

Molecular evolutionary and structural analysis of the cytosolic DNA sensor cGAS and STING

Xiaomei Wu^{1,*}, Fei-Hua Wu^{1,2,†}, Xiaoqiang Wang^{3,†}, Lilin Wang¹, James N. Siedow², Weiguo Zhang⁴ and Zhen-Ming Pei^{1,2,*}

¹College of Life and Environmental Sciences, Hangzhou Normal University, Hangzhou 310036, China, ²Department of Biology, Duke University, Durham, NC 27708, USA, ³Department of Biological Sciences, University of North Texas, TX 76203, USA and ⁴Department of Immunology, Duke University, Durham, NC 27710, USA

Received January 29, 2014; Revised June 11, 2014; Accepted June 12, 2014

ABSTRACT

Cyclic GMP-AMP (cGAMP) synthase (cGAS) is recently identified as a cytosolic DNA sensor and generates a non-canonical cGAMP that contains G(2',5')pA and A(3',5')pG phosphodiester linkages. cGAMP activates STING which triggers innate immune responses in mammals. However, the evolutionary functions and origins of cGAS and STING remain largely elusive. Here, we carried out comprehensive evolutionary analyses of the cGAS-STING pathway. Phylogenetic analysis of cGAS and STING families showed that their origins could be traced back to a choanoflagellate *Monosiga brevicollis*. Modern cGAS and STING may have acquired structural features, including zinc-ribbon domain and critical amino acid residues for DNA binding in cGAS as well as carboxy terminal tail domain for transducing signals in STING, only recently in vertebrates. In invertebrates, cGAS homologs may not act as DNA sensors. Both proteins cooperate extensively, have similar evolutionary characteristics, and thus may have co-evolved during metazoan evolution. cGAS homologs and a prokaryotic dinucleotide cyclase for canonical cGAMP share conserved secondary structures and catalytic residues. Therefore, non-mammalian cGAS may function as a nucleotidyl-transferase and could produce cGAMP and other cyclic dinucleotides. Taken together, assembling signaling components of the cGAS-STING pathway onto the eukaryotic evolutionary map illuminates the functions and origins of this innate immune pathway.

INTRODUCTION

Innate immune sensing of microbial infections represents a crucial element for host defense. Utilizing germ-line encoded pattern recognition receptors (PRRs), innate immunity examines extracellular, endosomal and cytosolic compartments for signs of infection and triggers type I interferon (IFN) induction and other proinflammatory cytokines when the pathway is activated. DNA has been known to stimulate immune responses for more than a century, long before it was shown to be the genetic material. Cytosolic DNA of pathogenic bacterial or viral origin, or leaking from the nucleus or mitochondria following cell stress can be sensed by eukaryotic cells as a danger signal or a sign of foreign invasion (1). The accumulation of self-DNA can also produce severe autoimmune diseases, such as systemic lupus erythematosus. Over the past several years, many PRRs for cytosolic DNA have been identified, including DNA-dependent activator of IFN-regulatory factors (DAI) (2), RNA polymerase III (3,4), DEAD box polypeptide 41 (DDX41) (5), absent in melanoma 2 (AIM2) (6–8) and IFN-inducible protein 16 (IFI16) (9). The presence of these multiple PRRs may reflect their functioning in a distinct cell-type- or DNA-sequence-specific manner (10), and no consensus has emerged until recently.

Detection of cytosolic DNA activates a stimulator of IFN genes (STING, also known as MITA, MPYS, ERIS or TMEM173), an endoplasmic reticulum (ER) translocon-associated transmembrane protein (11–14). STING in turn initiates a cascade of known events by first recruiting and activating the cytosolic kinases, I κ B kinase (IKK) and TANK-binding kinase 1 (TBK1), which phosphorylate and activate the transcription factors nuclear factor κ B (NF- κ B) and IFN regulatory factor 3 (IRF3), respectively. NF- κ B and IRF3 then enter the nucleus and function together to induce IFNs and other cytokines, and thereby trigger the host immune response (1). STING is a central player in the innate immune response to cytosolic nucleic acids

*To whom correspondence should be addressed. Tel: +1 919 613 8152; Fax: +1 919 660 7293; Email: zpei@duke.edu
Correspondence may also be addressed to Xiaomei Wu. Tel: +86 0571 28868542; Fax: +86 0571 28865333; Email: wuxm07@gmail.com

†The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.

(15). STING could also act as a direct PRR for cyclic dinucleotides, such as cyclic diguanylate monophosphate (c-di-GMP) and cyclic diadenylate monophosphate (c-di-AMP), which are conserved signaling molecules produced by bacteria that regulate bacterial motility and biofilm formation (16).

Recently, cyclic GMP-AMP (cGAMP) synthase (cGAS, also known as C6orf150 and MAB21D1) was reported in *Homo sapiens* (human) and *Mus musculus* (mouse) as a general (with broad specificity) cytosolic DNA sensor for activating type I IFN signaling pathway (17,18). cGAS binds DNA and catalyzes the synthesis of cGAMP from adenosine triphosphate (ATP) and guanosine triphosphate (GTP) in the presence of DNA. cGAMP, an endogenous second messenger structurally similar to c-di-GMP and c-di-AMP, binds and activates STING in the cytoplasm, illuminating how STING can stimulate type I IFN pathway in response to both cytosolic DNA and cyclic dinucleotides (18). Furthermore, cGAMP binds STING with an affinity of ~ 10 nM, which is significantly stronger than that of c-di-GMP, and induces a 'closed' conformation of STING that is important to activate the downstream of type I IFN (19–21).

The details of this cGAS-STING pathway were uncovered by a series of structural, biophysical and biochemical studies (19,21–25). Crystal structures of the nucleotidyltransferase domain of cGAS (23,26–27) established how cGAS functions as a DNA-sensing enzyme in a sequence-independent manner. cGAS interacts with the sugar-phosphate backbone along the minor groove of DNA by employing a positively charged surface as well as a zinc-ribbon domain insertion. A cGAS-DNA complex, harboring one cGAS molecule and one DNA molecule, was focused on in these studies. In contrast, a 2:2 complex that contains dimeric cGAS bound to two DNA molecules, was found most recently (24,25). Both of the two DNA binding surfaces and the dimer interface are critical to DNA binding. More interestingly, the endogenous cGAMP generated by cGAS contains a phosphodiester linkage between the 2'-OH of GMP and the 5'-phosphate of AMP and another between the 3'-OH of AMP and the 5'-phosphate of GMP. This specific isomer of cGAMP with 2'-5', 3'-5' linkages is termed 2'3'-cGAMP, and is distinguished from conventional cGAMP (with 3'-5', 3'-5' linkages and termed 3'3'-cGAMP) and other cyclic dinucleotides (such as c-di-AMP and c-di-GMP) of microbial origin. Both mouse (R231, with an Arg residue at site 231) and human (R232) STING proteins can be stimulated by 2'3'-cGAMP, 3'3'-cGAMP and bacterial 3'3'-c-di-GMP (20,21), but human STING with the H232 allele is only responsive to 2'3'-cGAMP (20–22).

Moreover, the Chen lab has provided genetic evidence that cGAS as well as STING are essential for the induction of type I IFN stimulated by foreign DNA (28). The cGAS-knockout mouse is strikingly similar to goldenticket mouse (loss of function of STING), both of which are susceptible to infection by DNA viruses (11,28). cGAS homologs are present in several vertebrate species and have structures similar to human or murine cGAS (17,26). In fish and pig, STING proteins also act as a mediator for activating different IFN genes (29–31). These results indicate that cGAS-STING signaling may be the major and non-

redundant method of DNA recognition in the innate immune system in mammals and even vertebrates.

Given the importance of cGAS-STING signaling in mammals, it is necessary to explore the evolutionary origins of this pathway. In this study, we aimed to present a comprehensive molecular evolutionary analysis of cGAS and STING proteins by applying a systematic homolog search on all eukaryotic genomes that have been fully sequenced. We identified the origins of mouse cGAS and STING in the choanoflagellate *Monosiga brevicollis*, the closest relative of metazoans. During metazoan evolution, both cGAS and STING were lost in nematodes and flatworms. We also utilized several methods to gain novel evolutionary insights into cGAS and STING structure–function relationships: (i) examining the evolutionary pattern of domain organization of cGAS and STING; (ii) mapping reported functionally critical residues on cGAS and STING through multiple sequence alignments across representative species and (iii) structural modeling of STING proteins from species other than human and obtaining a structural basis for understanding their binding with ligands. Furthermore, we presented the evolutionary distribution of the cGAS-STING signaling pathway in representative eukaryotic organisms, allowing us to explore the evolutionary history of this pathway.

MATERIALS AND METHODS

Eukaryotic species

A list of 190 fully sequenced eukaryotic species was derived from the Database of KEGG Organisms (updated on May 20, 2013) (http://www.genome.jp/kegg/catalog/org_list.html) (32). Chromosome, protein and mRNA sequences of these eukaryotic species were downloaded from National Center for Biotechnology Information (NCBI) (release 59) (<ftp://ftp.ncbi.nih.gov/refseq/release/>). The genome of Ctenophore *Mnemiopsis leidyi* that was evolved in early metazoans was fully sequenced recently (33). The genome data of *M. leidyi* was downloaded from the Mnemiopsis Genome Project Portal (<http://research.nhgri.nih.gov/mnemiopsis/>). The genome data of African clawed frog *Xenopus laevis* (JGI v6.0) was derived from Xenbase (<ftp://ftp.xenbase.org/pub/Genomics/JGI/Xenla6.0/>).

Identification of cGAS and STING homologs

Two rounds of searches on the protein and genomic sequences were carried out to detect putative cGAS and STING homologs in fully sequenced eukaryotic species. The illustration of procedures for the cGAS homolog searches is displayed in Supplementary Figure S1A. (i) In the first round of search based on the protein sequences, cGAS homologs were identified using a PSI-BLAST (34) search followed by reverse BLASTP. Mouse cGAS proteins were initially searched against the eukaryotic proteome data set via PSI-BLAST v2.2.26. We set the *E-value* threshold (-e) as 0.001, the *E-value* threshold for inclusion in the multipass model (-h) as 0.002 (default value), and the maximum running iterations (-j) as 5. Then, the hits satisfying

the thresholds were reversely aligned against the mouse proteome via BLASTP. The putative homologs were identified if the reverse BLASTP best hit was mouse cGAS. (ii) Considering that some cGAS homologs could not be found due to errors in genome annotation or the presence of some genes that have not been annotated, we carried out the second round of search that is based on the genomic sequences. For the species without cGAS homologs detected in the first round, all the candidate cGAS homologs were aligned against the assembled genomic sequences using TBLASTN (E -value ≤ 0.001 and coverage on query sequence $\geq 45\%$). The identified gene region in a target genome was extended by 1 kb on both sides, and then input to Fgenesh+ (HMM plus similar protein-based gene prediction) (35) to predict the gene structure and protein sequence. (iii) Finally, 52 cGAS homologs in 45 eukaryotic species were combined from the two rounds of searches that are based on protein and genomic sequences. A total of 48 putative STING homologs from 45 species were obtained in the same way (Supplementary Figure S1B). Detailed information on cGAS and STING homologs are listed in Supplementary Table S1 and their sequences in FASTA format are available in Supplementary Data S1.

Sequence analysis

The cGAS and STING homologs in platypus *Ornithorhynchus anatinus* are partial sequences. The cGAS homolog in western lowland gorilla lacks ~44% of the C-terminal fragment compared with the human sequence. To check whether there is any error in sequence annotation that leads to the incompleteness of the three proteins, we obtained the genomic regions of the three proteins and input them into Fgenesh+ by extending 3 kb on both sides. The predicted protein sequences are very similar to the original ones and are still incomplete. We did not include the three partial proteins in sequence alignments and construction of phylogenetic trees of cGAS and STING gene families. STING is known as an ER membrane protein in mammalian cells, containing four N-terminal transmembrane domains (TMs). We used HMMTOP v2.0 (36) to identify the TMs of the putative homologs. HMMTOP results were then checked manually and compared with human TMs. Four STING homologs in *Ixodes scapularis*, *Tribolium castaneum*, *Apis mellifera* and *Brachistostoma floridae* lacked TM at the N terminus. Considering that there may be some errors in sequence annotation of the four proteins, we obtained their genomic regions by extending 3 kb on both sides and then predicted gene structure and protein sequences using Fgenesh+. The predicted *B. floridae* protein sequence has two N-terminal TMs detected using HMMTOP and another software Phobius (37). Thus, the information of *B. floridae* STING homolog was updated in Supplementary Table S1 and Supplementary Data S1. The other three predicted sequences still lack TM. BLASTP result shows that at least 76% coverage on the three TM-lacking sequences in *I. scapularis*, *T. castaneum* and *A. mellifera* could be aligned with mouse STING protein with E -value $\leq 7e-10$ (Supplementary Table S2), which indicates that the three proteins do have similar sequences with mouse STING.

CTT domains of STING homologs were determined by two steps. Each STING homolog was first aligned with the human CTT region using BLASTP (E -value ≤ 0.01). Then, if the aligned hit on the STING homolog sequence had the two conserved residues, Ser 366 and Leu 374 in human STING which are important for IRF3 activation (38), the STING homolog was considered as containing a CTT domain. The multiple sequence alignment of the human CTT region on eukaryotic STING homologs was shown in Supplementary Figure S2. Amino acid sequences were aligned using PROMALS3D (39) and colored according to BLOSUM62 score in Jalview 2.8 (40). The secondary structure of human cGAS (residues from 164 to 513) was derived from the crystal structure (PDB ID: 4KM5). The secondary structure elements of all other cGAS homologs, *Vibrio cholerae* DncV and OAS1 proteins were identified using Jpred 3 (41).

Phylogenetic analysis

A species tree (Figure 1) was constructed for visualization of the distribution of cGAS and STING homologs across 61 metazoan species and the choanoflagellate *M. brevicollis*. The plant *Arabidopsis thaliana* and fungus *S. cerevisiae* were used as outgroup sequences to root the tree. First, 27 highly conserved proteins (Supplementary Table S3) were selected from the published list (42,43). Each set of homologs which were identified via best bi-directional BLASTP search is present in *A. thaliana*, *S. cerevisiae* and at least 58 (95% in 62) metazoan and *M. brevicollis* proteomes. Second, separate multiple sequence alignments for each set of homologs were built using MUSCLE v3.8.31 (44), with the maximum number of iterations set to 100. PhyML v3.1 (45) was used to derive the maximum likelihood (ML) trees with 100 bootstrap replicates by applying the JTT matrix (parameters set as -d aa -b 100 -m JTT -f e -v e -a e -quiet). Note that 100 bootstrapped gene trees were then constructed for each set of homologs. Each gene tree was rerooted with *A. thaliana* using the nw_reroot tool from the Newick Utilities package v1.6 (46). Third, bootstrapped trees for all genes were combined in one file and input to phybase R package v1.3 (47). From each set of bootstrap replicates, one STAR tree (48) was estimated (47). Based on the multispecies coalescent model, STAR uses average ranks of coalescences to estimate species trees from a set of rooted gene trees and constructs an NJ tree that is a consistent estimate of the species tree topology (48). STAR cannot estimate branch lengths. Finally, a consensus tree was constructed from the 100 bootstrapped STAR trees using the consense program in the phylip package v3.69 (49). Tree visualization was carried out in iTOL v2.1 (50) and then manually labeled and colored for clarity. This species tree is in accordance with trees published before (43,51–52) and NCBI taxonomy with the exception in the branch of mammals. *M. musculus* and *Rattus norvegicus* should be closer evolutionarily to *Macaca mulatta* than *Sus scrofa* and *Bos Taurus*.

The gene tree for depicting the phylogenetic relationship of cGAS gene family (Figure 2A) was constructed using PhyML based on the multiple sequence alignment of cGAS sequences via MUSCLE. JTT matrix and 100 bootstrap replicates were applied. The ML tree of STING gene family

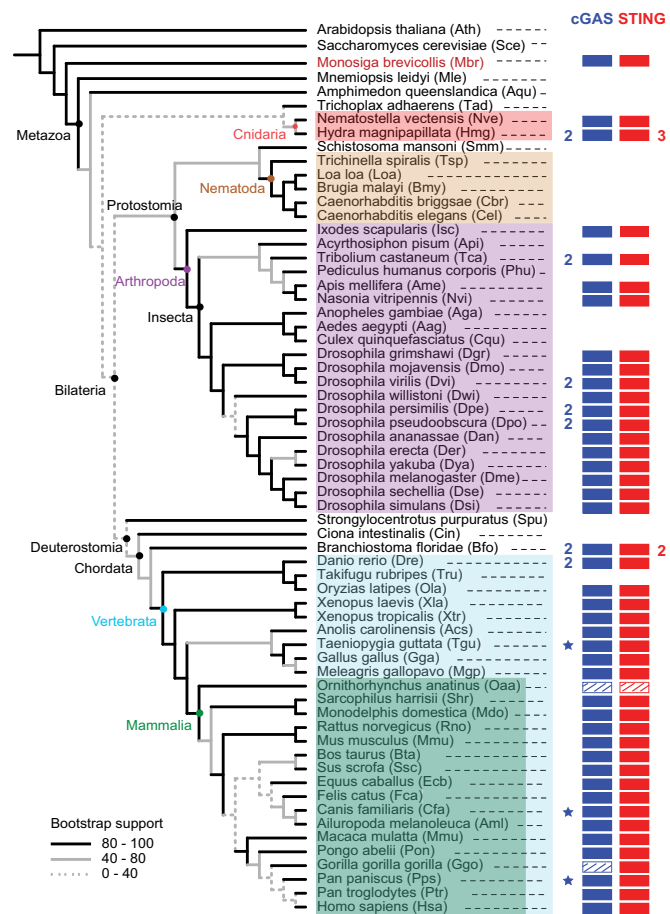


Figure 1. Distribution of cGAS and STING homologs across a choanoflagellate and 61 metazoan species. A species tree is shown in this figure. Branch lengths are not intended to be to scale. Plant *Arabidopsis thaliana* and fungus *Saccharomyces cerevisiae* were used as outgroups to root the species tree. The species phylogenies were inferred based on the summary statistics of coalescence times for 27 multilocus data sets. Support values were derived from 100 bootstrap replicates. Taxonomic branches are labeled in different colors. Pink, Cnidaria; brown, Nematoda; purple, Arthropoda; cyan, Vertebrata; green, Mammalia. Each leaf node is denoted as a standard species name followed by its three-letter abbreviation in brackets. cGAS and STING homologs are labeled with blue and red rectangles, respectively. The rectangles with oblique lines indicate that the genes only have partial sequences and so are excluded from the construction of gene trees in Figure 2. The species harboring multiple copies of cGAS and STING homologs are marked with the number of copies in blue and red, respectively. Sequences of three vertebrate cGAS homologs (marked with blue asterisks) have no insertion of zinc-ribbon DNA-binding domain (see Supplementary Text S1 for inspection of the three proteins). The sequence references, alignments and trees in this figure can be obtained from Supplementary Data S1.

(Figure 2B) was obtained in the same way. The phylogenetic tree showing the relationship of cGAS homologs, OAS1 homologs and *V. cholerae* DncV (Supplementary Figure S3) was built using PhyML based on the multiple sequence alignment of these protein sequences via Promals3D (39). All protein sequences, alignments and trees are available in Supplementary Data S1.

Calculation of Ka and Ks

One homolog pair was aligned at the protein sequence level using MUSCLE (44), and the codon multiple alignments of mRNA sequences were created from the protein alignments using PAL2NAL v14.0 (53). Ka and Ks values were then calculated with yn00 according to Yang and Nielsen (54), implemented in PAML v4.7 (55).

Molecular modeling of STING proteins

The comparative modeling program MODELLER (56) was used to generate models for STING homologs from other non-mammalian species. The structures of human STING (PDB ID: 4LOH) were used as templates in the structural modeling experiments. The 3D models of STING from other species were obtained by optimally satisfying spatial restraints derived from the sequence alignment based on the CLUSTALX results and 3D structures. The dimer structure models were assembled by superpositioning the two monomer models onto the human STING dimer structure architecture bound with cGAMP. All structural models were analyzed and minor manual adjustments of the modeling solution were made using a graphics program COOT (57). Figures were prepared with PyMOL (The PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC).

RESULTS

Presence of cGAS and STING in choanoflagellates and metazoans but not in nematodes

To elucidate the comprehensive evolutionary history of cGAS and STING proteins, we determined the distribution of these two proteins across 191 fully sequenced eukaryotic species. We used two rounds of searches, which are respectively based on the levels of protein and genomic sequences, to detect homologs of mouse cGAS and STING (see Materials and Methods, Supplementary Figure S1). There are 52 cGAS and 48 STING homologs present in 44 metazoan species and a unicellular eukaryotic organism, choanoflagellate *M. brevicollis* (Supplementary Table S1). *M. brevicollis* is considered as the closest living relative of animals. Except for the choanoflagellate homologs of cGAS and STING, we did not find any homolog of cGAS and STING in the other branches (fungi, plants or protists) of eukaryotes.

Figure 1 shows the visualization of the distribution of cGAS (in blue rectangles) and STING (in red rectangles) homologs on the 'life tree' of 61 metazoans and their close protozoan relatives. The species tree showing the phylogenetic relationship of metazoan species was inferred from 100 bootstrapped STAR trees (48) based on ML gene trees of 27 universal genes (see Materials and Methods). Within metazoans, homologs of both cGAS and STING are present as early as in cnidarians (sea anemone *Nematostella vectensis* and *Hydra magnipapillata*). Both proteins are distributed in all *Drosophila* species, several non-*Drosophila* arthropods and nearly all chordates except for torafugu *Takifugu rubripes*; however, they were lost together in the flatworm *Schistosoma mansoni* and nematodes. Interestingly, the fully

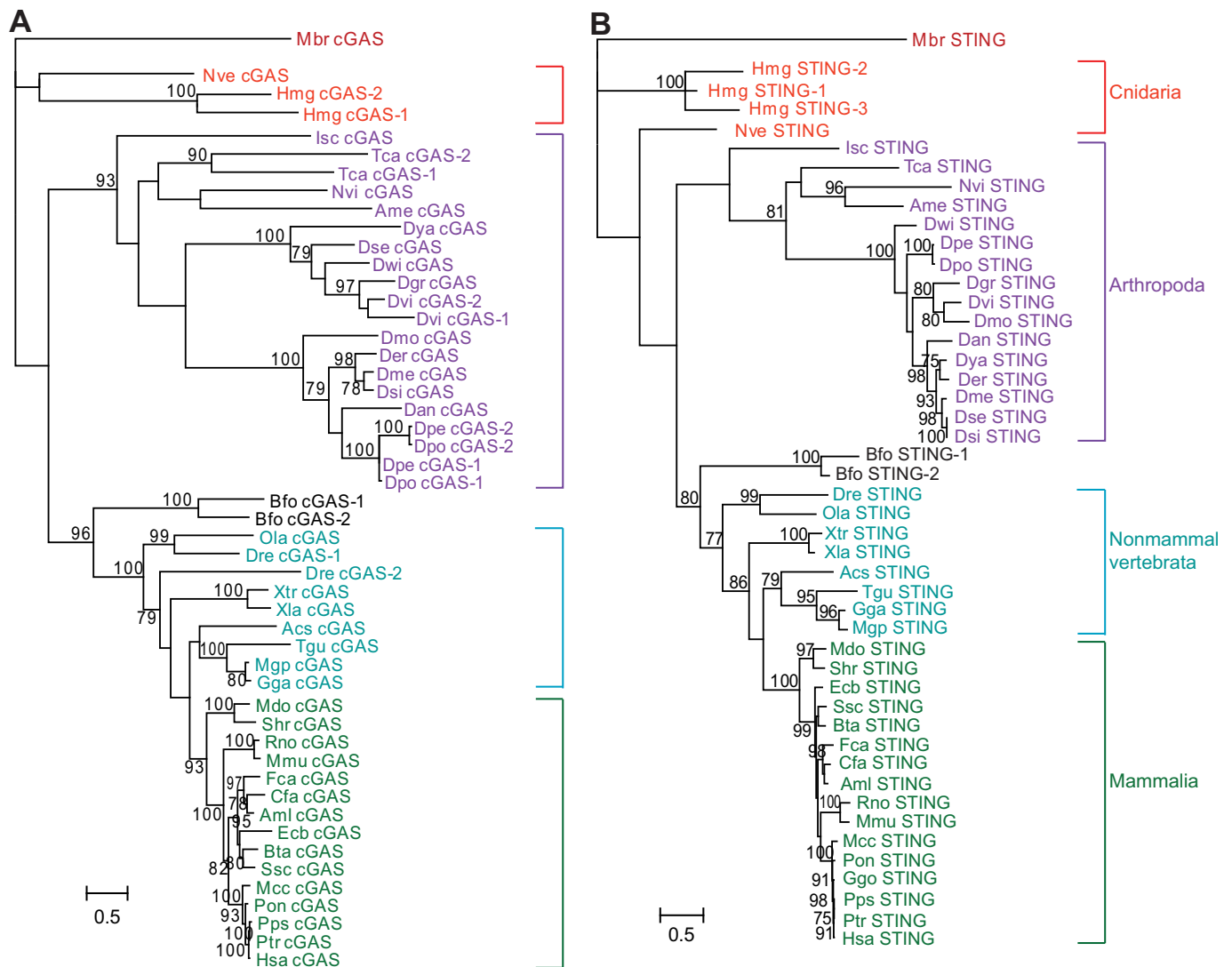


Figure 2. ML phylogenetic trees showing the relationships of cGAS (A) or STING (B) homologs. Three proteins with partial sequences, two cGAS homologs in platypus *Ornithorhynchus anatinus* and *Gorilla gorilla* and one STING homolog in *O. anatinus*, were not included in the construction of gene trees (see Materials and Methods). Bootstrap values equal to or larger than 75 are marked beside each node. Branch lengths indicate the number of amino acids substitution per site. The leaf nodes are colored according to the color scheme in Figure 1. Each leaf node is depicted as a three-letter abbreviation for species name, followed by cGAS or STING. The correspondence between the abbreviations and the standard species names can be found in Figure 1 and Supplementary Data S1. The sequence references, alignments and trees in this figure are in Supplementary Data S1, and accession numbers of proteins are noted in Supplementary Table S1.

sequenced species either contain homologs of both cGAS and STING or have lost both.

To explore phylogenetic relationships of cGAS or STING homologs, we built a ML gene tree for each family (Figure 2). All cGAS-containing species have one cGAS homolog except for seven species (labeled with copy numbers in blue in Figure 1). *H. magnipapillata*, *T. castaneum*, *Drosophila virilis*, *Drosophila persimilis*, *Drosophila pseudoobscura*, cephalochordate *B. floridae* (Florida lancelet) and *Danio rerio* (zebrafish) each of which has two cGAS homologs. The phylogenetic tree of the cGAS gene family (Figure 2A) shows that one round of cGAS homolog duplication may have occurred before the divergence of *D. persimilis* and *D. pseudoobscura*, and the cGAS homolog might have duplicated once in the other five genomes. It is

the same for the STING family. *H. magnipapillata* harbors three STING candidates and *B. floridae* has two STING candidates. Figure 2B shows that STING homologs also have experienced species-specific gene duplication.

The Ka/Ks ratio is an important index of functional constraints. Ka refers to the number of non-synonymous substitutions per non-synonymous site, while Ks represents the number of synonymous substitutions per synonymous site. The smaller the Ka/Ks ratio is, the stronger the functional constraints are (58). We listed Ka/Ks ratios of different proteins, cGAS, STING and TBK1 (a downstream protein activated by STING in IFN-pathway induction) within four metazoan lineages, namely, mammals (human and mouse), fish (*D. rerio* and *Oryzias latipes*), insects (*Drosophila grimshawi* and *Drosophila willistoni*, *Drosophila*

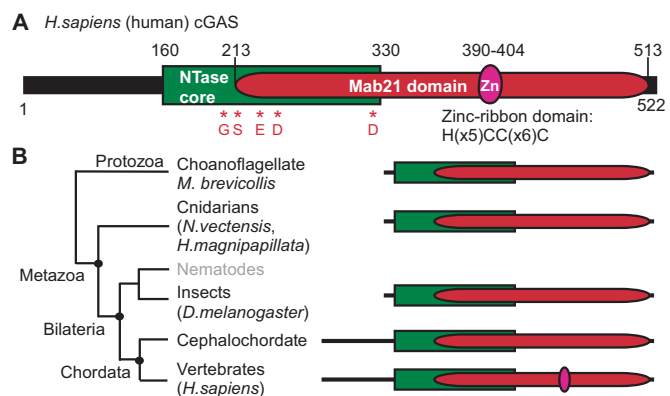


Figure 3. Evolution of functional domains in cGAS proteins. (A) A diagram of the domain organization of human cGAS. Five red asterisks indicate key catalytic residues (G212, S213, E225, D227 and D319) within the NTase fold of human cGAS. (B) Diagrams of the domain organization of cGAS homologs based on phylogeny of metazoans and *M. brevicollis*. The major metazoan groups followed by the representative species in brackets include cnidarians, such as sea anemone *N. vectensis* and *H. magnipapillata*, insects, such as *D. melanogaster* (fruit fly), cephalochordate *B. floridae* and vertebrates. The node of nematoda is in gray indicating that cGAS was lost within nematodes.

melanogaster and *Drosophila simulans*) and cnidarians (*H. magnipapillata* and *N. vectensis*) (see Table 1, Materials and Methods). The K_a/K_s ratios of all protein pairs are much smaller than 1.0, suggesting that these proteins are subject to purifying selections (functional constraints). cGAS and STING homologs show higher K_a/K_s ratios than TBK1, which reveals that TBK1 has undergone strong purifying selection, while the functional constraints on cGAS and STING, both of which are upstream proteins in the IFN induction pathway, may have been relaxed to some extent.

Evolutionary pattern of domain organization and critical residues of cGAS

Human cGAS is composed of an unstructured and poorly conserved N terminus (amino acid residues 1–160) and a highly conserved C terminus (160–513) (Figure 3A). The C-terminal fragment contains two highly conserved domains, nucleotidyltransferase (NTase) core domain (160–330) and Mab21 domain (213–513). In the NTase core domain, there are several conserved residues associated with active sites within the NTase superfamily: hG[G/S]X_{9–13}[D/E]h[D/E]h...h[D/E]h (h indicates a hydrophobic amino acid). The hG[G/S] pattern has a crucial role in docking substrates within active sites and the three conserved aspartate/glutamate residues are involved in coordination of divalent ions and activation of acceptor hydroxyl groups on the substrate (59). Inserted into the Mab21 region is a zinc-ribbon structural domain (390–405) typically defined as H(X₅)CC(X₆)C. The Mab21 domain was first identified in the nematode *Caenorhabditis elegans* Mab-21, a cell fate-determining gene (60). Human and mouse Mab-21-like proteins are homologous with the nematode Mab-21 (61,62). Although the mouse cGAS is also a Mab21-containing protein, the eukaryotic homologs of mouse cGAS identified in this study are more similar

to mouse cGAS than mouse Mab-21-like proteins (Supplementary Table S4).

Human cGAS shows significant structural similarity to the human oligoadenylate synthetase1 (OAS1) which polymerizes ATP into linear 2'-5'-linkage oligoadenylate upon stimulation by dsRNA. The structural similarity is especially striking in the catalytic domain fold for complex formation of cGAS with dsDNA and bound ligands (23,26–27,63–64). Furthermore, the first enzyme reported to synthesize cGAMP is a bacterial dinucleotide cyclase (DncV) in *V. cholerae*, which has no obvious primary sequence homology to human cGAS (17,65). Although these three kinds of enzymes, cGAS homologs, OAS1 homologs and DncV can be classified clearly according to the phylogenetic analysis (Supplementary Figure S3), they all belong to the nucleotidyltransferase superfamily and share similar patterns of secondary structural elements (Supplementary Figure S4).

To gain novel insights into the evolutionary perspective of the relationships between structure and function in cGAS, we used two complementary means: (i) placing the domain organization of cGAS homologs on the evolutionary map of metazoans and *M. brevicollis* (Figure 3B) and (ii) mapping functionally critical residues of cGAS reported in recent publications across these species (Figure 4, Supplementary Table S5). NTase core domain and Mab21 domain are conserved in these species, but *M. brevicollis* and invertebrate cGAS homologs do not have a zinc-ribbon domain at the C terminus (Figures 3B and 4). This zinc-ribbon domain is conserved among all vertebrate cGAS members except for the three homologs in *Pan paniscus* (bonobo), *Canis familiaris* (dog) and *Taeniopygia guttata* (zebra finch). Lack of a zinc-ribbon insertion in the three vertebrates may result from the different versions of genome assembly or genome annotation (Supplementary Text S1, Supplementary Figure S5 and Supplementary Table S6). The vertebrate-specific zinc-ribbon domain of cGAS is not in OAS1, nematode Mab-21 and mammalian Mab-21-like proteins (Figure 4). Furthermore, cGAS homologs across vertebrates as well as cephalochordate *B. floridae* have N-terminal fragments with an average length of 167 amino acids except for chicken *Gallus gallus* and turkey *Meleagris gallopavo*. In contrast, cnidarian and insect cGAS homologs contain a very short N-terminal fragment, ~70 amino acids in *N. vectensis* and fewer than 7 amino acids in the other species (Figures 3B and 4). Similarly, OAS1, nematode Mab-21 and mammalian Mab-21-like proteins also contain very short N-terminal tails (Figure 4). This could indicate that the ~167-amino-acid-long N-terminal tail has evolved in chordate/vertebrate lineage and seems to be a cGAS-specific adaptation.

We next looked at the conservation of key residues within the NTase fold of human cGAS (G212, S213, E225, D227 and D319) across OAS1, Mab-21-like proteins, nematode Mab-21 and the prokaryotic protein DncV in *V. cholerae* (Figure 4). *V. cholerae* DncV synthesizes conventional 3'/3'-cGAMP involved in bacterial chemotaxis and colonization while human cGAS synthesizes the specific 2'/3'-cGAMP involved in sensing dsDNA. In addition, early metazoan cGAS and DncV show very similar patterns of secondary structural elements (Supplementary Figure S4). Thus, we

Table 1. Ka, Ks and sequence divergence of protein homologs between two species

Metazoan branch	Species	Gene	Ks ^a	Ka ^b	Ka/Ks	Divergence ^c
Mammalians	<i>H.sapiens</i> and <i>M.musculus</i>	cGAS	0.63	0.28	0.45	0.44
		STING	0.57	0.21	0.36	0.31
		TBK1	0.60	0.03	0.06	0.06
Fishes	<i>D.rerio</i> ^e and <i>O.latipes</i>	cGAS-1	4.36	0.54	0.12	0.48
		cGAS-2	2.47	0.73	0.30	0.65
		STING	2.09	0.49	0.23	0.58
		TBK1	3.63	0.02	0.01	0.04
Insects	<i>D.grimshawi</i> and <i>D.willistoni</i>	cGAS	2.71	0.50	0.18	0.47
		STING	3.09	0.34	0.11	0.42
		TBK1	2.14	0.06	0.03	0.09
Insects	<i>D.melanogaster</i> and <i>D.simulans</i>	cGAS	0.14	0.06	0.46	0.13
		STING	0.11	0.04	0.36	0.09
		TBK1	0.15	0.02	0.11	0.04
Cnidarians ^d	<i>H.magnipapillata</i> ^e and <i>N.vectensis</i>	cGAS-1	3.96	0.88	0.22	0.73
		cGAS-2	3.99	0.87	0.22	0.76
		STING-1	3.66	0.64	0.17	0.61
		STING-2	4.01	0.74	0.18	0.69
		STING-3	4.00	1.11	0.28	0.72

^aKs, the number of synonymous substitutions per synonymous site.

^bKa, the number of non-synonymous substitutions per non-synonymous site.

^cSequence divergence, one minus identity of BLASTP alignment between two amino acid sequences.

^dTBK1 is absent from *H. magnipapillata* and so is not in the last group of this table.

^e*D. rerio* has two cGAS homologs (cGAS-1 and -2); *H. magnipapillata* has two cGAS homologs (cGAS-1 and -2) and three STING homologs (STING-1, -2 and -3).

wondered whether DncV and human cGAS share conserved NTase catalytical sites or not. Expectedly, the five aforementioned NTase active sites residues are completely conserved not only in cGAS homologs but also in DncV. OAS1 proteins share all five conserved NTase catalytic residues although they contain Asp in place of E225. Conversely, only one of the five residues, E225 is conserved in mammalian Mab-21-like proteins and nematode Mab-21, although another residue, D227, is conservatively replaced with a glutamate.

It has been recently reported that cGAS in human and mouse interact with DNA through two binding sites, forming a complex composed of dimeric cGAS bound to two DNA molecules. Both of the two DNA binding surfaces and the dimer interface play a critical role in DNA binding (24,25). We checked the conservation of five functionally important amino acid residues involved in 2'3'-cGAMP binding, three positively charged residues on the primary DNA binding surface and seven critical residues on the second DNA binding surface and dimer interface across eukaryotic cGAS homologs (Supplementary Table S5). Except that S434 of human cGAS is not conserved in mammals while Y436 is conserved in all species, the other three residues (K362, R376 and S378) involved in 2'3'-cGAMP binding are completely conserved in a cephalochordate and

in vertebrates but not in arthropods. Two lysine residues, K384 and K411, on the primary DNA binding surface are completely conserved in vertebrates but not in early-branching species, and the other one, K407, is highly conserved across all species except for cnidarians with Arg as the corresponding residue. Three critical residues, K347, K394 and K398, on the dimer interface are only conserved in vertebrates. Two residues, R236 and K254, on the second DNA binding surface are conserved in vertebrates except for amphibians (*Xenopus tropicalis* and *X. laevis*), but not in early-branching species, while the other two amino acids (K327 and R353) are not conserved residues. It seems reasonable that the above four residues (R236, K254, K327 and R353) are not conserved in vertebrates, because only double (R236E/K254E and K254E/K327E) and triple (R236E/K254E/K327E) mutations abrogated the ability of cGAS to stimulate IFN production (24), and the positively charged R353 of human cGAS is replaced by the other positively charged amino acid lysine in vertebrates. Strictly speaking, five functionally critical residues are not conserved in vertebrates while two residues are highly conserved in all species; the other eight residues are completely conserved in vertebrates but not in early-branching species.

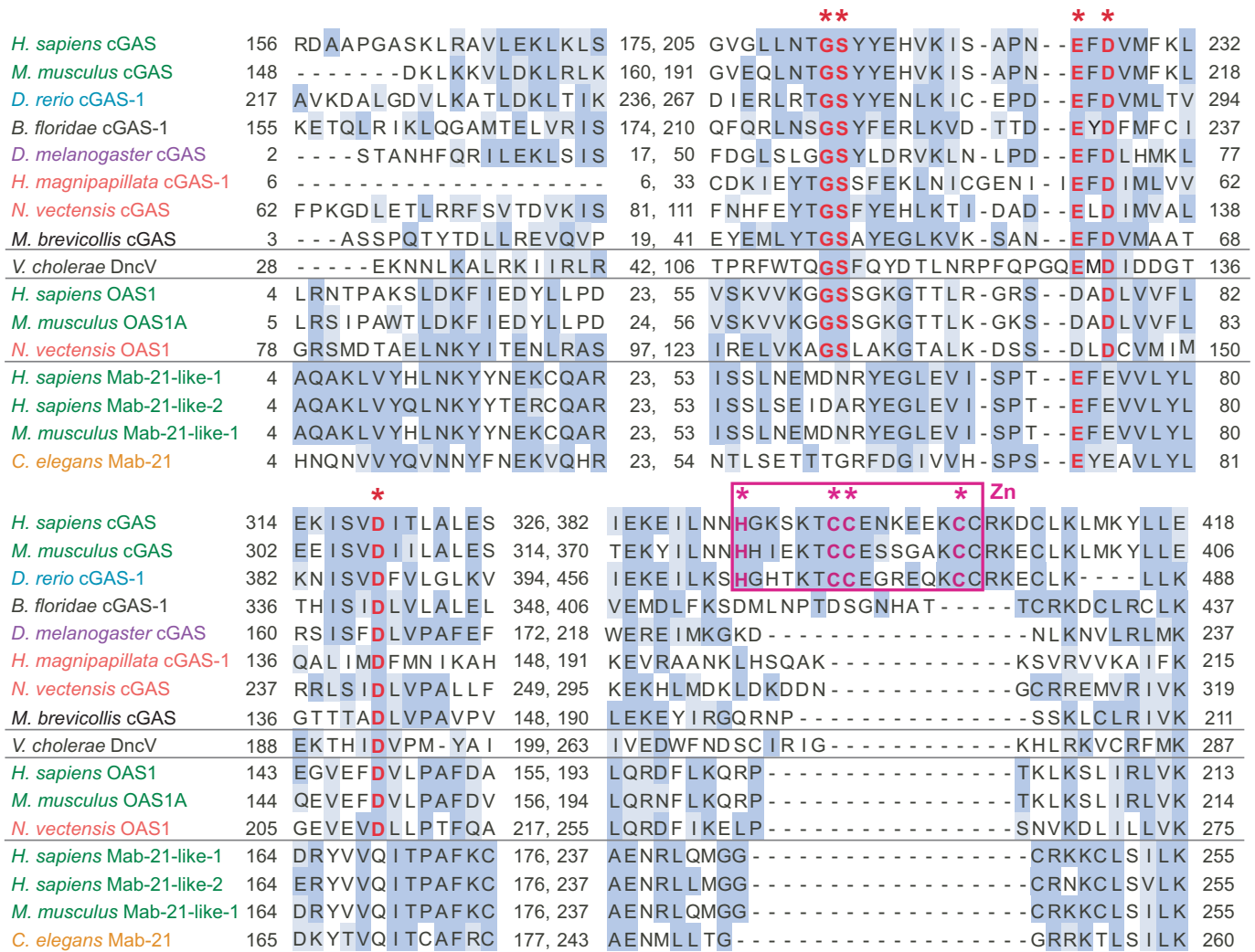


Figure 4. Multiple sequence alignment of DncV and the representative sequences of cGAS homologs, OAS1 and Mab-21-like proteins. DncV is a prokaryotic enzyme named dinucleotide cyclase in *Vibrio cholerae*. The listed species include mammals *H. sapiens* and *M. musculus*, zebrafish *D. rerio*, cephalochordate *B. floridae*, insect *D. melanogaster*, nematode *Caenorhabditis elegans* and cnidarians. If an organism harbors multiple cGAS homologs, only one homolog is listed. The Mab21 domain was first identified in the nematode cell fate-determining gene Mab-21 (60). Two human proteins (Mab-21-like 1 and Mab-21-like 2) and two mouse proteins (Mab-21-like 1 and Mab-21-like 2) are homologous with the nematode Mab-21 (61,62).

Evolutionary pattern of structural features of STING

Human STING consists of four N-terminal TMs (amino acid residues 21–136), a central c-di-GMP-binding domain (CBD, 153–340) and a C-terminal tail (CTT, 340–379) (Figure 5A). The CBD containing a dimerization domain (DD, 155–180), protrudes into the cytoplasm (15). The crystal structures based on the ~240-amino-acid-long globular carboxy-terminal domain (CBD+CTT or CTD, 138–379) have been reported to mediate binding to the bacterial second messenger c-di-GMP (66–70) and recently to 2'3'-cGAMP produced in mammalian cells (20,21). The CTT domain is important for STING to transduce signals (38,67).

Conserved structural domains in STING homologs are important for their proper biological roles. Detection of novel STING homologs enabled the detailed analysis of the evolutionary history of these conserved domains (Fig-

ure 5B). Human and mouse STING proteins reside exclusively on the ER membrane (14). Generally, four putative TMs (in red rectangles in Figure 5B) exist in protozoan *M. brevicollis* and metazoan STING homologs. Fewer than four TMs (in pink rectangles with dashed borders) are in arthropods except for jewel wasp *Nasonia vitripennis*, birds and the cephalochordate *B. floridae* (Figure 5B). Another three arthropod STING homologs in *I. scapularis*, *T. castaneum* and *A. mellifera* lack TM at the N terminus based on the current genome data (see Materials and Methods). All STING homologs have the conserved CBD and DD domains. However, the CTT is only observed in vertebrates. Overall, the modern STING proteins might have recently gained their structural domains during the early evolution of vertebrates, although the CTT domain is missing from STING homologs in amphibians *X. tropicalis* and *X. laevis*.

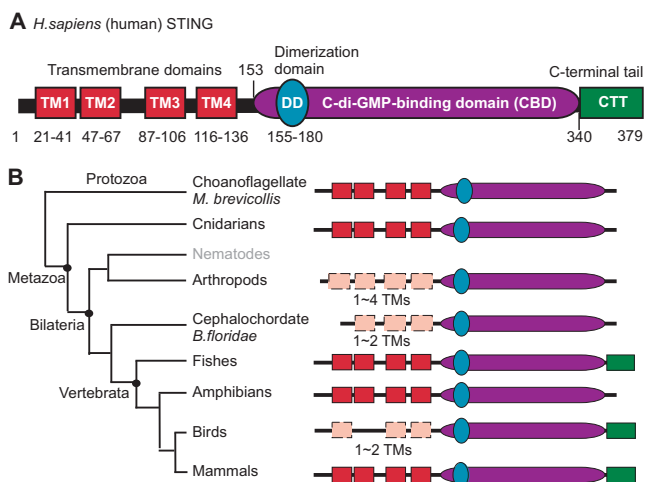


Figure 5. Evolution of functional domains in STING proteins. (A) A diagram of the domain organization of human STING. (B) Evolutionary pattern of the domain organization of STING homologs based on the phylogeny of metazoans and *M. brevicollis*. The node of nematoda is in gray indicating the absence of STING proteins. The TM and CTT domains were identified as described in the Materials and Methods.

As mentioned previously, replacing R232 with histidine in human STING was reported to affect its sensitivity particularly toward 2'-5'-linkage-containing cGAMP isomers. Recent structural studies found that residue 232 in human STING is part of the β sheet lid over the binding pocket of the STING-cGAMP complex, and the arginine residue interacts with the α -phosphate groups of cGAMP (20,21). The R232 allele is highly conserved in STING homologs except for *O. latipes* (Japanese medaka fish, Met 228), *H. magnipapillata* (Thr 230) and *M. brevicollis* (Ile 234) (Table 2). Furthermore, substitutions of several amino acid residues within the binding pocket of STING with Ala indicate that Y167, R238, Y240, N242, E260 and T263 are involved in the recognition of cGAMP isomers (20,21). Y167 and R238 are completely conserved in all STING homologs. E260 and T263 are highly conserved. The corresponding amino acid of E260 in mammal *S. scrofa* (pig) is a glycine, and serine replaces T263 in chicken *G. gallus*, turkey *M. gallopavo* and *M. brevicollis*. The other two sites Y240 and N242 are only fully conserved in mammals.

To initially obtain a structural basis for understanding the binding of STING homologs from the species other than human with 2'-3'-cGAMP, we generated homology models of STING homologs from three species, Japanese medaka *O. latipes*, chicken *G. gallus* and choanoflagellate *M. brevicollis*, according to the structure of human STING (PDB ID: 4LOH) (see Materials and Methods). Similar to the human STING structure reported recently (20), other STING homologs can also form a dimer (Figure 6A, Supplementary Figure S6A and C) and probably exhibit a closed conformation with a 2'-3'-cGAMP bound. In *O. latipes* STING (Figure 6B), Met 228 substitutes for Arg/His 232 of human STING. M228 is on the surface of molecule and in the entrance of the ligand binding pocket. M228 makes the pocket slightly more hydrophobic and would slightly affect ligand binding affinity (possibly a bit

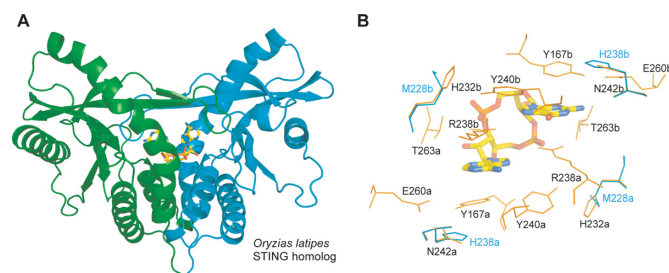


Figure 6. Structural modeling of STING from Japanese medaka *Oryzias latipes* binding with 2'-3'-cGAMP. (A) Modeled dimer structure of STING homolog from *O. latipes*. One monomer is shown in green and the other in cyan. 2'-3'-cGAMP is shown as thick bond model. (B) Comparison of 2'-3'-cGAMP binding pockets of human STING (orange) and *O. latipes* (cyan). Structural modeling of another two non-mammalian STING homologs in chicken *G. gallus* and choanoflagellate *M. brevicollis*, and their binding with 2'-3'-cGAMP are displayed in Supplementary Figure S6.

weaker) in *O. latipes* STING. M228 could also play a role during the open to closed conformational transition. A substitution by Ile 234 in *M. brevicollis* STING (Supplementary Figure S6D) would be similar to M228 in *O. latipes* STING, but Thr in *H. magnipapillata* makes the pocket and entrance slightly polar. His 238 in *O. latipes* STING (Figure 6B) and His 247 in *G. gallus* STING (Supplementary Figure S6B) are located on the corresponding position of N242 in human STING. Similar to N242 in human STING, these histidines can form a hydrogen bond with Y161 or Y172 (Y167 in human STING) which forms a stacking interaction with guanine or adenine ring of cGAMP. The corresponding residue Ile in *D. willistoni* would probably have weak interaction with the corresponding Tyr. In *G. gallus* and *M. brevicollis* STING homologs, Ser 268 and Ser 267 substitutes for T263 of human STING. However, these serines can also form a hydrogen bond with amino group on guanine ring of cGAMP, but are less restricted in ligand binding compared with T263 in human STING. In *M. brevicollis* STING, Met 242 replaces Y240 in human STING and provides similar hydrophobic environment for cGAMP binding. Our results showed that even though several key residues in human STING have mutations in STING homologs from other species, the STING from species other than human have structures similar to the human STING. They are still expected to exhibit functional levels of binding with 2'-3'-cGAMP, although the binding of 2'-3'-cGAMP with non-mammalian STING is expected to probably be slightly weaker than to human STING. We also tried STING structural modeling using 3'-3'-cGAMP instead of 2'-3'-cGAMP and found that the difference in binding between the two ligands is minor (data not shown).

Evolution of proteins in cGAS-STING signaling pathway

In infected cells, cGAS senses cytosolic dsDNA from diverse microbes and self-DNA in a sequence-independent manner and generates 2'-3'-cGAMP as an endogenous second messenger, which binds STING to trigger the signaling pathway that leads to the production of cytokines, such as type I IFN (Figure 7A). The activation of STING has

Table 2. Conservation analysis of key amino acid residues within the c-di-GMP binding domain (CBD) of human or mouse STING in eukaryotic STING family

Amino acid residues		Non-conserved in these species				
Mouse STING	Human STING	Mammals	Non-mammal vertebrates	Arthropods	Cnidarians	Protist (<i>M. brevicollis</i>)
R231	R232, H232 ^a	— ^c	Met (M) in <i>O. latipes</i> (Japanese medaka)	—	Thr (T) in <i>H. magnipapillata</i>	Ile (I)
Y166	Y167 ^b	—	—	—	—	—
R237	R238 ^b	—	—	—	—	—
Y239	Y240 ^b	—	Phe (F) in <i>A. carolinensis</i> (green anole), <i>X. tropicalis</i> and <i>X. laevis</i>	Phe (F) in most arthropods ^d	—	Met (M)
N241	N242 ^b	—	His (H) in <i>O. latipes</i> , <i>T. guttata</i> (zebra finch) and <i>G. gallus</i> (chicken); Gln (Q) in <i>M. gallopavo</i>	Asn (N) in <i>Ixodes scapularis</i> , <i>T. castaneum</i> , <i>A. mellifera</i> and <i>N. vitripennis</i> ; Ile (I) in <i>D. willistoni</i> ; His (H) in the others	His (H) in <i>N. vectensis</i> (sea anemone)	His (H)
E259	E260 ^b	Gly (G) in <i>S. scrofa</i> (pig)	—	—	—	—
T262	T263 ^b	—	Ser (S) in <i>G. gallus</i> and <i>M. gallopavo</i>	—	—	Ser (S)

^aR232/H232 in human and R231 in mouse are important for optimal response to 2'-5'-linkage-containing natural or unnatural cGAMP isomers (20,21).

^bAmino acid substitutions of Ala at these positions reduced or abolished cGAMP-isomer-dependent IFN-pathway activation (20).

^c— means the residue is conserved in all species.

^dSTING homologs in *D. willistoni*, *D. persimilis*, *D. pseudoobscura*, *Tribolium castaneum*, *Apis mellifera* and *Nasonia vitripennis* have the same amino acid Tyr (Y) as human STING.

been reported to be inhibited by ULK1 associated with AMPK α (71). To study how this cGAS-STING-dependent type I IFN induction evolved, we placed cGAS, STING and their downstream components onto the eukaryotic evolutionary map (Figure 7B). Fungus *S. cerevisiae* (budding yeast) was added to the representative species group as a contrast. AMPK α and ULK1 are present in nearly all representative species. cGAS and STING are present in choanoflagellate *M. brevicollis*. But the other four proteins (IKK, TBK1, NF- κ B and IRF3) involved in the activation of this pathway are absent from *M. brevicollis* and *S. cerevisiae*, suggesting they are probably metazoan-specific proteins. IKK ϵ and TBK1 have been reported to group together in early metazoans (72). Consistent with this, Figure 7B indicates that only a single homolog of IKK ϵ or TBK1 is present in the poriferan *Amphimedon queenslandica*, cnidarians, nematode *C. elegans* and insect *D. melanogaster*. At the origin of echinoderms, *Strongylocentrotus purpuratus* (purple sea urchin), IKK ϵ and TBK1 were diverged and present together in late-branching metazoans. NF- κ B is widely distributed in all metazoans except nematodes. IRF3 homologs occurred at the origin of fishes but are absent from all three birds gathered in this study, *G. gallus*, *M. gallopavo* and *T. guttata*. The bird genomes contain another member of the IRF family, IRF7, which is also a transcription factor that induces type I IFN and is grouped with IRF3 according to their evolutionary history (73). IRF7 is present in all vertebrates, which is consistent with previous studies (74,75).

DISCUSSION

Does cGAS co-evolve with STING in metazoans?

Using elegant biochemical and genetic experiments, Chen *et al.* identified that mammalian cGAS proteins act as the major and nonredundant cytosolic DNA sensor that generates the second-messenger product, 2'3'-cGAMP. 2'3'-cGAMP then activates STING, which further stimulates the downstream signaling pathway that leads to type I IFN production (17–18,28). The evolution of cGAS and STING was studied in two recent reviews (77,78). Schaap screened representative genomes of the major metazoan phyla and the choanoflagellate *M. brevicollis* using a modified best bidirectional BLASTP search, and found that STING homologs are distributed in all major animal phyla, except for porifera, and *M. brevicollis*. But cGAS is present as early as in cephalochordate *B. floridae*. In the other review, Wu and Chen obtained similar results of the origins of cGAS and STING. In our study, we applied a systematic and rigorous method to search homologs on all fully sequenced eukaryotic genomes, and the understanding of the evolutionary distribution of both cGAS and STING genes can be expanded across eukaryotic species. Besides STING, the emergence of cGAS could be traced back to the choanoflagellate *M. brevicollis*, which is the closest known relative of metazoans (79).

The cGAS family has several features in common with the STING family during metazoan evolution. First, both proteins are present early in several simple organisms, including cnidarians *N. vectensis* and *H. magnipapillata*.

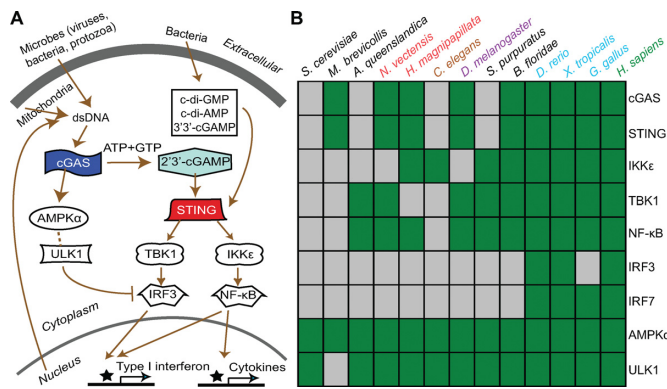


Figure 7. cGAS-STING signaling to trigger type I IFN and phylogenetic profiles of its molecular components. (A) Overview of cGAS-STING signaling pathway leading to the production of type I IFN adapted from recent studies (1,71,76). Cytosolic dsDNA from diverse microbes and self-DNA in infected cells are danger signals that are sensed by cGAS in a sequence-independent manner. cGAS generates 2'3'-cGAMP as an endogenous second messenger, which binds STING and induces a conformational change in STING. The activated STING recruits TBK1 and IKKε kinases, which in turn phosphorylate and activate IRF3 and NF-κB, respectively. IRF3 and NF-κB then translocate to the nucleus to induce type I IFN and other cytokines. Certain bacteria produce c-di-GMP, c-di-AMP and 3'3'-cGAMP, which could activate some STING alleles (such as R232 in human STING). The activation of STING is inhibited by ULK1 (71). (B) Distribution of each molecular component (on the right) of the signaling pathway in a representative group of species (on the top) with fully sequenced genomes. The transcription factor IRF7 was added in this part because just like IRF3, IRF7 is also considered a master regulator of type I IFN induction. Eukaryotic homologs of each molecular component were identified in a similar way to those of cGAS and STING (see Materials and Methods). The presence of a homolog of the molecular component in a particular species is in green and the absence in light gray. The species are placed according to the phylogenetic relationship in Figure 1. Most species have been indicated in Figure 4 except for yeast *S. cerevisiae*, sponge *Amphimedon queenslandica*, echinoderms *Strongylocentrotus purpuratus* (purple sea urchin), non-mammal vertebrates *X. tropicalis* (western clawed frog) and *G. gallus* (chicken).

However, both cGAS and STING were subsequently lost in nematodes and flatworms, from which other key components (NF-κB and IRF3) of the cGAS-STING signaling pathway are also absent (Figure 7B). It is possible that nematodes and flatworms may rely on different mechanisms to trigger innate immunity signaling. For example, even though *C. elegans* has homologs of Toll-like receptors which have well-established roles in innate immunity in mammals, these homologs do not function in response to infections in nematode (80).

Second, both cGAS and STING are subject to functional constraints that are relaxed to some extent compared with TBK1, a signaling component at the downstream end of this pathway, suggesting that cGAS and STING may be less conservative than TBK1 from insects to mammals (Table 1).

Third, the modern cGAS and STING proteins appear to have acquired their domain features early in the evolution of vertebrates. For cGAS proteins, the zinc-ribbon domain is not present in invertebrate cGAS (Figures 3B and 4). It has been reported that the zinc-ribbon domain is functionally important for a metal coordination and interaction with the major groove of DNA, suggesting that it serves as a molecular 'ruler' to scale the specificity of cGAS toward dsDNA

(26–27,76). Additionally, cGAS proteins in cnidarians and insects contain very short N-terminal domains (Figures 3B and 4). This highly positively charged N-terminal fragment of cGAS may play a role in DNA binding because this fragment can bind an immune stimulatory DNA (ISD, 45 bp) (17). Kranzusch *et al.* suspected that this N-terminal region also plays an important role in stabilization or autoinhibition like other nucleic acid sensors (RIG-I and AIM2) (27). Moreover, through investigating the conservation of 15 amino acid residues of human cGAS, which are critical for DNA binding or are involved in 2'3'-cGAMP binding, across eukaryotic cGAS homologs, we found that excluding five residues that are not conserved in vertebrates, most amino acids are completely conserved in vertebrates but not in arthropods, cnidarians or choanoflagellate *M. brevicollis* (Supplementary Table S5). Taken together, evolutionary analysis of key structural domains and critical residues shows that invertebrate cGAS homologs may not act as a DNA sensor by binding dsDNA. Similarly, the modern STING possibly began to acquire its CTT domain within the vertebrate lineage (Figure 5). The CTT domain plays an essential role in the biological function of STING. In the absence of ligand, CTT binds the STING CBD and maintains STING in an inactive and autoinhibited state. Binding of ligand relieves the autoinhibited state, exposes the CTT and stabilizes the STING dimers in complex with the ligand (67). The exposure of CTT also facilitates the interactions of CTT with TBK1 to promote activation of IRF3. Two residues, S366 and L374, in the CTT domain of human STING are important for IRF3 activation (38). Furthermore, Barber *et al.* studied the details of phosphorylation of STING and uncovered that the phosphorylated S366, induced by ULK1, may facilitate STING degradation to prevent sustained function (71). The ULK1-induced phosphorylation site S366 is within the CTT domain of human STING and conserved in all vertebrates (including fishes) except for amphibians *X. laevis* and *X. tropicalis* (Supplementary Figure S2). Amphibian STING homologs lack the CTT domain and probably depend on different mechanisms for activation and degradation. Thus, the function of cGAS and STING in sensing cytosolic DNA to trigger the innate immune response is possibly restricted to vertebrates, which is consistent with the finding that the IFN system functions only in jawed vertebrates (81).

In summary, cGAS and STING share three important evolutionary characteristics as discussed above. Interestingly, in mammals, cGAS and STING cooperate to bring about innate immune signaling. In that regard, cGAS represents a new class of cytosolic DNA sensor, while STING cooperates with cGAS and functions as a central adaptor molecule. The critical link is cGAMP which can activate STING to launch the type I IFN induction and also trigger negative-feedback control of STING activity (17,28,71). Therefore, contrary to the conclusion of Schaap that cGAS and STING did not evolve together (77), we speculate that cGAS has co-evolved with STING during the evolution of metazoans. Given that the critical functional zinc-ribbon domain and the most amino acid residues important for DNA binding in human cGAS are not conserved in early cGAS homologs, and even the CTT domain essential for STING activation and degradation through the ULK1-

induced phosphorylation site S366, is also not conserved in early STING homologs, cGAS-STING may possibly play other roles in early metazoans rather than activating an innate immunity response to cytosolic DNA.

Could cGAS produce cGAMP in early metazoans and *M. brevicollis*?

Even though we have suggested that cGAS may not function as a DNA sensor in invertebrates, it remains unknown whether cGAS in *M. brevicollis* and early metazoans can function as a cGAMP synthase. A series of mutations in mammalian cGAS at each position in the zinc-coordination site near the DNA-binding cleft abolished or severely impaired the activity of 2'3'-cGAMP synthesis *in vitro* and the production of type I IFN *in vivo* (26,27). A reason for this may be that the ability of cGAS to synthesize 2'3'-cGAMP requires the pronounced conformation changes taking place after the binding of cGAS to dsDNA, which triggers a repositioning of catalytic residues in the binding pockets of ATP and GTP (23–24,26,27,63–64). Invertebrate cGAS homologs do not contain the zinc-ribbon domain (Figures 3B and 4) and four of five amino acid residues involved in 2'3'-cGAMP in human cGAS are not conserved in arthropods, cnidarians or a choanoflagellate (Supplementary Table S5), but in the absence of any experimental evidence regarding invertebrate cGAS reactivity, we can not make an unambiguous conclusion that unlike the mammalian cGAS, cGAS homologs in invertebrates are not 2'3'-cGAMP synthase. However, the secondary structure pattern and the NTase-specific motif (hG[G/S]X9–13[D/E]h[D/E]h...h[D/E]h) are highly conserved in cGAS homologs throughout metazoan evolution, in the choanoflagellate *M. brevicollis* and in the *V. cholerae* DncV (Figure 4 and Supplementary Figure S4). Interestingly, DncV is capable of synthesizing 3'3'-cGAMP, c-di-AMP and c-di-GMP *in vitro* by incubating with the corresponding nucleoside triphosphate and a simple assay buffer. Therefore, we hypothesize that 'early' cGAS proteins may probably have the ability to produce cGAMP or other kinds of cyclic dinucleotides acting as a nucleotidyltransferase.

In a recent review, Schaap suspected that metazoan and their protist ancestors of STING could detect cyclic dinucleotides long before cGAS could synthesize 2'3'-cGAMP, through the conservation analysis of 10 residues in human STING that are required for binding cyclic dinucleotides across homologous sequences (77). Considering that cGAS and STING evolved together during animal evolution, the conclusion by Schaap probably provide complementary information for our hypothesis that 'early' cGAS homologs could synthesize cyclic dinucleotides. To provide more complementary information, we generated homology models of STING from three non-mammal species, fish *O. latipes* (Figure 6), chicken *G. gallus* and the choanoflagellate *M. brevicollis* (Supplementary Figure S6), according to the structure of human STING (PDB ID: 4LOH). Similar to the reported human STING structure (20), the three non-mammalian STING homologs can also form a dimer and probably exhibit functional levels of binding with cGAMP, although the binding would probably be slightly weaker than to human STING.

If cGAMP synthesis activity of cGAS evolved in *M. brevicollis* and early metazoans, another question arose naturally: Along the trajectory of human evolution, did two kinds of cGAMP, canonical 3'3'- and uncommon 2'3'-cGAMP, once exist together in the innate immune system? cGAMP and all other cyclic dinucleotides in bacterial cells are linked by 3'-5'-phosphodiester linkage. In the mammals, namely, human, mouse and pig, the endogenous cGAMP has a unique 2'-5'-phosphodiester linkage between GMP and AMP (19,21–23,27). Another nucleotidyltransferase, OAS1, can produce 2'-5'-phosphodiester linkages and polymerize ATP into 2'-5'-linked iso-RNA (2'-5'-oligoadenylate) instead of 3'-5'-linked RNA under dsRNA binding (63). Biochemical and evolutionary analysis concluded that the first OAS1 protein with 2'-5'-oligoadenylate synthesis activity is in *Geodia cydonium* (marine sponge), an earlier lineage than cnidaria in the kingdom metazoa. *G. cydonium* OAS1 can produce both 3'-5' and 2'-5' linkages but predominantly synthesizes the 2'-5' linkage (82,83). Nothing is yet known about the production of cGAS in invertebrates; however, we could get some clues from the STING family considering that cGAS and STING may have co-evolved in metazoans. For example, the H232 allele of human STING specifically responds to 2'3'-cGAMP but loses the ability to respond to 3'3'-cGAMP or bacterial c-di-GMP, while the R232 allele can respond to all these cyclic dinucleotides. This suggests that the responsiveness to bacterial cyclic dinucleotides was lost under a strong selective pressure during human evolution (22,76). Interestingly, the H232 STING allele only appears in humans while the R232 is highly conserved in most metazoans except for *O. latipes*, *H. magnipapillata* and *M. brevicollis* (Table 2). Therefore, although mammalian cGAS proteins synthesize the non-canonical 2'3'-cGAMP on sensing cytosolic DNA, we suspect that like OAS1, invertebrate cGAS might have had the ability to produce both 2'3'-cGAMP and 3'3'-cGAMP.

CONCLUSION

To conclude, cGAS and STING are already present in the unicellular eukaryotic organism *M. brevicollis*. But during the metazoan evolution that followed, both were lost in nematodes, flatworms. Because both proteins cooperate extensively in the stimulation of the innate immune pathway and display similar evolutionary characteristics, we hypothesize that cGAS and STING have co-evolved in *M. brevicollis* and metazoans. Given the critical functions of cGAS and STING in mammals, it is important to study the primitive biological functions controlled by cGAS and STING in *M. brevicollis* and early metazoans. Based on the evolutionary analysis of their structural organization, zinc-ribbon domain and long N-terminal fragment in cGAS as well as STING CTT domain, modern cGAS and STING proteins may have gained their functional domains early in the evolution of vertebrates. In addition, vertebrate cGAS homologs keep most of the amino acids residues in human cGAS that are important for DNA binding conserved, whereas invertebrate cGAS homologs have variations on these residues (Supplementary Table S5). Therefore, we hypothesize that cGAS and STING do not take part in the innate immunity response to cytosolic DNA in invertebrates. However, the

high conservation of secondary structures and key active residues between cGAS homologs and *V. cholerae* DncV indicates that cGAS may already have acquired the ability to synthesize cGAMP in *M. brevicollis* and early-branching metazoans. A question remaining is whether 3'3'- and 2'3'-cGAMP once existed together in the evolution of the innate immune system? We propose that cGAS might have been able to synthesize both 3'3'- and 2'3'-cGAMP during some stage of metazoan evolution. Although the specific physiological and biochemical roles of cGAS and STING homologs in invertebrates are remain uncertain, the conservation analysis of critical domains, secondary structural elements and amino acid residues provide novel insights into the relationships between structure and function in both proteins. cGAS and STING do not function in isolation; they activate cellular innate immune responses to cytosolic DNA through interactions with downstream molecules. The study of cGAS and STING combined with the other signaling components in an evolutionary perspective, which goes beyond the review by Schaap (77), may provide valuable molecular insights into the functions and origins of this type of pathway that initiates type I IFN.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank the three anonymous reviewers and the editor for their valuable comments as well as Benjamin Boyce, Yan Xue, Amelia Chen, Lynn Theprungsirikul, Allen Wojcik and Fang Yuan for illuminating discussions.

FUNDING

National Natural Science Foundation of China [31200625 and U1130304 to Xiaomei W.]; Zhejiang Provincial Natural Science Foundation of China [LY13C160014 to F.-H.W.]; Oklahoma Center for the Advancement of Science and Technology [PSB11-033 to Xiaoqiang W.]; NIH Grants [AI093717 and AI48674 to W.Z.]; National Science Foundation [IOS-0848263 to Z.-M.P.]. Funding for open access charge: National Natural Science Foundation of China [31200625].

Conflict of interest statement. None declared.

REFERENCES

- O'Neill, L.A. O'Neill, L.A. (2013) Sensing the dark side of DNA. *Science*, **339**, 763–764.
- Takaoka, A., Wang, Z., Choi, M.K., Yanai, H., Negishi, H., Ban, T., Lu, Y., Miyagishi, M., Kodama, T., and Honda, K. *et al.* Takaoka, A., Wang, Z., Choi, M.K., Yanai, H., Negishi, H., Ban, T., Lu, Y., Miyagishi, M., Kodama, T., and Honda, K. (2007) DAI (DLM-1/ZBP1) is a cytosolic DNA sensor and an activator of innate immune response. *Nature*, **448**, 501–505.
- Ablasser, A., Bauernfeind, F., Hartmann, G., Latz, E., Fitzgerald, K.A., and Hornung, V. (2009) RIG-I-dependent sensing of poly(dA:dT) through the induction of an RNA polymerase III-transcribed RNA intermediate. *Nat. Immunol.*, **10**, 1065–1072.
- Chiu, Y.H., Macmillan, J.B., and Chen, Z.J. Chiu, Y.H., Macmillan, J.B., and Chen, Z.J. (2009) RNA polymerase III detects cytosolic DNA and induces type I interferons through the RIG-I pathway. *Cell*, **138**, 576–591.
- Zhang, Z., Yuan, B., Bao, M., Lu, N., Kim, T., and Liu, Y.J. Zhang, Z., Yuan, B., Bao, M., Lu, N., Kim, T., and Liu, Y.J. (2011) The helicase DDX41 senses intracellular DNA mediated by the adaptor STING in dendritic cells. *Nat. Immunol.*, **12**, 959–965.
- Burckstummer, T., Baumann, C., Bluml, S., Dixit, E., Durnberger, G., Jahn, H., Planyavsky, M., Bilban, M., Colinge, J., and Bennett, K.L. *et al.* Burckstummer, T., Baumann, C., Bluml, S., Dixit, E., Durnberger, G., Jahn, H., Planyavsky, M., Bilban, M., Colinge, J., and Bennett, K.L. (2009) An orthogonal proteomic-genomic screen identifies AIM2 as a cytoplasmic DNA sensor for the inflammasome. *Nat. Immunol.*, **10**, 266–272.
- Fernandes-Alnemri, T., Yu, J.W., Datta, P., Wu, J., and Alnemri, E.S. Fernandes-Alnemri, T., Yu, J.W., Datta, P., Wu, J., and Alnemri, E.S. (2009) AIM2 activates the inflammasome and cell death in response to cytoplasmic DNA. *Nature*, **458**, 509–513.
- Hornung, V., Ablasser, A., Charrel-Dennis, M., Bauernfeind, F., Horvath, G., Caffrey, D.R., Latz, E., and Fitzgerald, K.A. Hornung, V., Ablasser, A., Charrel-Dennis, M., Bauernfeind, F., Horvath, G., Caffrey, D.R., Latz, E., and Fitzgerald, K.A. (2009) AIM2 recognizes cytosolic dsDNA and forms a caspase-1-activating inflammasome with ASC. *Nature*, **458**, 514–518.
- Unterholzner, L., Keating, S.E., Baran, M., Horan, K.A., Jensen, S.B., Sharma, S., Sirois, C.M., Jin, T., Latz, E., and Xiao, T.S. *et al.* Unterholzner, L., Keating, S.E., Baran, M., Horan, K.A., Jensen, S.B., Sharma, S., Sirois, C.M., Jin, T., Latz, E., and Xiao, T.S. (2010) IFI16 is an innate immune sensor for intracellular DNA. *Nat. Immunol.*, **11**, 997–1004.
- Keating, S.E., Baran, M., and Bowie, A.G. Keating, S.E., Baran, M., and Bowie, A.G. (2011) Cytosolic DNA sensors regulating type I interferon induction. *Trends Immunol.*, **32**, 574–581.
- Ishikawa, H. and Barber, G.N. Ishikawa, H. and Barber, G.N. (2008) STING is an endoplasmic reticulum adaptor that facilitates innate immune signalling. *Nature*, **455**, 674–678.
- Jin, L., Waterman, P.M., Jonscher, K.R., Short, C.M., Reisdorph, N.A., and Cambier, J.C. Jin, L., Waterman, P.M., Jonscher, K.R., Short, C.M., Reisdorph, N.A., and Cambier, J.C. (2008) MPYS, a novel membrane tetraspanner, is associated with major histocompatibility complex class II and mediates transduction of apoptotic signals. *Mol. Cell Biol.*, **28**, 5014–5026.
- Zhong, B., Yang, Y., Li, S., Wang, Y.Y., Li, Y., Diao, F., Lei, C., He, X., Zhang, L., and Tien, P. *et al.* Zhong, B., Yang, Y., Li, S., Wang, Y.Y., Li, Y., Diao, F., Lei, C., He, X., Zhang, L., and Tien, P. (2008) The adaptor protein MITA links virus-sensing receptors to IRF3 transcription factor activation. *Immunity*, **29**, 538–550.
- Sun, W., Li, Y., Chen, L., Chen, H., You, F., Zhou, X., Zhou, Y., Zhai, Z., Chen, D., and Jiang, Z. Sun, W., Li, Y., Chen, L., Chen, H., You, F., Zhou, X., Zhou, Y., Zhai, Z., Chen, D., and Jiang, Z. (2009) ERIS, an endoplasmic reticulum IFN stimulator, activates innate immune signaling through dimerization. *Proc. Natl. Acad. Sci. USA*, **106**, 8653–8658.
- Burdette, D.L. and Vance, R.E. Burdette, D.L. and Vance, R.E. (2013) STING and the innate immune response to nucleic acids in the cytosol. *Nat. Immunol.*, **14**, 19–26.
- Burdette, D.L., Monroe, K.M., Sotelo-Troha, K., Iwig, J.S., Eckert, B., Hyodo, M., Hayakawa, Y., and Vance, R.E. Burdette, D.L., Monroe, K.M., Sotelo-Troha, K., Iwig, J.S., Eckert, B., Hyodo, M., Hayakawa, Y., and Vance, R.E. (2011) STING is a direct innate immune sensor of cyclic di-GMP. *Nature*, **478**, 515–518.
- Sun, L., Wu, J., Du, F., Chen, X., and Chen, Z.J. Sun, L., Wu, J., Du, F., Chen, X., and Chen, Z.J. (2013) Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science*, **339**, 786–791.
- Wu, J., Sun, L., Chen, X., Du, F., Shi, H., Chen, C., and Chen, Z.J. Wu, J., Sun, L., Chen, X., Du, F., Shi, H., Chen, C., and Chen, Z.J. (2013) Cyclic GMP-AMP is an endogenous second messenger in innate immune signaling by cytosolic DNA. *Science*, **339**, 826–830.
- Ablasser, A., Goldeck, M., Cavlar, T., Deimling, T., Witte, G., Rohl, I., Hopfner, K.P., Ludwig, J., and Hornung, V. Ablasser, A., Goldeck, M., Cavlar, T., Deimling, T., Witte, G., Rohl, I., Hopfner, K.P., Ludwig, J., and Hornung, V. (2013) cGAS produces a 2'-5'-linked

- cyclic dinucleotide second messenger that activates STING. *Nature*, **498**, 380–384.
20. Gao, P., Ascano, M., Zillinger, T., Wang, W., Dai, P., Serganov, A.A., Gaffney, B.L., Shuman, S., Jones, R.A., and Deng, L. *et al.* Gao, P., Ascano, M., Zillinger, T., Wang, W., Dai, P., Serganov, A.A., Gaffney, B.L., Shuman, S., Jones, R.A., and Deng, L. (2013) Structure-function analysis of STING activation by c[G(2',5')pA(3',5')p] and targeting by antiviral DMXAA. *Cell*, **154**, 748–762.
 21. Zhang, X., Shi, H., Wu, J., Sun, L., Chen, C., and Chen, Z.J. Zhang, X., Shi, H., Wu, J., Sun, L., Chen, C., and Chen, Z.J. (2013) Cyclic GMP-AMP containing mixed phosphodiester linkages is an endogenous high-affinity ligand for STING. *Mol. Cell*, **51**, 226–235.
 22. Diner, E.J., Burdette, D.L., Wilson, S.C., Monroe, K.M., Kellenberger, C.A., Hyodo, M., Hayakawa, Y., Hammond, M.C., and Vance, R.E. Diner, E.J., Burdette, D.L., Wilson, S.C., Monroe, K.M., Kellenberger, C.A., Hyodo, M., Hayakawa, Y., Hammond, M.C., and Vance, R.E. (2013) The innate immune DNA sensor cGAS produces a noncanonical cyclic dinucleotide that activates human STING. *Cell Rep.*, **3**, 1355–1361.
 23. Gao, P., Ascano, M., Wu, Y., Barchet, W., Gaffney, B.L., Zillinger, T., Serganov, A.A., Liu, Y., Jones, R.A., and Hartmann, G. *et al.* Gao, P., Ascano, M., Wu, Y., Barchet, W., Gaffney, B.L., Zillinger, T., Serganov, A.A., Liu, Y., Jones, R.A., and Hartmann, G. (2013) Cyclic [G(2',5')pA(3',5')p] is the metazoan second messenger produced by DNA-activated cyclic GMP-AMP synthase. *Cell*, **153**, 1094–1107.
 24. Zhang, X., Wu, J., Du, F., Xu, H., Sun, L., Chen, Z., Brautigam, C.A., and Chen, Z.J. Zhang, X., Wu, J., Du, F., Xu, H., Sun, L., Chen, Z., Brautigam, C.A., and Chen, Z.J. (2014) The cytosolic DNA sensor cGAS forms an oligomeric complex with DNA and undergoes switch-like conformational changes in the activation loop. *Cell Rep.*, **6**, 421–430.
 25. Li, X., Shu, C., Yi, G., Chaton, C.T., Shelton, C.L., Diao, J., Zuo, X., Kao, C.C., Herr, A.B., and Li, P. Li, X., Shu, C., Yi, G., Chaton, C.T., Shelton, C.L., Diao, J., Zuo, X., Kao, C.C., Herr, A.B., and Li, P. (2013) Cyclic GMP-AMP synthase is activated by double-stranded DNA-induced oligomerization. *Immunity*, **39**, 1019–1031.
 26. Civril, F., Deimling, T., de Oliveira Mann, C.C., Ablasser, A., Moldt, M., Witte, G., Hornung, V., and Hopfner, K.P. Civril, F., Deimling, T., de Oliveira Mann, C.C., Ablasser, A., Moldt, M., Witte, G., Hornung, V., and Hopfner, K.P. (2013) Structural mechanism of cytosolic DNA sensing by cGAS. *Nature*, **498**, 332–337.
 27. Kranzusch, P.J., Lee, A.S., Berger, J.M., and Doudna, J.A. Kranzusch, P.J., Lee, A.S., Berger, J.M., and Doudna, J.A. (2013) Structure of human cGAS reveals a conserved family of second-messenger enzymes in innate immunity. *Cell Rep.*, **3**, 1362–1368.
 28. Li, X.D., Wu, J., Gao, D., Wang, H., Sun, L., and Chen, Z.J. Li, X.D., Wu, J., Gao, D., Wang, H., Sun, L., and Chen, Z.J. (2013) Pivotal roles of cGAS-cGAMP signaling in antiviral defense and immune adjuvant effects. *Science*, **341**, 1390–1394.
 29. Sun, F., Zhang, Y.B., Liu, T.K., Shi, J., Wang, B., and Gui, J.F. Sun, F., Zhang, Y.B., Liu, T.K., Shi, J., Wang, B., and Gui, J.F. (2011) Fish MITA serves as a mediator for distinct fish IFN gene activation dependent on IRF3 or IRF7. *J. Immunol.*, **187**, 2531–2539.
 30. Xie, L., Liu, M., Fang, L., Su, X., Cai, K., Wang, D., Chen, H., and Xiao, S. Xie, L., Liu, M., Fang, L., Su, X., Cai, K., Wang, D., Chen, H., and Xiao, S. (2010) Molecular cloning and functional characterization of porcine stimulator of interferon genes (STING). *Dev. Comp. Immunol.*, **34**, 847–854.
 31. Biacchesi, S., Merour, E., Lamoureux, A., Bernard, J., and Bremont, M. Biacchesi, S., Merour, E., Lamoureux, A., Bernard, J., and Bremont, M. (2012) Both STING and MAVS fish orthologs contribute to the induction of interferon mediated by RIG-I. *PLoS One*, **7**, e47737.
 32. Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., and Hirakawa, M. Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., and Hirakawa, M. (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.*, **34**, D354–D357.
 33. Ryan, J.F., Pang, K., Schnitzler, C.E., Nguyen, A.D., Moreland, R.T., Simmons, D.K., Koch, B.J., Francis, W.R., Havlak, P., and Smith, S.A. *et al.* Ryan, J.F., Pang, K., Schnitzler, C.E., Nguyen, A.D., Moreland, R.T., Simmons, D.K., Koch, B.J., Francis, W.R., Havlak, P., and Smith, S.A. (2013) The genome of the ctenophore *Mnemiopsis leidyi* and its implications for cell type evolution. *Science*, **342**, 1242592.
 34. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
 35. Salamov, A.A. and Solovyev, V.V. Salamov, A.A. and Solovyev, V.V. (2000) Ab initio gene finding in Drosophila genomic DNA. *Genome Res.*, **10**, 516–522.
 36. Tusnady, G.E. and Simon, I. Tusnady, G.E. and Simon, I. (2001) The HMMTOP transmembrane topology prediction server. *Bioinformatics*, **17**, 849–850.
 37. Kall, L., Krogh, A., and Sonnhammer, E.L. Kall, L., Krogh, A., and Sonnhammer, E.L. (2004) A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.*, **338**, 1027–1036.
 38. Tanaka, Y. and Chen, Z.J. Tanaka, Y. and Chen, Z.J. (2012) STING specifies IRF3 phosphorylation by TBK1 in the cytosolic DNA signaling pathway. *Sic. Signal.*, **5**, ra20.
 39. Pei, J., Kim, B.H., and Grishin, N.V. Pei, J., Kim, B.H., and Grishin, N.V. (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.*, **36**, 2295–2300.
 40. Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., and Barton, G.J. Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., and Barton, G.J. (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**, 1189–1191.
 41. Cole, C., Barber, J.D., and Barton, G.J. Cole, C., Barber, J.D., and Barton, G.J. (2008) The Jpred 3 secondary structure prediction server. *Nucleic Acids Res.*, **36**, W197–W201.
 42. Bick, A.G., Calvo, S.E., and Mootha, V.K. Bick, A.G., Calvo, S.E., and Mootha, V.K. (2012) Evolutionary diversity of the mitochondrial calcium uniporter. *Science*, **336**, 886.
 43. Ciccarelli, F.D., Doerks, T., von Mering, C., Creevey, C.J., Snel, B., and Bork, P. Ciccarelli, F.D., Doerks, T., von Mering, C., Creevey, C.J., Snel, B., and Bork, P. (2006) Toward automatic reconstruction of a highly resolved tree of life. *Science*, **311**, 1283–1287.
 44. Edgar, R.C. Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792–1797.
 45. Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.*, **59**, 307–321.
 46. Junier, T. and Zdobnov, E.M. Junier, T. and Zdobnov, E.M. (2010) The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics*, **26**, 1669–1670.
 47. Liu, L. and Yu, L. Liu, L. and Yu, L. (2010) Phybase: an R package for species tree analysis. *Bioinformatics*, **26**, 962–963.
 48. Liu, L., Yu, L., Pearl, D.K., and Edwards, S.V. Liu, L., Yu, L., Pearl, D.K., and Edwards, S.V. (2009) Estimating species phylogenies using coalescence times among sequences. *Syst. Biol.*, **58**, 468–477.
 49. Felsenstein, J. Felsenstein, J. (2005) PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author, Department of Genome Sciences, University of Washington, Seattle.
 50. Letunic, I. and Bork, P. Letunic, I. and Bork, P. (2011) Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res.*, **39**, W475–W478.
 51. Dunn, C.W., Hejnol, A., Matus, D.Q., Pang, K., Browne, W.E., Smith, S.A., Seaver, E., Rouse, G.W., Obst, M., and Edgecombe, G.D. *et al.* Dunn, C.W., Hejnol, A., Matus, D.Q., Pang, K., Browne, W.E., Smith, S.A., Seaver, E., Rouse, G.W., Obst, M., and Edgecombe, G.D. (2008) Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature*, **452**, 745–749.
 52. Obbard, D.J., Maclennan, J., Kim, K.W., Rambaut, A., O'Grady, P.M., and Jiggins, F.M. Obbard, D.J., Maclennan, J., Kim, K.W., Rambaut, A., O'Grady, P.M., and Jiggins, F.M. (2012) Estimating divergence dates and substitution rates in the Drosophila phylogeny. *Mol. Biol. Evol.*, **29**, 3459–3473.
 53. Suyama, M., Torrents, D., and Bork, P. Suyama, M., Torrents, D., and Bork, P. (2006) PAL2NAL: robust conversion of protein

- sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.*, **34**, W609–W612.
54. Yang, Z. and Nielsen, R. Yang, Z. and Nielsen, R. (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.*, **17**, 32–43.
 55. Yang, Z. Yang, Z. (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.*, **13**, 555–556.
 56. Fiser, A. and Sali, A. Fiser, A. and Sali, A. (2003) Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol.*, **374**, 461–491.
 57. Emsley, P. and Cowtan, K. Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Cryst.*, **D60**, 2126–2132.
 58. Zhang, P., Gu, Z., and Li, W.H. Zhang, P., Gu, Z., and Li, W.H. (2003) Different evolutionary patterns between young duplicate genes in the human genome. *Genome Biol.*, **4**, R56.
 59. Kuchta, K., Knizewski, L., Wyrwicz, L.S., Rychlewski, L., and Ginalski, K. Kuchta, K., Knizewski, L., Wyrwicz, L.S., Rychlewski, L., and Ginalski, K. (2009) Comprehensive classification of nucleotidyltransferase fold proteins: identification of novel families and their representatives in human. *Nucleic Acids Res.*, **37**, 7701–7714.
 60. Chow, K.L., Hall, D.H., and Emmons, S.W. Chow, K.L., Hall, D.H., and Emmons, S.W. (1995) The mab-21 gene of *Caenorhabditis elegans* encodes a novel protein required for choice of alternate cell fates. *Development*, **121**, 3615–3626.
 61. Margolis, R.L., Stine, O.C., McInnis, M.G., Ranen, N.G., Rubinsztein, D.C., Leggo, J., Brando, L.V., Kidwai, A.S., Loev, S.J., and Breschel, T.S. et al. Margolis, R.L., Stine, O.C., McInnis, M.G., Ranen, N.G., Rubinsztein, D.C., Leggo, J., Brando, L.V., Kidwai, A.S., Loev, S.J., and Breschel, T.S. (1996) cDNA cloning of a human homologue of the *Caenorhabditis elegans* cell fate-determining gene mab-21: expression, chromosomal localization and analysis of a highly polymorphic (CAG)_n trinucleotide repeat. *Hum. Mol. Genet.*, **5**, 607–616.
 62. Mariani, M., Baldessari, D., Francisconi, S., Viggiano, L., Rocchi, M., Zappavigna, V., Malgaretti, N., and Consalez, G.G. Mariani, M., Baldessari, D., Francisconi, S., Viggiano, L., Rocchi, M., Zappavigna, V., Malgaretti, N., and Consalez, G.G. (1999) Two murine and human homologs of mab-21, a cell fate determination gene involved in *Caenorhabditis elegans* neural development. *Hum. Mol. Genet.*, **8**, 2397–2406.
 63. Donovan, J., Dufner, M., and Korennykh, A. Donovan, J., Dufner, M., and Korennykh, A. (2013) Structural basis for cytosolic double-stranded RNA surveillance by human oligoadenylate synthetase. *Proc. Natl. Acad. Sci. USA*, **110**, 1652–1657.
 64. Kato, K., Ishii, R., Goto, E., Ishitani, R., Tokunaga, F., and Nureki, O. Kato, K., Ishii, R., Goto, E., Ishitani, R., Tokunaga, F., and Nureki, O. (2013) Structural and functional analyses of DNA-sensing and immune activation by human cGAS. *PLoS One*, **8**, e76983.
 65. Davies, B.W., Bogard, R.W., Young, T.S., and Mekalanos, J.J. Davies, B.W., Bogard, R.W., Young, T.S., and Mekalanos, J.J. (2012) Coordinated regulation of accessory genetic elements produces cyclic di-nucleotides for *V. cholerae* virulence. *Cell*, **149**, 358–370.
 66. Ouyang, S., Song, X., Wang, Y., Ru, H., Shaw, N., Jiang, Y., Niu, F., Zhu, Y., Qiu, W., and Parvatiyar, K. et al. Ouyang, S., Song, X., Wang, Y., Ru, H., Shaw, N., Jiang, Y., Niu, F., Zhu, Y., Qiu, W., and Parvatiyar, K. (2012) Structural analysis of the STING adaptor protein reveals a hydrophobic dimer interface and mode of cyclic di-GMP binding. *Immunity*, **36**, 1073–1086.
 67. Yin, Q., Tian, Y., Kabaleeswaran, V., Jiang, X., Tu, D., Eck, M.J., Chen, Z.J., and Wu, H. Yin, Q., Tian, Y., Kabaleeswaran, V., Jiang, X., Tu, D., Eck, M.J., Chen, Z.J., and Wu, H. (2012) Cyclic di-GMP sensing via the innate immune signaling protein STING. *Mol. Cell*, **46**, 735–745.
 68. Huang, Y.H., Liu, X.Y., Du, X.X., Jiang, Z.F., and Su, X.D. Huang, Y.H., Liu, X.Y., Du, X.X., Jiang, Z.F., and Su, X.D. (2012) The structural basis for the sensing and binding of cyclic di-GMP by STING. *Nat. Struct. Mol. Biol.*, **19**, 728–730.
 69. Shang, G., Zhu, D., Li, N., Zhang, J., Zhu, C., Lu, D., Liu, C., Yu, Q., Zhao, Y., and Xu, S. et al. Shang, G., Zhu, D., Li, N., Zhang, J., Zhu, C., Lu, D., Liu, C., Yu, Q., Zhao, Y., and Xu, S. (2012) Crystal structures of STING protein reveal basis for recognition of cyclic di-GMP. *Nat. Struct. Mol. Biol.*, **19**, 725–727.
 70. Shu, C., Yi, G., Watts, T., Kao, C.C., and Li, P. Shu, C., Yi, G., Watts, T., Kao, C.C., and Li, P. (2012) Structure of STING bound to cyclic di-GMP reveals the mechanism of cyclic dinucleotide recognition by the immune system. *Nat. Struct. Mol. Biol.*, **19**, 722–724.
 71. Konno, H., Konno, K., and Barber, G.N. Konno, H., Konno, K., and Barber, G.N. (2013) Cyclic dinucleotides trigger ULK1 (ATG1) phosphorylation of STING to prevent sustained innate immune signaling. *Cell*, **155**, 688–698.
 72. Gilmore, T.D. and Wolenski, F.S. Gilmore, T.D. and Wolenski, F.S. (2012) NF-kappaB: where did it come from and why?. *Immunol. Rev.*, **246**, 14–35.
 73. Nehyba, J., Hrdlickova, R., and Bose, H.R. Nehyba, J., Hrdlickova, R., and Bose, H.R. (2009) Dynamic evolution of immune system regulators: the history of the interferon regulatory factor family. *Mol. Biol. Evol.*, **26**, 2539–2550.
 74. Kasamatsu, J. Kasamatsu, J. (2013) Evolution of innate and adaptive immune systems in jawless vertebrates. *Microbiol. Immunol.*, **57**, 1–12.
 75. Huang, B., Qi, Z.T., Xu, Z., and Nie, P. Huang, B., Qi, Z.T., Xu, Z., and Nie, P. (2010) Global characterization of interferon regulatory factor (IRF) genes in vertebrates: glimpse of the diversification in evolution. *BMC Immunol.*, **11**, 22.
 76. Danilchanka, O. and Mekalanos, J.J. Danilchanka, O. and Mekalanos, J.J. (2013) Cyclic dinucleotides and the innate immune response. *Cell*, **154**, 962–970.
 77. Schaap, P. Schaap, P. (2013) Cyclic di-nucleotide signaling enters the eukaryote domain. *IUBMB Life*, **65**, 897–903.
 78. Wu, J. and Chen, Z.J. Wu, J. and Chen, Z.J. (2014) Innate immune sensing and signaling of cytosolic nucleic acids. *Annu. Rev. Immunol.*, **32**, 461–488.
 79. King, N., Westbrook, M.J., Young, S.L., Kuo, A., Abedin, M., Chapman, J., Fairclough, S., Hellsten, U., Isogai, Y., and Letunic, I. et al. King, N., Westbrook, M.J., Young, S.L., Kuo, A., Abedin, M., Chapman, J., Fairclough, S., Hellsten, U., Isogai, Y., and Letunic, I. (2008) The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature*, **451**, 783–788.
 80. Irazoqui, J.E., Urbach, J.M., and Ausubel, F.M. Irazoqui, J.E., Urbach, J.M., and Ausubel, F.M. (2010) Evolution of host innate defence: insights from *Caenorhabditis elegans* and primitive invertebrates. *Nat. Rev. Immunol.*, **10**, 47–58.
 81. Kaiser, P., Rothwell, L., Avery, S., and Balu, S. Kaiser, P., Rothwell, L., Avery, S., and Balu, S. (2004) Evolution of the interleukins. *Dev. Comp. Immunol.*, **28**, 375–394.
 82. Kjaer, K.H., Poulsen, J.B., Reintamm, T., Saby, E., Martensen, P.M., Kelve, M., and Justesen, J. Kjaer, K.H., Poulsen, J.B., Reintamm, T., Saby, E., Martensen, P.M., Kelve, M., and Justesen, J. (2009) Evolution of the 2'-5'-oligoadenylate synthetase family in eukaryotes and bacteria. *J. Mol. Evol.*, **69**, 612–624.
 83. Pari, M., Kuusksalu, A., Lopp, A., Reintamm, T., Justesen, J., and Kelve, M. Pari, M., Kuusksalu, A., Lopp, A., Reintamm, T., Justesen, J., and Kelve, M. (2007) Expression and characterization of recombinant 2',5'-oligoadenylate synthetase from the marine sponge *Geodia cydonium*. *FEBS J.*, **274**, 3462–3474.