Evolution of the *Cdk4/6–Cdkn2* system in invertebrates

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, Iwate University, Morioka, Japan

[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

The cell cycle is driven by cyclin-dependent kinases (Cdks). The decision whether the cell cycle proceeds is made during G1 phase, when Cdk4/6 functions. Cyclin-dependent kinase inhibitor 2 (Cdkn2) is a specific inhibitor of Cdk4/6, and their interaction depends on D84 in Cdkn2 and R24/31 in Cdk4/6. This knowledge is based mainly on studies in mammalian cells. Here, we comprehensively analyzed *Cdk4/6* and *Cdkn2* in invertebrates and found that *Cdk4/6* was present in most of the investigated phyla, but the distribution of *Cdkn2* was rather uneven among and within the phyla. The positive charge of R24/R31 in Cdk4/6 was conserved in all analyzed species in phyla with Cdkn2. The presence of *Cdkn2* and the conservation of the positive charge were statistically correlated. We also found that *Cdkn2* has been tightly linked to *Fas associated factor 1* (*Faf1*) during evolution. We discuss potential interactions between Cdkn2 and Cdk4/6 in evolution and the possible cause of the strong conservation of the microsynteny.

Introduction

Regulation of the cell cycle is closely associated with ontogeny, tissue regeneration, cell senescence, pathogenesis, especially carcinogenesis, and evolution. Cyclin-dependent kinases (Cdks) and their regulatory subunits, cyclins, promote cell cycle, while Cdk inhibitors (Cdkns) repress the Cdk function (reviewed in Weinberg, 2014). In mammalian cells, Cdk4 or Cdk6, and cyclin Ds drive the cell cycle in G1 phase, when cell proliferation is sensitive to extracellular signals. Fish have Cdk21, the third member of the Cdk4/6 subgroup of the Cdk family (Webster et al., 2018). Mammals have three types of structurally different Cdkns, Cdkn1–Cdkn3; the former includes Cdkn1a–Cdkn1c (aka $p21^{Cip1/Waf1}$, $p27^{Kip1}$, and $p57^{Kip2}$, respectively; reviewed in Mu□hleder *et al.*, 2021). Cdkn1s preferentially inhibit Cdk1 and Cdk2 but also Cdk4/6.

Cdk inhibitor 2s (Cdkn2s) are specific for Cdk4/6. Mammals have Cdkn2a–Cdkn2d (aka $p16^{Ink4a}$, $p15^{Ink4b}$, $p18^{Ink4c}$, and $p18^{Ink4d}$, respectively); Cdkn2a and Cdkn2b have four ankyrin repeats, and Cdkn2c and Cdkn2d have five (Baumgartner et al., 1998). Cdkn1s and Cdkn2s are known as tumor suppressors (El-Deiry et al., 1993; Jen *et al.*, 1994; Kamb et al., 1994; Chen et al., 1996; Zhang et al., 1997; Husemann et al., 1999; Gluick et al., 2013). In mammalian cells, Cdkn2s accumulate during cell senescence (Serrano et al., 1997; Kita et al., 2022; Liu et al., 2023). Electrostatic interaction between R24/R31 in Cdk4/6 and D84 in Cdkn2s is important for the formation of the Cdk4/6–Cdkn2 complex, which causes a structural change preventing cyclin and ATP binding to Cdk4/6 and its following activation (Brotherton et al., 1998; Russo et al., 1998; reviewed in Baker & Reddy 2013)*. Mutations R24C and R24H in CDK4 cause familial melanoma, suggesting that strong positive charge (R or possibly K, not H) at

the residue is necessary for the complex formation (Wölfel et al., 1995; Zuo et al., 1996; Soufir et al., 1998; Tsao et al., 1998; Molven et al., 2005; Veinalde et al., 2013). This interaction mechanism is used in a cell immortalization protocol (Shiomi et al., 2011).

Phosphatase Cdkn3 (aka Cdk2-associated dual specificity phosphatase, Kap or Cip2) specifically inhibits Cdk2 (reviewed in Patterson et al., 2009). Unlike Cdkn1s and Cdkn2s, which act as tumor suppressors, Cdkn3 has been suggested to act as an oncogene (Liu et al., 2019). No direct interaction between Cdk4/6 and Cdkn3 has been reported.

Unlike the molecular mechanisms of the cell cycle in mammals, those in invertebrates (except *Caenorhabditis elegans* and *Drosophila melanogaster*) are poorly studied (reviewed in Kipreos & van den Heuvel, 2019). These two species have no Cdkn2, and Cdkn1 is thought to subrogate the Cdk4/6 function in G1 phase (Kipreos & van den Heuvel, 2019). On the other hand, *Cdkn1* compensates for the lack of *Cdkn2* in the embryos of *Ciona* (Urochordata; Kobayashi et al., 2022). Planarians lack Cdkn1s, and Cdh1 is important for exiting the cell cycle (Sato et al., 2022).

A few pioneering small-scale genome studies reported the existence of *Ckd4/6* genes in invertebrates, namely 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*) (Cao et al., 2014; Espinal-Centeno et al., 2020). No *Ckd4/6* gene was found in Porifera (*Amphimedon queenslandica*) nor *Cdkn2* gene in Platyhelminthes (*S. mediterranea*) or Cnidaria (*H. vulgaris*) (Cao et al., 2014; Espinal-Centeno et al., 2020). Considering the diversity of invertebrates, more systematic and comprehensive investigation of the regulatory

mechanisms of the cell cycle is needed.

Here, we investigated the presence of *Cdkn2* and *Cdk4/6* genes in invertebrates on a large scale (61 Cdkn2 sequences from 61 invertebrate species, and 1152 Cdk4/6 sequences from 1172 invertebrate species).

***Footnote**

The numbering of the amino acids is based on human proteins. D84 in human CDKN2A corresponds to D86 in CDKN2B, D76 in CDKN2C, and D80 in CDKN2D. For simplicity, we use D84 for D84/D86/D76/D80 in this paper.

Results

Presence of Cdkn2 in invertebrates

First, we systematically searched for *Cdkn2* in the genomes of all 24 invertebrate phyla whose genome data are available (Figure 1). During the initial keyword and blastp searches and the following more thorough tblastn searches in whole-genome shotgun contig databases, we detected *Cdkn2* only in seven phyla (Figures 1, S1, Tables 1, S1, S2). Among invertebrate Deuterostomia, we detected *Cdkn2* in Cephalochordata, Hemichordata, and Echinodermata, but not in Urochordata or Xenacoelomorpha. Among Lophotrochozoa, we detected *Cdkn2* only in Mollusca, Annelida, and Brachiopoda/Phoronida; among Mollusca, we detected *Cdkn2* in Gastropoda (except Heterobranchia) and Cephalopoda (except Coleoidea), but not in Bivalvia. Among Ecdysozoa, we found *Cdkn2* only in Onychophora. We did not find it in the basal invertebrates (non-Bilateria).

During tblastn searches, we noticed sporadic "weakly positive" sequences in phyla where our searches mostly suggested the lack *Cdkn2* (Table S2b). Some of them apparently originated from contaminating DNA of other species. We used three additional criteria to clarify whether the "weakly positive" sequences other than obvious contaminating DNA might be *bona fide Cdkn2*. In all Cdkn2 candidate sequences analyzed in this study, the exon junction between Q50 and V51 was conserved (Table S1). We also determined the amino acid residues conserved among invertebrate Cdkn2s (Figure S1c). The "weakly positive" sequences failed both criteria. In the phylogenetic analysis of the Cdkn2 candidates and the human homologs of other ankyrin repeat–containing sequences with human CDKN1A as an outgroup (Tables S1 and S2), the Cdkn2 candidates formed one clade (Figure S1d; Table S3). On the basis

of these data, we concluded that the "weakly positive" sequences are not *bona fide* Cdkn2 (for details, see notes in Table S2b).

Some invertebrate sequences were potentially misannotated as Cdkn2 in the NCBI Protein database (Table S4). Some of them appeared to be Cdkn1, and others were also not *bona fide* Cdkn2 but other ankyrin repeat–containing proteins on the basis of the criteria mentioned above (for details, see notes in Table S4).

We found at least one *Cdkn1* gene in all invertebrate phyla except Dicyemida and Orthonectida (Figure 1; Table S5). Among Platyhelminthes, we found *Cdkn1* in Cestoda, but not in Rhabditophora or Trematoda.

Analysis of residues corresponding to D84 in Cdkn2

In vertebrates (reference), the residue corresponding to Cdkn2 D84 was conserved in almost all Cdkn2s (1852/1853, 99.9%; Figure S2; Table S6). The negative charge was conserved in most of the detected invertebrate *Cdkn2* (57/60, 95.0%; Figures 2, S1b). This residue was substituted in two Crinoid species (D84Q in *Anneissia japonica* and D84N in *Nesometra* (*Dorometra*) *sesokonis*) from Echinodermata and in a Brachiopoda/Phoronida species (D84N in *Phoronis ovalis*), resulting in a loss of the negative charge (Figures 2, S1b).

Microsynteny around Cdkn2 and Faf1 loci

We analyzed the microsynteny around the *Cdkn2* locus in invertebrates and found that all analyzed invertebrate *Cdkn2* genes were linked to *Fas associated factor 1* (*Faf1*), similar to vertebrate *Cdkn2c* (Figure 3, Tables S1, S7). Cephalochordate *Cdkn2* genes were linked to *methylthioadenosine phosphorylase* (*Mtap*), similar to vertebrate

Cdkn2a/*b* (Figure 3).

To obtain clues on the origin and loss of *Cdkn2*, we analyzed microsynteny around *Faf1* in the species without *Cdkn2*. We could not detect any evolutionary conservation of linkages comparable to that between *Cdkn2* and *Faf1*. No "weakly positive" sequences mentioned above was not linked with *Faf1*, which is another evidence that they are not *bona fide Cdkn2*. The linkage between *Faf1* and *KN Motif and Ankyrin Repeat Domains 1 (Kank1*) was conserved among some species in Anthozoa of Cnidaria (Figure 4, Table S8).

Presence of Cdk4/6 in invertebrates

Our systematic search for *Cdk4/6* in invertebrate genome databases with gene names and blastp/tblast revealed 1296 candidate sequences from 1265 species among 21 phyla (Table S9). We could not detect Cdk4/6 candidate sequences in Porifera (Cao et al., 2014), Dicyemida and Orthonectida (Tables 1, S9). To confirm the identity of the sequences as Cdk4/6, we performed a phylogenetic analysis. Because of the various lengths of the 1296 sequences, we selected 1172 sequences, from 1152 species among 20 phyla, which contain the protein kinase domain (PKD) without apparent deletions for the analysis to obtain a reliable result. The phylum excluded by this selection was Priapulida. Although the candidate sequence (XP_014677898.1) is annotated as a Cdk6-like and the blastp analysis supported it, we did not include the phylogenetic analysis because a part of the PKD was missing due to gaps of the contigs. The result with the PKD of the 1172 sequences as well as the ones of human representative Cdk members (CDK1, 4, 5, 6, 7, 8, 9, 11B, and 20; Malumbres, 2014) showed that all the Cdk4/6 candidate sequences formed one clade (Figure S3, Table S10). We used the 1172

sequences for further analyses.

Analysis of the residues corresponding to R24/R31 in Cdk4/6

In vertebrate Cdk4/6/21 from 572 species, the strong positive charge of the residues corresponding to R24/R31 was conserved (99.7%; 1248/1252; Figure S5, Table S11). In invertebrate the charge was conserved in 1048 sequences (89.4%), including 198 sequences (16.9%) in which the R was substituted with K (Tables 2, S9). The strong positive charge was lost in 124 sequences (10.6%); among invertebrate Deuterostomia, two *Ciona* species in Urochordata lacked the positive charge (2/31, 6.5%) (Figures 5, S4a). In contrast, in *Corella inflata* and *Ascidia aspersa*, which belong to the same order (Phlebobranchia) as *Ciona*, the R was substituted with K, thus, the charge was conserved (Shito & Hotta, personal communication). The strong positive charge was lost in 37.7% (43/114) of the sequences investigated in Lophotrochozoa (Table 2, Figure S4b, c), in 7.7% (75/969) in Ecdysozoa (Table 2, Figure S4d, e) and in 6.9% (4/58) in basal invertebrates (Table 2, Figure S4f). The corresponding residue was mostly T in 97.5% (39/40) of Platyhelminthes species (Lophotrochozoa) and in 62.1% (41/66) of the Nematoda species (Ecdysozoa). It is interesting that many species in the evolutionarily distant Platyhelminthes and Nematoda, both of which contain many parasites, had T instead of R24/R31, although the importance is unknown.

Relationships between the presence of Cdkn2 and the positive charge of the residue corresponding to R24/R31 in Cdk4/6 among phyla

While species in most of animal phyla have *Cdk4/6*, the extent of the evolutionary conservation of the residue corresponding to R24/R31 greatly differed among phyla.

We speculated that the presence of *Cdkn2* might affect the extent, we compared them in each phylum (Figure 1). In all analyzed species in phyla with *Cdkn2* gene(s), the strong positive charge of the residue corresponding to the R24/R31 of Cdk4/6 homologs was conserved, even in the Crinoid *A. japonica* where the corresponding negative charge of D84 in Cdkn2 was lost. In phyla without *Cdkn2*, high diversity was observed. The strong positive charge was lacking in a few species (Cnidaria, 7.1%; Arthropoda, 3.7%; and Urochordata, 18.2%), many species (Nematoda, 62.1%), or almost all species (Platyhelminthes, 97.5%).

Finally, we analyzed the statistical significance of correlation between the presence of *Cdkn2* and the conservation of a strong positive charge corresponding to R24/R31 in Cdk4/6. We analyzed nine phyla (Figure S6) in which at least five species were included in this study (Figure 1). The Spearman's rank coefficient was 0.816 (P = 0.00735), suggesting that the presence of *Cdkn2* and the conservation of the positive charge were strongly correlated with statistical significance.

Discussion

In this study, we comprehensively investigated the presence of *Cdkn2* and *Cdk4/6* in invertebrate genomes and the evolutionary conservation of the key charged residues; we also investigated the *Cdkn2* synteny. The distribution of *Cdkn2* was uneven among and within invertebrate phyla while most of the phyla contained *Cdk4/6* (Table 1). The lack of *Cdk4/6* in Dicyemida and Orthonectida may be caused by the genome size shrinkage (Lu et al., 2019; Slyusarev et al., 2020).

Evolution of Cdkn2 gene

On the basis of the data on the *Cdkn2* presence in invertebrates, we propose a scenario of *Cdkn2* evolution (Figure 6a). The ancestral *Cdkn2* gene first originated from an ankyrin repeat–encoding gene during the early evolution of Bilateria. During the diversification of Bilateria, this gene has been lost in some phyla, although to a different extent among Ecdysozoa, Lophotrochozoa, and Deuterostomia.

The uneven distribution could indicate that the loss of *Cdkn2* might be neutral or non-harmful in some cases, and its presence might be beneficial for survival or reproduction in other cases. Because Cdkn2s regulate the cell cycle and are known as tumor suppressors as well as cell senescence markers in mammals, their beneficial role in invertebrates might be related to cell proliferation, cell differentiation, cell aging, or prevention of cancer. However, it should be noted that even in deuterostome invertebrates, the mechanisms of cell aging seem to differ from those in mammals (reviewed in Amir et al., 2020). Species in the invertebrate phyla without *Cdkn2*, like those in Platyhelminthes, Nemertea, Bryozoa, and Urochordata as well as non-Bilateria, are known for their extreme regenerative ability. However, species in the invertebrate

phyla with *Cdkn2*, like those in Phoronida, Annelida, and Echinodermata, also have this ability (reviewed in Bely & Nyberg, 2010). Thus, *Cdkn2* is not another example of genes that are highly linked to extreme regeneration ability, such as the JmjC domain–encoding gene (Cao et al., 2019). It seems that longevity may not be related to the presence of *Cdkn2* in invertebrates. Even if we exclude non-Bilateria, which include species with extreme longevity or theoretical immortality (Brien, 1953; Piraino et al., 1996; Jochum et al., 2012), both a sea urchin, *Mesocentrotus* (*Strongylocentrotus*) *franciscanus*, which has a *Cdkn2* gene (Figure S1), and a clam, *Arctica islandica* (Bivalvia), without it (Table 1), are known for to their longevity, although the nuclear genome of *A. islandica* has not yet been published (Ebert & Southon, 2003; Scho□ne et al., 2005). Thus, we have currently no compelling clear-cut hypothesis to explain the uneven distribution of *Cdkn2*.

Strong evolutionary conservation of the microsynteny in the Cdkn2–Faf1 locus

Our synteny analysis revealed a strong linkage between *Cdkn2* and *Faf1*. On the basis of these data, we propose a scenario of the evolution of the linkage around the *Cdkn2* locus in invertebrates (Figure 6b). When the ancestral *Cdkn2* gene first appeared, it was linked to *Faf1*. In an ancestral chordate, the *Cdkn2* gene became linked also to *Mtap*. During the double duplication of the genome in an ancestral vertebrate, a *Cdkn2* copy (*Cdkn2c*) remained linked to *Faf1*, and another one (*Cdkn2b*) remained linked to *Mtap*. One of the four *Cdkn2* copies became *Cdkn2d*, and the last one, "*Cdkn2e*," was lost. Later, the duplication of *Cdkn2b* produced *Cdkn2a*. Human *Faf1*, *Mtap*, and *Cdkn2* are tumor suppressors, and the genetic linkage between *Cdkn2* and *Faf1* can be traced back to the *Cdkn2* origin (Jen et al., 1994; Kamb et al., 1994; Husemann et al., 1999;

Christopher et al., 2002; Menges et al., 2009; Gluick et al., 2013). Since *Faf1* is linked with *Kank1* encoding another ankyrin repeat–containing protein in Anthozoa of Cnidaria, *Cdkn2* in Bilateria and *Kank1* in Anthozoa might share the origin in ancient metazoans.

Why has the microsynteny in the *Cdkn2–Faf1* locus been conserved during bilaterian evolution? 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., 2012; Irimia et al., 2012; Simakov & Kawashima, 2017; Cai et al., 2010). Although both *Cdkn2*s and *Faf1* are tumor suppressor genes in mammals, they function in different cellular systems (Cdkn2s in the cell cycle and Faf1 in apoptosis), and their physical interaction has not been reported. Thus, *cis*-regulatory constraints between them might maintain the microsynteny. In this context, it is interesting that the expression of a lncRNA gene, which is on the immediate opposite side of the *C1orf185* gene from *Cdkn2c*, can affect *Faf1* expression in humans, although the linkage between *C1orf185* and the *Cdkn2c–Faf1* locus is not evolutionarily conserved even in vertebrates (Wan et al., 2014; Araki, unpublished data). A conservation of microsynteny might also reflect a hidden physical interaction (Moretti et al., 2017).

Potential interactions between Cdkn2 and Cdk4/6 evolution

Our comparison between the presence of *Cdkn2* and the differential conservation of the strong positive charge at the residue corresponding to R24/R31 in Cdk4/6 among phyla suggests potential interactions between *Cdkn2* and *Cdk4/6* evolution. All species in which we detected both *Cdkn2* and *Cdk4/6* in Annelida, Mollusca, and Echinodermata

retain the charge, while all investigated species in Platyhelminthes and Nematoda lack *Cdkn2*, and a high percentage of them also lack the positive charge, implying that the presence of *Cdkn2* might constrain the residue corresponding to R24/R31 as the interface from a substitution or loss when the presence of *Cdkn2* is beneficial, or its loss is harmful to the organism.

On the other hand, most of the species in Cnidaria, Arthropoda, and Urochordata retain the charge but lack *Cdkn2*. The Crinoid *A. japonica* lacks the negative charge in Cdkn2 at the residue corresponding to D84, but retains R24/R31 in Cdk4/6s. In this species, the Cdk4/6 sequence might lose the positive charge in the future. Alternatively, this positive charge in the Cdk4/6 homologs might be conserved to maintain the three-dimensional structure and/or interactions with proteins other than Cdkn2.

Cell cycle regulation without Cdkn2

What does substitute Cdkn2 in species without it? In mammalian cells, Cdkn1 preferentially inhibits Cdk1/2 but can also inhibit Cdk4/6, although the structure of the Cdk4/6–Cdkn1 complex has not yet been solved by X-ray crystallography (reviewed in Wood & Endicott, 2018, and Mu□hleder et al., 2021). It is easy to speculate that Cdkn1 substitutes for Cdkn2 in the species without *Cdkn2*, as reported for some invertebrates including *Ciona*, *D. melanogaster* and *C. elegans* (Kobayashi et al., 2022; reviewed in Kipreos & van den Heuvel, 2019). However, caution is needed in extrapolating this mechanism to other urochordates or species lacking *Cdkn2* in other phyla, because *Ciona* is the only urochordate investigated that lacks the positive charge at the residue corresponding to R24/R31 in Cdk4/6. Since the extent of the evolutionary conservation of the strong positive charge greatly differs among phyla lacking *Cdkn2*, the conserved

positive charge in Cdk4/6 of Cnidaria, Arthropoda, and Urochordata might not be involved in the interaction with Cdkn1 and might be necessary for the interactions with other proteins.

Conclusion

We believe that the results in this study not only widen our knowledge on the mechanisms of cell cycle regulation and their evolution in metazoans, but also provide fundamentals to design a cell immortalization protocol in invertebrates for basic sciences as well as drug discovery and clinical research.

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. The evolutionary position of Xenacoelomorpha is controversial (Cannon et al., 2016; Philippe et al., 2019); we tentatively categorized it as a diverged deuterostome.

Database searches

To analyze the presence of *Cdkn2* and *Cdk4/6* genes and to obtain the corresponding amino acid sequences, we used keyword and blastp searches in the NCBI Protein database (https://www.ncbi.nlm.nih.gov/protein/), Ensembl Metazoa (http://metazoa.ensembl.org/index.html), Ensembl Rapid Release (https://rapid.ensembl.org/index.html), Ghost Database: *Ciona intestinalis* genomic and cDNA resources (http://ghost.zool.kyoto-u.ac.jp/SearchGenomekh.html), Aniseed - Ascidian Network for In Situ Expression and Embryological Data (https://aniseed.fr/), PlanMine (https://planmine.mpinat.mpg.de/planmine/begin.do), and OIST Marine Genomics Unit Genome Browser (https://marinegenomics.oist.jp/gallery) (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 searches in the NCBI Whole-genome shotgun contigs database. To confirm the identity of the sequences obtained, we used blastp searches against vertebrate sequences. We also used vertebrate *Cdkn2* and *Cdk4/6/21* sequences as references and noticed that some vertebrate genes were misannotated 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 suggested that it is Cdkn2a; thus, we treated it as Cdkn2a, not Cdkn2b. As the paralogs Cdkn2a–Cdkn2d are specific to vertebrates, such misannotations do not affect this study.

In this analysis, we included 61 Cdkn2 sequences from 61 invertebrate species, and 1172 Cdk4/6 sequences from 1152 invertebrate species. As references, we used 1853 Cdkn2 sequences and 1252 Cdk4/6/21 sequences from 572 vertebrates. The database IDs of the analyzed sequences are listed in Tables S1 (invertebrate Cdkn2), S6 (vertebrate Cdkn2), S9 (invertebrate Cdk4/6), S11 (vertebrate Cdk4/6/21).

Multiple alignment of protein sequences

We used NCBI COBALT (Constraint-based Multiple Alignment Tool) to align protein sequences (Papadopoulos & Agarwala, 2007). Questionable alignments were corrected 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.

Phylogenetic tree construction

The phylogenetic trees (Figures S1d and S4) were constructed with NGPhylogeny.fr (https://ngphylogeny.fr/) (Edgar, 2004; Capella-Gutiérrez et al., 2009; Junier & Zdobnov, 2010; Katoh & Standley, 2013; Goloboff & Catalano, 2016; Lemoine et al., 2019). To construct the *Cdkn2* phylogenetic tree, we used Cdkn2 sequences without apparent truncation at the N– or C– terminal (47 of a total 61 invertebrate sequences [77.0%] and 4 human sequences; the 47 sequences were from all the phyla that we

judged *Cdkn2*-positive), sequences for 73 human ankyrin repeat–containing proteins and human CDKN1A sequence, and the following programs with default settings: MAFFT, trimAI, TNT, Newick Display. To construct the Cdk4/6 phylogenetic tree, we used the PKDs (PF00069) without apparent deletions (1172 out of total 1296 (90.4%) invertebrate sequences and 2 human sequences) and 7 human sequences of other representative Cdk family members, and the following programs with default settings: MUSCLE, TNT, Newick Display.

Synteny analysis

We analyzed protein−coding genes only in the synteny analysis. We used the NCBI gene database for the sequences with an NCBI gene ID (Brown et al., 2015). For long genomic sequences without ID, we analyzed the annotation, if available, or used a blastp search against the flanking sequences. If the surrounding genes have been annotated, we used blastp against vertebrates to confirm the identities. We checked GeneCards (https://www.genecards.org) for the abbreviated gene names.

Statistical analysis

We used the statistical software EZR to determine Spearman's rank correlation coefficient and create the scatter plot (Kanda, 2013). To determine the Spearman's rank correlation coefficient, we used the percentage of the number of species with both Cdkn2 and Cdk4/6 to that with Cdk4/6, and the percentage of the number of species with the conserved strong positive charge at Cdk4/6 R24/31.

Acknowledgements

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.

Capella-Gutiérrez, S., Silla-Martinez, J. M. & Gabaldòn, T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*, 1972-1973. https://doi.org/10.1093/bioinformatics/btp348.

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.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*, 1792–1797. https://doi.org/10.1093/nar/gkh340.

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.

Goloboff, P. A. & Catalano, S. A. (2016) TNT version 1.5, including a full implementation of phylogenetic morphometrics. *Cladistics*, *32*, 221-238. https://doi.org/10.1111/cla.12160.

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. & Mu□ller, 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.

Junier, T. & Zdobnov, E. M. (2010). The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics*, 26, 1669-1670.

https://doi.org/10.1093/bioinformatics/btq243.

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.

Kanda, Y. (2013). Investigation of the freely available easy-to-use software 'EZR' for medical statistics. *Bone Marrow Transplantation*, *48*, 452–458. https://doi.org/10.1038/bmt.2012.244.

Katoh, K. & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, *30*, 772–780. https://doi.org/10.1093/molbev/mst010.

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.

Kipreos, E. T. & van den Heuvel, S. (2019). Developmental Control of the Cell Cycle: Insights from *Caenorhabditis elegans*. *Genetics*, *211*, 797-829. https://doi.org/10.1534/genetics.118.301643.

Kita, A., Saito, Y., Miura, N., Miyajima, M., Yamamoto, S., Sato, T., ... Chikenji, T. S. (2022). Altered regulation of mesenchymal cell senescence in adipose tissue promotes pathological changes associated with diabetic wound healing. *Communications Biology*, *5*, 310. https://doi.org/10.1038/s42003-022-03266-3.

Kobayashi, K., Tokuoka, M., Sato, H., Ariyoshi, M., Kawahara, S., Fujiwara, S., ... Satou, Y. (2022). Regulators specifying cell fate activate cell cycle regulator genes to determine cell numbers in ascidian larval tissues. *Development*, *149*, dev201218.

https://doi.org/10.1242/dev.201218.

Lemoine, F., Correia, D., Lefort, V, Doppelt-Azeroual, O., Mareuil, F., Cohen-Boulakia, S. & Gascuel, O. (2019). NGPhylogeny.fr: new generation phylogenetic services for non-specialists. *Nucleic acids research*, *47*, W260-W265.

Liu, J., Min, L., Zhu, S., Guo, Q., Li, H., Zhang, Z., ... Zhang, S. (2019). Cyclin-Dependent Kinase Inhibitor 3 Promoted Cell Proliferation by Driving Cell Cycle from G1 to S Phase in Esophageal Squamous Cell Carcinoma. *Journal of Cancer*, *10*, 1915-1922. 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.

Malumbres, M. (2014). Cyclin-dependent kinases. *Genome Biology*, *15*, 122. https://doi.org/10.1186/gb4184.

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.

Moretti, A. I. S., Pavanelli, J.C., Nolasco, P., Leisegang, M. S., Tanaka, L. Y., Fernandes, C.G., ... Laurindo, F. R. M. (2017). Conserved Gene Microsynteny Unveils Functional Interaction Between Protein Disulfide Isomerase and Rho Guanine-Dissociation Inhibitor Families. *Scientific Reports*, *7*, 17262. https://doi.org/10.1038/s41598-017-16947-5.

Mühleder, S., Fernández-Chacón, M., Garcia-Gonzalez, I. & Benedito, R. (2021). Endothelial sprouting, proliferation, or senescence: tipping the balance from physiology to pathology. *Cellular and Molecular Life Sciences*, *78*, 1329-1354. https://doi.org/10.1007/s00018-020-03664-y.

Papadopoulos, J. S, & Agarwala, R. (2007). COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics*, *23*, 1073-1079.

Patterson, K. I., Brummer, T., O'Brien, P. M. & Daly, R. J. (2009). Dual-specificity phosphatases: critical regulators with diverse cellular targets. *Biochemical Journal*, *418*, 475-489. https://doi.org/10.1042/bj20082234.

Philippe, H., Poustka, A. J., Chiodin, M., Hoff, K. J., Dessimoz, C., Tomiczek, B., ... Telford, M. J. (2019). Mitigating Anticipated Effects of Systematic Errors Supports Sister-Group Relationship between Xenacoelomorpha and Ambulacraria. *Current 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.

Scho□ne, 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.

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.

Shinzato, C., Takeuchi, T., Yoshioka, Y., Tada, I., Kanda, M., Broussard, C., ... Inoue, M. (2021). Whole-Genome Sequencing Highlights Conservative Genomic Strategies of a Stress-Tolerant, Long-Lived Scleractinian Coral, *Porites australiensis* Vaughan, 1918. *Genome Biology and Evolution*, *13*, evab270. https://doi.org/10.1093/gbe/evab270.

Shiomi, K., Kiyono, T., Okamura, K., Uezumi, M., Goto, Y., Yasumoto, S., ... Hashimoto, N. (2011). CDK4 and cyclin D1 allow human myogenic cells to recapture growth property without compromising differentiation potential. *Gene Therapy*, *18*, 857-866. https://doi.org/10.1038/gt.2011.44.

Simakov, O., Kawashima, T., Marlétaz, F., Jenkins, J., Koyanagi, R., Mitros, T., ... Gerhart, J. (2015). Hemichordate genomes and deuterostome origins. *Nature*, *527*, 459-465. https://doi.org/10.1038/nature16150.

Simakov, O. & Kawashima, T. (2017). Independent evolution of genomic characters during major metazoan transitions. *Developmental Biology*, *427*, 179-192. https://doi.org/10.1016/j.ydbio.2016.11.012.

Slyusarev, G. S., Starunov, V. V., Bondarenko, A. S., Zorina, N. A. & Bondarenko, N. I. (2020). Extreme Genome and Nervous System Streamlining in the Invertebrate 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. *Nucleic Acids Research*, *50*, D996–D1003. 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.

Figure legends

Figure 1 Distribution of *Cdkn2*, *Cdk4/6* and *Cdkn1* and conservation of the key charged amino acids in all invertebrate phyla. The phylogenetic tree was modified from Telford et al. (2015). The rows for phyla in which at least five species were analyzed are in yellow. * We did not include the Priapulida candidate sequence (XP_014677898.1) for the phylogenetic analysis because of the lack of a part of the PKD due to the gaps of the contig (see the main text). ND, not determined because of the lack of genome data. NA, not applicable.

Figure 2 Amino acid alignment around the residue corresponding to Cdkn2 D84 in representative invertebrate species from the seven phyla where *Cdkn2* was detected. The residues corresponding to Cdkn2 D84 are in green if negatively charged or in magenta if not. Human CDKN2s are aligned as references. **a** and **b** represent the amino acids of α-helices and a β-turn in CDKN2A, respectively (Baumgartner et al., 1998).

Figure 3 The microsynteny analysis revealed that *Cdkn2* and *Faf1* are tightly linked during evolution. Arrows show the direction of transcription. The chromosome (Chr) number or the linkage group (LG) is shown when available. Microsynteny around human *CDKN2* loci is shown as a reference. Genes encoding ankyrin repeat–containing proteins other than Cdkn2 are in red.

Figure 4 Microsynteny around the *Faf1* locus in representative invertebrates without *Cdkn2*. Arrows show the direction of transcription. The chromosome (Chr) number or

the linkage group (LG) is shown when available. A gene encoding an ankyrin repeat–containing protein other than Cdkn2 is in red.

Figure 5 Amino acid alignment around the residue corresponding to Cdk4/6 R24/31 in representative invertebrate Deuterostomia species. The residues corresponding to Cdk4/6 R24/31 are green if positively charged or in magenta if not. Human CDK4 and CDK6 are aligned as references.

Figure 6 Hypothesis on the evolution of the *Cdkn2* locus. (a) Loss of *Cdkn2* occurred multiple times. (b) Evolution of microsynteny at the *Cdkn2* locus. *Cdkn2e* is a hypothetical transient gene.

Supplementary materials

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

Table S2 Results of blast searches for invertebrate *Cdkn2*. (a) blastp on refseq_protein, (b) tblastn on wgs database.

Table S3 The list of the invertebrate Cdkn2 candidates and the human homologs of other ankyrin repeat–containing sequences as well as human CDKN1A as an outgroup for the phylogenetic analysis in FASTA format.

Table S4 The list of invertebrate genes possibly misannotated as *Cdkn2*.

Table S5 The list of the identified *Cdkn1*s (a) in invertebrates other than Arthropoda and (b) in Arthropoda.

Table S6 The list of the analyzed vertebrate *Cdkn2*s.

Table S7 Microsynteny around the *Cdkn2* locus in invertebrates.

Table S8 Microsynteny around the *Faf1* locus in (a) invertebrates other than Arthropoda and (b) Arthropoda.

Table S9 The list of the analyzed invertebrate *Cdk4/6*s. (a) deuterostome invertebrates, (b) Platyhelminthes, (c) other Lophotrochozoa, (d) Nematoda, (e) other Ecdysozoa, and (f) basal invertebrates.

Table S10 The list of the analyzed vertebrate *Cdk4/6*s.

Table S11 The list of the invertebrate Cdk4/6 candidate sequences and human homolog sequences of other representative Cdk members in FASTA format.

Figure S1 Amino acid alignment of invertebrate Cdkn2s. (a) N-terminal half of Cdkn2. (b) C-terminal half. The residue corresponding to D84 is in green if negatively charged, or in magenta if not. (c) Consensus residues in invertebrate Cdkn2. **a** and **b** represent

amino acids of α-helices and β-turns in *Homo sapiens* CDKN2s, respectively (Baumgartner et al., 1998). (d) Phylogenetic analysis of invertebrate Cdkn2 candidate sequences.

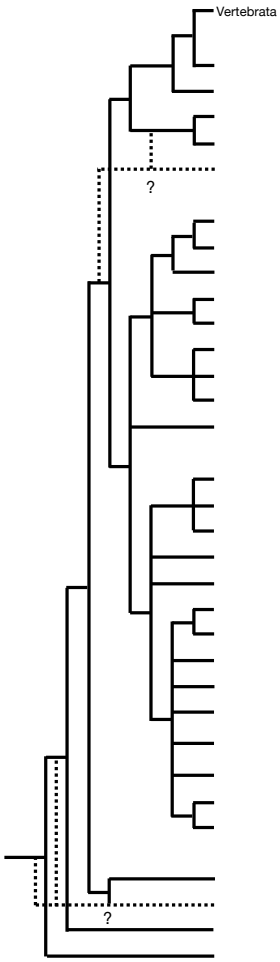
Figure S2 Cdkn2 amino acid alignment around the residue corresponding to D84 in vertebrates. Cdkn2a in (a) mammals and (b) non-mammalian vertebrates; Cdkn2b in (c) mammals, (d) non-mammalian/non-fish vertebrates, and (e) fish; Cdkn2c in (f) mammals, (g) non-mammalian/non-fish vertebrates, and (h) fish; Cdkn2d in (i) mammals and (j) non-mammalian vertebrates. The residue corresponding to Cdkn2 D84 is in green if negatively charged or in magenta, if not.

Figure S3 Phylogenetic analysis of invertebrate Cdk4/6 candidate sequences and human representative Cdk members (CDK1, 4, 5, 6, 7, 8, 9, 11B, and 20).

Figure S4 Cdk4/6 amino acid alignment around the residue corresponding to R24/R31 in invertebrates. (a) deuterostome invertebrates, (b) Nematoda, (c) other Ecdysozoa, (d) Platyhelminthes, (e) other Lophotrochozoa, and (f) basal invertebrates. The residue corresponding to R24/31 in Cdk4/6 is in green if positively charged and in magenta, if not.

Figure S5 Cdk4/6/21 amino acid alignment around the residue corresponding to R24/R31 in vertebrates. Cdk4 in (a) mammals, (b) non-mammalian/non-fish vertebrates, and (c) fish; Cdk6 in (d) mammals, (e) non-mammalian/non-fish vertebrates, and (f) fish; Cdk21 in (g) fish. The residue corresponding to Cdk4/6 R24/31 is in green if the strong positive charge is conserved and in magenta, if not.

Figure S6 Determination of the Spearman's rank correlation coefficient. (a) The data used and (b) the scatter plot.



	Cdkn2	Cdkn2 D84	*Cdk4/6*	percentage of the number of species without R or K at Cdk4/6 R24/31	*Cdkn1*
Deuterostome invertebrates					
Urochordata	not detected	NA	detected	18.2%	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	NA	detected	0.0%	detected
Ecdysozoa					
Arthropoda	not detected	NA	detected	3.7%	detected
Onychophora	detected	conserved	detected	0.0%	detected
Tardigrada	not detected	NA	detected	50.0%	detected
Nematomorpha	not detected	NA	detected	0.0%	detected
Nematoda	not detected	NA	detected	62.1%	detected
Priapulida	not detected	NA	(detected)*	NA	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	NA	detected	28.6%	detected
Dicyemida	not detected	NA	not detected	NA	not detected
Orthonectida	not detected	NA	not detected	NA	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	NA	detected	100%	detected
Brachiopoda/Phoronida	detected	conserved except in *P. ovalis*	detected	0%	detected
Bryozoa	not detected	NA	detected	0%	detected
Platyhelminthes	not detected	NA	detected	97.5%	detected only in Cestoda
Gastrotricha	ND	ND	ND	ND	ND
Basal invertebrates					
Cnidaria	not detected	NA	detected	7.1%	detected
Ctenophora	not detected	NA	detected	0.0%	detected
Placozoa	not detected	NA	detected	0.0%	detected
Porifera	not detected	NA	not detected	NA	detected

Figure 1 Distribution of *Cdkn2*, *Cdk4/6* and *Cdkn1* and conservation of the key charged amino acids in all invertebrate phyla. The phylogenetic tree was modified from Telford et al. (2015). The rows for phyla in which at least five species were analyzed are in yellow. * We did not include the Priapulida candidate sequence (XP_014677898.1) for the phylogenetic analysis because of the lack of a part of the PKD due to the gaps of the contig (see the main text). ND, not determined because of the lack of genome data. NA, not applicable.

```
                                                                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 the seven phyla where *Cdkn2* was detected. The residues corresponding to Cdkn2 D84 are in green if negatively charged or in magenta if not. Human CDKN2s are aligned as references. **a** and **b** represent the amino acids of α-helices and a β-turn in CDKN2A, respectively (Baumgartner et al., 1998).

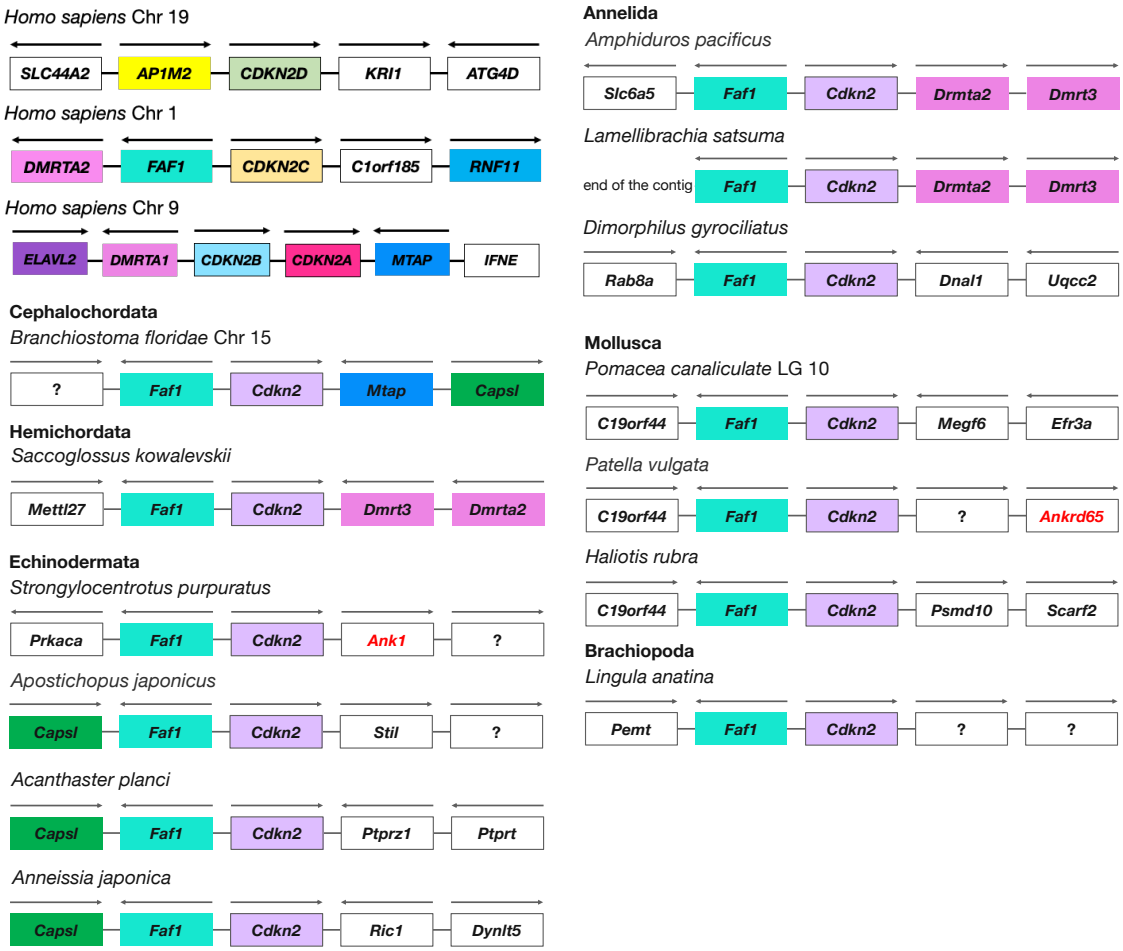


Figure 3 The microsynteny analysis revealed that *Cdkn2* and *Faf1* are tightly linked during evolution. Arrows show the direction of transcription. The chromosome (Chr) number or the linkage group (LG) is shown when available. Microsynteny around human *CDKN2* loci is shown as a reference. Genes encoding ankyrin repeat–containing proteins other than Cdkn2 are in red.

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

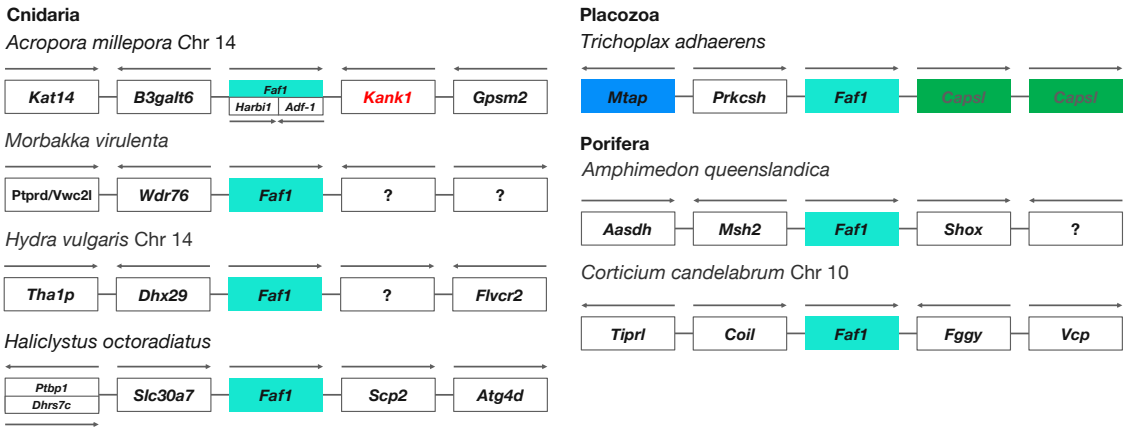


Figure 4 Microsynteny around the *Faf1* locus in representative invertebrates without *Cdkn2*. Arrows show the direction of transcription. The chromosome (Chr) number or the linkage group (LG) is shown when available. A gene encoding an ankyrin repeat–containing protein other than Cdkn2 is in red.

```
CDK4
Homo sapiens                    RYEPVAEIGVGAYGTVYKARDPHSG---HFVALKSVRVPN
CDK6
Homo sapiens                    QYECVAEIGEGAYGKVFKARDLKNG--GRFVALKRVRVQT

Cdk4/6
Urochordata
Ciona intestinalis              RYHTLSEIGVGAYGVVFKAQDMQSDDENKFVAIKCVRVEN
Ciona savignyi                  RYHTLTEIGVGAYGVVFKAQDRHSEEENKFVAMKCIRVEN
Corella inflata                 RYKPVEKIGVGAYGVVFKAKDLRKD--GEFVAMKCLRVEN
Halocynthia roretzi             NYEMVAEIGVGAYGIVYRARDLEND--NRYVALKCLRVQN
Molgula occulta                 KYEKDEEIGVGAYGTVYRGKDLANL--GKYVALKCLRVQT
Botrylloides leachii            NYEMVAEIGVGAYGVVYRARDLEND--RKLVALKCLRVQN
Oikopleura dioica               DWIDMEEIGGGAYGVVFRARDRQHD--NRVVALKRLRVQA
Cephalochordata
Branchiostoma belcheri          EYEKVAKIGTGAYGTVYKARDLKNG--GEFVALKRVRVQT
Hemichordata
Saccoglossus kowalevskii        KYEEVAEIGTGAYGTVYKARDLETG---RYVALKKVRVQT
Echinodermata
Strongylocentrotus purpuratus EYVQAAEIGSGAYGIVYKARDTETG---HFVALKSVRIPI
Acanthaster planci              NYAVISEIGTGAYGTVYKARDLETG---KVVALKKVRLHT
Anneissia japonica              DFQKIEKIGEGTYGVVYKARDKISG---EVVALKKIRLDT
Xenacoelomorpha
Hofstenia miamia                HYSVLEKVGVGAYGEVYRAKDPRTG---GHVALKRVRVFD
```

Figure 5 Amino acid alignment around the residue corresponding to Cdk4/6 R24/31 in representative invertebrate Deuterostomia species. The residues corresponding to Cdk4/6 R24/31 are green if positively charged or in magenta if not. Human CDK4 and CDK6 are aligned as references.

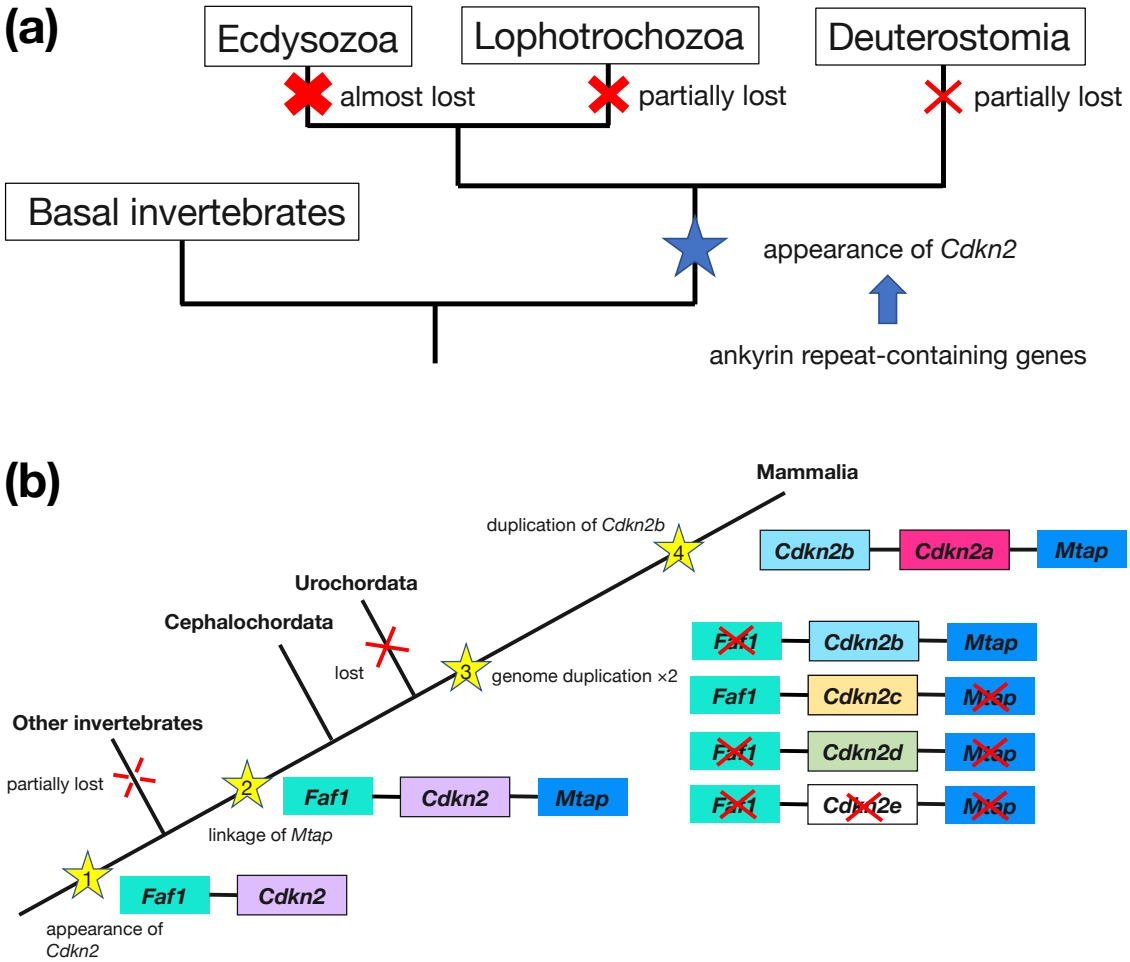


Figure 6 Hypothesis on the evolution of the *Cdkn2* locus. (a) Loss of *Cdkn2* occurred multiple times. (b) Evolution of microsynteny at the *Cdkn2* locus. *Cdkn2e* is a hypothetical transient gene.

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	3	4
Hemichordata	2	1	2
Echinodermata	21	21	21
Echinozoa (incl. Holothuroidea)	10	7	10
Asterozoa (incl. Ophiuroidea)	9	6	9
Pelmatozoa	2	1	2
Onychophora	2	1	2
Annelida	9	7	6
Polychaeta (incl. Pogonophora)	7	6	5
Sipuncula	1	0	1
Clitellata	1	1	0
Mollusca	62	54	22
Bivalvia	29	29	0
Gastropoda	30	23	21
Caenogastropoda	11	5	10
Heterobranchia	6	6	0
Neomphalina	2	1	1
Patellogastropoda	4	4	3
Vetigastropoda	7	7	7
Cephalopoda	2	2	1
Coleoidea	2	2	0
Nautiloidea	1	1	1
Brachiopoda/Phoronida	4	2	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	**31**	**26**	**3**									**2**					**6.5%**
Urochordata	11	7	2									2					18.2%
Cephalochordata	3	3															0.0%
Hemichordata	1	1															0.0%
Echinodermata	15	15															0.0%
Xenacoelomorpha	1		1														0.0%
Ecdysozoa	**969**	**710**	**184**	**2**	**1**	**1**	**2**	**2**	**52**	**1**	**5**	**1**	**2**			**6**	**7.7%**
Arthropoda	898	693	172	2	1			2	17		4		1			6	3.7%
Onychophora	1	1															0.0%
Tardigrada	2	1				1											50.0%
Nematomorpha	2	2															0.0%
Nematoda	66	13	12				2		35	1	1	1	1				62.1%
Lophotrochozoa	**114**	**63**	**8**			**1**	**5**	**3**	**31**	**1**				**1**	**1**		**37.7%**
Rotifera	7	5					1							1			28.6%
Annelida	7	6	1														0.0%
Mollusca	55	49	6														0.0%
Nemertea	2						2										100.0%
Brachiopoda /Phoronida	2	2															0.0%
Bryozoa	1	1															0.0%
Platyhelminthes	40		1			1	2	3	31	1					1		97.5%
Basal Invertebrates	**58**	**51**	**3**			**1**	**1**								**2**		**6.9%**
Cnidaria	56	49	3			1	1								2		7.1%
Ctenophora	1	1															0.0%
Placozoa	1	1															0.0%