Okuno[5,**], Hiroyuki Tanaka[5], Rei Kajitani[5], Atsushi Toyoda[6,7], Takehiko Itoh[5], Kazuho Ikeo[2], Takenori Sasaki[2,8], and Davin H. E. Setiamarga[3,8,#]

[1]Marine Biological Science Section, Education and Research Center for Biological Resources, Faculty of Life and Environmental Science, Shimane University, Okinoshima, Shimane 685-0024, Japan

[2]Graduate School of Science, The University of Tokyo, Bunkyo-ku, Tokyo, 113-8654, Japan

[3]Department of Applied Chemistry and Biochemistry, National Institute of Technology (KOSEN), Wakayama College, Gobo, Wakayama, 644-0012 Japan

[4]Center for Information Biology, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan

[5]School of Life Science and Technology, Tokyo Institute of Technology, Meguro-ku, Tokyo 152-8550, Japan

[6]Comparative Genomics Laboratory, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan

[7]Advanced Genomics Center, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan

[8]The University Museum, The University of Tokyo, Bunkyo-ku, Tokyo, 113-0033, Japan

[#]Corresponding authors:

MAY mayoshida@life.shimane-u.ac.jp; DHES davin@wakayama.kosen-ac.jp

[*]Present address: Pelagic Fish Resources Division, Fisheries Stock Assessment Center,

Fisheries Resources Institute, Japan Fisheries Research and Education Agency, Yokohama, Kanagawa 236-8648, Japan

**Present address: Division of Microbiology, Department of Infectious Medicine, Kurume University School of Medicine, Kurume, Fukuoka, 830-0011, Japan

**Keywords;** Comparative genomics, biomineralization, concerted evolution, Hox cluster

## Abstract

The paper nautilus, *Argonauta argo*, also known as the greater argonaut, is a species of octopods distinctly characterized by its pelagic lifestyle and by the presence of a spiral-shaped shell-like eggcase in females. The eggcase functions by protecting the eggs laid inside it, and by building and keeping air intakes for buoyancy. To reveal the genomic background of the species' adaptation to pelagic lifestyle and the acquisition of its shell-like eggcase, we sequenced the draft genome sequence of the species. The genome size was 1.1 Gb, which is the smallest among the cephalopods known to date, with the top 215 scaffolds (average length 5,064,479 bp) covering 81% (1.09 Gb) of the total assembly. A total of 26,433 protein-coding genes were predicted from 16,802 assembled scaffolds. From these, we identified nearly intact HOX, Parahox, Wnt clusters and some gene clusters probably related to the pelagic lifestyle, such as *reflectin*, *tyrosinase*, and *opsin*. For example, *opsin* might have undergone an extensive duplication in order to adapt to the pelagic lifestyle, as opposed to other octopuses, which are mostly the benthic. Our gene models also discovered several genes homologous to those related to calcified shell formation in Conchiferan Mollusks, such as Pif-like, SOD, and TRX. Interestingly, comparative genomics analysis revealed that the homologous genes for such genes were also found in the genome of the octopus, which does not have a shell, as well as the basal cephalopods *Nautilus*. Therefore, the draft genome

51 sequence of *A. argo* we presented here had not only helped us to gain further insights into the
52 genetic background of the dynamic recruitment and dismissal of genes for the formation of an
53 important, converging extended phenotypic structure such as the shell and the shell-like
54 eggcase, but also the evolution of lifestyles in Cephalopods and the octopods, from benthic to
55 pelagic.
56

## Introduction

The paper nautilus, or the argonaut *Argonauta argo* is a member of Argonautoidea, a superfamily of octopods (Cephalopoda, Octopodiformes), but has specialized characters not found in other octopus species. It is a cosmopolitan species distributed in the global tropical and subtropical open seas (Norman, 2000). Phylogenetic analyses have placed *A. argo* together with its congener (e.g. *A. hians*), forming a monophyletic Argonautidae, and then, form a sister relationship with the blanket octopuses (e.g. *Tremoctopus*), and thus further forming the superfamily Argonautoidea (Hirota et al., 2021; Strugnell et al., 2006; Sanchez et al., 2018). Distinct synapomorphies of this superfamily, which could also be found in *A. argo*, are the extreme female-biased sexual size dimorphism, a comparatively large and entirely transformed hectocotylus that is coiled in a pouch below the eye until maturity, and the transferring of spermatophores to the female mantle cavity by hectocotylus detachment (Naef, 1923; Bello, 2012). Although the consensus phylogeny also suggested that Argonautoidea split from benthic ancestral octopods, members of the superfamily including *A. argo* are fully adapted to the holopelagic lifestyle and thus does not need to have any contact with the seafloor throughout its lifecycle. Several studies have suggested that the holopelagic lifestyle was probably achieved by evolutionary acquisitions of distinct characters enabling members of Argonautoidea to keep afloat in midwater and to egg brooding away from the sea floor (cf. Naef 1923; Packard & Wurtz 1994; Young 1985; Bizikov 2004). Extreme adaptations of this group to their midwater habitat have masked their evolutionary origins. Buoyancy in argonauts was probably obtained after their ancestors had already become pelagic, potentially via the pelagic paralarval or juvenile stages found in many benthic octopuses with small eggs (Finn and Norman 2010).

One conspicuous character separating Argonautidae, a family which includes all argonauts (genus *Argonauta*), with the rest of Argonautoidea is the presence of a biomineralized

eggcase in females, which external morphology mimics the spirally-wound shells of *Nautilus* and the extinct ammonites (Scales, 2015; Stevens et al., 2015). The eggcase is thought to protect the eggs laid inside or, as well as taking in air for maintaining buoyancy (Finn and Norman 2010). As such, the re-acquisition of this shell-like structure was probably important, because it helps *Argonauta* to maintain its holopelagic lifestyle. Previous observations have maintained that the "shell-like" eggcase is not a "true" shell (the Conchiferan shell) (Naef, 1923).

The evolutionary story of shell formation and loss in Cephalopods is interesting in itself. Although being classified as a member of the Conchifera, a subphylum of Mollusks composed of members with external shells biomineralized with calcium carbonate, except for the basally diverged Nautiloids (Setiamarga et al., 2021a), extant Cephalopods mostly degenerated their shells, resulting in the complete shell loss in octopods, and vestigial shells in some decapods (squids and cuttlefishes) (Kröger et al., 2011). True Conchiferan shells are formed through the secretion of proteins from the mantle tissue, made from aragonite and calcite, have the nacreous layer and intricate microstructures (Jackson et al., 2009; Kocot et al., 2016; Jackson et al., 2017), which has evolved since at least in late Ordovician (Vendrasco et al., 2011; 2013). An extant member of the basal Cephalopods, the Nautiloid *Nautilus pompilius* apparently also forms their shells this way (Marie et al., 2009; Setiamarga et al., 2021a). However, despite convergence in their general external morphology, the eggcase of *Argonauta* is not considered as a true Conchiferan shell but an evolutionary innovation of the genus (Naef, 1923; Scales, 2015). It is formed through the secretion of related proteins from their arms (Naef, 1923; Scales, 2015) and has different biomineralization and microstructural profiles (Revelle and Fairbridge 1957; Mitchell *et al*. 1994; Nixon and Young 2003; Saul and Stadum 2005; Oudot *et al*. 2020). Previously, we conducted an extensive multi-omics analysis on the eggcase of two argonaut species, *A. argo*

and *A. hians*, which samples were obtained from the Sea of Japan (Setiamarga et al., 2021b). Two important points relevant to our present study could be taken from our previous one: (1) almost no Conchiferan homologous SMP, including those of the basal Cephalopoda *Nautilus pompilius* (Setiamarga et al., 2021a) was present in the eggcase matrix of the two argonaut species, and (2) Conchiferan SMP homologs (or homologous domains) were also found in the genome of the shell-less octopods, *Octopus bimaculoides.* These points thus indicate that our result was in agreement with the result of morphological observations, which maintains that the eggcase is not a homologous structure of the shell. However, the observations have also caused other questions: Are the SMP genes not used in the eggcase formation still retained in the genomes of the argonauts? Comparative genome analyses across Cephalopoda, and among different representative species of Conchifera, are thus needed to answer this question. Such genomic level comparative studies would also give important insights on the evolution of holopelagic lifestyle at the genetic level.

Until very recently, the lack of genome data had prevented us from understanding the genetic basis of Cephalopod biology, and even molluscan biology. This was remedied by recent reports of various Cephalopod genomes, such as the genomes of *O. bimaculoides* (Albertin et al., 2019), *Euprymna scolopes* (Belcaid et al., 2019), *Architeuthis dux* (da Forsa et al. 2020), and the basal Cephalopod *Nautilus pompilius* (Zhang et al., 2021; Huang et al., 2021). Comparative genomics studies of these genomes have allowed us to identify notable characteristics of Cephalopod genomes except *Nautilus*, such as: (1) the average genome size of around 3 Gigabases (Gb), which is slightly bigger than that of other molluscan species (Gregory 2021), (2) highly rearranged genome with transposable element expansion, which have caused the genomes to be highly repetitive in nature (Albertin et al., 2015; da Fonseca et al., 2020), (3) lineage specific duplication of certain types of genes (Yoshida et al., 2011), and (4) whole transcript-wide adenosine to inosine (A-to-I) RNA editing (Alon et al. 2015;

Liscovitch-Brauer et al. 2017). These genomic characteristics have thus suggested that the coleoid Cephalopods have intriguingly different genomes from "standard" metazoan genomes. Another interesting point is that such differences were apparently evolutionarily acquired in ancestral Coleoids, which members show similar body plans and morphology in general (Young 1971) despite their ancient divergences (Decapodiformes (squid and cuttlefishes) vs. Octopodiformes (vampire squid and octopuses) = 242 ± 38 million years ago (Mya); Nautiloidea vs. Coleoidea = 415 ± 60 Mya) (Kröger et al., 2011; Vinther et al., 2012; Sanchez et al., 2016). Therefore, additional genomic data, especially of the Octopodiformes, will allow us to trace the ancestral chromosomes of Cephalopods and their transition within Mollusks, which in the end might help to unravel the evolutionary origin of these "genomic idiosyncrasies" These are major obstacles to tracing the ancestral chromosomes of cephalopods and their transition within the Mollusks.

Here, we report a high-quality draft genome assembly of the greater paper nautilus / greater argonaut *Argonauta argo*. We found that this species has an exceptionally small genome size, making the species an ideal species for genomic studies. Although studies targeting argonauts have not progressed because it is difficult to keep in aquaria, we have access to the location in the Sea of Japan, where fresh and living samples of this species could be easily obtained by fixed nets from June to August (Sakurai and Kono, 2010). Using obtained genome data, we focusedly discussed the evolution of some interesting genomic features such as those related to shell evolution, eggcase formation, and color vision, in order to gain insights to the genomic basis of the adaptation to the open-ocean holopelagic lifestyle of this species in particular, and the evolution of Cephalopods in general.

## Results and Discussion

### *The draft genome assembly of* A. argo

We generated a draft genome from a single individual of a female argonaut obtained from a fixed net set on the coasts of Oki Island, Shimane Prefecture, Japan (Figure 1). A total of 1.34 Gb was assembled from input sequences obtained from genome sequencing with 201× coverage (107× PE , 24× 3 kb MP, 24× 6 kb MP, 24× 10kb MP, and 24× 15kb MP). 57,036 scaffolds of various lengths were assembled, with the top 215 scaffolds larger than 1000 kb (average length 5,064,479 bp), covering 81% (1.09 Gb) of the total assembly. Half of assembled scaffolds (N50) were of 6.18 Megabases (Mb) or longer, reflecting high contiguity. These statistics (Table S1) thus showed that our *A. argo* draft genome sequence ranks among the top quality draft genomes of Molluscs, and the most comprehensive for Cephalopods. For example, our N50 indicates that our *A. argo* draft genome is twice as long as that of the Hawaiian bobtail squid *E. scolopes* (Belcaid et al. 2019).

The discrepancy between GenomeScope estimation (1.1Gb, Figure S1) and our actual assembly size (1.34Gb; 1.25Gb non-gap regions) might be caused by the presence of bacterial contamination and/or heterogeneities caused by large insertion and deletion between haploid genomes. However, we only found a very minute amount of bacterial genome contamination in our assembly, indicating that the latter was most likely the main cause of the discrepancy. To assess the completeness of the gene space of the assembly, an analysis using BUSCO v3.0.2 (genome mode) (Simão et al. 2015) was performed by using the provided metazoan data set (metazoa_odb9, n=978), resulting in the recovery of 91.1% of the predicted gene sequences (Table S2). Krait analysis showed that the microsatellite regions account for 4.6% of the genome, with dimer and trimer regions accounting for more than 85% (Figure S2, Table S3).

The high-quality and relatively high level of completeness of our genome assembly, as shown by the statistics we presented above will allow us to address some lingering questions on Cephalopod biology and evolution at the genetic and genomic levels. For example, future

182 studies might utilize the microsatellite regions, which compose a part of the repeat regions in
183 the genome, as individual markers because of the large polymorphisms within individuals, or
184 as markers for paternity analysis of egg-masses.
185
186 ***Ancient gene clusters in the cephalopods: HOX, Parahox, and Wnt genes***
187 The improved contiguity of our genome assembly confirmed the presence of a Hox
188 cluster. A large Hox cluster of nine Hox genes on four separate scaffolds was recovered in
189 the *A. argo* genome, totaling to a length of at least 18 Mb (Figure 2). Three of the nine Hox
190 genes are not presumed to be gene models, but we have used the Homeobox domain
191 sequence to confirm that they are indeed present on the scaffold and that they are indeed hox
192 genes compared to other Lophotrochozoa genes (Figure S3, Figure S4). *Hox2*/*proboscipedia*
193 (*pb*) was not found, as in squid genomes (Belcaid et al., 2019; da Fonseca et al. 2020) except
194 *Nautilus* (Zhang et al. 2021). We also could not find *Hox4*/*Deformed* (*Dfd*), which is similar
195 to *O. bimaculoides* (Albertin et al., 2015), and thus probably a common feature in benthic
196 octopods (Figure S3). There are at least 10 ORFs inserted among several different Hox genes
197 (3 between *Scr* and *Antp*, 7 between *Lox4* and *Post2*). Interestingly, no homolog was found in
198 other organisms, including even the giant squid *A. dux* (da Fonseca et al., 2020), for any of
199 these ORFs (Table S4).
200 Hox clusters are usually found in contigs of about 100 kb in vertebrates and >1,000 kb in
201 invertebrates (Powers et al., 2000; Wagner et al., 2003). Meanwhile, the octopus Hox gene
202 cluster is apparently fragmented, and the genes are present separately on its genome one by
203 one (Albertin et al., 2015), unlike most other bilaterian genomes (Duboule, 2007). Intuitively,
204 this finding seems to be in accordance with the staggered, non-colinear expression pattern of
205 Hox genes in Cephalopods (Lee et al., 2003; Wollesen et al., 2018). However, our finding of
206 a Hox cluster in the genome of *A. argo*, albeit incomplete, suggests that fragmentation of the

cluster is probably a feature limited to benthic octopods (or even a possible artifact of the genome assembly process of *O. bimaculoides*).

The presence of ORFs located among several Hox genes in the genome of *A. argo* is also intriguing, since it might indicate that the Hox cluster is actually breaking at the place where the intervening genes are located. The Patellogastropod limpet *Lottia gigantea*, another member of the shelled mollusks (Conchifera), was found to have a typical invertebrate Hox cluster spanning 471 kb with no intervening ORFs among any of its Hox genes (Simakov et al. 2012). Meanwhile, recent findings indicate that although the genome of the basal Cephalopod *N. pompilius* contains a complete set of molluscan Hox genes, they are not located together in a cluster, but are divided in 7 contigs (Zhang et al. 2021). On the other hand, the Hox genes in another Cephalopod, the giant squid *A. dux*, are apparently arranged into a disorganized cluster with insertions of intervening non-Hox genes among cluster members (da Fonseca et al., 2020). However, we found no apparent homology or synteny between any of the intervening ORFs of *A. argo* and those of *A. dux* (Table S4). The acquisition of putative ORFs inside the Hox cluster of *A. argo* is probably an indication of a situation not dissimilar to what was proposed for the fruitfly *Drosophila melanogaster*, which Hox cluster is split into two complexes, with the presence of non-homeotic genes in between (Von Allmen et al., 1996; Wagner et al., 2003; Robertson and Mahaffey, 2017), although *Drosophila* still maintained its colinear expression pattern (Graham et al., 1989; Gaunt, 2015). However, the Hox cluster break in *Drosophila* is most likely a lineage-specific feature, since another model insect, the genome of the beetle *Tribolium castaneum* are intact (Von Allmen et al., 1996; Tribolium Genome Sequencing Consortium, 2008; Shippy et al., 2008). When considered altogether, it seems that the splits and breaking offs of Hox cluster could be a symplesiomorphic feature of the Cephalopod genome, but with the actual “Hox de-clustering” processes happened lineage-specifically. This might explain why Cephalopods do

not exhibit typical invertebrate Hox cluster arrangement seen in, for example, the limpet *L. gigantea*.

The Extended Hox complex (*Hox* genes plus *Evx*, *Mox*, and possibly *Dlx*) is also a common feature in bilaterian genomes (Montavon, 2015). In vertebrate genomes, the complex is shown to be linked to the EHGbox (*En*, *Hb9*, and *Gbx*) and NKL gene groups (*Msx*, *Emx*, etc.) and form a supercluster (Garcia-Fernàndez, 2005). In the *A. argo* genome, we found, probably for the first time in Spiralia, a linkage among *Dlx*, *Engrailed* (*En*), and the Hox genes (Figure 2). In the genome of the giant squid *A. dux*, *Dlx* and *En* were found in different scaffolds with no linkage to the Hox cluster whatsoever (Table S4). However in *A. argo, Dlx* was found to be located anterior to Scr relative to their positions to the Hox genes (Hox cognate group4), while *E*n was found to be located posterior to *Post1* (Figure 2), and thus reversing the presumed ancestral state (Garcia-Fernàndez, 2005). Although the possibility of their reinsertions into the Hox group cannot be ruled out, this may indicate that the presence of the Extended Hox group is probably conserved in modern cephalopods, although the constraint to preserve gene order is probably relatively weak. The weak constraint in preserving gene order could also explain the "Hox de-clustering", which characteristics include insertions of ORFs in intergenic regions, observed in Cephalopods.

We also found the presence of the ParaHox cluster, an evolutionary sister complex of the Hox cluster, in the genome of *A. argo* (Figure 2). The ParaHox cluster, which consists of the *Gsx*, *Xlox*, and *Cdx* gene families, are transcription factors involved in the anterior-posterior development during early embryogenesis of bilaterians (Brooke et al, 1998; Garstang and Ferrier, 2013). The ParaHox cluster is usually found intact in the genomes of Deuterostomes except sea urchin and Ascidians (Garstang and Ferrier, 2013)). However, in Lophotrochozoans, such as the annelid *Platynereis dumerilii* and the limpet *L. gigantea*, only *Gsx* and *Xlox* are clustered together, with *Cdx* broken off and thus unlinked in the genome.

The ParaHox cluster of *A. argo* were found to conserve the structure of a typical Lophothrocozoan cluster, similar to those reported in *Nautilus* (Huang et al. 2021) and the octopus (Li et al., 2020). Although further study is still needed, the highly conserved nature of the ParaHox clusters among Cephalopoda, mollusks, and even Lophotrochozoans, indicates a possible presence of an evolutionary constraint to conserve the cluster's presence and arrangement in the genome, after the breakage of *Cdx* from *Gsx* and *Xlox*.

Similarly, an older gene cluster that is widespread in the animal kingdom is Wnt. Most genomes of bilaterians have a common cluster, *wnt9-wnt1-wnt6-wnt10*, or parts of this cluster (Huang et al., 2021). This ancestral cluster of *wnt*s is thought to originate in the evolution of the common ancestor of cnidarians and bilaterians (Janssen et al., 2010; Holstein 2012). In other shelled mollusks (i.e. Conchifera) such as the rock oyster *Crassostrea gigas* and the Japanese pearl oyster *P. fucata*, the limpet *L. gigantea*, and also in *O. bimaculoides*, the *wnt1-wnt6-wnt10* cluster was conserved (Du et al., 2018a), with *L. gigantea* and *P. fucata* seemingly retaining some of the basal lophotrochozoan / protostome *wnt* paralogs (Cho et al., 2010; Setiamarga et al., 2013). In this study, we also confirmed the linkage of *wnt6-wnt9* in *A. argo* (Figure 2), besides the standard Conchiferan cluster. This suggests that *A. argo* probably also derived this arrangement from the basal metazoan form of Wnt gene orientation and clustering (*wnt9-wnt1-wnt6-wnt10*). Meanwhile, we also observed the lack of *wnt3* and *wnt8*, which seems to be lost in the ancestral protostomes / lophotrochozoans and in ancestral Conchiferans, respectively (Janssen et al., 2010; Setiamarga et al., 2013; Liu et al., 2018; Bai et al., 2020; Wang et al., 2021).

***Tandem gene duplications of gene clusters related to pelagic lifestyle***

Our *A. argo* genome assembly, which is of sufficiently better quality than those of previous octopods, allowed us to investigate the existence of tandem gene arrangements. Our

searches found two gene clusters, the Reflectins and the Tyrosinases (Figure 3, 4). Both are highly expressed in the first arms of the organism (Figure 4).

Four tandemly arranged gene models of the octopus *reflectin*/*tbc1* domain family (Aargo020153-6) and one possible ORF recovered by a BLAST search in a single scaffold were found in *A. argo* genome (Figure 3). Phylogenetic analysis showed that the three gene models are monophyletic in *A. argo*, and form a monophyletic clade together with sequences of *E. scolopes*, which also formed a monophyletic clade (Figure S5). The translated sequences of three of the four gene models have at least five of the so-called "Reflectin motifs" ($M/FD(X)_5MD(X)_5MDX_{3/4}$) (Levenson et al. 2017, Figure S6). With only 23 nucleotide substitutions, regardless of codon positions, the CDS of the tandemly duplicated *reflectin* genes in *A. argo* match each other sequences at 97%, covering 760 bp (Figure S7). Meanwhile, it is also enticing to suggest that the only gene model with a different sequence, Aargo020154, was inverted to the rest of the genes.

There are two possible causes to explain which duplicated genes are conserved to form gene clusters: either high level expression are favored and thus retaining duplicated genes would help to increase transcript number, or the multiple copies are conserved under different selection pressures as a result of subfunctionalization (Lynch and Force, 2000; Hahn, 2009; Morel et al., 2015; Hallin and Landry, 2019; Song et al., 2020; Ascencio et al., 2021). It has also been pointed out that the duration of concerted evolution can be influenced by selection for a certain dosage of a gene product, as gene conversion leading to highly similar sequence retentions can be advantageous when there is a selection for higher expression level of that particular gene product, or disadvantageous when divergent gene duplicates are advantageous (Sugino and Innan, 2006). Transcriptome analysis shows that in *A. argo*, *reflectin* is very highly expressed in the 1st arm and eye, and it seems to be transcribed by the three genes (Figure 3B). Therefore, this could be evidence supporting the hypothesis that the cause of

307 gene retention was to have a high level of expression. This concerted evolution may also be
308 the reason why the Cephalopod reflectin formed monophyletic clades with members of the
309 clusters within each species (Figure S5).

310 Then, what would be the function of the highly expressed *reflectin*? *reflectin* is found only
311 in Cephalopods, and the function of the protein products were shown to be related to
312 camouflage by reflecting and refracting light in the surrounding environment (DeMartini et
313 al., 2015). Expressed proteins fill the lamellae of intracellular Bragg reflectors, allowing
314 individuals to exhibit dynamic iris and structural color changes (Crookes et al. 2004). Several
315 tandemly-arranged *reflectin* gene clusters have been found in the genome of *E. scolopes*, with
316 the dominant *reflectin* transcripts are almost exclusively expressed in the light organ, eyes,
317 and skin, and thus probably consistent with the development of symbiotic fluorescent organs
318 specifically evolved in this lineage (Belcaid et al. 2019). However, although in *E. scolopes*,
319 the symbiotic luminous organs are important for countershading and survival, no such organ
320 have been found in any of the argonauts. As a defence mechanism, pelagic cephalopods blend
321 into their surroundings by camouflaging, which is done either through translucence or cryptic
322 coloration. The first arm membranes of the argonauts are always wrapped around the shell,
323 and reflect light by iridescent chromatophores, causing it to look like a mirror. Meanwhile,
324 the giant squid *A. dux* has seven *reflectin* genes and three *reflectin-like* genes on its genome,
325 all except one are clustered on the same scaffold (da Fonseca *et al*. 2020). This non-
326 luminescent deep-sea species has a mirror-like light-reflecting skin for cryptic coloration.
327 These observations probably indicate that the abundantly expressed Reflectin might help the
328 animals to have light-reflecting mirror-like surfaces, which might then play a role in the
329 ability of these species to blend into their surroundings in the open ocean.

330 A similar pattern of possible gene conversion was observed in the *tyrosinase* gene cluster.
331 Of the nine *tyrosinase* gene models predicted in *A. argo* genome, eight were of the

extracellular or secreted (alpha) type, of which four (Aargo001559-62) were found to be tandemly arranged in a single scaffold (Figure 4, Figure S8). Of the four gene models, excluding unaligned regions, similarities of amino acid sequences of the first two (Aargo001559-60) and the last two (Aargo001561-62) are very high, but only 75% similarities between the two gene pairs. However, the four genes shared an almost exact match in a region on the second half of the gene, at around the 520th - 680th aa (Figure S9). The coding DNA sequence (CDS) match rate for this region is 97% with only ca. 60 substitutions, regardless of codon positions (Figure S10). These two pairs of tyrosinases are orthologous to closely related molluscan taxa including the octopus, and form monophyletic groups (Figure S8). This thus suggests that the four *tyrosinase* copies probably underwent gene conversions in two pairs (between Aargo001559 and 1560, between Aargo001661 and 1662) with some partial recombinations among the four genes. Gene expression analysis using Stringtie shows that the four have a common gene expression profile, with high levels of expression in the arms and mantle. Meanwhile, the phylogenetic tree also indicates that the two pairs of the *tyrosinase* genes are apparently orthologous to those found as shell matrix protein-coding genes in Conchiferan mollusks. This finding, i.e., the genes expressed only in the arms belong to different gene clusters than those of other Tyrosinase-coding genes, might indicate that novel gene paralogs originated from previously existing endogenous *tyrosinase* genes were being duplicated and obtained high expression in the arms, possibly used for the calcified eggcase formation, which helps *A. argo* and other argonaut octopods to attain buoyancy and thus their pelagic lifestyle.

***Opsin duplications and change in absorption wavelength related to pelagic lifestyle***

Changes in number and sequences of Opsin are thought to be involved in adaptation to visually-guided behavior. We found that *A. argo* possesses five visual pigment genes in its

357 genome: two noncanonical *r-opsins*, one canonical *r-opsins*, two *xenopsin*, and one
358 *rgr/peropsins/retinochromes* (Figure S11-13) (Ramirez et al., 2016). In previous studies on
359 the pygmy squid *Idiosepius paradoxus*, two *r-opsins*, one *xenopsin*, and two *retinochromes*
360 were identified (Yoshida et al. 2015; Ramirez et al. 2016). We also checked if the expression
361 of the opsins are tissue specific, in order to see whether there is any functional differentiation
362 among the duplicated opsins. However, at present, we were unable to confirm such
363 specificity, at least in the organs we examined because the data is present, such as the eye and
364 skin. In fact, almost no study has been conducted on the functions of *xenopsins* and non-
365 canonical *r-opsins* in the Cephalopods. In the future, a thorough gene expression analysis of
366 these genes in different tissues should probably be conducted to resolve this issue.
367 The presence of two xenopsins in the genome of *A. argo* was apparently not an artifact or
368 assembly error, which means that *A. argo* has an extra copy of *xenopsin* than *I. paradoxus*.
369 The gene models for *xenopsin* in *A. argo* (Aargo004635 and Aargo004636) exist in tandem in
370 the scaffold, albeit with the amino acid sequence being too short to be considered full-length.
371 If we also assume that the two exons of the neighboring Aargo004633 are shared among the
372 three gene models, we can obtain two putative complete Xenopsin proteins. In other words, it
373 makes sense to think that the two xenopsins were probably splicing variants with alternative
374 promoters and shared two exons, which then duplicated and subfunctionalized (Force et al.,
375 1999; Hahn, 2009). *xenopsin* is found to be widespread but exclusively only in protostomes,
376 co-expressed together with *r-opsin* mostly in their ciliary photoreceptor cells (Passamaneck et
377 al., 2011; Vöcking et al., 2017; Rawlinson et al., 2019). Functional studies on *xenopsin* are
378 lacking and we therefore cannot decisively predict its function in *A. argo*. Phylogenetically,
379 *xenopsin* and *c-opsin* are apparently spread exclusively from each other, with *c-opsin* being
380 found exclusively in the photoreceptor cells of deuterostomes / vertebrates, suggesting that
381 *xenopsin*, similar to its deuterostomian counterpart *c-opsin*. is probably involved mainly in

382 phototactic responses and visual functions in protostomes (Döring et al., 2020), including *A.*
383 *argo*, *I. paradoxus*, and probably, other Cephalopods.
384 The two copies of non-canonical *r-opsins* in the genome of *A. argo* is most likely due to
385 the duplication of heterogeneous regions in the assembly, since the sequences matched
386 perfectly. This suggests that there is only a single non-canonical *r-opsin* in the genome of *A.*
387 *argo*, which is thus similar to *I. paradoxus*, as mentioned previously (Yoshida et al. 2015). At
388 present, the function of this Opsin homolog is still unknown, although previous studies
389 suggest that it's probably unrelated to visions, although apparently still related to
390 photoreception (Ramirez and Oakley, 2015; Ramirez et al. 2016; Bonade et al., 2020) . We
391 found two amino acid substitutions (T118S and Y178F) in the amino acid sequence of the
392 non-canonical R-Opsin of *A. argo* when compared to bovine rhodopsin. T118S was found in
393 both the benthic *O. bimaculoides* and *A. argo*, while Y178F was found only in the latter.
394 Prediction of light absorption wavelength of the non-canonical R-Opsin of *A. argo* indicates
395 that photoreceptions in *A. argo* are probably adapted more to red light than that of the benthic
396 octopus. The extra amino acid substitutions are thus consistent with the ecology of *A. argo*,
397 which lived closer to the sea surface than other cephalopods, indicating that the red-shift may
398 be an adaptation to shallow water light environment.
399
400 ***The evolution of shell and eggcase matrix proteins through independent recruitments, losses,***
401 ***and domain changes allows* A. argo *to obtain its eggcase and thus its pelagic lifestyle***
402 In this study, we found all of the eggcase matrix protein-coding genes in the genome of *A.*
403 *argo* (Table S5) as identified by our previous multi-omics study to survey and identify major
404 proteins of the eggcase matrices of two congeneric argonaut octopods, *A. argo* and *A. hians*
405 (Setiamarga et al., 2021b). Exactly congruent to our previous result, most of the proteins are
406 apparently not shared with the shell matrix proteins of Conchiferans, including those of the

basal Cephalopoda *Nautilus* (Setiamarga et al., 2021a; Huang et al., 2021), although the genes / proteins themselves are present in the genomes of the Conchiferan mollusks such as the limpet *L. gigantea* (Simakov et al., 2013), the true oyster *C. gigas* (Peñaloza et al., 2021), and the Japanese pearl oyster *P. fucata* (Takeuchi et al., 2016). Meanwhile, the Conchiferan shell matrix protein-coding genes were also mostly found in the genomes of *A. argo* and the shell-less benthic octopod *O. bimaculoides* (Albertin et al., 2015), indicating their retention despite shell loss in the octopod lineage. Interestingly, the genes for eggcase matrix proteins were also found in the genome of *O. bimaculoides*, and thus, when considered altogether, supported our hypothesis suggested previously, saying the argonaut octopods recruited many proteins unrelated to the shell formation and used them for their eggcase (Setiamarga et al., 2021b). However, very interestingly, some proteins related to calcification such as the Pif-like LamG3, seemed to be used at least by *A. hians* (Setiamarga et al., 2021b).

In that previous study, we arbitrarily categorized the Pif-like proteins mostly identified as Conchiferan shell matrix proteins into three paralogous groups, based on three monophyletic clades recovered in the phylogeny (see Figure 5 in Setiamarga et al., 2021b), which were also recovered in this study (Figure 5). We arbitrarily named them Blue Mussel Shell Protein (BMSP), Laminin G3 (LamG3), and Pif, and called them altogether the BMSP/LamG3/Pif proteins. These proteins could be distinguished by their domain combinations. BMSP was first identified as SMPs in the blue mussel *Mytilus galloprovincialis* and *L. gigantea*, respectively (Suzuki et al. 2011; Marie et al. 2017). The protein has one Chitin-Binding (ChtBd) and multiple (three or four) von Willebrand factor type A (VWA) domains. BMSP is present throughout the nacreous layer with dense localization in the myostracum, suggesting its possible role in Conchiferan nacreous layer formation (Suzuki et al. 2011). Meanwhile, Pif proteins, which was originally found in the nacre of *P. fucata*, usually have two types of domains, one VWA and two ChtBd domains, but with a different domain compositions and

arrangements (Figure 5) (Suzuki et al. 2009; Setiamarga et al., 2021a, b). *In vitro* functional analysis has shown that it is involved in calcium crystallization (Suzuki et al. 2013).

We did not detect any of the homologs of Pif and BMSP in the eggcase matrix of the argonauts in our previous multi-omics eggcase matrix protein study (Setiamarga et al., 2021b), nor in the shell matrix protein study of *Nautilus* (Setiamarga et al., 2021a). LamG3 was first identified by Marie et al. (2017) as one of the two Pif-like isoforms composed of the one VWA, three ChtBd, and one LamG domains. In both cephalopods, we instead found the last type of Pif homologs (*sensu* Setiamarga et al., 2021b), the LamG3 protein, in both the eggcase matrix of *A. hians* (but not in the eggcase matrix proteome and transcriptome of *A. argo*), and the shell matrix of *Nautilus*. However, very interestingly, differing with the results of our previous multi-omics study (Setiamarga et al., 2021b), in this study, we found the presence of *lamG3* in the genome of *A. argo* (Aargo013232) (Figure 5). Further studies must thus be conducted to assess if the absence of any transcript/protein product of *lamG3* in *A. argo*, despite its presence on the genome, is an artifact caused by the possible non-exhaustiveness of our previous multi-omics study, or if it is not used in the eggcase matrix of *A. argo*, making the eggcases of the two congeneric species different in nature.

At present, however, we are working under the hypothesis that this protein is a key protein for the formation of calcified eggcases in argonaut octopods, because it is one of the putative paralogs of the BMSP/LamG3/Pif-like proteins, which members have been identified as major component of the shell of Conchiferan mollusks. Intriguingly, however, although in the previous multi-omics study we did not find any sequence of *pif* or *BMSP* both in the transcriptome data of all tissues studied and the proteome data of *A. argo*, we found the presence of an intact coding sequence of *pif* in the genome of the species (Aargo018021). More interestingly, an intact *pif* sequence was also found in the genome of *N. pompilius* (Huang et al. 2021), although the sequence was not found in the transcriptome and proteome

457 data of our recent shell matrix proteins multi-omics study of the species (Setiamarga et al.,
458 2021a). The exon-intron structures of each cluster are different and are located at different
459 positions in the genome of both *A. argo* (Figure 5b) and *Nautilus*. Meanwhile, LamG3 has
460 also been shown to be associated with the biomineralization of shells in the pond snail
461 *Lymnaea stagnalis*, although no BMSP nor Pif were apparently found in its shell matrix
462 proteome and transcriptome, although additional studies involving genome analysis of the
463 species is still needed to confirm this observation (Ishikawa et al. 2020). These results seem
464 to thus indicate that the two Pif homologs (*pif* and *lamG3*) were probably already present
465 separately at least in the basal Conchiferan mollusks. However, ancestral Cephalopods even
466 more basal than the *Nautilus* probably lost *bmsp*, retained *lamG3* and *pif*, but use only *lamG3*
467 for the formation of biomineralized shells. Although further confirmation is still needed,
468 *lamG3* was probably recruited independently as an SMP, independently in each lineage
469 leading to terrestrial gastropods and cephalopods. The presence of *pif* in the genomes of
470 *Nautilus* and *A. argo*, and *lamG3* in the genome of *O. bimaculoides*, even though they are not
471 involved in the formation of shells or shell-like structures, is probably because they acquired
472 new functions unrelated to shell formation.

473 The lack of a typical LamG3 domain in BMSP and Pif have been reported (Suzuki et al.,
474 2013), and domain searches using various tools such as SMART (http://smart.embl-
475 heidelberg.de/, accessed in June 2021), InterProScan
476 (https://www.ebi.ac.uk/interpro/search/sequence/, accessed in June 2021), and Pfam
477 ( http://pfam.xfam.org/, accessed in June 2021) seemed to support this notion. In both BMSP
478 and Pif, no domain was detected. Some searches would detect only the repetitive low-
479 complexity domains (RLCD) which designates a possibility that the particular region used to
480 have a domain, but has probably degraded down and thus only recognizable partially at the
481 sequence level (Suzuki et al., 2017; Setiamarga et al., 2021a; b). We also predicted the

482 stereostructures of some representatives of BMSP (*Mytilus galloprovincialis*), Pif (*A. argo*,
483 *N. pompilius*, *P. fucata*, *L. gigantea*), and LamG3 (*A. argo*, *N. pompilius*, *O. bimaculoides*, *P.*
484 *fucata*, *L. gigantea, C. gigas*) using Alphafold2 (Jumper et al., 2021), and compared their 3D
485 structures (Figure 5). AlphaFold2 predicts accurate protein 3D structural models. Structural
486 comparisons using such models would allow us to obtain surprising information about the
487 function and evolution of the proteins, unattainable only through sequence comparison and
488 comparative genomics. The models of BMSP/LamG3/Pif proteins predicted by AlphaFold2
489 indicate that the region where no domain was detected in the proteins (except for the LamG3
490 proteins) actually still retains enough of its LamG3 domain characteristics (Figure 5). LamG3
491 domain is a receptor for various extracellular matrix proteins, which function is mediated by
492 the calcium ion (Tryggvason 1993; Yurchenco et al. 1993; Yu and Talts 2003; Klees et al.,
493 2008; Suzuki et al., 2017). This thus might explain the usefulness of the domain not only for
494 the formation of calcified structures but also for other functions, while at the same time also
495 suggests the unnecessity of the organisms compared to retain or use all of the
496 BMSP/LamG3/Pif protein homologs for the same function, which might thus also explain
497 why Cephalopods (*Nautilus*) only use LamG3 for their shell formation, and why the
498 argonauts also re-recruited this protein to form their eggcase.
499

500 **Conclusion**

501 Until very recently, studies on the evolution of Cephalopoda lacked insights from genomic
502 perspectives. However, recent genome data publications of various species have remedied
503 this. In this study, we present a genome assembly of *Argonauta argo*, which provides
504 significant insight into the genetic and evolutionary background of the adaptation to the
505 pelagic environment, such as the evolution of the visual proteins Opsin and Reflectin, and the
506 shell matrix protein Tyrosinase. The improved quality of the genome assembly also allowed

us to identify the presence of sexually highly polymorphic regions, which would be useful in future studies aiming at the elucidation of the genetic underpinnings of extreme male-female dimorphisms in the species. The pronounced sexual dimorphism probably evolved as an adaptation to holopelagic life in the open ocean with few male-female encounters. Besides that, the improved contiguity of the genome assembly confirmed the presence of several gene clusters including both deeply conserved ones, such as Hox, ParaHox, and Wnt, and unique ones that might be involved in evolutionary novelty.

The newly obtained draft genome sequence also allowed us to hypothesize about the evolution of some major shell matrix proteins related to calcification, seemingly re-recruited in the formation of the eggcase, which was impossible to do in our previous multi-omics-based studies. We also were able to corroborate our previous report based on a multi-omics study on the eggcase matrix proteins. In this study, we found all of the eggcase matrix proteins previously identified, while at the same time, also found the presence of LamG3 in the genome of *A. argo* (and *O. bimaculoides*), which was found as one of the eggcase matrix proteins of *A. hians* but not in *A. argo* in our previous multi-omics study. We also found an ortholog of the Pif coding gene in the genome of *A. argo*, besides in the recently published genome of *N. pompilius*. Combined with the protein structure prediction using Alphafold2, we thus were able to build a hypothesis about how BMSP/LamG3/Pif proteins evolved. In our hypothesis, the BMSP/LamG3/Pif proteins are key proteins for the formation of calcified external structures, including the eggcase. Therefore, the presence of *pif* in the genomes of *Nautilus* and *A. argo* and *lamG3* in the genome of *O. bimaculoides* might explain the usefulness of LamG3 domain for the formation of calcified structures, which might thus explain why the argonauts also re-recruited LamG3 protein, although not necessarily Pif and BMSP, to form their eggcases.

## Materials and Methods

### *Sampling, Sequencing, and Genome Size Estimation*

The *A. argo* DNA used for sequencing was derived from a single female provided by bycatch caught in the fixed nets set along the coast in Oki Island Town, Shimane Prefecture, Japan (36°17'20.6"N 133°12'46.4"E). Pieces of the gonad (ovary) were collected from an individual female specimen collected in 2018. The shell is registered as a collection of The University Museum, The University of Tokyo in Tokyo, Japan (Voucher No. RM33391). Genomic DNA was extracted from the ovary using the QIAGEN Genomic-tip kit. Pooled DNA was used for the preparation of three paired-end and three mate-pair (3, 6, 10, and 15 kbp insert size) libraries, that were sequenced on an Illumina HiSeq 2500 at the National Institute of Genetics, Japan with supports by Platform for Advanced Genome Science (PAGS) (Table S6, S7).

Pieces of the mantle, arm membrane of the first arm, and 2nd arm tip were obtained from the same single individual to genomic DNA. Eyes, hearts, and gill hearts were sampled from different individuals of *A. argo*. Six transcriptomes of *A. argo* were obtained and raw data statistics are provided in Table S6. Total RNA was extracted from the tissue samples using Trizol (Invitrogen) followed by an on-column DNaseI treatment using the RNeasy mini kit (Qiagen). The RNA acquoliot was stored at –80ºC until further transcriptome analyses.

The *A. argo* genome size and heterozygosity were assessed with GenomeScope v2.0 (Ranallo-Benavidez et al. 2020), based on the quality-filtered Illumina reads. A heterozygosity rate of 1.44% was estimated from the 32-mer-based assessment of the *A. argo* genome (Supplementary Figure S1). Complete microsatellite sequences were estimated and visualized with Krait v1.3.3 (Du et al. 2018b).

Raw read sequence data will be available in the DNA Data Bank of Japan (DDBJ). We are willing to share our raw data before the publication of the original paper on the assumption

that it will be done as a collaborative research.

*De Novo Genome Assembly and Annotation*

Using the predicted 1.1 Gb genome size estimate of the *A. argo*, the total raw sequence coverage of Illumina reads was 201× (pair-end reads, 3 kb, 6 kb, 10kb, and 15 kb mate-paired libraries). To reconstruct the mitogenome, we performed contig assembly (-n 200) with Platanus v1.2.4 (Kajitani et al. 2014) using the paired-end data. Contigs annotated as mitochondrial sequences were extracted by using the mitogenome data of a closely related species, *A. hians* (NC_036354), as the query for BLASTn homology search. After assembling the contigs, both ends of the resulting single contig were manually confirmed to overlap, and redundant parts were removed to complete the full circular mitogenome.

The pair-end sequence reads (PE600) after adapter trimming were assembled using De Bruijn graph assembler, Platanus-allee v. 2.2.2 (Kajitani et al. 2019). The basic algorithm of the Platanus-allee v2.2.2 is based on the arrangement of two independently assembled sequences derived from each haplotype of the corresponding two homologous chromosomes. Contig assembly was performed using only the PE library, and then scaffolding and gap closure were performed using all libraries. Assembly statistics by Platanus v222 was shown in Table S8.

Gene prediction models were generated using custom-made annotation pipeline as in (Inoue et al. 2021). In brief, this pipeline combines RNA-seq-based prediction results, homology-based prediction results for related species, and ab initio prediction results using in-house dynamic program. RNA-seq based prediction utilized both the assembly-first method and the mapping-first method. For the assembly-first method, RNA-seq data were assembled using Trinity (Grabherr et al. 2011) and Oases (Schulz et al. 2012). Then, assembled contigs were splice-mapped with GMAP (Wu et al. 2005). For the mapping-first

582 method, RNA-seq data were mapped to genome scaffolds and genes were predicted using
583 HISAT2 (Kim et al. 2019) and StringTie (Petea et al. 2016). In terms of homology-based
584 prediction, amino acid sequences of *Octopus vulgaris* (Zarrella *et al*. 2019), *Octopus*
585 *bimaculoides* (Albertin *et al*. 2015), *Architeuthis dux* (da Fonseca *et al*. 2020), *Crassostrea*
586 *giga* (Zhang *et al*. 2012), and *Mizuhopecten yessoensis* (Wang *et al.* 2017), were spliced-
587 mapped to genome scaffolds using Spaln62, and gene sets were predicted. For ab initio
588 prediction, raining sets were selected from RNA-seq based prediction results and
589 AUGUSTUS (Stanke et al. 2003) and SNAP (Korf et al. 2004) were trained and used for
590 prediction. Predicted results of each tool are shown in Table S9 and as a final result, 20,293
591 protein coding genes were predicted (Table S9). Predicted genes were evaluated using
592 BUSCO v3.0.2 (protein mode) (Simão et al. 2015) and resulted in 97.0% complete gene
593 marked, suggesting high accuracy of the annotation (Table S10). This goes beyond the
594 cephalopod genomes sequenced so far, and is comparable to high quality mollusc genomes
595 (Table S11).
596
597 *Phylogenetic Analysis*
598 Phylogenetic analyses were conducted on a total of five gene families obtained in this
599 study (*Hox, reflectin, tyrosinase, opsin, bmsp/lamg3/pif proteins*). To build single-gene trees
600 based on orthologs, we performed webBLAST search using *A. argo* protein sequences
601 translated from the gene sequences. Sequences for the phylogenetic tree were collected from
602 Genbank to cover the whole Lophotrochozoan clade. To perform multiple alignments of
603 protein sequences, we utilized the online version of MAFFT v7.487 (Katoh et al. 2002;
604 https://mafft.cbrc.jp/alignment/software/; accessed in August 2021), followed by the removal
605 of ambiguously aligned sites using the online version of trimAl_v1.4beta (automated option)
606 (Capella-Gutiérrez et al. 2009; http://phylemon2.bioinfo.cipf.es/index.html; accessed in

August 2021). Maximum likelihood phylogenetic inferences were executed on the software RAxMLGUI v2.0.5 (Silvestro et al. 2012; Stamatakis 2006) the rapid tree search setting with 1000 bootstrap replications under the best fit models (BMSP/LamG3/Pif proteins = WAG + Γ, Hox = LG + Γ + I, Reflectin = JTT + Γ + F, Tyrosinase = LG + Γ + I). The best fit models were inferred using MEGA X (Kumar et al. 2018). Obtained trees were visualized with FigTree v1.4.2 (Rambaut 2009).

For Opsin, sequences from other metazoans were collected from GenBank and Ensembl databases. Multiple sequence alignments of protein sequences were also performed by MAFFT. The best fit models were inferred using Modeltest (Darriba et al. 2020). Maximum likelihood phylogenetic inferences were executed on the software IQ-TREE (ref) the tree search setting with 1000 bootstrap replications under the best fit models (LG+G4: Best-fit model according to Bayesian Information Criterion (BIC) for c-opsins, LG+F+I+G4: BIC for r-opsins). The trees were visualized with FigTree v1.4.2 (Rambaut 2009).

## Acknowledgments

Samples were obtained with the courtesy of Mr. Minoru Yoshida and his colleagues of Yoshida Suisan, Oki Islands. Drs. Noriyoshi Sato of Tokai University and Hiroki Ono of Shimane University contributed to sample collections. Computations were partially performed on the NIG supercomputer at ROIS National Institute of Genetics, Shizuoka, Japan. We are grateful to Drs. Kazutoshi Yoshitake and Takeshi Kawashima for their help on how to use the Singularity module on the NIG supercomputer system. DHES, HK, and TS would like to thank members of Setiamarga Lab at National Institute of Technology, Wakayama College, and Sasaki Lab at The University Museum of The University of Tokyo, for their constant support during the commencement of the study.

The work was supported by Human Frontier Science Program grant (RGP0060/2017) to

632 MAY, Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number
633 18K06363 (Grant-in-Aid for Scientific Research (C))] to MAY and DHES and 16H06279
634 (PAGS) to AT and TI. DHES was also supported by the FY2016 Research Grant for
635 Chemistry and Life Sciences (The Asahi Glass Foundation), and partially supported by the
636 FY2017 Research Grant for Zoology (Fujiwara Natural History Research Foundation) and
637 the FY2019 Grant-in-Aid for Challenging (Exploratory) Research (Grant number:
638 19K21646) (awarded to TS (PI) and DHES (Co-PI)). KI was partially supported by the
639 Platform Project for Supporting Drug Discovery and Life Science Research from the
640 Japanese Agency for Medical Research and Development (AMED). We also thank the
641 faculty of Life and Environmental Science at Shimane University for help in financial
642 support for publishing this report.

643

644 **References**

645 Addadi, L., Joester, D., Nudelman, F. & Weiner, S. Mollusk shell formation: a source of new
646 concepts for understanding biomineralization processes. Chemistry 12, 980–987 (2006)
647 Albertin, C. B. et al. The octopus genome and the evolution of cephalopod neural and
648 morphological novelties. Nature 524, 220–224 (2015)
649 Alon, S. et al. The majority of transcripts in the squid nervous system are extensively recoded
650 by A-to-I RNA editing. Elife 4, (2015)
651 Ascencio, D. et al. Expression attenuation as a mechanism of robustness against gene
652 duplication. Proc. Natl. Acad. Sci. U. S. A. 118, (2021)
653 Bai, Y., Nie, H., Wang, Z. & Yan, X. Genome-wide identification and transcriptome-based
654 expression profiling of Wnt gene family in Ruditapes philippinarum. Comp. Biochem.
655 Physiol. Part D Genomics Proteomics 35, 100709 (2020)
656 Belcaid, M. et al. Symbiotic organs shaped by distinct modes of genome evolution in

657 cephalopods. Proc. Natl. Acad. Sci. U. S. A. 116, 3030–3035 (2019)
658 Bello, G. Exaptations in Argonautoidea (Cephalopoda: Coleoidea: Octopoda). Neues
659 Jahrbuch für Geologie und Paläontologie - Abhandlungen 266, 85–92 (2012)
660 Bizikov, V. A. The shell in Vampyropoda (Cephalopoda): Morphology, functional role and
661 evolution. 3, 1–88 (2004)
662 Bonadè, M., Ogura, A., Corre, E., Bassaglia, Y. & Bonnaud-Ponticelli, L. Diversity of Light
663 Sensing Molecules and Their Expression During the Embryogenesis of the Cuttlefish
664 (Sepia officinalis). Front. Physiol. 11, 521989 (2020)
665 Brain, P. F. The Brain and Lives of Cephalopods. By Marion Nixon and John Z. Young, 2003
666 (Oxford: Oxford University Press) [xiv + 392 p with numerous bw photographs,
667 photomicrographs and line drawings]. Price £175 (hbk). ISBN 0-19-852761-6. J. Nat.
668 Hist. 39, 863–863 (2005)
669 Brooke, N. M., Garcia-Fernàndez, J. & Holland, P. W. The ParaHox gene cluster is an
670 evolutionary sister of the Hox gene cluster. Nature 392, 920–922 (1998)
671 Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated
672 alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973
673 (2009)
674 Chiu, Y.-W., Chang, C.-W., Lin, H.-D. & Shen, K.-N. The complete mitogenome of the
675 winged argonaut Argonauta hians and its phylogenetic relationships in Octopoda. Conserv.
676 Genet. Resour. 10, 359–362 (2018)
677 Cho, S.-J., Vallès, Y., Giani, V. C., Jr, Seaver, E. C. & Weisblat, D. A. Evolutionary
678 dynamics of the wnt gene family: a lophotrochozoan perspective. Mol. Biol. Evol. 27,
679 1645–1658 (2010)
680 Crookes, W. J. et al. Reflectins: The Unusual Proteins of Squid Reflective Tissues. Science
681 (2004) doi:10.1126/science.1091288

682 da Fonseca, R. R. et al. A draft genome sequence of the elusive giant squid, Architeuthis dux.
683 Gigascience 9, (2020)
684 Darriba, D. et al. ModelTest-NG: A New and Scalable Tool for the Selection of DNA and
685 Protein Evolutionary Models. Mol. Biol. Evol. 37, 291–294 (2020)
686 DeMartini, D. G., Izumi, M., Weaver, A. T., Pandolfi, E. & Morse, D. E. Structures,
687 Organization, and Function of Reflectin Proteins in Dynamically Tunable Reflective Cells.
688 J. Biol. Chem. 290, 15238–15249 (2015)
689 Du, J. et al. Wnt gene family members and their expression profiling in Litopenaeus
690 vannamei. Fish Shellfish Immunol. 77, 233–243 (2018)
691 Du, L. et al. Krait: an ultrafast tool for genome-wide survey of microsatellites and primer
692 design. Bioinformatics 34, 681–683 (2018)
693 Dyachuk, V. Extracellular matrix components in Bivalvia: Shell and ECM components in
694 developmental and adult tissues. Fish. Aquac. J. 09, (2018)
695 Döring, C. C., Kumar, S., Tumu, S. C., Kourtesis, I. & Hausen, H. The visual pigment
696 xenopsin is widespread in protostome eyes and impacts the view on eye evolution. Elife 9,
697 (2020)
698 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high
699 throughput. Nucleic Acids Res. 32, 1792–1797 (2004)
700 Feng, D., Li, Q., Yu, H., Kong, L. & Du, S. Identification of conserved proteins from diverse
701 shell matrix proteome in Crassostrea gigas: characterization of genetic bases regulating
702 shell formation. Sci. Rep. 7, 45754 (2017)
703 Finn, J. K. & Norman, M. D. The argonaut shell: gas-mediated buoyancy control in a pelagic
704 octopus. Proc. Biol. Sci. 277, 2967–2971 (2010)
705 Force, A. et al. Preservation of duplicate genes by complementary, degenerative mutations.
706 Genetics 151, 1531–1545 (1999)

Garstang, M. & Ferrier, D. E. K. Time is of the essence for ParaHox homeobox gene clustering. BMC Biol. 11, 72 (2013)

Gaunt, S. J. The significance of Hox gene collinearity. Int. J. Dev. Biol. 59, 159–170 (2015)

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

Graham, A., Papalopulu, N. & Krumlauf, R. The murine and Drosophila homeobox gene complexes have common features of organization and expression. Cell 57, 367–378 (1989)

Gregory, T.R. (2021). Animal Genome Size Database. http://www.genomesize.com.

Hallin, J. & Landry, C. R. Regulation plays a multifaceted role in the retention of gene duplicates. PLoS biology vol. 17 e3000519 (2019)

Hahn, M. W. Distinguishing among evolutionary models for the maintenance of gene duplicates. *J. Hered.* **100**, 605–617 (2009)

Hirota, K., Yoshida, M.-A., Itoh, T., Toyoda, A. & Setiamarga, D. H. E. The full mitochondrial genome sequence of the greater argonaut Argonauta argo (Cephalopoda, Argonautoidea) and its phylogenetic position in Octopodiformes. Mitochondrial DNA B Resour 6, 1451–1453 (2021)

Holstein, T. W. The evolution of the Wnt pathway. Cold Spring Harb. Perspect. Biol. 4, a007922 (2012)

Huang, Z. et al. Genomic insights into the adaptation and evolution of the nautilus, an ancient but evolving 'living fossil'. Mol. Ecol. Resour. (2021) doi:10.1111/1755-0998.13439

Inoue, K. et al. Genomics and Transcriptomics of the green mussel explain the durability of its byssus. Sci. Rep. 11, 5992 (2021)

Ishikawa, A. et al. Functional shell matrix proteins tentatively identified by asymmetric snail shell morphology. Sci. Rep. 10, 9768 (2020)

732 Jackson, D. J. et al. Parallel evolution of nacre building gene sets in molluscs. Mol. Biol.
733 Evol. 27, 591–608 (2010)
734 Jackson, D. J. et al. Variation in Orthologous Shell-Forming Proteins Contribute to
735 Molluscan Shell Diversity. Mol. Biol. Evol. 34, 2959–2969 (2017)
736 Janssen, R. et al. Conservation, loss, and redeployment of Wnt ligands in protostomes:
737 implications for understanding the evolution of segment formation. BMC Evol. Biol. 10,
738 374 (2010)
739 Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. Nature 596,
740 583–589 (2021)
741 Kajitani, R. et al. Platanus-allee is a de novo haplotype assembler enabling a comprehensive
742 access to divergent heterozygous regions. Nat. Commun. 10, 1702 (2019)
743 Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment
744 and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. 37, 907–915 (2019
745 Klees, R. F. et al. Dissection of the osteogenic effects of laminin-332 utilizing specific LG
746 domains: LG3 induces osteogenic differentiation, but not mineralization. Exp. Cell Res.
747 314, 763–773 (2008)
748 Kocot, K. M., Aguilera, F., McDougall, C., Jackson, D. J. & Degnan, B. M. Sea shell
749 diversity and rapidly evolving secretomes: insights into the evolution of biomineralization.
750 Front. Zool. 13, 23 (2016)
751 Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: a fast,
752 scalable and user-friendly tool for maximum likelihood phylogenetic inference.
753 Bioinformatics 35, 4453–4455 (2019)
754 Kröger, B., Vinther, J. & Fuchs, D. Cephalopod origin and evolution: A congruent picture
755 emerging from fossils, development and molecules: Extant cephalopods are younger than
756 previously realised and were under major selection to become agile, shell-less predators.

757 Bioessays 33, 602–613 (2011)

758 Levenson, R., DeMartini, D. G. & Morse, D. E. Molecular mechanism of reflectin's tunable
759 biophotonic control: Opportunities and limitations for new optoelectronics. APL Materials
760 5, 104801 (2017)

761 Li, F. et al. Chromosome-level genome assembly of the East Asian common octopus
762 (Octopus sinensis) using PacBio sequencing and Hi-C technology. Mol. Ecol. Resour. 20,
763 172–1582 (2020)

764 Liscovitch-Brauer, N. et al. Trade-off between Transcriptome Plasticity and Genome
765 Evolution in Cephalopods. Cell 169, 191–202.e11 (2017)

766 Liu, J., Xu, F., Ji, P., Li, L. & Zhang, G. Evolutionary dynamics of the Wnt gene family:
767 implications for lophotrochozoans. Journal of Oceanology and Limnology 36, 1720–1730
768 (2018)

769 Lowenstam, H. A. & Weiner, S. On Biomineralization. in On Biomineralization (Oxford
770 University Press, 2020). doi:10.1093/oso/9780195049770.001.0001

771 Lynch, M. & Force, A. The probability of duplicate gene preservation by
772 subfunctionalization. Genetics 154, 459–473 (2000)

773 Mann, K., Edsinger-Gonzales, E. & Mann, M. In-depth proteomic analysis of a mollusc shell:
774 acid-soluble and acid-insoluble matrix of the limpet Lottia gigantea. Proteome Sci. 10, 28
775 (2012)

776 Marie, B. et al. Different secretory repertoires control the biomineralization processes of
777 prism and nacre deposition of the pearl oyster shell. Proc. Natl. Acad. Sci. U. S. A. 109,
778 20986–20991 (2012)

779 Marie, B. et al. Deep conservation of bivalve nacre proteins highlighted by shell matrix
780 proteomics of the Unionoida Elliptio complanata and Villosa lienosa. J. R. Soc. Interface
781 14, (2017)

782 Marie, B. et al. Evolution of nacre: biochemistry and proteomics of the shell organic matrix
783 of the cephalopod Nautilus macromphalus. Chembiochem 10, 1495–1506 (2009)
784 Marin, F. et al. Skeletal Organic Matrices in Molluscs: Origin, Evolution, Diagenesis. in
785 Biomineralization 325–332 (Springer Singapore, 2018). doi:10.1007/978-981-13-1002-
786 7_34
787 McDougall, C., Aguilera, F. & Degnan, B. M. Rapid evolution of pearl oyster shell matrix
788 proteins with repetitive, low-complexity domains. J. R. Soc. Interface 10, 20130041
789 (2013)
790 Mitchell, P. R., Phakey, P. P. & Rachinger, W. A. Ultrastructural Observations of the
791 Argonaut Shell. Scanning Microsc. 8, 4 (1994)
792 Miyamoto, H. et al. The diversity of shell matrix proteins: genome-wide investigation of the
793 pearl oyster, Pinctada fucata. Zoolog. Sci. 30, 801–816 (2013)
794 Montavon, T. HoxGenes: Embryonic Development. eLS 1–8 (2015)
795 doi:10.1002/9780470015902.a0005046.pub2
796 Morel, G. et al. Differential gene retention as an evolutionary mechanism to generate
797 biodiversity and adaptation in yeasts. Sci. Rep. 5, 11571 (2015)
798 Naef A, 1923. Cephalopoda. Fauna e Flora del Golfo di Napoli (Translated from German by
799 the Israel Program for Scientific Translations, Jerusalem, 1972). Monograph 35.
800 Norman M. 2000. Cephalopods a world guide Octopuses, Argonauts, Cuttlefish, Squid,
801 Nautilus. 320pp. ConchBooks
802 Oudot, M. et al. The shell matrix and microstructure of the Ram's Horn squid: Molecular and
803 structural characterization. J. Struct. Biol. 211, 107507 (2020)
804 Packard, A. & Wurtz, M. An octopus, Ocythoe, with a swimbladder and triple jets. Philos.
805 Trans. R. Soc. Lond. B Biol. Sci. 344, 261–275 (1994)
806 Passamaneck, Y. J., Furchheim, N., Hejnol, A., Martindale, M. Q. & Lüter, C. Ciliary

807 photoreceptors in the cerebral eyes of a protostome larva. Evodevo 2, 6 (2011)
808 Peñaloza, C. et al. A chromosome-level genome assembly for the Pacific oyster Crassostrea
809 gigas. Gigascience 10, (2021)
810 Powers, T. P. et al. Characterization of the Hox cluster from the mosquito Anopheles
811 gambiae (Diptera: Culicidae). Evol. Dev. 2, 311–325 (2000)
812 Ramirez, M. D. et al. The Last Common Ancestor of Most Bilaterian Animals Possessed at
813 Least Nine Opsins. Genome Biol. Evol. 8, 3640–3652 (2016)
814 Ramirez, M. D. & Oakley, T. H. Eye-independent, light-activated chromatophore expansion
815 (LACE) and expression of phototransduction genes in the skin of Octopus bimaculoides. J.
816 Exp. Biol. 218, 1513–1520 (2015)
817 Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and Smudgeplot
818 for reference-free profiling of polyploid genomes. Nat. Commun. 11, 1432 (2020)
819 Rawlinson, K. A. et al. Extraocular, rod-like photoreceptors in a flatworm express xenopsin
820 photopigment. Elife 8, (2019)
821 Revelle, R. & Fairbridge, R. W. Carbonates and carbon dioxide. (Geological Society of
822 America, 1957).
823 Robertson, L. K. & Mahaffey, J. W. Insect Homeotic Complex Genes and Development,
824 Lessons From Drosophila and Beyond☆. in Reference Module in Life Sciences (Elsevier,
825 2017). doi:10.1016/B978-0-12-809633-8.04008-5
826 Sakurai, T. & Kawano, S. Argonautidae (Cephalopoda) obtained from set nets off the
827 Shimane Peninsula, southwestern part of the Japan Sea during summer, 2009. 8, 41–46
828 (2010)
829 Sanchez, G. et al. Genus-level phylogeny of cephalopods using molecular markers: current
830 status and problematic areas. PeerJ 6, e4331 (2018)
831 Saul, L. R. & Stadum, C. J. Fossil argonauts (Mollusca: Cephalopoda: Octopodida) from Late

832 Miocene Siltstones of the Los Angeles Basin, California. J. Paleontol. 79, 520–531 (2005)
833 Scales, H. Spirals in Time: The Secret Life and Curious Afterlife of Seashells. (Bloomsbury
834 Publishing, 2015).
835 Schulz, M. H., Zerbino, D. R., Vingron, M. & Birney, E. Oases: robust de novo RNA-seq
836 assembly across the dynamic range of expression levels. Bioinformatics 28, 1086–1092
837 (2012)
838 Setiamarga, D. H. E. et al. Hydrophilic Shell Matrix Proteins of Nautilus pompilius and The
839 Identification of a Core Set of Conchiferan Domains. bioRxiv 2020.11.14.382804 (2020)
840 doi:10.1101/2020.11.14.382804
841 Setiamarga, D. H. E. et al. Independent adoptions of a set of proteins found in the matrix of
842 the mineralized shell-like eggcase of Argonaut octopuses. bioRxiv 2021.07.10.451900
843 (2021) doi:10.1101/2021.07.10.451900
844 Setiamarga, D. H. E. et al. An in-silico genomic survey to annotate genes coding for early
845 development-relevant signaling molecules in the pearl oyster, Pinctada fucata. Zoolog. Sci.
846 30, 877–888 (2013)
847 Shippy, T. D. et al. Analysis of the Tribolium homeotic complex: insights into mechanisms
848 constraining insect Hox clusters. Dev. Genes Evol. 218, 127–139 (2008)
849 Simakov, O. et al. Insights into bilaterian evolution from three spiralian genomes. Nature
850 493, 526–531 (2013)
851 Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M.
852 BUSCO: assessing genome assembly and annotation completeness with single-copy
853 orthologs. Bioinformatics 31, 3210–3212 (2015)
854 Song, X.-M. et al. Preferential gene retention increases the robustness of cold regulation in
855 Brassicaceae and other plants after polyploidization. Hortic Res 7, 20 (2020)
856 Stanke, M. & Waack, S. Gene prediction with a hidden Markov model and a new intron

857 submodel. Bioinformatics 19 Suppl 2, ii215–25 (2003)

858 Stevens, K., Iba, Y., Suzuki, A. & Mutterlose, J. Biological and environmental signals
859 recorded in shells of Argonauta argo (Cephalopoda, Octobrachia) from the Sea of Japan.
860 Mar. Biol. 162, (2015)

861 Strugnell, J., Jackson, J., Drummond, A. J. & Cooper, A. Divergence time estimates for
862 major cephalopod groups: evidence from multiple genes. Cladistics 22, 89–96 (2006)

863 Sugino, R. P. & Innan, H. Selection for more of the same product as a force to enhance
864 concerted evolution of duplicated genes. Trends Genet. 22, 642–644 (2006)

865 Suzuki, M., Iwashima, A., Kimura, M., Kogure, T. & Nagasawa, H. The molecular evolution
866 of the pif family proteins in various species of mollusks. Mar. Biotechnol. 15, 145–158
867 (2013)

868 Suzuki, M. et al. Identification and characterisation of a calcium carbonate-binding protein,
869 blue mussel shell protein (BMSP), from the nacreous layer. Chembiochem 12, 2478–2487
870 (2011)

871 Suzuki, M., Kogure, T. & Nagasawa, H. Studies on the chemical structures of organic
872 matrices and their functions in the biomineralization processes of molluscan shells. AGri-
873 Biosci. Monogr. 7, 25–39 (2017)

874 Suzuki, M. et al. An Acidic Matrix Protein, Pif, Is a Key Macromolecule for Nacre
875 Formation. Science 325, 1388–1390 (2009)

876 Takeuchi, T. et al. Bivalve-specific gene expansion in the pearl oyster genome: implications
877 of adaptation to a sessile lifestyle. Zoological Lett 2, 3 (2016)

878 Tribolium Genome Sequencing Consortium et al. The genome of the model beetle and pest
879 Tribolium castaneum. Nature 452, 949–955 (2008)

880 Tryggvason, K. The laminin family. Curr. Opin. Cell Biol. 5, 877–882 (1993)

881 Vendrasco, M. J., Checa, A., Heimbrock, W. P. & Baumann, S. D. J. Nacre in Molluscs from

the Ordovician of the Midwestern United States. Geosci. J. 3, 1–29 (2013)

Vendrasco, M. J., Checa, A. G. & Kouchinsky, A. V. Shell microstructure of the early bivalve Pojetaia and the independent origin of nacre within the mollusca. Palaeontology 54, 825–850 (2011)

Vinther, J., Sperling, E. A., Briggs, D. E. G. & Peterson, K. J. A molecular palaeobiological hypothesis for the origin of aplacophoran molluscs and their derivation from chiton-like ancestors. Proc. Biol. Sci. 279, 1259–1268 (2012)

Von Allmen, G. et al. Splits in fruitfly Hox gene complexes. Nature 380, 116 (1996)

Vöcking, O., Kourtesis, I., Tumu, S. C. & Hausen, H. Co-expression of xenopsin and rhabdomeric opsin in photoreceptors bearing microvilli and cilia. Elife 6, (2017)

Wagner, G. P., Amemiya, C. & Ruddle, F. Hox cluster duplications and the opportunity for evolutionary novelties. Proc. Natl. Acad. Sci. U. S. A. 100, 14603–14606 (2003)

Wang, S. et al. Scallop genome provides insights into evolution of bilaterian karyotype and development. Nat Ecol Evol 1, 120 (2017)

Wang, C. et al. Characterization of Wnt Genes in Argopecten Scallops and Their Involvement in Responses to Different Temperature Stresses in Bohai Red Scallops. Research Square (2021) doi:10.21203/rs.3.rs-829136/v1

Wu, T. D. & Watanabe, C. K. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics 21, 1859–1875 (2005)

Yoshida, M. A. et al. Molecular Evidence for Convergence and Parallelism in Evolution of Complex Brains of Cephalopod Molluscs: Insights from Visual Systems. Integr. Comp. Biol. 55, 1070–1083 (2015)

Yoshida, M.-A. et al. Genome structure analysis of molluscs revealed whole genome duplication and lineage specific repeat variation. Gene 483, 63–71 (2011)

Young, J. Z. Cephalopods and Neuroscience. Biol. Bull. 168, 153–158 (1985)

907 Young, J. Z. The anatomy of the nervous system of Octopus vulgaris,. (Clarendon Press,
908 1971).
909 Yu, H. & Talts, J. F. Beta1 integrin and alpha-dystroglycan binding sites are localized to
910 different laminin-G-domain-like (LG) modules within the laminin alpha5 chain G domain.
911 Biochem. J 371, 289–299 (2003)
912 Yurchenco, P. D., Sung, U., Ward, Yamada, Y. & O'Rear, J. J. Recombinant laminin G
913 domain mediates myoblast adhesion and heparin binding. J. Biol. Chem. 268, 8356–8365
914 (1993)
915 Zarrella, I. et al. The survey and reference assisted assembly of the Octopus vulgaris genome.
916 Sci Data 6, 13 (2019)
917 Zhang, G. et al. The oyster genome reveals stress adaptation and complexity of shell
918 formation. Nature 490, 49–54 (2012)
919

## Figure legends and Tables

Figure 1 The Argonaut octopuses. A. The shell-like eggcase of *Argonauta argo*. B. The shell-like eggcase of *A. hians.* C. Collect location.

Figure 2 Schematic representations of Hox/Parahox/Wnt clusters. A. Simplified classification of the Hox cluster genomic organization of the cephalopods with the genome sequenced. Scaffold number and length are shown for the *A. argo* genome. The gene model IDs of each gene are shown above each box. The sequences of the homeobox region were confirmed from scaffold for those gene IDs not listed. Hox2/pb and Hox4/Dfd were also not found in the *A. argo* genome as in the *O. bimaculoides* genome. B. Simplified classification of the Parahox and Wnt cluster genomic organizations of molluscs with the genome sequenced. Scaffold number and length are shown for the *A. argo* genome.

Figure 3 Schematic representations of reflectin clusters. A. Reflectin clusters of the octopuses. B. Gene expression levels of *A. argo* reflectins.

Figure 4 Schematic representations of tyrosinase clusters.

Figure 5 Phylogenetic relationships of Pif/Pif-like/BMSPs of Molluscs and representative 3D protein models. The maximum likelihood tree was estimated under the best fit models (WAG + Γ). Numbers on the nodes are Bootstrap Support (BS) values. BS lower than 41% are shown as "--", while 100% support is not written. Representative structures of the proteins of the sequences included in the analyses, shown as SMART protein domains, are shown below the trees. Four 3D structural models (PIF; Aargo018021 [*Argonauta argo*], pfu_aug2_0_956_1_21296_t1 [*Pinctada fucata*], LamG3; Aargo013232 [*Argonauta argo*], Ocbimv22010162m_p [*Octopus bimaculoides*]) were estimated with AlphaFold2. Schematic representation of domain structure and 3D structural model were colored each domain characteristic: Signal peptide, red; VWA, pale orange; 1st ChtBd, green; 2nd ChtBd, blue; 3rd ChtBd, yellow; LamG and RLCD, pink.

945 Figure S1 GenomeScope result.

946 Figure S2 Microsatellite types found in the Argonaut genome.

947 Figure S3 Molecular phylogenetic tree of the Hox genes. The maximum likelihood
948 phylogenetic tree inferred under the LG + Γ + I model with 1000 bootstrap replicates. Hox
949 genes of *Argonauta argo* are marked with a black arrow. Abbreviations: Nuctum: *Nucula*
950 *tumidula*, Cragig: *Crassostrea gigas*, Pecmax: *Pecten maximus*, Gibvar: *Gibbula varia*,
951 Lotgig: *Lottia gigantea*, Apcal: *Aplysia californica*, Eupsco: *Euprymna scolopes*, Octbim:
952 *Octopu bimaculoides*, Naupom: *Nautilus pompilius*, Acacri: *Acanthochitona crinite*,
953 Antent: *Antalis entails*, Glympell: *Gymnomenia pellucida*, Alivir: *Alitta virens*, Linana:
954 *Lingula anatine*: Dromel: *Drosophila melanogaster*, Braflo: *Branchiostoma floridae*:
955 Caeele, *Caenorhabditis elegans*.

956 Figure S4 Alignment of Hox genes recovered in the scaffolds but not in the gene models.

957 Figure S5 Reflectin phylogenetic tree

958 Figure S6 Reflectin alignment with reference to repetitive reflectin motifs

959 Figure S7 Reflectin alignment to show gene conversion

960 Figure S8 Tyrosinase phylogenetic tree. The maximum likelihood phylogenetic tree inferred
961 under the LG + Γ + I model with 1000 bootstrap replicates. Numbers on the nodes are
962 Bootstrap Support (BS) values. BS lower than 41% are not shown, while 100% support is
963 shown as a black square. Tyrosinase of *Argonauta argo* are marked with underlined. Three
964 type of tyrosinase are shown secreted (α), cytosolic (β) and membrane-bound (γ)
965 subclasses.

966 Figure S9 Tyrosinase alignment at amino acid level

967 Figure S10 Tyrosinase alignment to show gene conversion

968 Figure S11 Opsin alignment and unique amino acid changes of argonaut based on bovine
969 rhodopsin

970 Figure S12 Phylogenetic tree of RGR and rhabdomeric opsins

971 Figure S13 Xenopsin phylogenetic tree

972 Figure S14-16 Protein structure of Pif/Pif-like/BMSPs. The schematic representation of three
973 proteins of Pif/Pif-like/BMSPs are shown within the box frame. Conserved domains within
974 each protein are predicted in SMART, and the 3D structural models were estimated with
975 AlphaFold2. The domains regions, which distinguished by the domain prediction and its
976 conserved alignment regions, are marked different color: Signal peptide (red), 1st VWA
977 (deep orange), 2nd VWA (ocher), 3rd VWA (pale orange), 4th VWA (orange), 1st ChtBd
978 (green), 2nd ChtBd (blue), 3rd ChtBd (yellow), LamG and RLCD (pink).

979

980 Table S1 Assembly comparison among molluscan genomes

981 Table S2 Assembly comparison based on BUSCO scores (genome mode, metazoa_odb9,
982 n=978)

983 Table S3 Krait estimation of the microsatellite regions

984 Table S4 Comparison of ORFs in the Hox cluster between blue mussels and giant squid

985 Table S5 Proteome list

986 Table S6 List of *A. argo* genome sequencing data

987 Table S7 List of *A. argo* RNA-seq sequencing data

988 Table S8 Assembly statistics by Platanus v222

989 Table S9 Gene prediction models using custom-made annotation pipeline with transcriptomic
990 data

991 Table S10 Gene model comparison based on BUSCO scores (protein mode, metazoa_odb9,
992 n=978)

993 Table S11 Gene model comparison among molluscan genomes



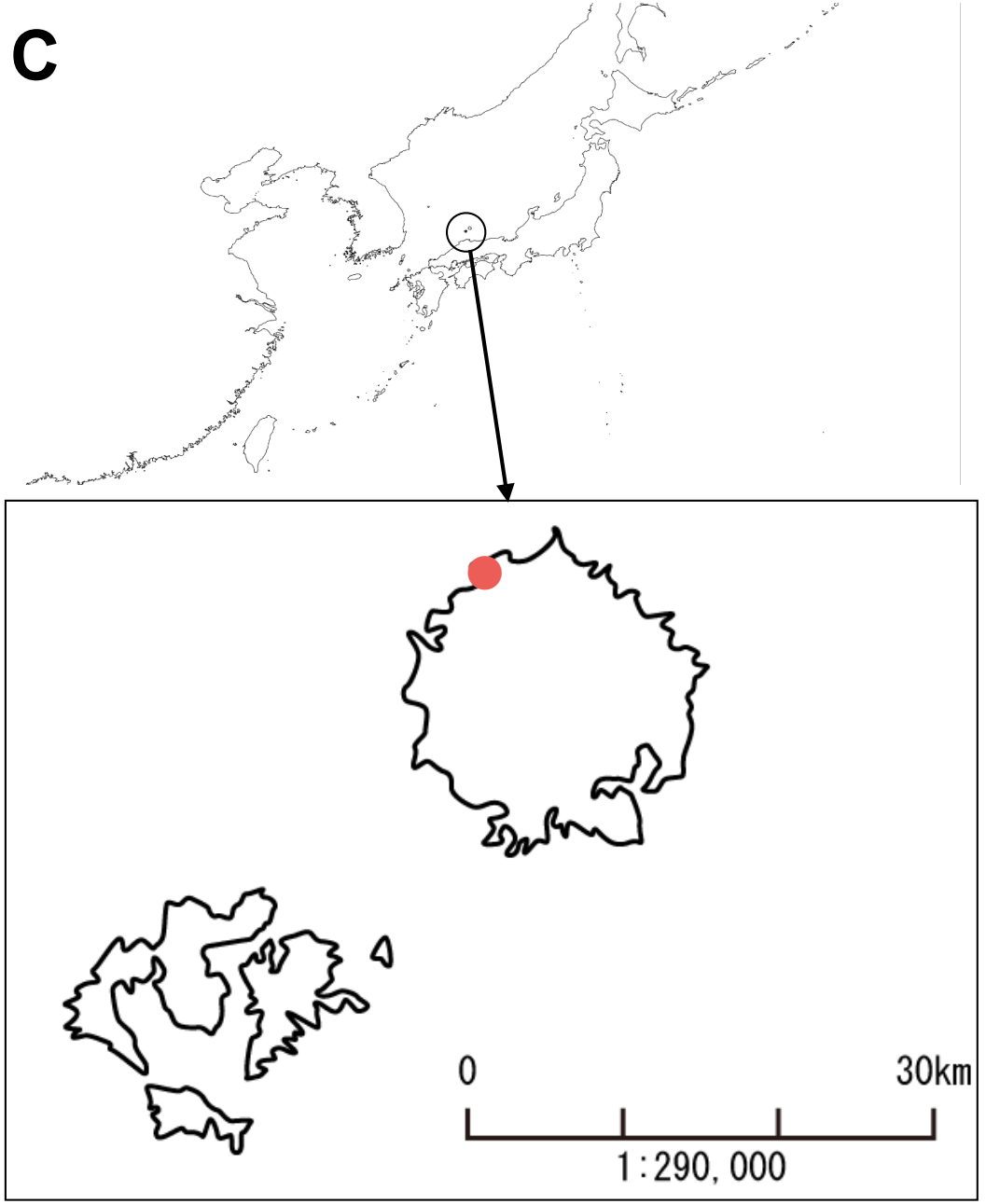



Fig. 1

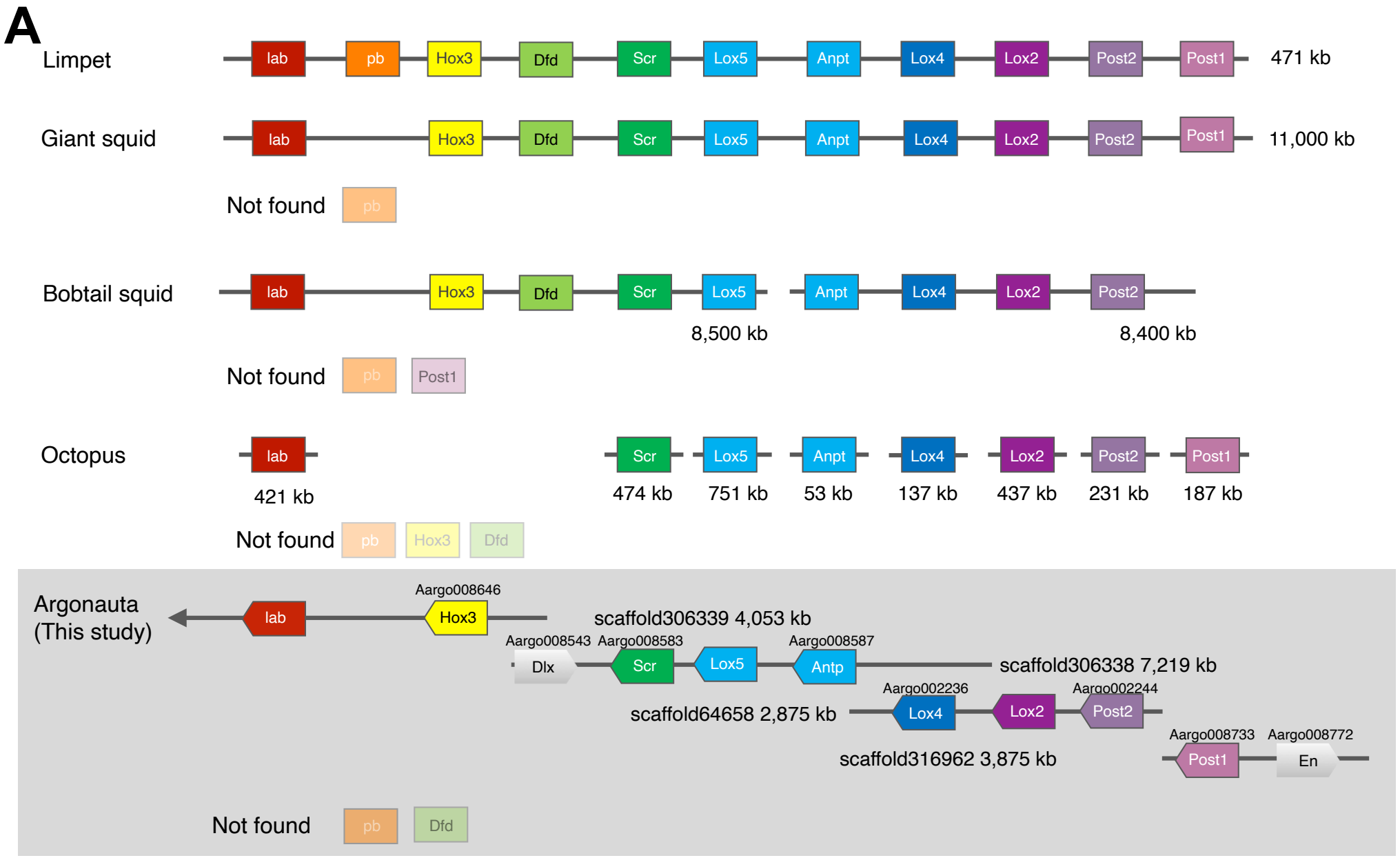

A
Limpet
lab
pb
Hox3
Dfd
Scr
Lox5
Anpt
Lox4
Lox2
Post2
Post1
471 kb
Giant squid
11,000 kb
Not found
Bobtail squid
8,500 kb
8,400 kb
Octopus
421 kb
474 kb
751 kb
53 kb
137 kb
437 kb
231 kb
187 kb
Argonauta
(This study)
Aargo008646
scaffold306339 4,053 kb
Aargo008543
Aargo008583
Aargo008587
Dlx
Antp
scaffold306338 7,219 kb
Aargo002236
Aargo002244
scaffold64658 2,875 kb
Aargo008733
Aargo008772
scaffold316962 3,875 kb
En

Fig.2

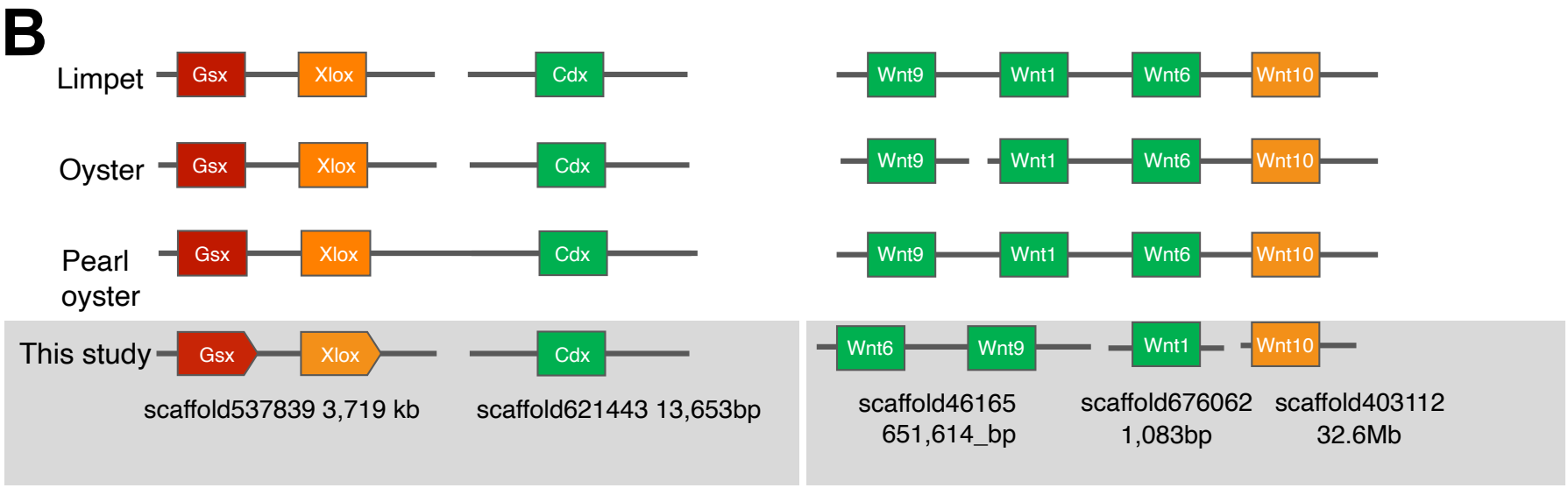
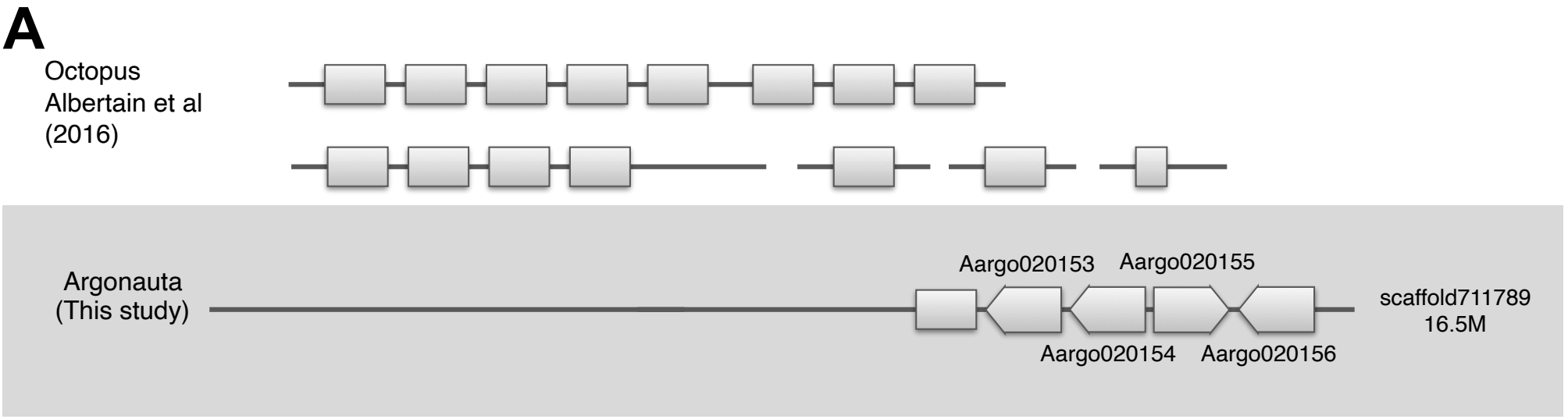

B
Limpet
Gsx
Xlox
Cdx
Wnt9
Wnt1
Wnt6
Wnt10
Oyster
Gsx
Xlox
Cdx
Wnt9
Wnt1
Wnt6
Wnt10
Pearl oyster
Gsx
Xlox
Cdx
Wnt9
Wnt1
Wnt6
Wnt10
This study
Gsx
Xlox
Cdx
Wnt6
Wnt9
Wnt1
Wnt10
scaffold537839 3,719 kb
scaffold621443 13,653bp
scaffold46165 651,614_bp
scaffold676062 1,083bp
scaffold403112 32.6Mb

Fig.2b


A
Octopus
Albertain et al
(2016)
Argonauta
(This study)
Aargo020153
Aargo020155
Aargo020154
Aargo020156
scaffold711789
16.5M

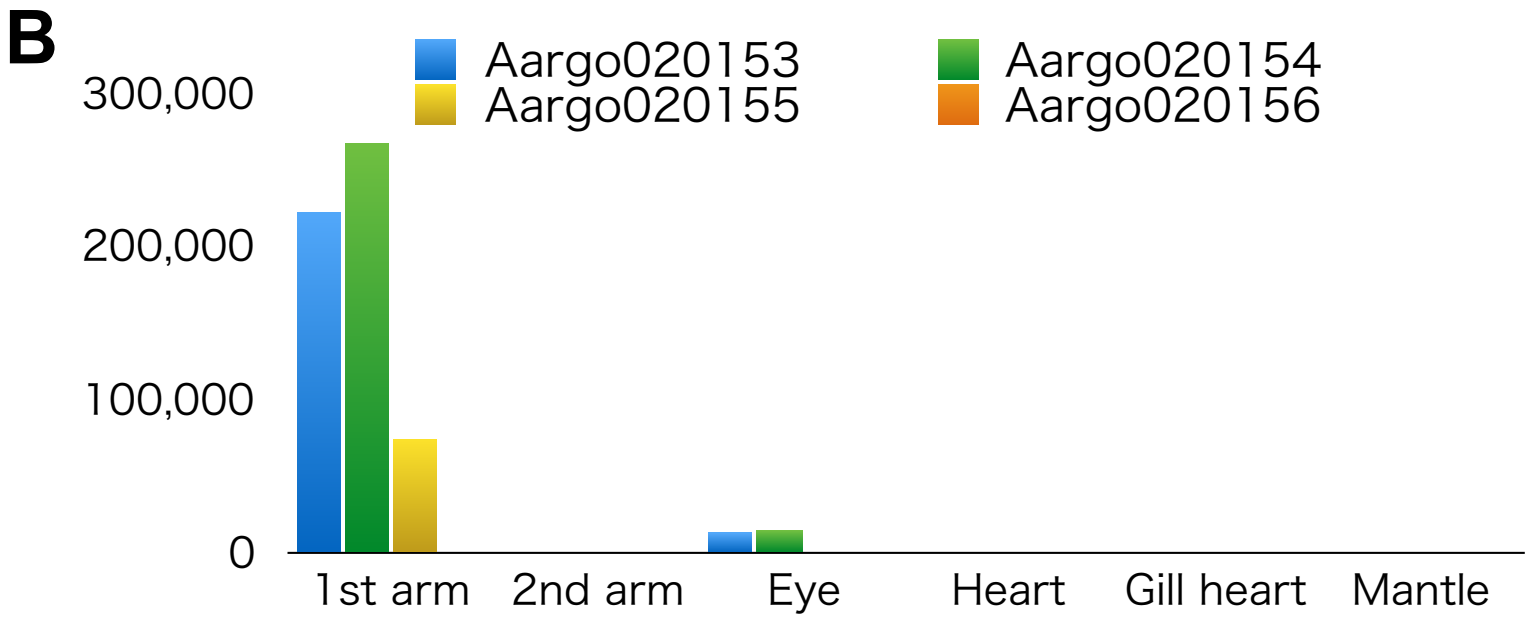

B
Aargo020153
Aargo020155
Aargo020154
Aargo020156
300,000
200,000
100,000
0
1st arm
2nd arm
Eye
Heart
Gill heart
Mantle

Fig.3

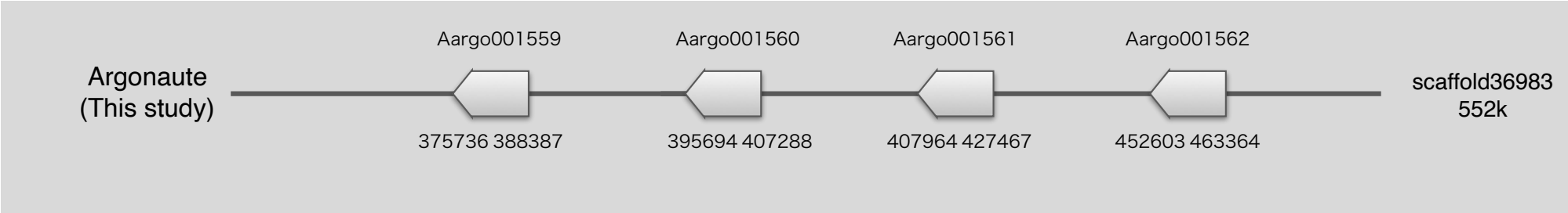

Argonaute
(This study)
Aargo001559
375736 388387
Aargo001560
395694 407288
Aargo001561
407964 427467
Aargo001562
452603 463364
scaffold36983
552k

Fig.4

# Fig. 5