Evolution of invertebrate *Cdk4/6-Cdkn2* system

Shiori Yuki[1], Shunsuke Sasaki[2], Yuta Yamamoto[2], Fumika Murakami[1], Kazumi Sakata[1,2] and Isato Araki[1,2]* (1: Graduate School of Arts and Sciences, 2: Faculty of Science and Engineering, Iwate University, Morioka, Japan)

* Corresponding author

Isato Araki

Faculty of Science & Engineering, Iwate University

3-18-33 Ueda, Morioka, Iwate 020-8550, Japan

phone/fax: +81-(0)19-621-6909

e-mail: iaraki@iwate-u.ac.jp

Short title: Evolution of invertebrate Cdk4/6-Cdkn2

Keywords: cell cycle regulation, comparative genomics, cyclin-dependent kinase, cyclin-dependent kinase inhibitor, evolution, invertebrates, synteny, tumor suppressor

Abstract

Cell cycle is driven by cyclin-dependent kinase (Cdk). The decision whether the cell cycle proceeds is made during G1 phase when Cdk4/6 functions. Cyclin-dependent kinase inhibitor 2 (Cdkn2) is the specific inhibitor for Cdk4/6, and the interaction depends on D84 in Cdkn2 and R24/31 in Cdk4/6. These knowledge have been obtained mainly from studies in mammalian cells. Here, we comprehensively analyzed *Cdk4/6* and *Cdkn2* in invertebrates, and found that while *Cdk4/6* exist in most of investigated phyla, there is presence bias of *Cdkn2*. Interestingly, the positive charge in Cdk4/6 R24/R31 was conserved in all analyzed species in phyla with Cdkn2, and the presence of *Cdkn2* and the conservation of the positive charge were statistically correlated. We also found that *Cdkn2* has been tightly linked with *Fas associated factor 1* (*Faf1*) during evolution. These genome-wide data suggest that 1) *Cdkn2* might have been lost multiple times during evolution, 2) the interaction between Cdkn2 D84 and Cdk4/6 R24/31 might impose constraint on the evolution of cell cycle regulation mechanisms, 3) the loss of *Cdkn2* might loosen the constraint, and 4) there might be potential interaction between *Cdkn2* and *Faf1*.

Introduction

Regulation of cell cycle is closely associated with ontogeny, tissue regeneration, cell senescence, pathogenesis, especially carcinogenesis and evolution. Cyclin-dependent kinases (Cdk's) and their regulatory subunits, cyclins, promote cell cycle while Cdk inhibitors (Cdkn's) inhibit the cycle by repressing the function of Cdk's (Reviewed in Weinberg, 2014). In mammalian cells Cdk4 or Cdk6, and cyclin D's drive cell cycle in G1 phase, when cells are sensitive to extracellular regulatory signals on the proliferation. Fish have Cdk21, the third member of Cdk4/6 subgroup of the Cdk family (Webster *et al*., 2018). Mammals have three types of structurally different Cdkn's, Cdkn1, Cdkn2 and Cdkn3. They possess three Cdkn1's, Cdkn1a, Cdkn1b and Cdkn1c (aka $p21^{Cip1/Waf1}$, $p27^{Kip1}$ and $p57^{Kip2}$, respectively; reviewed in Mühleder *et al*., 2021). Cdkn1's are preferential inhibitors for Cdk1 and Cdk2, but can also inhibit Cdk4/6. Cdkn1's are tumor suppressors or the candidate (El-Deiry *et al*., 1993; Chen *et al*., 1996; Zhang *et al*., 1997).

Cdk inhibitor 2's (Cdkn2's) are specific inhibitors for Cdk4/6. Mammals possess four Cdkn2's, Cdkn2a, Cdkn2b, Cdkn2c and Cdkn2d (aka $p16^{Ink4a}$, $p15^{Ink4b}$, $p18^{Ink4c}$ and $p18^{Ink4d}$, respectively). Mammalian Cdkn2a and Cdkn2b possess four ankyrin repeats while Cdkn2c and Cdkn2d do five (Baumgartner *et al*., 1998). Cdkn2's are known as tumor

suppressors (Kamb *et al*., 1994; Jen *et al*. 1994; Husemann *et al*., 1999; Gluick *et al*., 2013) as Cdkn1's. It has been also shown that Cdkn2's accumulate in mammalian cells during cell senescence (Serrano *et al*., 1997; Kita *et al*., 2022; Liu *et al*., 2023). It has been thought that electrostatic interaction between R24/R31 of Cdk4/6 and D84 of Cdkn2's plays a significant role in the formation of Cdk4/6-Cdkn2 complex, and the Cdkn2 binding to R24/R31 causes structural change preventing cyclin and ATP binding to Cdk4/6 and its following activation (Russo *et al*., 1998; Brotherton *et al*., 1998; reviewed in Baker & Reddy 2013) *. R24C and R24H mutations in CDK4 has been reported to cause familial melanoma, suggesting that strong positive charge at the residue is necessary for the complex formation (Wölfel *et al*., 1995; Zuo *et al*., 1996; Tsao *et al*., 1998; Soufir *et al*., 1998; Molven *et al*., 2005; Veinalde *et al*., 2013). This interaction mechanism is also applied to a cell immortalization protocol, K4DT method (Shiomi *et al*., 2011).

Cdkn3 (aka Cdk2-associated dual specificity phosphatase, Kap or Cip2) is a phosphatase and specifically inhibits Cdk2 (reviewed in Patterson *et al*., 2009). It has been suggested that Cdkn3 could act an oncogene, not a tumor suppressor gene, as opposed to Cdkn1's and Cdkn2's (Liu *et al*., 2019). No direct interaction between Cdk4/6 and Cdkn3 has been reported.

In contrast to mammals the molecular mechanisms of cell cycle have been poorly studied especially in invertebrates other than *Caenorhabditis elegans* and *Drosophila melanogaster* (reviewed in Kipreos & van den Heuvel, 2019). *C. elegans* and *D. melanogaster* have no Cdkn2 gene; instead, it is thought that the Cdkn1 subrogates the Cdk4/6 function in G1 phase. A few pioneering small-scale genome studies reported the existence of *Ckd4/6* genes in invertebrates Chordata (*Ciona intestinalis*, and *Branchiostoma floridae)*, Echinodermata (*Strongylocentrotus purpuratus*), Arthropoda (*Drosophila melanogaster*), Platyhelminthes (*Schmidtea mediterranea*), Cnidaria (*Nematostella vectensis* and *Hydra vulgaris*), and Placozoa (*Trichoplax adhaerens*) while no *Ckd4/6* gene in Porifera (*Amphimedon queenslandica*) nor *Cdkn2* gene in Platyhelminthes (*Schmidtea mediterranea*) and Cnidaria (*Hydra vulgaris*) (Cao *et al*., 2014; Espinal-Centeno *et al*., 2020). However, considering the diversity of invertebrates more systematic and comprehensive investigation on the regulatory mechanisms of the cell cycle has been needed. From this point of view, we performed a large-scale investigation on the existence of *Cdk4/6* and *Cdkn2* genes in invertebrates.

Footnote

The numbering of the amino acids is based on the human proteins. D84 in human CDKN2A correspond to D86 in CDKN2B, D76 in CDKN2C and D80 in CDKN2D,

respectively. To avoid complexities, we use D84 for the D84/D86/D76/D80 in this paper.

Results

The presence of Cdkn2 in invertebrate was biased

First, we systematically searched *Cdkn2* in the genomes of all invertebrate phyla. Even in the initial keyword and blastp searches followed by more thorough tblastn searches on whole-genome shotgun contigs databases, we noticed that the presence bias of *Cdkn2* among invertebrates. We concluded that *Cdkn2* was detected only in seven phyla (Figures 1, 2, S1, Tables S1, S2). The species in the most of phyla in invertebrate Deuterostomia, except in Urochordata and Xenacoelomorpha, possess *Cdkn2*. Some phyla in Lophotrochozoa also possess *Cdkn2*; Mollusca, Annelida, and Brachiopoda/Phoronida. In contrast we could not find *Cdkn2* in Ecdysozoa except in Onychophora. We could not find the *Cdkn2* in the basal invertebrates, neither. We also found that the *Cdkn2* presence in Mollusca is biased. For example, we could not detect *Cdkn2* in Bivalvia, Heterobranchia, and Coleoidea (Table 1).

During in the tblastn search we noticed that there were sporadic "weakly positive" sequences in phyla where majority of the search result suggest that the phylum may lack *Cdkn2* (shown in red in Table S2b). Some of them apparently originated from contaminated DNA of other species. To obtain more standards other than blast search to distinguish whether the "weakly positive" sequences, excluding evident contaminated

DNA, are *bona fide* Cdkn2 or genes for other ankyrin repeat containing-proteins, we first analyzed the gene structure, and found that the exon junction between Q50 and V51 is conserved in all species analyzed in this study. We also extracted conserved amino acid residues among invertebrate Cdkn2's (Figure S1c). Based on the conserved exon junction, presence of the conserved residue, and linkage to *Faf1* (see below) as well as blast search results we concluded that the "weakly positive" sequences are not *bona fide* Cdkn2 (for details, see the notes in Table S2b).

We also noticed that there are invertebrate sequences potentially misannotated as Cdkn2 in the NCBI Protein database (Table S4). Some of them were apparently Cdkn1. We judged that the sequences other than Cdkn1 homologs are not *bona fide* Cdkn2 but other ankyrin repeat containing proteins based on the criteria mentioned above (for details, see the notes in Table S4).

The corresponding residue to the D84 in Cdkn2 was conserved in most of invertebrates, but not in Crinoid

Next, we investigated the evolutionary conservation of the corresponding residue to Cdkn2 D84 in invertebrates. We found that it was conserved in most of (95.1%, 58/61) the detected invertebrate *Cdkn2*; however, Cdkn2 sequences from two Crinoid species

(D84Q and D84N) as well as one of the two Cdkn2 in *Phoronis ovalis* (D84N) lacked the negative charge (Figures 2, S1b), although their Cdk4/6's retained R24/R31 (see below). In vertebrates as a control, the negative charge was conserved in almost all Cdkn2 sequences (99.9% 1849/1850; Figure S3).

Cdkn2 gene has been strongly linked to Faf1 gene during evolution

We also analyzed the microsynteny around the *Cdkn2* locus in invertebrates. The results showed that all analyzed invertebrate *Cdkn2* genes are linked with *Faf1* (*Fas associated factor 1*) as same as vertebrate *Cdkn2c* (Figure 3, Tables S1, S5). In addition, cephalochordate *Cdkn2* genes are linked with *Mtap* (*methylthioadenosine phosphorylase*) as same as vertebrate *Cdkn2a/b* (Figure 3).

Microsynteny analysis around Faf1 locus in the species lacking Cdkn2

To obtain clues on the origin and the loss of *Cdkn2* we analyzed microsynteny around *Faf1* in the species without *Cdkn2*. While we could not detect any evolutionary conservation of linkages comparable to the one between *Cdkn2* and *Faf1*, potentially interesting is that the linkage between *Faf1* and *Kank1* (*KN Motif and Ankyrin Repeat Domains 1*) is conserved among some species in Anthozoa of Cnidaria (Figure 4, Table

S6). Considering the evolutionary position of Cnidaria, *Cdkn2* in Bilateria and *Kank1* in Anthozoa might share the origin in ancient metazoans.

Analysis of the presence of Cdkn1 in invertebrate phyla

It is thought that the Cdkn1 subrogates the Cdk4/6 function in G1 phase in *C. elegans* and *D. melanogaster* (reviewed in Kipreos & van den Heuvel, 2019). A recent study suggested that *Cdkn1* compensates the lack of *Cdkn2* also in an Urochordata *Ciona* embryos (Kobayashi *et al.*, 2022), while another study reported that planarians lack Cdkn1's and Cdh1 is important for exiting cell cycle (Sato *et al.*, 2022). To more precise distribution of Cdkn1 among invertebrates we performed comprehensive search on invertebrate genomes. We could find at least one *Cdkn1* gene in all invertebrate phyla except Dicyemida and Orthonectida (Table 1; Table S7). In Platyhelminthes we could find Cdkn1 in Cestoda, but not in Rhabditophora nor Trematoda.

All investigated invertebrates except Porifera, Dicyemida and Orthonectida *possessed Cdk4/6 gene*

Then, we systematically searched *Cdk4/6* in invertebrate genomes. We could find *Cdk4/6* in all invertebrate phyla except in Porifera, as previously reported (Cao *et al.*,

2014), Dicyemida and Orthonectida (Table 1, Table S8). The lack of *Cdk4/6*, as well as *Cdkn1* mentioned above, in Dicyemida and Orthonectida may be due to the shrunk genome size (Lu *et al*., 2019; Slyusarev *et al*., 2020). Thirty species including the ones in Arachnida whose genome analyses suggested an ancient whole-genome duplication event possess more than one *Cdk4/6* (Figure S3e, Table S8; Schwager *et al*., 2017).

The extent of the evolutionary conservation of the corresponding residue to the R24/R31 in Cdk4/6 varied among phyla

Next, we investigated the evolutionary conservation of the corresponding residue to the R24/R31 in invertebrates. Among 1296 investigated sequences of Cdk4/6 from 1262 invertebrate species in 20 phyla the positive charge in R24/R31 was conserved in 1148 sequences (88.6%) (Tables 2, S8). In 215 sequences (16.6%) the arginine was substituted with lysine while the strong positive charge is lost in other 148 sequences (11.4%). In invertebrate Deuterostomia two *Ciona* species in Urochordata and one species in Xenacoelomorpha lacked the strong positive charge at the residue corresponding to the Cdk4/6 R24/R31 (3/43, 7.0%) (Figures 5, S3a). This was especially in contrast to *Corella inflata* and *Ascidia aspersa*, species belonging to the same order Phlebobranchia as *Ciona*, in which the charge was conserved, although it has been

substituted with lysine (Shito & Hotta, personal communication). In Lophotrochozoa 39.1% (52/133) of the sequences investigated the strong positive charge was not conserved (Table 2, Figure S3b, c). Strikingly, 97.9% (46/47) of the species in Platyhelminthes lacked the charge and the corresponding residues are mostly threonine's. In Ecdysozoa 8.0% (85/1058) of the sequences investigated the charge was not conserved (Table 2, Figure S3d, e). A significant ratio of species (63.9%, 46/72) in Nematoda also lacked the charge and the corresponding residues are mostly, again, threonine's. In basal invertebrates 9.7% (6/62) of the sequences investigated the strong charge was not conserved (Table 2, Figure S3f). These data in invertebrate Cdk4/6 are contrast to the one in vertebrate Cdk4/6/21, in which the charge is conserved in 1249 out of 1252 investigated sequences (99.8%) from 572 species (Figure S4, Table S9).

Comparison between the presence bias of Cdkn2 and the differential conservation of the corresponding residue to the R24/R31 in Cdk4/6 among phyla

While species in most of animal phyla possess *Cdk4/6*, the extent of the evolutionary conservation of the corresponding residue to the R24/R31 greatly differed among phyla. Since we speculated that the presence of *Cdkn2* might affect the extent, we compared them in each phylum, and summarized the result in Figure 1. In all analyzed species in

phyla with *Cdkn2* gene(s) the strong positive charge in the corresponding residue to the R24/R31 of Cdk4/6 homologs was conserved, even in the two species in Crinoid where the negative charge in the corresponding residue to Cdkn2 D84 has been lost. On the other hand, the evolutionary conservation of the strong positive charge in the Cdk4/6 homologs in phyla where we could detect no *Cdkn2* gene showed extreme difference. For example, while only small number of species in Cnidaria, Arthropoda, and Urochordata lacked the strong positive charge (10.0%, 3.9%, 14.3%, respectively), many or almost all species in Nematoda, Rotifera and Platyhelminthes seemed to lack the strong positive charge (63.9%, 44.4%, 97.9%, respectively). It is interesting that many species in the evolutionary distant two phyla, Platyhelminthes and Nematoda, both of which contains large numbers of parasites, share a threonine instead of R24/R31, although the significance is unknown.

Finally, we analyzed the statistical correlation between the presence of *Cdkn2* and the conservation of the strong positive charge in the corresponding residue at the corresponding residue to the R24/R31 in Cdk4/6. We performed the analysis on the nine phyla in which more than five species were included in this study (highlighted yellow in the Figure 1). We used the percentage of the number of species in which both Cdkn2 and Cdk4/6 to the number of species in which Cdk4/6 was detected ("**[2]/[1]** in %" in Figure

S5) , and the percentage of number of species in which the strong positive charge was conserved in each of the phyla ("% of -R, K at Cdk4/6 R24/31" in Figure S5) to determine the Spearman's rank correlation coefficient. The coefficient was 0.816 (P value = 0.00735), suggesting that the presence of *Cdkn2* and the conservation of the positive charge were strongly correlated with statistical significance.

Discussion

Evolution of Cdkn2 gene

Based on the data on the presence bias of *Cdkn2* in invertebrate genomes, we postulate a scenario of the evolution of *Cdkn2* (Figure 6a). In this scenario the ancestral *Cdkn2* gene first appears from one of the ankyrin repeat-containing genes during the early evolution of Bilateria. During the diversification of Bilateria *Cdkn2* gene has been lost in some phyla, although the extent differs among Ecdysozoa, Lophotrochozoa, and Deuterostomia.

The observation that some invertebrates possess *Cdkn2* while the other not might be obscurely explained that each of the strategies are beneficial to survive or reproduce more efficiently in each group. Since Cdkn2's regulate the cell cycle in mammals, the cause of the presence bias of *Cdkn2* in invertebrates might be related to cell proliferation, cell differentiation or cell aging. However, it should be noted that even in deuterostomal invertebrates the mechanisms of cell aging seem to be different from the ones in mammals (reviewed in Amir *et al*., 2020). Species in phyla without *Cdkn2*, like Platyhelminthes, Nemertea, Bryzoa and Urochordata as well as non-Bilateria, are known to their extreme regenerative ability; however, species in phyla with *Cdkn2*, like Phoronida, Annelida and Echinodermata, are also the same (reviewed in Bely & Nyberg, 2010). Thus, *Cdkn2* is not another example of genes like JmjC domain encoding gene which has been shown to

be highly linked to extreme regeneration ability (Cao *et al*., 2019). It seems that longevity may not the cause of the presence bias of *Cdkn2* in invertebrates because even if you exclude non-Bilateria which has species with ultra-extreme longevity or theoretical immortality (Jochum *et al*., 2012; Piraino *et al*., 1996; Brien, 1953), both a sea urchin, *Mesocentrotus franciscanus* (aka *Strongylocentrotus franciscanus*), in which we could detect a *Cdkn2* gene (Figure S1), and a clam, *Arctica islandica*, which belong to Bivalvia where we could not detect *Cdkn2* (Table 1), are known to their longevity, although the latter nuclear genome has not yet been published (Ebert & Southon, 2003; Schöne *et al*., 2005). Thus, we have currently no compelling clear-cut hypothesis to explain this issue.

Strong evolutionary conservation of the microsynteny in Cdkn2-Faf1 locus

Our synteny analysis revealed that *Cdkn2* has been strongly linked with *Faf1*. Based on the data we postulate a scenario of the evolution of the linkage around Cdkn2 locus in invertebrate (Figure 6b). In this scenario the ancestral *Cdkn2* gene first appeared linking with *Faf1*. In the ancestral chordate the *Cdkn2* gene has begun to link also with *Mtap*. During the twice duplication of the genome in the ancestral vertebrate one of the *Cdkn2*, i.e., *Cdkn2c*, kept linked with *Faf1* while another one, i.e., *Cdkn2b*, kept linked with *Mtap*. One of other *Cdkn2*'s became *Cdkn2d*, and the other one "*Cdkn2e*" was lost. Later, the

duplication of *Cdkn2b* created *Cdkn2a*. It is interesting that in human *Faf1*, *Mtap* and *Cdkn2* are all known as tumor suppressors, and the genetic linkage between *Cdkn2* and *Faf1* among them can be traced back into the appearance of *Cdkn2* (Christopher *et al*., 2002; Menges *et al*., 2009; Kamb *et al*., 1994; Jen *et al*., 1994; Husemann *et al*., 1999; Gluick *et al*., 2013).

Why has the microsynteny in *Cdkn2-Faf1* locus been conserved during bilaterian evolution? It has been suggested that the microsynteny in a locus with multiple genes may be conserved by *cis*-regulatory constraints between the genes and/or the functional interdependency of the gene products (Imai *et al*., 2011; Irimia *et al*., 2012; Simakov & Kawashima, 2017; Cai *et al*., 2010). Although both *Cdkn2*'s and *Faf1* are categorized to tumor suppressor genes in mammals as mentioned above, they function in different cellular systems, i.e., Cdkn2's regulate cell cycle while Faf1 does apoptosis, and their physical interaction has not yet been reported. Thus, *cis*-regulatory constraints between them might keep the microsynteny. In that context it is interesting that the expression of a lncRNA gene, which is on the immediate opposite side of *C1orf185* gene from *Cdkn2c*, can affect *Faf1* expression in *Homo sapiens*, although the linkage between *C1orf185* and *Cdkn2c-Faf1* locus is not evolutionary conserved even in vertebrates (Wan *et al*., 2014; Araki, unpublished data). However, it is also noteworthy that a conservation of

microsynteny might reflect a covert physical interaction (Moretti *et al*., 2017).

Potential interactions between Cdkn2 and Cdk4/6 evolutions

The results of the comparative analysis between the presence bias of *Cdkn2* and the differential conservation of the strong positive charge at the corresponding residue to the R24/R31 in Cdk4/6 among phyla suggest potential interactions between *Cdkn2* and *Cdk4/6* evolutions. The fact all investigated species in Annelida, Mollusca and Echinodermata keep the charge and possess *Cdkn2* while all investigated species in Platyhelminthes, Rotifera and Nematoda lack *Cdkn2* and high percentage of the species also lack the charge imply that the existence of *Cdkn2* might constrain the corresponding residue to the R24/R31 from amino acid substitution or loss.

On the other hand, most of the species in Cnidaria, Arthropoda and Urochordata keep the charge while they lack *Cdkn2*. Furthermore, Cdkn2 sequences from two Crinoid species lack the negative charge in the corresponding residue to D84, although their Cdk4/6's retain R24/R31. In these species the Cdk4/6 sequences might be still on the way to amino acid substitution. Alternatively, the strong positive charge in the Cdk4/6 homologs may be conserved for the maintenance of the three-dimensional structure, and/or protein interaction with other than Cdkn2.

Cell cycle regulation without Cdkn2

What substitutes for Cdkn2 in the species without *Cdkn2*? It has been shown that Cdkn1 is a preferential inhibitor for Cdk1/2, but can also inhibit Cdk4/6 in mammalian cells, although the interaction between Cdk4/6 and Cdkn1 has not yet been solved by X-ray crystallography (reviewed in Wood & Endicott, 2018, and Mühleder *et al*., 2021). Thus, it is easy to speculate that Cdkn1 substitutes for Cdkn2 in the species without *Cdkn2*, and, in fact, it has been reported that Cdkn1 can substitute for Cdkn2 in some invertebrates including *Ciona* as mentioned above (Kobayashi *et al*., 2022; reviewed in Kipreos & van den Heuvel, 2019). However, it might be necessary to consider cautious to extrapolate this mechanism to other urochordates, or, species lacking *Cdkn2* in other phyla because *Ciona* is the only urochordates investigated lacking the strong positive charge at the residue corresponding to Cdk4/6 R24/R31. Since the extent of the evolutionary conservation of the positive charge greatly differs among phyla lacking *Cdkn2*, the conserved strong positive charge in Cnidaria, Arthropoda and Urochordata might not be involved in the potential interaction with Cdkn1, and the positive charge might be necessary for interaction with other than Cdkn2 as mentioned above.

1 *Conclusion*

2 In this research we performed the first comprehensive investigations on the distribution
3 of *Cdk4/6* and *Cdkn2* in invertebrate genomes and the evolutionary conservation of the
4 key charged residue as well as the synteny analysis on *Cdkn2*. We believe that the result
5 should widen our knowledge on the mechanisms of the cell cycle regulation and their
6 evolution in Metazoan.

7

Experimental procedures

Taxonomy

We used the taxonomy of invertebrates by Telford *et al*. (2015) for phyla, and the one in the NCBI Taxonomy database for subclassification in this study. There has been controversy on the evolutionary position on Xenacoelomorpha (Cannon *et al*., 2016; Philippe *et al*., 2019). In the paper we tentatively categorize it as a diverged deuterostome.

Database search

To analyze the presence of *Cdkn2* and *Cdk4/6* genes, and to obtain their amino acid sequences, we performed keyword and blastp search in the NCBI Protein database, Ensembl Metazoa, Ensembl Rapid Release, Ghost Database: *Ciona intestinalis* genomic and cDNA resources, Aniseed - Ascidian Network for In Situ Expression and Embryological Data, PlanMine and OIST Marine Genomics Unit Genome Browser (Sayers *et al*., 2024; Yates *et al*., 2022; Satou *et al*., 2005; Dardaillon *et al*., 2020; Rozanski *et al*., 2019; Shinzato *et al*., 2020, 2021; Yoshioka *et al*., 2022; Khalturin *et al*., 2019; Luo *et al*., 2018; Takeuchi *et al*., 2022; Arimoto *et al*., 2019; Simakov *et al*., 2015). We also performed tblastn search in the NCBI Whole-genome shotgun contigs Database. To confirm the identity of the sequences obtained as results, we performed

blastp search against vertebrate sequences. We also collected vertebrate *Cdkn2* and *Cdk4/6/21* sequences as controls. During the process we noticed that there are misannotations on the vertebrate genes in the NCBI databases. For example, a platypus Cdkn2 (XP_028910528.1) is annotated as "cyclin-dependent kinase 4 inhibitor B-like", but the microsynteny suggests that it is Cdkn2a; thus, we treated it as Cdkn2a in this paper. Since 1) the paralogs are specific to vertebrates, thus, the misannotation does not affect the results in this study, and 2) it is out of scope in this paper, we will publish the details of the analysis elsewhere. The database ID of the analyzed sequences were tabulated in the Tables S1, S2, S8, S9).

As the result of the search 61 Cdkn2 sequences from 61 invertebrate species, and 1330 Cdk4/6 sequences from 1264 invertebrate species were included in this analysis. In addition, 1850 Cdkn2 sequences and 1252 Cdk4/6/21 sequences from 572 vertebrate species were included as controls in the analysis.

Protein sequence multiple alignment

We used NCBI COBALT (Constraint-based Multiple Alignment Tool) for protein sequence multiple alignment (Papadopoulos & Agarwala, 2007). When the alignment was questionable, we corrected it manually. We used WebLogo

(https://weblogo.berkeley.edu/logo.cgi) to create a consensus sequence logo, and MOTIF Search (https://www.genome.jp/tools/motif/) to identify protein motives.

Synteny analysis

We used the NCBI gene database for synteny analysis if the sequence had a NCBI gene ID (Brown *et al*., 2015). If the sequence had no gene ID, and the long enough genomic sequence was available, we analyzed the annotation, if available, or performed blastp search against the surrounding sequence. If the surrounding genes have been annotated, we performed blastp against vertebrates to confirm the identities. We looked up GeneCards (https://www.genecards.org) for abbreviation of gene name.

Statistical analysis

We used a statistical software EZR to determine Spearman's rank correlation coefficient and create the scatter plot (Kanda, 2013).

Acknowledgement

We thank Drs. Yutaka Satou (Kyoto Univ) and Takeshi Kawashima (Inuyama Bio) for critically reading the manuscript, and Takuya Minokawa (Tohoku Univ), Kazuya Kobayashi (Hirosaki Univ), Hiroshi Wada (Tsukuba Univ), and Koji Tamura (Tohoku Univ) for helpful discussion. We also thank Drs. Takumi Shito and Kohji Hotta (Keio Univ) for sharing their unpublished data. This research was financially supported by Iwate University.

References

Amir, Y., Insler, M., Giller, A., Gutman, D. & Atzmon, G. (2020). Senescence and Longevity of Sea Urchins. *Genes* (Basel), *11*, 573. https://doi.org/10.3390/genes11050573.

Arimoto, A., Hikosaka-Katayama, T., Hikosaka, A., Tagawa, K., Inoue, T., Ueki, T., ... Satoh, N. (2019). A draft nuclear-genome assembly of the acoel flatworm *Praesagittifera naikaiensis*. *Gigascience*, 8, 1-8. https://doi.org/10.1093/gigascience/giz023.

Baker, S. J. & Reddy, E. P. (2013). CDK4: A Key Player in the Cell Cycle, Development, and Cancer. *Genes and cancer*, *3,* 658-669. https://doi.org/10.1177/1947601913478972.

Baumgartner, R., Fernandez-Catalan, C., Winoto, A., Huber, R., Engh, R. A. & Holak, T. A. (1998). Structure of human cyclin-dependent kinase inhibitor $p19^{INK4d}$: comparison to known ankyrin-repeat-containing structures and implications for the dysfunction of tumor suppressor $p16^{INK4a}$. *Structure*, *6*, 1279-1290. https://doi.org/10.1016/s0969-2126(98)00128-2.

Bely, A. E. & Nyberg, K. G. (2010). Evolution of animal regeneration: re-emergence of a field. *Trends in ecology and evolution*, *25*, 161-170.

https://doi.org/10.1016/j.tree.2009.08.005.

Brien, P. (1953). La pérennité somatique. *Biological Reviews*, *28*, 308-349. https://doi.org/10.1111/j.1469-185x.1953.tb01381.x.

Brotherton, D. H., Dhanaraj, V., Wick, S., Brizuela, L., Domaille, P. J., Volyanik, E., ... Laue, E. D. (1998). Crystal structure of the complex of the cyclin D-dependent kinase Cdk6 bound to the cell-cycle inhibitor $p19^{INK4d}$. *Nature*, *395*, 244-250. https://doi.org/10.1038/26164.

Brown, G. R., Hem, V., Katz, K. S., Ovetsky, M., Wallin, C., Ermolaeva, O., ... Murphy, T. D. (2015). Gene: a gene-centered information resource at NCBI. *Nucleic Acids Res*, 43, D36-42. https://doi.org/10.1093/nar/gku1055.

Cai, W. W., Wang, L. & Chen, Y. (2010). Aspartyl aminopeptidase, encoded by an evolutionarily conserved syntenic gene, is colocalized with its cluster in secretory granules of pancreatic islet cells. *Bioscience, Biotechnology, and Biochemistry*, *74*, 2050-2055. https://doi.org/10.1271/bbb.100349.

Cannon, J. T., Vellutini, B. C., Smith, 3rd, J., Ronquist, F., Jondelius, U. & Hejnol, A. (2016). Xenacoelomorpha is the sister group to Nephrozoa. *Nature*, *530*, 89-93. https://doi.org/10.1038/nature16520.

Cao, L., Chen, F., Yang, X., Xu, W., Xie, J. & Yu, L. (2014). Phylogenetic analysis of

CDK and cyclin proteins in premetazoan lineages. *BMC Ecology and Evolution*, *14*, 10. https://doi.org/10.1186/1471-2148-14-10.

Cao, P.-L., Kumagai, N., Inoue, T., Agata, K. & Makino, T. (2019). JmjC Domain-Encoding Genes Are Conserved in Highly Regenerative Metazoans and Are Associated with Planarian Whole-Body Regeneration. Genome Biology and Evolution, 11, 552-564. https://doi.org/10.1093/gbe/evz021.

Chen, J., Willingham, T., Shuford, M. & Nisen, P. D. (1996). Tumor suppression and inhibition of aneuploid cell accumulation in human brain tumor cells by ectopic overexpression of the cyclin-dependent kinase inhibitor $p27^{KIP1}$. *Journal of Clinical Investigation*, *97*, 1983-1988. https://doi.org/10.1172/JCI118631.

Christopher, S. A., Diegelman, P., Porter, C. W. & Kruger, W. D. (2002). Methylthioadenosine phosphorylase, a gene frequently codeleted with $p16^{cdkN2a/ARF}$, acts as a tumor suppressor in a breast cancer cell line. Cancer Research, 62, 6639-6644.

Dardaillon, J., Dauga, D., Simion, P., Faure, E., Onuma, T. A., DeBiasse, M. B., ... Lemaire, P. (2020). ANISEED 2019: 4D exploration of genetic data for an extended range of tunicates. *Nucleic Acids Res*, 48, D668-D675. https://doi.org/10.1093/nar/gkz955.

Ebert, T. A. & Southon, J. R. (2003). Red sea urchins (*Strongylocentrotus franciscanus*)

can live over 100 years: Confirmation with A-bomb ^{14}carbon. *Fishery Bulletin*, *101*, 915-922.

El-Deiry, W. S., Tokino, T., Velculescu, V. E., Levy, D. B., Parsons, R., Trent, J. M., ... Vogelstein, B. (1993). *WAF1*, a potential mediator of p53 tumor suppression. *Cell*, *75*, 817-825. https://doi.org/10.1016/0092-8674(93)90500-p.

Espinal-Centeno, A., Dipp-Álvarez, M., Saldaña, C., Bako, L. & Cruz-Ramírez, A. (2020). Conservation analysis of core cell cycle regulators and their transcriptional behavior during limb regeneration in *Ambystoma mexicanum*. *Mechanisms of Development*, *164*, 103651. https://doi.org/10.1016/j.mod.2020.103651.

Gluick, T., Yuan, Z., Libutti, S. K. & Marx, S. J. (2013). Mutations in *CDKN2C* (*p18*) and *CDKN2D* (*p19*) may cause sporadic parathyroid adenoma. *Endocrine-Related Cancer*, *20*, L27-29. https://doi.org/10.1530/ERC-13-0445.

Husemann, K., Wolter, M., Büschges, R., Boström, J., Sabel, M. & Reifenberger, G. (1999). Identification of two distinct deleted regions on the short arm of chromosome 1 and rare mutation of the *CDKN2C* gene from 1p32 in oligodendroglial tumors. *Journal of Neuropathology & Experimental Neurology*, *58*, 1041-1050.

Imai, K. S., Daido, Y., Kusakabe, T. G. & Satou Y. (2012) Cis-acting transcriptional repression establishes a sharp boundary in chordate embryos. *Science*, 337, 964-967.

https://doi.org/10.1126/science.1222488.

Irimia, M., Tena, J. J., Alexis, M. S., Fernandez-Miñan, A., Maeso, I., Bogdanovic, O., ... Fraser, H. B. (2012) Extensive conservation of ancient microsynteny across metazoans due to cis-regulatory constraints. *Genome Research*, *22*, 2356-2367. https://doi.org/10.1101/gr.139725.112.

Jen, J., Harper, J. W., Bigner, S. H., Bigner, D. D., Papadopoulos, N., Markowitz, S., ... Vogelstein, B. (1994). Deletion of *p16* and *p15* genes in brain tumors. *Cancer Research*, *54*, 6353-6358.

Jochum, K. P., Wang, X., Vennemann, T. W., Sinha, B. & Müller, W. E. G. (2012) Siliceous deep-sea sponge *Monorhaphis chuni*: A potential paleoclimate archive in ancient animals. *Chemical Geology*, *300-301*, 143–151. https://doi.org/10.1016/j.chemgeo.2012.01.009.

Khalturin, K., Shinzato, C., Khalturina, M., Hamada, M., Fujie, M. Koyanagi, R., ... Satoh, N. (2019). Medusozoan genomes inform the evolution of the jellyfish body plan. *Nature Ecology & Evolution*, 3, 811-822. https://doi.org/10.1038/s41559-019-0853-y.

Kamb, A., Gruis, N. A., Weaver-Feldhaus, J., Liu, Q., Harshman, K., Tavtigian, S. V., ... Skolnick, M. H. (1994). A cell cycle regulator potentially involved in genesis of many tumor types. *Science*, *264*, 436-440. https://doi.org/10.1126/science.8153634.

1 Kanda, Y. (2013). Investigation of the freely available easy-to-use software 'EZR' for
2 medical statistics. *Bone Marrow Transplantation*, 48, 452–458.
3 https://doi.org/10.1038/bmt.2012.244.
4 Kipreos, E. T. & van den Heuvel, S. (2019). Developmental Control of the Cell Cycle:
5 Insights from *Caenorhabditis elegans*. *Genetics*, *211*, 797-829.
6 https://doi.org/10.1534/genetics.118.301643.
7 Kita, A., Saito, Y., Miura, N., Miyajima, M., Yamamoto, S., Sato, T., ... Chikenji, T. S.
8 (2022). Altered regulation of mesenchymal cell senescence in adipose tissue promotes
9 pathological changes associated with diabetic wound healing. *Communications*
10 *Biology*, *5*, 310. https://doi.org/10.1038/s42003-022-03266-3.
11 Kobayashi, K., Tokuoka, M., Sato, H., Ariyoshi, M., Kawahara, S., Fujiwara, S., ... Satou,
12 Y. (2022). Regulators specifying cell fate activate cell cycle regulator genes to
13 determine cell numbers in ascidian larval tissues. *Development*, *149*, dev201218.
14 https://doi.org/10.1242/dev.201218.
15 Liu, J., Min, L., Zhu, S., Guo, Q., Li, H., Zhang, Z., ... Zhang, S. (2019). Cyclin-
16 Dependent Kinase Inhibitor 3 Promoted Cell Proliferation by Driving Cell Cycle from
17 G1 to S Phase in Esophageal Squamous Cell Carcinoma. *Journal of Cancer*, *10*, 1915-
18 1922. https://doi.org/10.7150/jca.27053.

Liu, J., Lin, X., McDavid, A., Yang, Y., Zhang, H., Boyce, B. F. & Xing, L. (2023). Molecular signatures distinguish senescent cells from inflammatory cells in aged mouse callus stromal cells. *Frontiers in Endocrinology* (Lausanne), *14*, 1090049. https://doi.org/10.3389/fendo.2023.1090049.

Lu, T.-M., Kanda, M., Furuya, H. & Satoh, N. (2019). Dicyemid Mesozoans: A Unique Parasitic Lifestyle and a Reduced Genome. *Genome Biology and Evolution*, *11*, 2232-2243. https://doi.org/10.1093/gbe/evz157.

Luo, Y.-J., Kanda, M., Koyanagi, R., Hisata, K., Akiyama, T., Sakamoto, H., ... Satoh, N. (2018). Nemertean and phoronid genomes reveal lophotrochozoan evolution and the origin of bilaterian heads. *Nature Ecology & Evolution*, 2, 141-151. https://doi.org/10.1038/s41559-017-0389-y.

Menges, C. W., Altomare, D. A. & Testa, J.R. (2009). FAS-associated factor 1 (FAF1): diverse functions and implications for oncogenesis. *Cell Cycle*, *8*, 2528-2534. https://doi.org/10.4161/cc.8.16.9280.

Molven, A., Grimstvedt, M. B., Steine, S. J., Harland, M., Avril, M.-F., Hayward, N. K. & Akslen, L. A. (2005). A large Norwegian family with inherited malignant melanoma, multiple atypical nevi, and *CDK4* mutation. *Genes, Chromosomes & Cancer*, *44*, 10-18. https://doi.org/10.1002/gcc.20202.

1 Moretti, A. I. S., Pavanelli, J.C., Nolasco, P., Leisegang, M. S., Tanaka, L. Y., Fernandes,
2 C.G., ... Laurindo, F. R. M. (2017). Conserved Gene Microsynteny Unveils Functional
3 Interaction Between Protein Disulfide Isomerase and Rho Guanine-Dissociation
4 Inhibitor Families. *Scientific Reports*, *7*, 17262. https://doi.org/10.1038/s41598-017-
5 16947-5.
6 Mühleder, S., Fernández-Chacón, M., Garcia-Gonzalez, I. & Benedito, R. (2021).
7 Endothelial sprouting, proliferation, or senescence: tipping the balance from
8 physiology to pathology. *Cellular and Molecular Life Sciences*, *78*, 1329-1354.
9 https://doi.org/10.1007/s00018-020-03664-y.
10 Papadopoulos, J. S, & Agarwala, R. (2007). COBALT: constraint-based alignment tool
11 for multiple protein sequences. *Bioinformatics*, 23, 1073-1079.
12 Patterson, K. I., Brummer, T., O'Brien, P. M. & Daly, R. J. (2009). Dual-specificity
13 phosphatases: critical regulators with diverse cellular targets. *Biochemical Journal*,
14 *418*, 475-489. https://doi.org/10.1042/bj20082234.
15 Philippe, H., Poustka, A. J., Chiodin, M., Hoff, K. J., Dessimoz, C., Tomiczek, B., ...
16 Telford, M. J. (2019). Mitigating Anticipated Effects of Systematic Errors Supports
17 Sister-Group Relationship between Xenacoelomorpha and Ambulacraria. *Current*
18 *Biology*, *29*, 1818-1826.e6. https://doi.org/10.1016/j.cub.2019.04.009.

Piraino, S., Boero, F., Aeschbach, B. & Schmid, V. (1996). Reversing the Life Cycle: Medusae Transforming into Polyps and Cell Transdifferentiation in *Turritopsis nutricula* (Cnidaria, Hydrozoa). *Biology Bulletin*, *190*, 302-312. https://doi.org/10.2307/1543022.

Rozanski, A., Moon, H., Brandl, H., Martín-Durán, J. M., Grohme, M. A., Hüttner, K., ... Rink, J. C. (2019). PlanMine 3.0-improvements to a mineable resource of flatworm biology and biodiversity. *Nucleic Acids Res*, 47, D812-D820. https://doi.org/10.1093/nar/gky1070.

Russo, A. A., Tong, L., Lee, J. O., Jeffrey, P. D. & Pavletich, N. P. (1998). Structural basis for inhibition of the cyclin-dependent kinase Cdk6 by the tumour suppressor $p16^{INK4a}$. *Nature*, *395*, 237-243. https://doi.org/10.1038/26155.

Sato, Y., Umesono, Y., Kuroki, Y., Agata, K. & Hashimoto, C. (2022). Proliferation maintains the undifferentiated status of stem cells: The role of the planarian cell cycle regulator Cdh1. Developmental Biology, 482, 55–66. https://doi.org/10.1016/j.ydbio.2021.12.006

Satou, Y., Kawashima, T., Shoguchi, E., Nakayama, A. & Satoh, N. (2005). An Integrated Database of the Ascidian, *Ciona intestinalis*: Towards Functional Genomics. *Zool Sci*, 22, 837-843. https://doi.org/10.2108/zsj.22.837.

Sayers, E. W., Beck, J., Bolton, E. E., Brister, J. R., Chan, J., Comeau, D. C., ... Sherry, S. T. (2024). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*, 52, D33-D43. https://doi.org/10.1093/nar/gkad1044.

Schöne, B. R., Fiebig, J., Pfeiffer, M., Gless, R., Hickson, J., Johnson, A. L. A., ... Oschmann, W. (2005). Climate records from a bivalved Methuselah (*Arctica islandica*, mollusca; Iceland). *Palaeogeography, Palaeoclimatology, Palaeoecology*, *228*, 130-148. https://doi.org/10.1016/j.palaeo.2005.03.049.

Schwager, E. E., Sharma, P. P., Clarke, T., Leite, D. J., Wierschin, T., Pechmann, M., ... McGregor, A. P. (2017). The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. *BMC Biology*, *15*, 62. https://doi.org/10.1186/s12915-017-0399-x.

Serrano, M., Lin, A. W., McCurrach, M. E., Beach, D. & Lowe, S. W. (1997). Oncogenic *ras* provokes premature cell senescence associated with accumulation of p53 and $p16^{INK4a}$. *Cell*, *88*, 593-602. https://doi.org/10.1016/s0092-8674(00)81902-9.

Shinzato, C., Khalturin, K., Inoue, J., Zayasu, Y., Kanda, M., Kawamitsu, M., ... Satoh, N. (2020). Eighteen coral genomes reveal the evolutionary origin of Acropora strategies to accommodate environmental changes. *Molecular Biology and Evolution*, 38, 16-30. https://doi.org/10.1093/molbev/msaa216.

1 Shinzato, C., Takeuchi, T., Yoshioka, Y., Tada, I., Kanda, M., Broussard, C., ... Inoue,
2 M. (2021). Whole-Genome Sequencing Highlights Conservative Genomic Strategies
3 of a Stress-Tolerant, Long-Lived Scleractinian Coral, *Porites australiensis* Vaughan,
4 1918. *Genome Biology and Evolution*, 13, evab270.
5 https://doi.org/10.1093/gbe/evab270.
6 Shiomi, K., Kiyono, T., Okamura, K., Uezumi, M., Goto, Y., Yasumoto, S., ... Hashimoto,
7 N. (2011). CDK4 and cyclin D1 allow human myogenic cells to recapture growth
8 property without compromising differentiation potential. *Gene Therapy*, *18*, 857-866.
9 https://doi.org/10.1038/gt.2011.44.
10 Simakov, O., Kawashima, T., Marlétaz, F., Jenkins, J., Koyanagi, R., Mitros, T., ...
11 Gerhart, J. (2015). Hemichordate genomes and deuterostome origins. *Nature*, 527, 459-
12 465. https://doi.org/10.1038/nature16150.
13 Simakov, O. & Kawashima, T. (2017). Independent evolution of genomic characters
14 during major metazoan transitions. *Developmental Biology*, *427*, 179-192.
15 https://doi.org/10.1016/j.ydbio.2016.11.012.
16 Slyusarev, G. S., Starunov, V. V., Bondarenko, A. S., Zorina, N. A. & Bondarenko, N. I.
17 (2020). Extreme Genome and Nervous System Streamlining in the Invertebrate
18 Parasite *Intoshia variabili*. *Current Biology*, 30, 1292-1298. doi:

10.1016/j.cub.2020.01.061.

Soufir, N., Avril, M. F., Chompret, A., Demenais, F., Bombled, J., Spatz, A., ... Bressac-de Paillerets, B. (1998). Prevalence of *p16* and *CDK4* germline mutations in 48 melanoma-prone families in France. The French Familial Melanoma Study Group. *Human Molecular Genetics*, *7*, 209-216. https://doi.org/10.1093/hmg/7.2.209.

Takeuchi, T., Suzuki, Y., Watabe, S., Nagai, K., Masaoka, T., Fujie, M., ... Myers, E. W. (2022). A high-quality, haplotype-phased genome reconstruction reveals unexpected haplotype diversity in a pearl oyster. *DNA Research*, 29, 1-13. https://doi.org/10.1093/dnares/dsac035.

Telford, M. J., Budd, G. E., & Philippe, H. (2015). Phylogenomic Insights into Animal Evolution. *Current Biology*, *25*, R876-887. https://doi.org/10.1016/j.cub.2015.07.060.

Tsao, H., Benoit, E., Sober, A. J., Thiele, C. & Haluska, F. G. (1998). Novel mutations in the *p16/CDKN2A* binding region of the *cyclin-dependent kinase-4* gene. *Cancer Research*, 58, 109-113.

Veinalde, R., Ozola, A., Azarjana, K., Molven, A., Akslen, L. A., Doniņa, S., ... Pjanova, D. (2013). Analysis of Latvian familial melanoma patients shows novel variants in the noncoding regions of *CDKN2A* and that the *CDK4* mutation R24H is a founder

mutation. *Melanoma Research*, *23*, 221-226. https://doi.org/10.1097/CMR.0b013e3283608695.

Wan, Z.-Y., Song, F., Sun, Z., Chen, Y.-F., Zhang, W.-L., Samartzis, D., ... Luo, Z.-J. (2014). Aberrantly expressed long noncoding RNAs in human intervertebral disc degeneration: a microarray related study. *Arthritis Research and Therapy*, *16*, 465. https://doi.org/10.1186/s13075-014-0465-5.

Webster, K. A., Henke, K., Ingalls, D. M., Nahrin, A., Harris, M. P. & Siegfried, K. R. (2018). Cyclin-dependent kinase 21 is a novel regulator of proliferation and meiosis in the male germline of zebrafish. *Reproduction*, *157*, 383-398. https://doi.org/10.1530/REP-18-0386.

Weinberg, R. A. (2014). *The Biology of Cancer*, *2nd edition*. New York City: Garland Science.

Wölfel, T., Hauer, M., Schneider, J., Serrano, M., Wölfel, C., Klehmann-Hieb, E., ... Beach, D. (1995). A $p16^{INK4a}$-insensitive CDK4 mutant targeted by cytolytic T lymphocytes in a human melanoma. *Science*, *269*, 1281-1284. https://doi.org/10.1126/science.7652577.

Wood, D. J. & Endicott, J. A. (2018). Structural insights into the functional diversity of the CDK-cyclin family. *Open Biology*, *8*, 180112. https://doi.org/ 10.1098/rsob.180112.

Yates, A. D., Allen, J., Amode, R. M., Azov, A. G., Barba, M., Becerra, A., ... Flicek, P. (2022). Ensembl Genomes 2022: an expanding genome resource for non-vertebrates. https://doi.org/10.1093/nar/gkab1007.

Yoshioka, Y., Suzuki, G., Zayasu, Y., Yamashita, H. & Shinzato, C. (2022). Comparative genomics highlight the importance of lineage-specific gene families in evolutionary divergence of the coral genus, *Montipora*. *BMC Ecology and Evolution*, 22, 71. https://doi.org/10.1186/s12862-022-02023-8.

Zhang, P., Liégeois, N. J., Wong, C., Finegold, M., Hou, H., Thompson, J. C., ... Elledge, S. J. (1997). Altered cell differentiation and proliferation in mice lacking $p57^{KIP2}$ indicates a role in Beckwith-Wiedemann syndrome. *Nature*, *387*, 151-158. https://doi.org/10.1038/387151a0.

Zuo, L., Weger, J., Yang, Q., Goldstein, A. M., Tucker, M. A., Walker, G. J., ... Dracopoli, N. C. (1996). Germline mutations in the $p16^{INK4a}$ binding domain of CDK4 in familial melanoma. *Nature Genetics*, *12*, 97-99. https://doi.org/10.1038/ng0196-97.

1 **Figure legends**

2 **FIGURE 1** Distribution of *Cdkn2*, *Cdk4/6* and *Cdkn1*, and the extent of the

3 conservation of the key charged amino acids in all invertebrate phyla. The phylogenetic

4 tree was modified from Telford *et al*. (2015). The rows for phyla in which more than

5 five species were included in this analysis are highlighted with yellow. *The N-terminal

6 sequence including the corresponding residue to R24/31 in the contig was lacking. ND:

7 not determined because of the lack of the genome data. N/A: not applied.

8

9 **FIGURE 2** Amino acid alignment around the residue corresponding to Cdkn2 D84 in

10 representative invertebrate species from seven phyla where *Cdkn2* was detected. The

11 residue corresponding to Cdkn2 D84 is colored with light green when the negative

12 charge is conserved. If not, it is colored with magenta. Human CDKN2's are aligned as

13 references. **a** and **b** represent amino acids of α-helix and β-turn in CDKN2A,

14 respectively (Baumgartner *et al*., 1998).

15

16 **FIGURE 3** *Cdkn2* and *Faf1* are tightly linked during evolution. Arrows show the

17 direction of transcription. The number of the chromosome or the linkage group is shown

18 when available. Microsynteny around human *CDKN2* loci are shown as references.

Genes encoding ankyrin repeat-containing proteins other than Cdkn2 are represented in red.

FIGURE 4 Microsynteny around the *Faf1* locus in representative invertebrates without *Cdkn2*. Arrows show the direction of transcription. The number of the chromosome or the linkage group is shown when available. Genes encoding ankyrin repeat-containing proteins other than Cdkn2 are represented in red.

FIGURE 5 Amino acid alignment around the residue corresponding to Cdk4/6 R24/31 in representative invertebrate species of Deuterostomata. The residue corresponding to Cdk4/6 R24/31 is colored with light green when the strong positive charge was conserved. If not, it is colored with magenta. Human CDK4 and CDK6 are aligned as references.

FIGURE 6 Hypothesis on the evolution of *Cdkn2* locus. (a) Loss of *Cdkn2* occurred multiple times. (b) Evolution of microsynteny at *Cdkn2* locus. *Cdkn2e* is a hypothetical transient gene. For details, see the main text.

1

2 **Supplementary material**

3 **Table S1** The list of the analyzed invertebrate *Cdkn2*'s.

4 **Table S2** Results of blast search on invertebrate Cdkn2. (a) Result of blastp on

5 refseq_protein with Cdkn2. (b) Results of tblastn on wgs database with Cdkn2. (a) and

6 (b) are put together in one file with the worksheet tab.

7 **Table S3** The list of the analyzed vertebrate *Cdkn2*'s.

8 **Table S4** Microsynteny around *Cdkn2* locus in invertebrates.

9 **Table S5** Microsynteny around *Faf1* locus in invertebrates. (a) Microsynteny around *Faf1*

10 in invertebrates other than Arthropoda. (b) Microsynteny around *Faf1* in Arthropoda. (a)

11 and (b) are put together in one file with the worksheet tab.

12 **Table S6** The list of the identified invertebrate *Cdkn1*'s. (a) The list of the *Cdkn1* in the

13 species other than Arthropoda. (b) The list of the *Cdkn1* in Arthropoda. (a) and (b) are

14 put together in one file with the worksheet tab.

15 **Table S7** The list of the analyzed invertebrate *Cdk4/6*'s. The tables contain the database

16 ID number and the scientific name of the species in the deuterostome invertebrates (A),

17 Platyhelminthes (B), other Lophotrochozoa (C), Nematoda (D), other Ecdysozoa (E), and

18 basal invertebrates (F).

1 **Table S8** The list of the analyzed vertebrate *Cdk4/6*'s.

2 **Table S9** The list of possibly misannotated invertebrate genes as *Cdkn2*.

3 **Figure S1** Amino acid alignment of invertebrate Cdkn2's. (a) The alignment of the N-

4 terminal half of Cdkn2. (b) The alignment of the C-terminal half. The residue

5 corresponding to Cdkn2 D84 is colored with light green when the negative charge is

6 conserved. If not, it is colored with magenta. (c) Consensus residues in invertebrate

7 Cdkn2. **a** and **b** represent amino acids of α-helix and β-turn in *Homo sapiens* CDKN2's,

8 respectively (Baumgartner *et al*., 1998).

9 **Figure S2** Cdkn2 amino acid alignment around the residue corresponding to D84 in

10 vertebrates; Cdkn2a in mammals (a), and in non-mammalian vertebrates (b); Cdkn2b in

11 mammals (c), in non-mammalian/non-fish vertebrates (d), and in fish (e); Cdkn2c in

12 mammals (f), in non-mammalian/non-fish vertebrates (g), and in fish (h); Cdkn2d in

13 mammals (i) and in non-mammalian vertebrates (j). The residue corresponding to

14 Cdkn2 D84 is colored with light green when the negative charge is conserved. If not, it

15 is colored with magenta.

16 **Figure S3** Cdk4/6 amino acid alignment around the residue corresponding to R24/R31

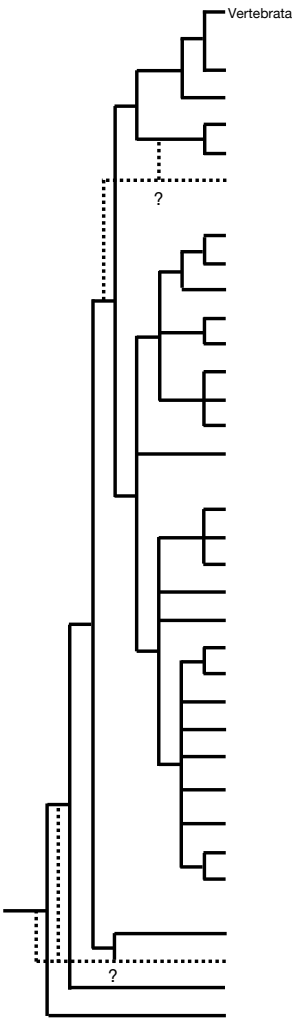
17 in invertebrates; deuterostome invertebrates (a), Nematoda (b), other Ecdysozoa (c),

18 Platyhelminthes (d), other Lophotrochozoa (e), and basal invertebrates (f). The residue

corresponding to Cdk4/6 R24/31 is colored with light green when the strong positive charge is conserved. If not, it is colored with magenta.

Figure S4 Cdk4 amino acid alignment around the residue corresponding to R24/R31 in vertebrates; Cdk4 in mammals (a), in non-mammalian/non-fish vertebrates (b), and in fish (c); Cdk6 in mammals (d), in non-mammalian/non-fish vertebrates (e), and in fish (f); Cdk21 in fish (g). The residue corresponding to Cdk4/6 R24/31 is colored with light green when the strong positive charge is conserved. If not, it is colored with magenta.

Figure S5 The data to determine the Spearman's rank correlation coefficient. (a) "**[2]/[1]** in %" and "% of -R, K at Cdk4/6 R24/31" were used to determine the Spearman's rank correlation coefficient. (b) The scatter plot of "**[2]/[1]** in %" and " of -R, K at Cdk4/6 R24/31".



	Cdkn2	Cdkn2 D84	*Cdk4/6*	% of non-R, K at Cdk4/6 R24/31	*Cdkn1*
Deuterostome Invertebrates					
Urochordata	not detected	N/A	detected	14.3%	detected
Cephalochordata	detected	conserved	detected	0.0%	detected
Hemichordata	detected	conserved	detected	0.0%	detected
Echinodermata	detected	conserved except Crinoid	detected	0.0%	detected
Xenacoelomorpha	not detected	N/A	detected	50.0%	detected
Ecdysozoa					
Arthropoda	not detected	N/A	detected	3.9%	detected
Onychophora	detected	conserved	detected	0.0%	detected
Tardigrada	not detected	N/A	detected	50.0%	detected
Nematomorpha	not detected	N/A	detected	0.0%	detected
Nematoda	not detected	N/A	detected	63.9%	detected
Priapulida	not detected	N/A	detected*	N/A	detected
Loricifera	ND	ND	ND	ND	ND
Kinorhyncha	ND	ND	ND	ND	ND
Chaetognatha	ND	ND	ND	ND	ND
Lophotrochozoa					
Micrognathozoa	ND	ND	ND	ND	ND
Gnathostomulida	ND	ND	ND	ND	ND
Rotifera	not detected	N/A	detected	44.4%	detected
Dicyemida	not detected	N/A	not detected	N/A	not detected
Orthonectida	not detected	N/A	not detected	ND	not detected
Cycliophora	ND	ND	ND	ND	ND
Entoprocta	ND	ND	ND	ND	ND
Annelida	detected	conserved	detected	0.0%	detected
Mollusca	detected	conserved	detected	0.0%	detected
Nemertea	not detected	N/A	detected	100%	detected
Brachiopoda/Phoronida	detected	conserved except in *P. ovalis*	detected	0%	detected
Bryozoa	not detected	N/A	detected	0%	detected
Platyhelminthes	not detected	N/A	detected	97.9%	detected only in Cestoda
Gastrotricha	ND	ND	ND	ND	ND
Basal Invertebrates					
Cnidaria	not detected	N/A	detected	10.0%	detected
Ctenophora	not detected	N/A	detected	0.0%	detected
Placozoa	not detected	N/A	detected	0.0%	detected
Porifera	not detected	N/A	not detected	N/A	detected

FIGURE 1 Distribution of *Cdkn2*, *Cdk4/6* and *Cdkn1*, and the extent of the conservation of the key charged amino acids in all invertebrate phyla.

```
                                                                ankyrin repeat 3
                                  aa         b          bb  aaaaaaaa  aaaaaaaaaa
Homo sapiens CDKN2A               LHGAEPNCADPA----------TLTRPVHDAAREGFLDTLVVLHRAGARL
Homo sapiens CDKN2B               LHGAEPNCADPA----------TLTRPVHDAAREGFLDTLVVLHRAGARL
Homo sapiens CDKN2C               LRGANPDLKDRT-----------GFAVIHDAARAGFLDTLQTLLEFQADV
Homo sapiens CDKN2D               KQGASPNVQDTS-----------GTSPVHDAARTGFLDTLKVLVEHGADV

Cephalochrdata
Branchiostoma belcheri            ENGADPNVRDPGR---------KGQYPVHDAARGGWDEALKLLFDYHAEL
Hemichordata
Ptychodera flava                  RRGADPNVADST-----------GRAPIHDAAEGGYVNTVQVFVEYGGDL
Saccoglossus kowalevskii          RRGADPNVADST-----------GRAPIHDAAEGGYVNTVQVFVEYGGDL
Echinodarmata
Strongylocentrotus purpuratus     DHGADLDRPDPS----------VQRTILHDIAERGCIETLNLLLEHGANA
Apostichopus japonicus            QKGANPNVYDMNP---DSDHRPFLRTPLHDALHHGNEDSAILLIQHGADL
Acanthaster planci                KSGADPNVRDRD----------VGRTPLHDAAEHGHEDTVRLLLKHGADP
Ophioderma brevispina             DEGADPNVRDPD----------VGRTPLHDAVEYGIPSTVMCLLQGGADV
Anneissia japonica                EYGANPNLKEPV----------MGGIPLHQAAREGATATLELMLQFGANV
Nesometra (Dorometra) sesokonis   EYGANPHLPEPV----------MGGIPLHNAAREGATTTLELLLQYGSNV
Onychophora
Epiperipatus broadwayi            KAGANVNVSDV-----------TLRTPVHDAAEAGFLDVVQNLIAYGADI
Euperipatoides rowelli            SAGANPNICDR-----------TLRTPVHDAAEAGFLDVLQCLLANGGVI
Annelida
Lamellibrachia satsuma            RNGADPNVRDRS--YP--GHQDTLVTPAHDVARAGYTDVLKCLLRYGADV
Dimorphilus gyrociliatus          KAGANPNVQDRG--IG--NGRCLLLTPGHDAARVGYVRTLKVLLKYGADT
Sipunculus nudus                  EAGADPNVVDHS--KN--NKKDTLLTPAHDVARQGYLDTLKVLLWYGADI
Mollusca
Pomacea canaliculata              EAGGDPNALDIG--KR--TDTSTLRTVAHDVAASGYTDVLRCLLRHGADI
Nodopelta heminoda                RAGADPNVKDVS--LRQDGGSSTFRTICHDVAESGALSTLRCLLRHGADI
Patella vulgata                   HSGANPNNQDIS--TRDTDCLSTYRTICHDVAASGYLETLHVLLKYGTDT
Haliotis rufescens                CAGADPNVKDIS--HRHIGGTSTYRTICHDVAESGSLPTLQCLLRHGADI
Nautilus pompilius                RAGADPNLMEGD----------TRRTICHDVAESGYLDTLRCLLKHGADL
Brachiopoda/Phoronida
Lingula anatina                   KHGADPNIQDKH--FQD-LHLSSYRTPAHDIAEQGFLDTMKVLLKHGADV
Phoronis ovalis                   EHGADPNLKERSPHHEADPSRSLLRTPAHNTCCEGFLDTLKTLLRCGANI
Phoronis pallida                  MRGADPNIQDT----------STLRTILHDVCEFGYMDTLQCLISFGANL
```

FIGURE 2 Amino acid alignment around the residue corresponding to Cdkn2 D84 in representative invertebrate species from seven phyla where *Cdkn2* was detected.

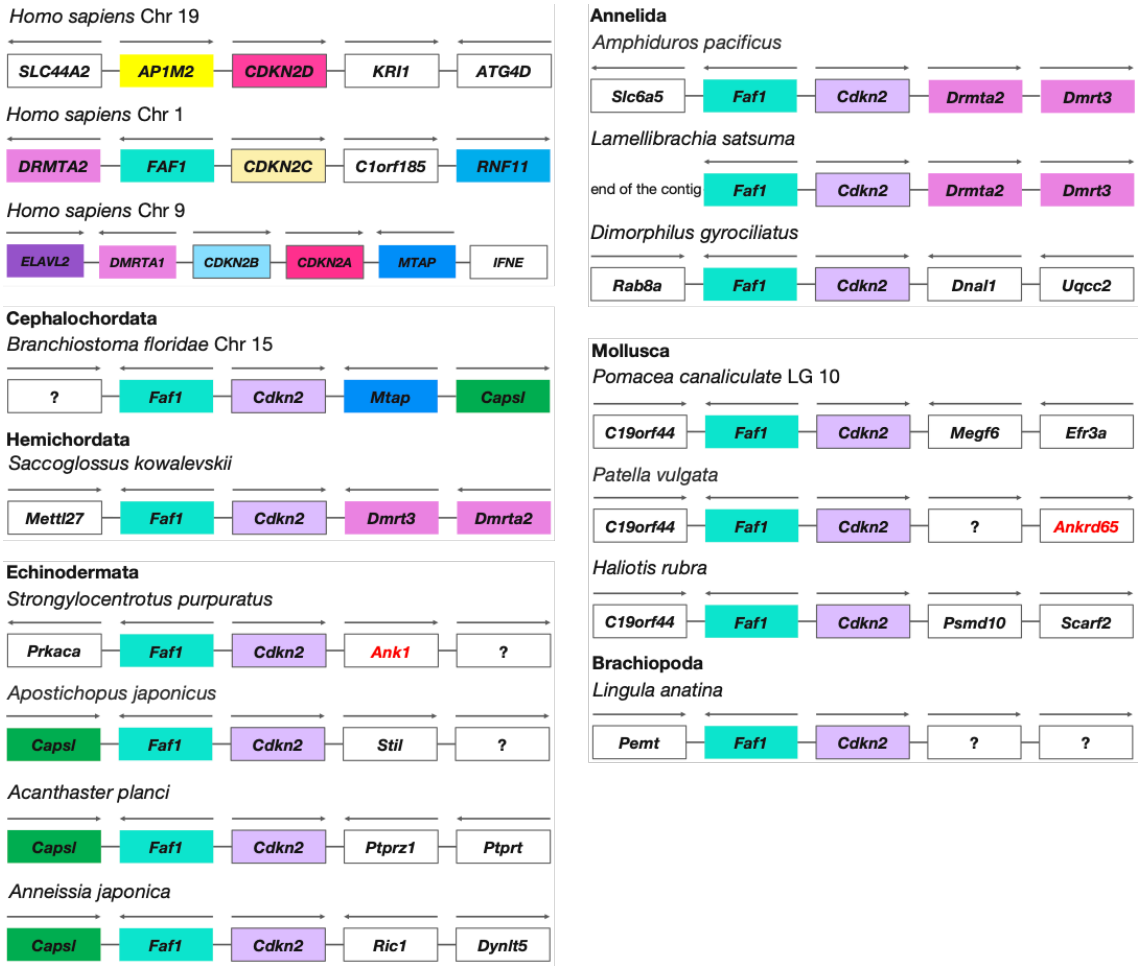


FIGURE 3 *Cdkn2* and *Faf1* are tightly linked during evolution.

Urochordata
Ciona intestinalis
Pelp1 Blnk Faf1 Slc43a3 Yipf1
Styela clava
Zbed6 K02A2.6 Faf1 ? ?
Xenacoelomorpha
Praesagittifera naikaiensis
Upb1 Capn15 Faf1 end of the contig
Arthropoda
Argiope bruennichi Chr 7
Snx18 si:dkey-109l4.3 Faf1 Rnf170 Tmem222
Daphnia magna LG10
Klhl2 Slc39a14 Faf1 Ssuh2 ?
Tribolium castaneum LG7
Pabpn1 Sgf29 Faf1 Grik2 Grid2/Grid1
Drosophila melanogaster Chr 2R
Egflam C16orf89 Faf1 Frrs1 Emc7
casp
Tardigrada
Paramacrobiotus metropolitanus
Apod Bphl Faf1 Apod Borcs5
Priapulida
Priapulus caudatus
Isca1 Oxct1 Faf1 ? ?
Nematomorpha
Gordionus sp. m RMFG-2023 Chr 4
Eapp Aifm1 Faf1 Nfic Rorb
Nematoda
Caenorhabditis elegans Chr II
Slc36a1 Ing3 Faf1 ? chs-2
ubxn-3
Chr X
Exoc6 Serpinb5 Faf1 Ube2k Zfp64
ubc-23
Brugia malayi
Cdc14b Rcc1 Faf1 Incenp Epx
Strongyloides ratti Chr 2
Trim16 Nap1l4 Faf1 ? ?
Rotifera
Adineta ricciae
Atf7ip Nphp1 Faf1 ? Myo7a
Caps2 ? Faf1 Dvl2 Eri1
Annelida
Helobdella robusta
RT from mobile element ? Faf1 Baat ?
Owenia fusiformis Chr 1
Cyp1a1 Nsmce2 Faf1 Wdr36 Sh2d5
Mollusca
Pecten maximus Chr 4
Dmrta2 Dmrta2 Faf1 Tep1 Pde4d
Aplysia californica
? Alg2 Faf1 C19orf44 Lrrc40
Gigantopelta aegis Chr 3
? Pes1 Faf1 Ift20 Kctd10
Octopus sinensis LG4
Stpg1 Nrsn1 Faf1 Apc Pdss2
Nemertea
Notospermus geniculatus
Ptk2 Cox5b Faf1 Aspg ?
Bryozoa
Bugula neritina
? Mrpl40 Faf1 end of the contig
Platyhelminthes
Schistosoma mansoni Chr 6
Limd1 Ctsa Faf1 Znf786 Mfsd14b
Mesocestoides corti
C5orf49 Mapk9 Faf1 Zftraf1 ?
Schmidtea mediterranea
Sub1 ? Faf1 ? Grm8

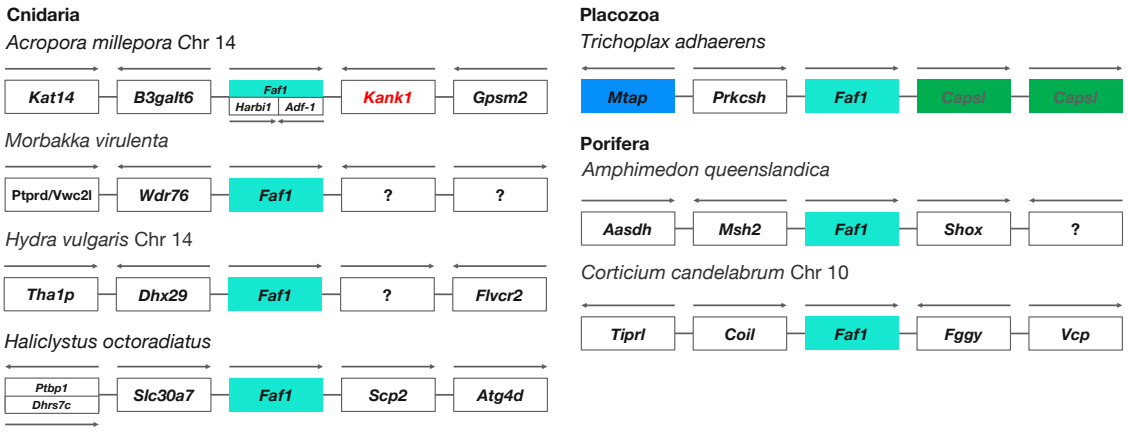


FIGURE 4 Microsynteny around the *Faf1* locus in representative invertebrates without *Cdkn2*.

Cdk4	
Homo sapiens	RYEPVAEIGVGAYGTVYKARDPHSG---HFVALKSVRVPN
Cdk6	
Homo sapiens	QYECVAEIGEGAYGKVFKARDLKNG--GRFVALKRVRVQT
Cdk4/6	
Urochordata	
Phallusia fumigata	RYMPLEEIGVGAYGIVYKARDTSPDTDHGVVAMKCLRVEN
Ciona intestinalis	RYHTLSEIGVGAYGVVFKAQDMQSDDENKFVAIKCVRVEN
Ciona savignyi	RYHTLTEIGVGAYGVVFKAQDRHSEEENKFVAMKCIRVEN
Corella inflata	RYKPVEKIGVGAYGVVFKAKDLRKD--GEFVAMKCLRVEN
Molgula occulta	KYEKDEEIGVGAYGTVYRGKDLANL--GKYVALKCLRVQT
Botrylloides schlosseri	NYEMVAEIGVGAYGVVYRARDLEND--RKLVALKCLRVQN
Styela clava	NYEMVAEIGVGAYGIVYRARDLENN--GRFVALKCLRVQN
Halocynthia roretzi	NYEMVAEIGVGAYGIVYRARDLEND--NRYVALKCLRVQN
Oikopleura dioica	DWIDMEEIGGGAYGVVFRARDRQHD--NRVVALKRLRVQA
Cephalochordata	
Branchiostoma belcheri	EYEKVAKIGTGAYGTVYKARDLKNG--GEFVALKRVRVQT
Asymmetron lucayanum	QYEKVAKIGTGAYGTVYKARDLKNG--GEFVALKRVRVQT
Hemichordata	
Ptychodera flava	RYEELAEIGTGAYGIVYKARDVESG---RIVALKKVRVQT
Saccoglossus kowalevskii	KYEEVAEIGTGAYGTVYKARDLETG---RYVALKKVRVQT
Echinodermata	
Strongylocentrotus purpuratus	EYVQAAEIGSGAYGIVYKARDTETG---HFVALKSVRIPI
Apostichopus japonicus	EFSIIEEIGTGAYGTVYKARSKATR---EFVALKKIIVSH
Acanthaster planci	NYAVISEIGTGAYGTVYKARDLETG---KVVALKKVRLHT
Ophionereis fasciata	NYSVMAEIGTGAYGTVFKARDKTNG---HIVALKKVRIQT
Anneissia japonica	DFQKIEKIGEGTYGVVYKARDKISG---EVVALKKIRLDT
Xenacoelomorpha	
Hofstenia miamia	HYSVLEKVGVGAYGEVYRAKDPRTG---GHVALKRVRVFD
Praesagittifera naikaiensis	NYSIVESVGKGAYAKVYKAVNPQNPT--EIVALKQINIAS

FIGURE 5 Amino acid alignment around the residue corresponding to Cdk4/6 R24/31 in representative invertebrate species of Deuterostomata.

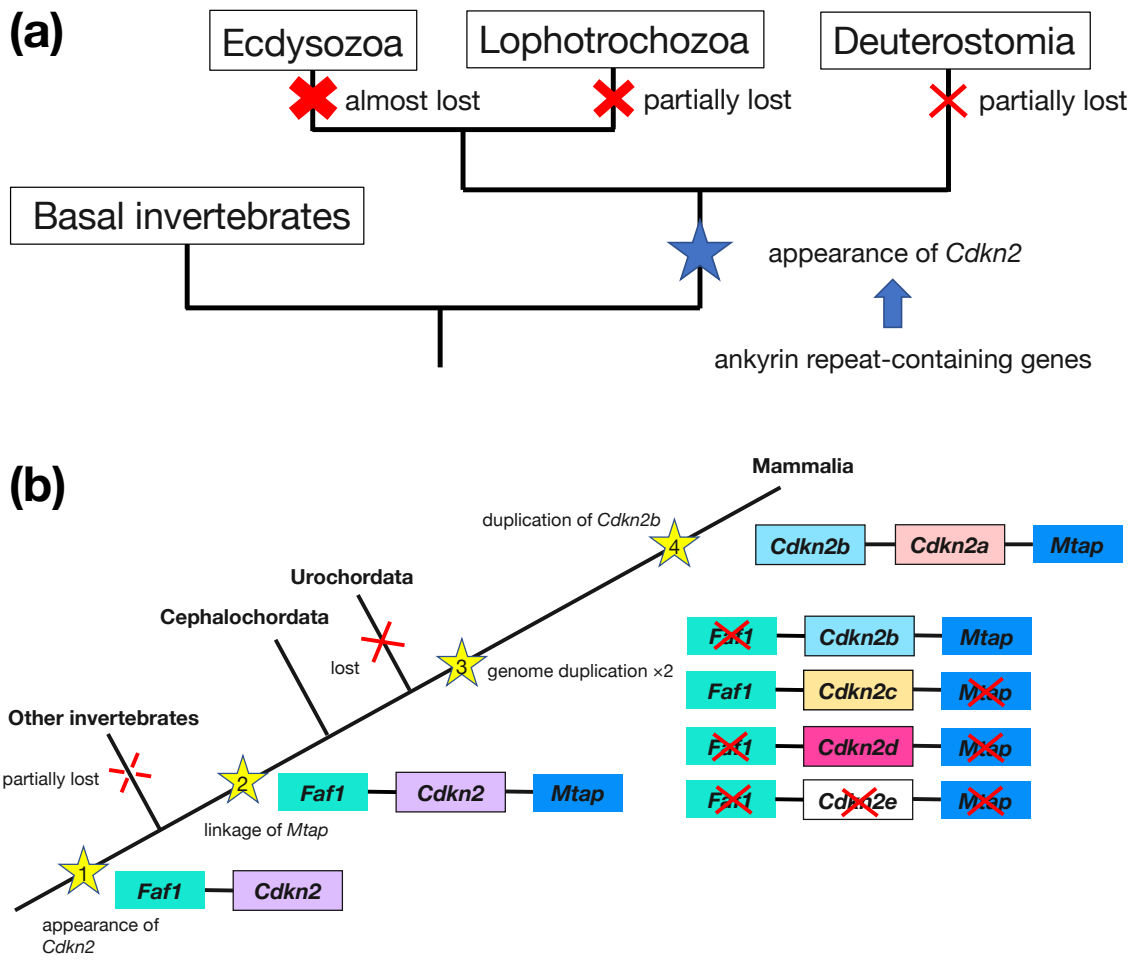


FIGURE 6 Hypothesis on the evolution of *Cdkn2* locus.

TABLE 1 Detailed distribution of Cdkn2 and Cdk4/6 in the subtaxons of the seven phyla where Cdkn2 was detected.

Phyla/subtaxons	Total number of analyzed species	The number of species where *Cdk4/6* was detected	The number of species where *Cdkn2* was detected
Cephalochordata	4	4	3
Hemichordata	2	2	2
Echinodermata	21	21	21
Echinozoa (incl. Holothuroidea)	9	9	9
Asterozoa (incl. Ophiuroidea)	9	9	9
Pelmatozoa	2	2	2
Onychophora	2	1	2
Annelida	13	7	6
Polychaeta (incl. Pogonophora)	9	6	5
Sipuncula	1	0	1
Clitellata	1	1	0
Mollusca	65	63	22
Bivalvia	30	30	0
Gastropoda	30	29	21
Caenogastropoda	11	10	10
Heterobranchia	6	6	0
Neomphalina	2	2	1
Patellogastropoda	4	4	3
Vetigastropoda	7	7	7
Cephalopoda	5	5	1
Coleoidea	4	4	0
Nautiloidea	1	1	1
Brachiopoda/Phoronida	4	4	4

TABLE 2 Number of species with each type of substituted amino acids for CDK4/6 R24/R31.

Phyla	Total number of analyzed sequences	R	K	H	A	V	I	L	T	S	N	Q	E	Y	F	C	% of non-R, K
Deuterostome Invertebrates	**43**	**36**	**4**			**1**						**2**					**7.0%**
Urochordata	14	9	3									2					14.3%
Cephalochordata	4	4															0.0%
Hemichordata	2	2															0.0%
Echinodermata	21	21															0.0%
Xenacoelomorpha	2		1			1											50.0%
Ecdysozoa	**1058**	**772**	**199**	**3**	**1**	**2**	**3**	**2**	**56**	**2**	**5**	**2**	**2**			**7**	**8.0%**
Arthropoda	979	753	186	3	1	1	1	2	17		4	1	1			7	3.9%
Onychophora	2	2															0.0%
Tardigrada	2	1				1											50.0%
Nematomorpha	3	3															0.0%
Nematoda	72	13	13				2		39	2	1	1	1				63.9%
Lophotrochozoa	**133**	**72**	**9**	**1**		**2**	**5**	**3**	**35**	**1**		**1**		**3**	**1**		**39.1%**
Rotifera	9	5					1							3			44.4%
Annelida	7	6	1														0.0%
Mollusca	63	56	7														0.0%
Nemertea	2						2										100.0%
Brachiopoda /Phoronida	4	4															0.0%
Bryozoa	1	1															0.0%
Platyhelminthes	47		1	1		2	2	3	35	1		1			1		97.9%
Basal Invertebrates	**62**	**53**	**3**			**2**	**1**	**1**							**2**		**9.7%**
Cnidaria	60	51	3			2	1	1							2		10.0%
Ctenophora	1	1															0.0%
Placozoa	1	1															0.0%