# *The widespread IS200/IS605 transposon family encodes diverse programmable RNA-guided endonucleases*

HHS Public Access

Author manuscript

*Science*. Author manuscript; available in PMC 2022 March 17.

Published in final edited form as:

*Science*. 2021 October ; 374(6563): 57–65. doi:10.1126/science.abj6856.

# The widespread IS200/605 transposon family encodes diverse programmable RNA-guided endonucleases

**Han Altae-Tran**[1,2,3,4,5,*], **Soumya Kannan**[1,2,3,4,5,*], **F. Esra Demircioglu**[1,2,3,4,5], **Rachel Oshiro**[1,2,3,4,5], **Suchita P. Nety**[1,2,3,4,5], **Luke J. McKay**[6,7,8], **Mensur Dlakić**[9], **William P. Inskeep**[6,7], **Kira S. Makarova**[10], **Rhiannon K. Macrae**[1,2,3,4,5], **Eugene V. Koonin**[10], **Feng Zhang**[1,2,3,4,5,†]

[(1)]Howard Hughes Medical Institute, Cambridge, MA 02139, USA;

[(2)]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA;

[(3)]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA;

[(4)]Department of Brain and Cognitive Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA;

[(5)]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA;

[(6)]Department of Land Resources and Environmental Sciences, Montana State University, Bozeman, MT 59717, USA;

[(7)]Thermal Biology Institute, Montana State University, Bozeman, MT 59717, USA;

[(8)]Center for Biofilm Engineering, Montana State University, Bozeman, MT 59717;

[(9)]Department of Microbiology and Cell Biology, Montana State University, Bozeman, MT 59717, USA;

[(10)]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA.

## Abstract

[†]Correspondence should be addressed to zhang@broadinstitute.org (F.Z.).

**Author contributions:** H.A.-T., S.K., and F.Z. conceived of the project. H.A.-T., S.K., E.D., R.O., S.P.N., K.S.M., E.V.K., and F.Z. designed and performed experiments. L.M., M.D., and W.I. collected metagenomic data. F.Z. supervised the research and experimental design with support from R.M. H.A.-T., S.K., R.M., E.V. K., and F.Z. wrote the manuscript with input from all authors.

[*]These authors contributed equally to this work.

**Competing interests:** H. A.-T., S.K., E.D., S.P.N., and F.Z. are co-inventors on U.S. provisional patent applications filed by the Broad Institute related to this work. F.Z. is a cofounder of Editas Medicine, Beam Therapeutics, Pairwise Plants, Arbor Biotechnologies, and Sherlock Biosciences.

Supplementary Materials
Materials and Methods
Supplementary Text
Fig. S1 – S36
Table S1 – S6
Data S1 – S4
References (25 – 65)

IscB proteins are putative nucleases encoded in a distinct family of IS200/IS605 transposons and are likely ancestors of the RNA-guided endonuclease Cas9, but the functions of IscB and its interactions with any RNA remain uncharacterized. Using evolutionary analysis, RNA-seq, and biochemical experiments, we reconstruct the evolution of CRISPR-Cas9 systems from IS200/IS605 transposons. We show that IscB utilizes a single non-coding RNA for RNA-guided cleavage of double-stranded DNA and can be harnessed for genome editing in human cells. We also demonstrate the RNA-guided nuclease activity of TnpB, another IS200/605 transposon-encoded protein and the likely ancestor of Cas12 endonucleases. This work reveals a widespread class of transposon-encoded RNA-guided nucleases, which we name OMEGA (Obligate Mobile Element Guided Activity), with strong potential for developing as biotechnologies.

## ONE-SENTENCE SUMMARY:

IS200/IS605 family transposons associate with a distinct non-coding RNA and encode widespread RNA-guided nucleases that likely gave rise to the CRISPR endonucleases Cas9s and Cas12s.

---

The prokaryotic RNA-guided defense system CRISPR-Cas9 (type II CRISPR-Cas), which has been adopted for genome editing in eukaryotic cells (1, 2), is thought to have evolved from IscB proteins (3). Despite its wide distribution across prokaryotes and shared domain composition and architecture with Cas9, the function of IscB remains unknown (Fig. S1). Moreover, given that IscB has not been reported to be associated with non-coding RNA (ncRNA) or CRISPR arrays, the evolutionary origins of the RNA-guided activity in Cas9 systems are unclear. IscB is encoded by a distinct subset of IS200/605 superfamily transposons that also include transposons encoding *tnpB*, a putative endonuclease distantly related to *iscB* and thought to be the ancestor of Cas12, the type V CRISPR effector (3–5). Using phylogenetic analysis, RNA-seq, and biochemical experiments, we sought to elucidate the functions of these proteins and the origin of RNA-guided activity in class 2 CRISPR systems.

## IscB is associated with an evolutionarily conserved non-coding RNA

IscB is ~400 amino acids (aa) long and contains a RuvC endonuclease domain split by the insertion of a bridge helix (BH) and an HNH endonuclease domain, an architecture that is shared with Cas9 (Fig. 1A) (3). We performed a comprehensive search for proteins containing an HNH or a split RuvC endonuclease domain and found that Cas9 and IscB were the only proteins that contained both domains (Data S1). This search also showed that IscB contains a previously unidentified N-terminus that lacks clear homology to known domains and is absent in Cas9, which we denoted PLMP after its conserved sequence motifs (Fig. 1A, Fig. S2). Clustering and phylogenetic analysis of the combined RuvC, BH, and HNH domains strongly suggests that all extant Cas9s descended from a single ancestral IscB (Fig. 1B, Data S2, S3). We searched for CRISPR arrays adjacent to *iscB* genes from each cluster and found 6 distinct groups of IscB, containing 16 clusters (of 603 total), that were CRISPR-associated, contrary to previous observations (3). CRISPR-associated IscBs were scattered around the IscB phylogenetic tree, suggesting they evolved independently, with one association event leading to the Cas9 lineage (Fig. 1B). In total we identified 31 unique CRISPR-associated *iscB* loci (of 2811 total).

Given their association with CRISPR arrays, we suspected that the rarely occurring CRISPR-associated IscBs may be RNA-guided nucleases. We first examined a cluster of CRISPR-associated IscBs similar to non-CRISPR associated IscBs (at ~50% aa identity). We heterologously expressed a representative locus from this clade in *E. coli* and performed small RNA-seq, which showed expression of not only the CRISPR array, but also a 329-bp intergenic region between the CRISPR array and the IscB open reading frame (ORF) (Fig. 1C). We purified the IscB protein and sequenced the co-purified RNA, demonstrating that this protein interacts with a single ncRNA component, encompassing both the CRISPR array and this intergenic region (Fig. 1C).

Given its interaction with a ncRNA that includes the CRISPR direct repeat (DR) and spacer, as well as its similar domain architecture to Cas9, we tested this IscB for RNA-guided endonuclease activity. Using a previously established protospacer adjacent motif (PAM)-discovery assay (Table S1) (6), we observed depletion of specific PAM sequences (Fig. 1D, Fig. S3), indicating that CRISPR-associated IscBs are reprogrammable RNA-guided nucleases. We confirmed this enzymatic activity with an *in vitro* cleavage assay using recombinant ribonucleoprotein (RNP) complexes (Fig. 1E).

Our finding that IscB functionally associated with CRISPR at least once, and likely on additional occasions, suggested that IscB systems more generally share a core ancestral ncRNA gene that is prone to evolving into a CRISPR array and in some cases a separate trans-acting *tracrRNA* (7). To test this hypothesis, we aligned 563 non-redundant *iscB* loci and searched for conserved nucleotide (nt) sequences either upstream or downstream of the *iscB* ORF. This analysis revealed a highly conserved intergenic region ~300 bp in length upstream of the ORF with a drop in conservation at the 5’ end, which corresponds to an IS200/605 transposon end. Secondary structure predictions for individual sequences revealed the presence of multiple G:U pairs (Fig. S4), suggesting that the conserved region encodes an ncRNA containing functionally important hairpins, which we named ωRNA. Small RNA-seq on a sample of *Ktedonobacter racemifer* strain SOSP1-21, a soil bacterium that harbors 49 IscB loci in its genome (3), demonstrated expression of the predicted ωRNA in many of these loci (Fig. 1F, Fig. S5, S6A). Moreover, we observed that the transcripts consistently extended beyond the conservation boundary at the 5’ end.

An RFAM search for potential homologs of the ωRNA showed that the conserved region of the ωRNA partially matched the previously reported HEARO RNA, a ncRNA that was found upstream of HNH domain-containing proteins, which at the time were thought to be homing endonucleases (8, 9). However, the RFAM search did not provide any clues about the nature of the 5’-terminal non-conserved portion of these transcripts. Comparison of the consensus CRISPR-associated IscB ncRNA and the covariance folded ωRNA secondary structures revealed high degrees of structural and sequence similarity, particularly in shared multi-stem regions and pseudoknots (Fig. 1G, Fig. S7, Supplementary Text). Most importantly, we inferred that the 5’-most non-conserved sequence in the ωRNA might function as a guide sequence, because the sequence immediately downstream was predicted to form hairpins that structurally resembled the hairpins formed by the DR/anti-repeat duplex in the CRISPR-associated IscB ncRNA (Fig. 1G).

## IscB is a reprogrammable RNA-guided DNA endonuclease

To test whether IscB was capable of cleaving DNA complementary to the putative ωRNA guide, we performed an *in vitro* plasmid cleavage assay with KraIscB-1 using an *in vitro* transcription/translation (IVTT) expression system (Fig. 2A, B). We found that KraIscB-1 cleaved the target in an ωRNA-dependent manner, with an ATAAA 3' target-adjacent motif (TAM) (Fig. 2C). Retargeting of KraIscB-1 using a different guide (Fn guide) (6) also mediated cleavage of the cognate target (Fig. 2C, Fig. S6B), implying that IscB is a reprogrammable RNA-guided nuclease.

Next, we biochemically characterized IscB *in vitro*. We identified activity in 57/86 (66%) selected phylogenetically diverse systems (Table S2) as determined by the identification of a TAM (Fig. S8). Of these 57 functional IscBs, 5 could be reconstituted with the respective ωRNA *in vitro* to achieve efficient target cleavage, and from those, we selected AwaIscB (from *Allochromatium warmingii*) for detailed biochemical characterization (Fig. 2D–G).

We confirmed the ability of recombinant AwaIscB to cleave multiple dsDNA targets in a programmable manner (Fig. 2E) and showed that the activity of AwaIscB is magnesium-dependent with a temperature optimum from 35–40°C (Fig. S9A, B). Appreciable activity was observed *in vitro* with guide lengths between 15 and 45 nt (Fig. S9D). Mutation of the catalytic RuvC-II residue (E157A) abolished the nucleolytic activity on the non-target DNA strand, whereas the HNH domain catalytic mutant H212A abolished the nucleolytic activity on the target strand (Fig. 2F). Combination of the E157A and H212A mutations (dAwaIscB) abolished all dsDNA nucleolytic activity (Fig. 2F) (10, 11). Sequencing of the cleavage products showed that AwaIscB cleaves the target strand 3 nt upstream of the TAM, similar to Cas9s (12). Cleavage of the non-target strand occurred 8 or 12 nt upstream of the TAM, generating 5- or 9-nt long 5' overhangs (Fig. 2G, Fig. S10). Exonuclease III mapping of a target substrate engaged by the dAwaIscB-ωRNA RNP showed that the RNP hindered exonuclease III treatment 19 nt upstream of the TAM on the target strand and 6 nt downstream of the targeted sequence on the non-target strand (Fig. S11) (13). We also found that truncation of more than 4 aa of the PLMP domain of AwaIscB abolished cleavage activity (Fig. S12).

## IscB employ multiple guide-encoding mechanisms

A distinct advantage of RNA-guided systems is that they allow an effector to target many substrates by simply reprogramming the RNA guide. One way IscB evolved to use multiple guides is association with CRISPR arrays (Fig. 3A). However, given that *iscB* loci typically encode a single ωRNA, it is unclear how or even whether these systems achieve such modularity in general. By searching for ωRNAs not directly adjacent to *iscB* ORFs, we uncovered three additional potential mechanisms for guide encoding and switching: ωRNA arrays, transposon expansion, and standalone, *trans*-acting ωRNAs (Fig. 3A). ωRNA arrays consist of multiple ωRNAs, each encompassing a distinct guide, separated by up to 200 bp, and are found in 15/3356 unique IscB/IsrB loci (0.4%). Transposon expansion involves the insertion of nearly identical IS200/605 superfamily transposons in multiple locations, resulting in multiple loci per genome, each capable of expressing a nearly identical ωRNA

scaffold with a unique guide (Fig. S13). By contrast, standalone ωRNAs, which show no detectable genomic associations with *iscB,* were more common and were found in multiple copies in some genomes (Table S3). *Cis* ωRNAs from 95/3356 (2.8%) unique IscB/IsrB loci were nearly identical (≥ 95% sequence identity) to distally encoded standalone ωRNAs (Fig. S14), implying that these standalone ωRNAs could encode guides used by *trans*-encoded IscBs.

We tested this possibility by examining 10 standalone ωRNAs in the *K. racemifer* genome, (Fig. 3B), 9 of which were found to be expressed (Fig. 3C, Fig. S15). Of the 6 standalone ωRNAs tested, we found that 5 could mediate RNA-guided DNA cleavage with a distally encoded IscB from the same genome (Fig. 3D), demonstrating that a single IscB can use multiple *trans*-encoded ωRNAs. Guides from many ωRNAs, both IscB-adjacent and *trans*-encoded, mostly target prokaryotic genomic sequences (61.5% genomic, 0.7% plasmid, 2.0% phage, 35.8% unmatched, $N$=36323), suggesting a non-defense function for IscB systems (Fig. S14, Table S3). In particular, we found that more than a third of the ωRNAs (34.1%) targeted the same locus without the IS200/605 transposon insertion (Table S3, Fig. S16).

## Evolution and diversity of IscB systems

We next investigated the evolutionary relationships between IscB, Cas9, and other homologous proteins to gain a broader insight into the evolution of RNA-guided mechanisms. In our search for proteins containing split RuvC domains, we detected another group of shorter, ~350 aa IscB homologs that are also encoded in IS200/605 superfamily transposons. These proteins contain a PLMP domain and split RuvC but lack the HNH domain. We renamed these proteins IsrB (Insertion sequence RuvC-like OrfB) to emphasize their distinct domain architecture, replacing the previous designation, IscB1 (3). In addition to IscB and IsrB, we identified a family of even smaller (~180 aa) proteins that only contained the PLMP domain and HNH domain but no RuvC domain, which we named IshB (Insertion sequence HNH-like OrfB).

To investigate the relationships between these proteins, we built a maximum likelihood (ML) tree from a multiple alignment of the split RuvC nuclease and BH domains using IQ-TREE 2 (Fig. 4A, Fig. S17–18, Data S2, S3, Table S4) (14). The topology of the resulting tree was supported by several additional ML and Bayesian phylogenetic and robustness analyses (Fig. S17–25, Data S2, S3; see Supplementary Text for details). In the resulting tree, IsrB, IscB, and Cas9 formed distinct, strongly supported clades, suggesting that each of these nucleases originated from a unique evolutionary event (Fig. 4A, S20C, D, S21, S22A, C, S23, Supplementary Text). We then analyzed the associations between each protein cluster and IS200/605 *tnpA* genes (3), ωRNAs, CRISPR-Cas adaptation genes (*cas1*, *cas2*, *cas4*, and *csn2*), CRISPR arrays upstream and downstream of the respective ORF, and CRISPR anti-repeats (Fig. 4A). As discussed above, *IscB* and *isrB* were rarely associated with CRISPR arrays and were not found to be associated with CRISPR-Cas adaptation genes. The *isrB*s are associated with structurally distinct ωRNAs. The *iscB*s are flanked by transposon ends similar to those mobilized by TnpA (3), but are only found near *tnpA* in 56/2811 of unique IscB loci (2.0%) (Fig. 4A, Fig. S26D).

Additionally, we identified two distinct groups of Cas9s. The first is a new subtype, II-D, a group of relatively small *cas9*s (~700aa) that are not associated with any other known *cas* genes (15). The second is a distinct clade branching from within the II-C subtype, which includes exceptionally large *cas9*s (>1700aa) that are associated with *tnpA* (Fig. 4A, Fig. S26). The *tnpA*-associated II-C loci often encompass unusually long DRs (more than 42bp in length) and in some cases encode HIRAN domain proteins between the *cas9* and other *cas* genes (Fig. 4A, Fig. S27). Predicted transposon ends surround various combinations of the *tnpA, cas* acquisition genes, and CRISPR arrays in these loci.

These phylogenetic and association analyses confirm that IS200/605 transposon-encoded IscBs and IsrBs share a common evolutionary history with Cas9 (Supplementary Text). Given the deep position of the IsrB clade in the tree (Fig. 4A) and the lack of the HNH domain, IsrBs likely represent the ancestral state, probably having evolved from the compact RuvC endonuclease (16). Almost all *isrB*s are associated with an ωRNA, suggesting that these systems became RNA-guided at an early stage of evolution, concomitantly with the insertions in the RuvC-like domain that are likely to be involved in complex formation with ωRNA. IsrB subsequently gained the HNH domain, possibly through insertion of another mobile element or recombination with a gene encoding an IshB-like protein, founding the IscB family (turquoise squares, Fig. 4A, B, Supplementary Text).

CRISPR arrays emerged within IscB systems on multiple, independent occasions (black circles, Fig. 4A, B). These short arrays consist of repeats that could have evolved by duplication of segments of the ancestral ωRNA. The resulting systems encompass a hybrid CRISPR-ωRNA that consists of a CRISPR array preceding a partial ωRNA. These CRISPR-associated IscB proteins likely also gained REC-like insertions between the RuvC-I and RuvC-II subdomains on a number of occasions, often contemporaneously with or shortly after the CRISPR association (white squares, Fig. 4A, B, Fig. S28). In particular, one CRISPR-associated IscB cluster (cluster 2089) apparently founded the Cas9 family (Fig. S23) upon the loss of the hallmark PLMP domain (gray square, Fig. 4A, B, S28). Moreover, the tracrRNAs of Subtype II-D, a deep branch in the Cas9 subtree (ML branch support: ≥ 97/100, Bayesian posterior probability: 100%, Fig. S20B–D, Fig. S23), shows significant similarity to IscB ωRNAs (E-value 4.1e-8), suggesting that the Cas9 tracrRNA originally evolved from ωRNA (Fig. S29). The continued evolution of Cas9 apparently involved the gain of additional REC-like insertions between the bridge helix and the RuvC-II domains resulting in increased protein size (Fig. S28). Finally, upon the association with the CRISPR adaptation machinery (*cas1*, *cas2*, and possibly *cas4*) (light blue circles, Fig. 4A, B), a burst of Cas9 diversification and widespread dispersion among bacteria via horizontal gene transfer followed, resulting in the evolution of multiple type II CRISPR subtypes.

We also explored the evolutionary history of ωRNAs. By iteratively building a set of ωRNA profiles that spanned all major groups of ωRNAs associated with *iscB*s and *isrB*s, we found that diverse ωRNAs are associated with almost all *iscB*s and *isrB*s. Moreover, different IsrB and IscB clades are associated with distinct ωRNA structures (Fig. 4A, C, Fig. S18A, S24A, S30). The transition from *isrB* to *iscB* was likely accompanied by loss of a second pseudoknot, the adaptor pseudoknot, between the transposon end region and the multi-stem loop in *isrB*-associated ωRNAs (yellow square, Fig. 4A, B, C). The inverse relationship

between the complexity of the ωRNA structure and the associated protein size is also reflected by the simplified ωRNA structures associated with clades of large IscBs and the even smaller tracrRNAs associated with large Cas9s (Fig. 4C, Fig. S30).

## IS200/IS605 elements encode diverse RNA-guided nucleases

In addition to the distinct succession of evolutionary events that yielded the abundant and diverse type II CRISPR systems, our phylogenetic analysis revealed several other events in the evolution of IscB and related proteins that led to the extant diversity, which we sought to experimentally explore.

First, we searched for IscB homologs in eukaryotic genomes and identified multiple *iscB* loci in the chloroplast genome of *Ignatius tetrasporus* UTEX B 2012, a terrestrial green alga (Fig. 5A, B, Fig. S31). Although the ORF is disrupted by multiple stop codons in most of these loci, one locus encodes an intact IscB (~50% aa identity to related prokaryotic IscBs) and a transcriptionally active ωRNA (Fig. 5C). This eukaryotic IscB cleaves DNA with a minimal NNG TAM (Fig. 5D), which differs from other characterized IscB TAMs (Fig. S8).

Second, we investigated the clade of large IscBs, which contain a BH domain that is split in two by REC domain-like insertions (white squares, Fig. 4A, 5A). We hypothesized that these insertions might enhance DNA unwinding, similarly to the REC lobe of Cas9 (17) and would therefore facilitate genome editing in the complex landscape of eukaryotic chromatin structure. We screened 6 large IscB proteins, using a pool of 12 guides each, for their ability to generate insertions/deletions (indels) in HEK293FT cells (see Methods, Table S5); one (OgeuIscB) produced appreciable indels (Fig. 5E, F, Fig. S32A). To further examine OgeuIscB activity, we tested a range of guide lengths targeting 3 loci in the human genome and found that OgeuIscB achieved the maximum indel rate with a 16 nt guide (Fig. S32B). On a panel of 46 sites in the human genome, we found that OgeuIscB induced indels at 28 of these sites with varying efficiency up to 4.4% (Fig. 5G, Fig. S32C, Table S5). Thus, OgeuIscB seems a promising candidate for further development of IscB-based genome editing tools.

Third, we experimentally characterized the putative nuclease activity of IsrB, the apparent ancestor of IscB (Fig. 5A). *K. racemifer* contains 5 *isrB*s associated with ωRNAs that are natively expressed (Fig. 5H, Fig. S33). We found that the IsrB-ωRNA RNP nicks the non-target strand of a dsDNA substrate in a guide- and TAM-specific manner (Fig. 5I, J, Fig. S34), which is analogous to the activity of IscB upon inactivation of the HNH domain (Fig. 2F).

Finally, we sought to determine if IS200/605 transposons in general harbor RNA-guided nucleases. In addition to the distinct IscB and IsrB families, most IS200/IS605 transposons encode RuvC-like endonucleases of another family, TnpB, which is thought to be the ancestor of Cas12s, the type V CRISPR effectors (Fig. 5A) (5). Additionally, TnpB is the likely ancestor of larger proteins, Fanzors, encoded in diverse eukaryotic transposons (Fig. 5A) (18). The TnpB family, including Fanzor, is an order of magnitude more diverse than

the IscB family; an HMMER search identified more than a million *tnpB* loci in publicly available prokaryotic genomes.

We identified conserved non-coding regions immediately downstream of the CDS of many *tnpB*s, suggesting the presence of associated ncRNAs that could function as RNA guides (Fig. S35). Previous work has identified ncRNAs overlapping the 3'-end of *tnpB* genes in archaea and bacteria (19, 20), but the function of these ncRNAs has not been characterized. Small RNA-seq of *K. racemifer* revealed native expression of a ncRNA overlapping the 3' end of the associated *tnpB* ORF (Fig. 5K), which we classified as a distinct group of ωRNAs. The reverse complement of the KraTnpB ωRNA 3' end is nearly identical to the 5' of the ωRNA associated with some KraIscBs, a region that corresponds to the predicted transposon end in each locus (Fig. 5L).

Analysis of non-redundant loci containing *tnpB* genes that clustered with KraTnpB showed a drop of sequence conservation at the 3' end of the loci (Fig. S35), corresponding to the IS200/605 transposon end. Comparison to the small RNA-seq trace revealed expression beyond the conservation drop, indicating possible presence of a guide sequence in the transcript (Fig. 5M). *In vitro* plasmid cleavage assays for multiple TnpB proteins from this cluster using a reprogrammed guide demonstrated RNA-guided cleavage with a 5' TAM (Fig. 5N, Fig. S36). We recombinantly purified a TnpB from *Alicyclobacillus macrosporangiidus* (AmaTnpB) and confirmed its reprogrammable RNA-guided dsDNA endonuclease activity (Fig. 5O, Fig. S36). We also observed that AmaTnpB robustly cleaved target-containing ssDNA substrates (Fig. 5P) and non-specifically cleaved a collateral substrate upon recognition of dsDNA or ssDNA substrates (Fig. 5Q).

## Discussion

Naturally programmable biological systems offer an efficient solution for diverse organisms to achieve scalable complexity via modularity of their components. RNA-guided defense and regulatory systems, which are widespread in prokaryotes and eukaryotes, are a prominent case in point, and have served as the basis of numerous biotechnology applications thanks to the ease with which they can be engineered and reprogrammed (21–23).

Here, through the exploration of Cas9 evolution, we discovered the programmable RNA-guided mechanism of 3 highly abundant but previously uncharacterized transposon-encoded nucleases: IscB, IsrB, and TnpB, which we collectively refer to as Ω (OMEGA: Obligate Mobile Element Guided Activity) (Fig. 6) because the mobile element localization and movement likely determines the identity of their guides. Although the biological functions of Ω systems remain unknown, several hypotheses are compatible with the available evidence, including roles in facilitating TnpA-catalyzed, RNA-guided transposition, or acting as a toxin, with the transposon acting as the antitoxin, securing maintenance of IS200/605 insertions (Supplementary Text).

The broad distribution of the Ω systems characterized here indicates that RNA-guided mechanisms are more widespread in prokaryotes than previously suspected and suggests that

RNA-guided activities are likely ancient and evolved on multiple, independent occasions, of which only the most common ones have likely been identified so far. The TnpB family is far more abundant and diverse than the IscB family; indeed, we identified more than a million putative *tnpB* loci in bacterial and archaeal genomes, making it one of the most common prokaryotic genes altogether. These TnpBs might represent an untapped wealth of diverse RNA-guided mechanisms present not only in prokaryotes, but also in eukaryotes. Combined with our identification of a chloroplast-encoded IscB, these findings suggest that the expansion of RNA-guided systems into eukaryotic genomes could be a general phenomenon, and more broadly, that RNA-guided systems are functionally diverse and permeate all domains of life.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGEMENTS:

We thank J. Strecker, S. Hirano, and D. Strebinger for advice regarding biochemistry experiments, G. Faure for advice regarding computational analyses, and all members of the Zhang lab for helpful discussions. We are grateful to the following individuals for generously providing access to their metagenomic data (IMG accessions provided in parenthesis): B. Campbell (IMG3300025818 and IMG3300007960), A. Buchan (IMG3300017968), and E. Edwards (IMG3300020812 and IMG3300023203). We appreciate assistance in DNA extraction and troubleshooting from M. Forbes for the Yellowstone Lake metagenomes. Yellowstone Lake samples were collected with support from the National Park Service - Yellowstone National Park (Research Permit YELL-2016/17-SCI-7018).

**Funding:**

National Science Foundation Integrated Earth Systems grant subaward A101357 (L.M. and W.I.)

NSF Division of Environmental Biology grant 1950770 (M.D. and W.I.)

Department of Energy - Joint Genome Institute grant CSP 1675 (W.I.)

National Library of Medicine (K.M.S. and E.V.K.)

National Institutes of Health grant 1R01-HG009761 (FZ)

National Institutes of Health grant 1DP1-HL141201 (FZ)

Howard Hughes Medical Institute (FZ)

Open Philanthropy Project (FZ)

Harold G. and Leila Mathers Foundation (FZ)

Edward Mallinckrodt, Jr. Foundation (FZ)

Poitras Center for Psychiatric Disorders Research at MIT (FZ)

Hock E. Tan and K. Lisa Yang Center for Autism Research at MIT (FZ)

Yang-Tan Center for Molecular Therapeutics at MIT (FZ)

Phillips family (FZ)

R. Metcalfe (FZ)

J. and P. Poitras (FZ)

## Data and materials availability:

Sequences of genes used in the experimental studies will be made available via online sequence repositories and expression plasmids will be available from Addgene under a uniform biological material transfer agreement. Raw reads from microbial small RNA-seq are available on SRA under BioProject PRJNA744508. Scripts for data analysis and visualization are available at Zenodo (24). Additional information available via the Zhang Lab website (https://zhanglab.bio).

## References and Notes


1. Zhang F, Development of CRISPR-Cas systems for genome editing and beyond. Quarterly Reviews of Biophysics. 52 (2019).
2. Hille F, Richter H, Wong SP, Bratovič M, Ressel S, Charpentier E, The biology of CRISPR-Cas: Backward and forward. Cell. 172, 1239–1259 (2018). [PubMed: 29522745]
3. Kapitonov VV, Makarova KS, Koonin EV, ISC, a novel group of bacterial and Archaeal DNA transposons that encode Cas9 homologs. J. Bacteriol 198, 797–807 (2015). [PubMed: 26712934]
4. Siguier P, Gourbeyre E, Chandler M, Bacterial insertion sequences: their genomic impact and diversity. FEMS Microbiol. Rev 38, 865–891 (2014). [PubMed: 24499397]
5. Shmakov S, Smargon A, Scott D, Cox D, Pyzocha N, Yan W, Abudayyeh OO, Gootenberg JS, Makarova KS, Wolf YI, Severinov K, Zhang F, Koonin EV, Diversity and evolution of class 2 CRISPR-Cas systems. Nat. Rev. Microbiol 15, 169–182 (2017). [PubMed: 28111461]
6. Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, Volz SE, Joung J, van der Oost J, Regev A, Koonin EV, Zhang F, Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell. 163, 759–771 (2015). [PubMed: 26422227]
7. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E, CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature. 471, 602–607 (2011). [PubMed: 21455174]
8. Kalvari I, Nawrocki EP, Ontiveros-Palacios N, Argasinska J, Lamkiewicz K, Marz M, Griffiths-Jones S, Toffano-Nioche C, Gautheret D, Weinberg Z, Rivas E, Eddy SR, Finn RD, Bateman A, Petrov AI, Rfam 14: expanded coverage of metagenomic, viral and microRNA families. Nucleic Acids Res. 49, D192–D200 (2021). [PubMed: 33211869]
9. Weinberg Z, Perreault J, Meyer MM, Breaker RR, Exceptional structured noncoding RNAs revealed by bacterial metagenome analysis. Nature. 462, 656–659 (2009). [PubMed: 19956260]
10. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E, A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 337, 816–821 (2012). [PubMed: 22745249]
11. Gasiunas G, Barrangou R, Horvath P, Siksnys V, Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc. Natl. Acad. Sci. U. S. A 109, E2579–86 (2012). [PubMed: 22949671]
12. Gasiunas G, Young JK, Karvelis T, Kazlauskas D, Urbaitis T, Jasnauskaite M, Grusyte MM, Paulraj S, Wang P-H, Hou Z, Dooley SK, Cigan M, Alarcon C, Chilcoat ND, Bigelyte G, Curcuru JL, Mabuchi M, Sun Z, Fuchs RT, Schildkraut E, Weigele PR, Jack WE, Robb GB, Venclovas Č, Siksnys V, A catalogue of biochemically diverse CRISPR-Cas9 orthologs. Nat. Commun 11, 5512 (2020). [PubMed: 33139742]
13. Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E, Anders C, Hauer M, Zhou K, Lin S, Kaplan M, Iavarone AT, Charpentier E, Nogales E, Doudna JA, Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. Science. 343, 1247997 (2014). [PubMed: 24505130]
14. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R, IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol. Biol. Evol 37, 1530–1534 (2020). [PubMed: 32011700]

15. Makarova KS, Wolf YI, Iranzo J, Shmakov SA, Alkhnbashi OS, Brouns SJJ, Charpentier E, Cheng D, Haft DH, Horvath P, Moineau S, Mojica FJM, Scott D, Shah SA, Siksnys V, Terns MP, Venclovas Č, White MF, Yakunin AF, Yan W, Zhang F, Garrett RA, Backofen R, van der Oost J, Barrangou R, Koonin EV, Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. Nat. Rev. Microbiol 18, 67–83 (2020). [PubMed: 31857715]
16. Majorek KA, Dunin-Horkawicz S, Steczkiewicz K, Muszewska A, Nowotny M, Ginalski K, Bujnicki JM, The RNase H-like superfamily: new members, comparative structural analysis and evolutionary classification. Nucleic Acids Res. 42, 4160–4179 (2014). [PubMed: 24464998]
17. Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F, Nureki O, Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell. 156, 935–949 (2014). [PubMed: 24529477]
18. Bao W, Jurka J, Homologues of bacterial TnpB_IS605 are widespread in diverse eukaryotic transposable elements. Mob. DNA 4, 12 (2013). [PubMed: 23548000]
19. Gomes-Filho JV, Zaramela LS, da S VC. Italiani, N. S. Baliga, R. Z. N. Vêncio, T. Koide, Sense overlapping transcripts in IS1341-type transposase genes are functional non-coding RNAs in archaea. RNA Biol. 12, 490–500 (2015). [PubMed: 25806405]
20. Weinberg Z, Lünse CE, Corbino KA, Ames TD, Nelson JW, Roth A, Perkins KR, Sherlock ME, Breaker RR, Detection of 224 candidate structured RNAs by comparative analysis of specific subsets of intergenic regions. Nucleic Acids Res. 45, 10811–10823 (2017). [PubMed: 28977401]
21. Hüttenhofer A, Schattner P, The principles of guiding by RNA: chimeric RNA-protein enzymes. Nat. Rev. Genet 7, 475–482 (2006). [PubMed: 16622413]
22. Schneider A, A short history of guide RNAs: The intricate path that led to the discovery of a basic biological concept. EMBO Rep. 21, e51918 (2020). [PubMed: 33289251]
23. Koonin EV, Evolution of RNA- and DNA-guided antivirus defense systems in prokaryotes and eukaryotes: common ancestry vs convergence. Biol. Direct 12 (2017).
24. Altae-Tran H, Kannan S, Zhang F, Code and processed data for: The widespread IS200/605 transposon family encodes diverse programmable RNA-guided endonucleases, Version 1.0, Zenodo (2021); <link to be inserted after official acceptance of manuscript>.
25. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL, BLAST+: architecture and applications. BMC Bioinformatics. 10, 421 (2009). [PubMed: 20003500]
26. Katoh K, Standley DM, MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol 30, 772–780 (2013). [PubMed: 23329690]
27. Steinegger M, Söding J, MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat. Biotechnol 35, 1026–1028 (2017). [PubMed: 29035372]
28. Steinegger M, Meier M, Mirdita M, Vöhringer H, Haunsberger SJ, Söding J, HH-suite3 for fast remote homology detection and deep protein annotation. BMC Bioinformatics. 20, 473 (2019). [PubMed: 31521110]
29. Eddy SR, Accelerated Profile HMM Searches. PLoS Comput. Biol 7, e1002195 (2011). [PubMed: 22039361]
30. Shmakov SA, Makarova KS, Wolf YI, Severinov KV, Koonin EV, Systematic prediction of genes functionally linked to CRISPR-Cas systems by gene neighborhood analysis. Proc. Natl. Acad. Sci. U. S. A 115, E5307–E5316 (2018). [PubMed: 29784811]
31. Crawley AB, Henriksen JR, Barrangou R, CRISPRdisco: An Automated Pipeline for the Discovery and Analysis of CRISPR-Cas Systems. CRISPR J. 1, 171–181 (2018). [PubMed: 31021201]
32. Haft DH, Loftus BJ, Richardson DL, Yang F, Eisen JA, Paulsen IT, White O, TIGRFAMs: a protein family resource for the functional identification of proteins. Nucleic Acids Res. 29, 41–43 (2001). [PubMed: 11125044]
33. Delcher A, Improved microbial gene identification with GLIMMER. Nucleic Acids Research. 27 (1999), pp. 4636–4641. [PubMed: 10556321]
34. Naser-Khdour S, Minh BQ, Zhang W, Stone EA, Lanfear R, The Prevalence and Impact of Model Violations in Phylogenetic Analysis. Genome Biol. Evol 11, 3341–3352 (2019). [PubMed: 31536115]

35. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS, ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589 (2017). [PubMed: 28481363]
36. Price MN, Dehal PS, Arkin AP, FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One. 5, e9490 (2010). [PubMed: 20224823]
37. Altekar G, Dwarkadas S, Huelsenbeck JP, Ronquist F, Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. Bioinformatics. 20, 407–415 (2004). [PubMed: 14960467]
38. Stamatakis A, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30, 1312–1313 (2014). [PubMed: 24451623]
39. Lorenz R, Bernhart SH, zu Siederdissen CH, Tafer H, Flamm C, Stadler PF, Hofacker IL, ViennaRNA Package 2.0. Algorithms Mol. Biol 6, 1–14 (2011). [PubMed: 21235792]
40. Rivas E, Clements J, Eddy SR, A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. Nat. Methods 14, 45–48 (2017). [PubMed: 27819659]
41. Nawrocki EP, Eddy SR, Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics. 29, 2933–2935 (2013). [PubMed: 24008419]
42. Weinberg Z, Breaker RR, R2R--software to speed the depiction of aesthetic consensus RNA secondary structures. BMC Bioinformatics. 12, 3 (2011). [PubMed: 21205310]
43. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, Finn RD, Bateman A, Pfam: The protein families database in 2021. Nucleic Acids Res. 49, D412–D419 (2021). [PubMed: 33125078]
44. Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyrpides NC, Hugenholtz P, CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. BMC Bioinformatics. 8, 209 (2007). [PubMed: 17577412]
45. Asnicar F, Weingart G, Tickle TL, Huttenhower C, Segata N, Compact graphical representation of phylogenetic data and metadata with GraPhlAn. PeerJ. 3, e1029 (2015). [PubMed: 26157614]
46. Maaten L, Hinton G, Visualizing Data using t-SNE. Journal of machine learning research. 9, 2579–2605 (2008).
47. Poličar PG, Stražar M, Zupan B, Embedding to reference t-SNE space addresses batch effects in single-cell classification. bioRxiv (2019), , doi:10.1101/671404.
48. Rousseeuw PJ, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math 20, 53–65 (1987).
49. Campello RJGB, Moulavi D, Sander J, in Advances in Knowledge Discovery and Data Mining (Springer Berlin Heidelberg, Berlin, Heidelberg, 2013), Lecture notes in computer science, pp. 160–172.
50. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW, CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25, 1043–1055 (2015). [PubMed: 25977477]
51. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH, GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics (2019), doi:10.1093/bioinformatics/btz848.
52. Frey S, Görlich D, A new set of highly efficient, tag-cleaving proteases for purifying recombinant proteins. J. Chromatogr. A 1337, 95–105 (2014). [PubMed: 24636565]
53. Martin M, Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, 10–12 (2011).
54. Langmead B, Salzberg SL, Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359 (2012). [PubMed: 22388286]
55. Sriramachandran AM, Petrosino G, Méndez-Lago M, Schäfer AJ, Batista-Nascimento LS, Zilio N, Ulrich HD, Genome-wide Nucleotide-Resolution Mapping of DNA Replication Patterns, Single-Strand Breaks, and Lesions by GLOE-Seq. Mol. Cell 78, 975–985.e7 (2020). [PubMed: 32320643]
56. Clement K, Rees H, Canver MC, Gehrke JM, Farouni R, Hsu JY, Cole MA, Liu DR, Joung JK, Bauer DE, Pinello L, CRISPResso2 provides accurate and rapid genome editing sequence analysis. Nat. Biotechnol 37, 224–226 (2019). [PubMed: 30809026]

57. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol 19, 455–477 (2012). [PubMed: 22506599]
58. Turmel M, Otis C, Lemieux C, Divergent copies of the large inverted repeat in the chloroplast genomes of ulvophycean green algae. Sci. Rep 7, 994 (2017). [PubMed: 28428552]
59. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS, UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol. Biol. Evol 35, 518–522 (2018). [PubMed: 29077904]
60. Lockhart PJ, Larkum AW, Steel M, Waddell PJ, Penny D, Evolution of chlorophyll and bacteriochlorophyll: the problem of invariant sites in sequence analysis. Proc. Natl. Acad. Sci. U. S. A 93, 1930–1934 (1996). [PubMed: 8700861]
61. Criscuolo A, Gribaldo S, BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol. Biol 10, 210 (2010). [PubMed: 20626897]
62. Strecker J, Ladha A, Gardner Z, Schmid-Burgk JL, Makarova KS, Koonin EV, Zhang F, RNA-guided DNA insertion with CRISPR-associated transposases. Science. 365, 48–53 (2019). [PubMed: 31171706]
63. Klompe SE, Vo PLH, Halpin-Healy TS, Sternberg SH, Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration. Nature. 571, 219–225 (2019). [PubMed: 31189177]
64. Krupovic M, Makarova KS, Forterre P, Prangishvili D, Koonin EV, Casposons: a new superfamily of self-synthesizing DNA transposons at the origin of prokaryotic CRISPR-Cas immunity. BMC Biol. 12, 1–12 (2014). [PubMed: 24417977]
65. Crotty SM, Minh BQ, Bean NG, Holland BR, Tuke J, Jermiin LS, Haeseler AV, GHOST: Recovering Historical Signal from Heterotachously Evolved Sequence Alignments. Syst. Biol 69, 249–264 (2020). [PubMed: 31364711]

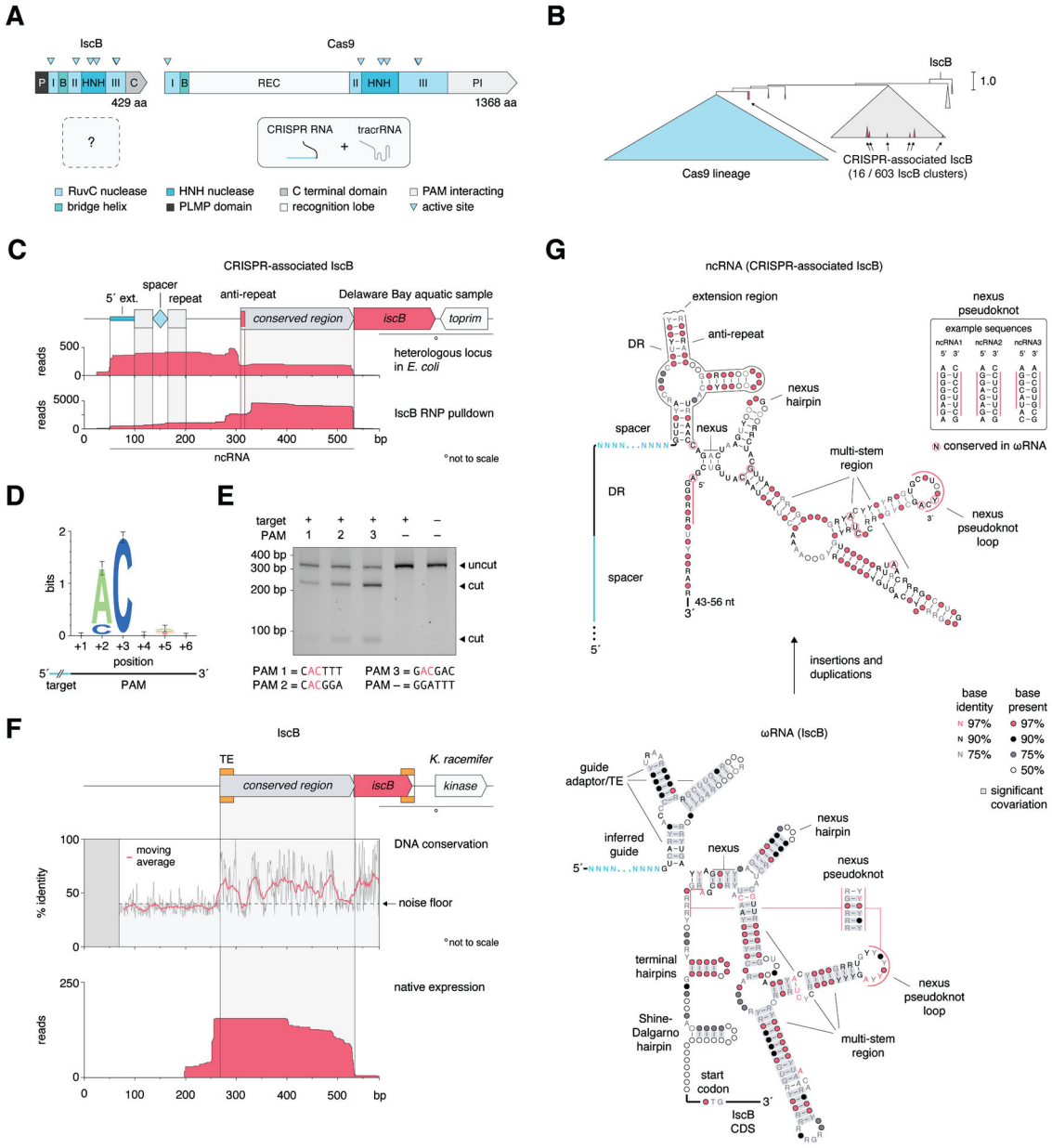


**Fig. 1. IscBs are associated with ncRNAs of unknown function.**

**(A)** Comparison of IscB and Cas9 domains and previously described ncRNAs.

**(B)** Phylogenetic analysis of the RuvC, BH, and HNH domains of Cas9 and IscB clusters using IQ-Tree 2. Genomic association shows 16/603 IscB clusters have strong association to CRISPR, occurring independently in multiple clades.

**(C)** Small RNA-seq of a heterologously expressed CRISPR-associated IscB locus (top) and RNP pulldown (bottom).

**(D)** Sequence logo for the PAM as determined by a plasmid depletion assay.

**(E)** *In vitro* cleavage by IscB-single guide RNA RNP complex.

**(F)** (Top) Conservation analysis of regions upstream of $N$=563 non-redundant IscB loci. (Bottom) Small RNA-seq of an IscB locus in *K. racemifer* strain SOSP1-21.

**(G)** Secondary structure predictions of CRISPR-associated IscB ncRNA and IscB ωRNA. Guiding function of ωRNAs was inferred by comparison of the two structures. TE: transposon end.

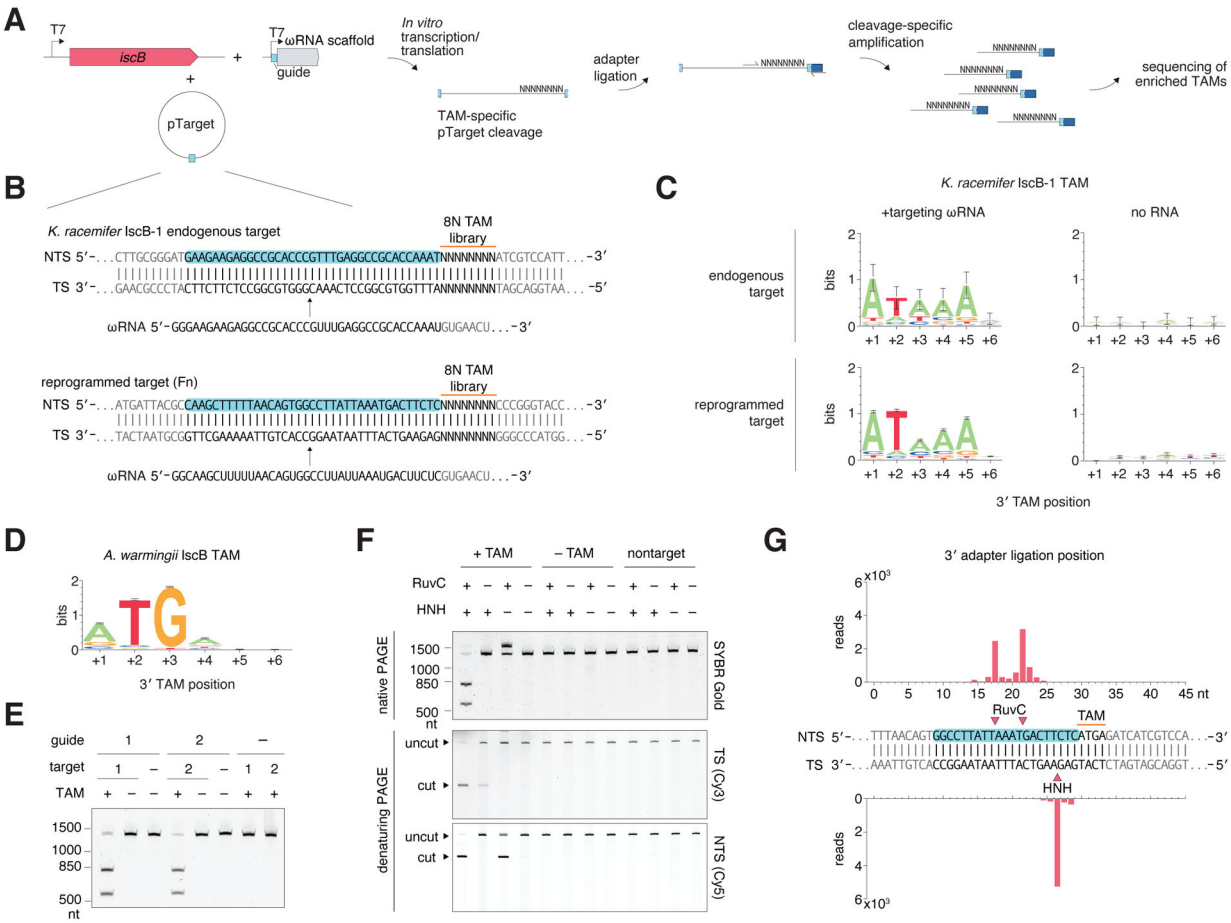


**Fig. 2. IscB is an RNA-guided DNA endonuclease.**

**(A)** Design of an IVTT-based TAM screen.

**(B)** KraIscB-1 endogenous target and reprogrammed target sequences used in IVTT TAM screens.

**(C)** dsDNA cleavage by KraIscB-1 and ωRNA targeting sequence flanked by ATAAA 3' TAM. **(D)** dsDNA clevage by AwaIscB and ωRNA targeting sequence flanked by ATGA 3' TAM.

**(E)** *In vitro*-reconstituted AwaIscB-ωRNA RNP cleavage of dsDNA substrates in the presence or absence of a target and/or TAM. TS: target strand; NTS: non-target strand; nt: nucleotides.

**(F)** *In vitro* dsDNA cleavage of AwaIscB with selectively inactivated nuclease domains.

**(G)** Sequencing of cleavage products generated by AwaIscB.

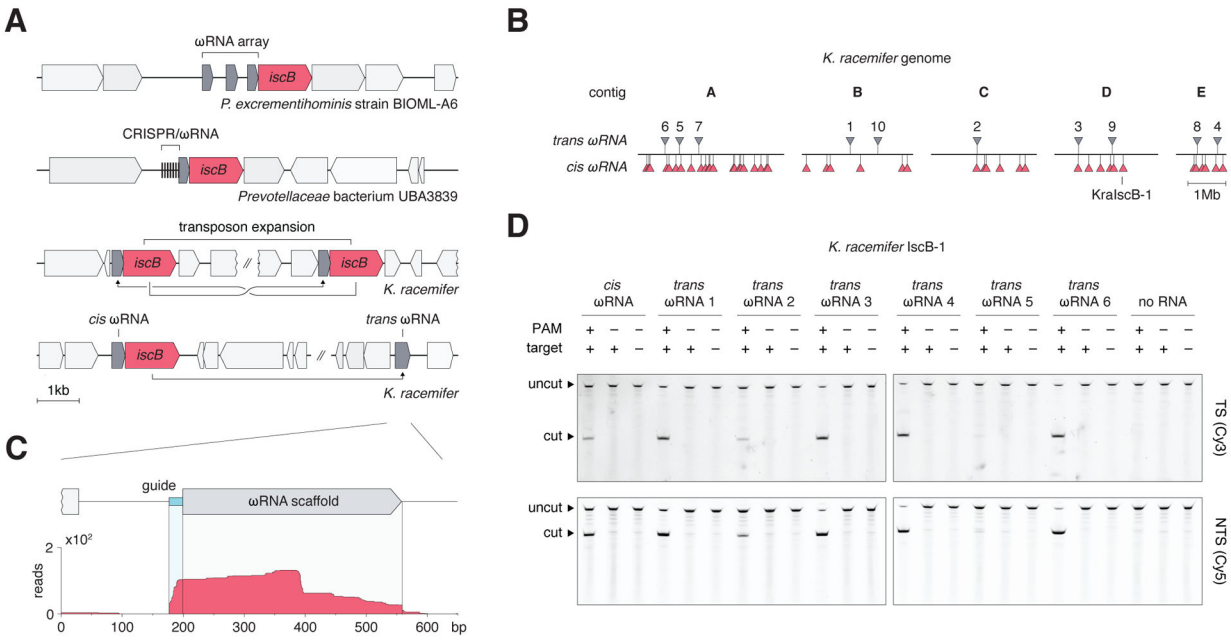


**Fig. 3. Guide-encoding mechanisms of IscB**

**(A)** Example loci for each major mechanism of encoding multiple guides: entire ωRNAs arrays associate with IscB, ωRNAs duplicate or insert into CRISPRs, transposition expansion results in multiple nearly identical loci that each express different guides, and standalone *trans*-acting ωRNAs form independently of adjacent IscBs.

**(B)** *K. racemifer* encodes 48 IscB loci with *cis* ωRNAs and 10 standalone *trans*-acting ωRNAs.

**(C)** Small RNA-seq of a standalone ωRNA locus in *K. racemifer*.

**(D)** KraIscB-1, in complex with *cis* or *trans* ωRNAs with the same guide sequence, mediate cleavage of dsDNA in a TAM- and target-dependent manner. Reactions were performed in IVTT using 5' strand-specific labeled linear targets. TS: Target strand; NTS: non-target strand. Contig accession and position information for all displayed loci are listed in Table S6.

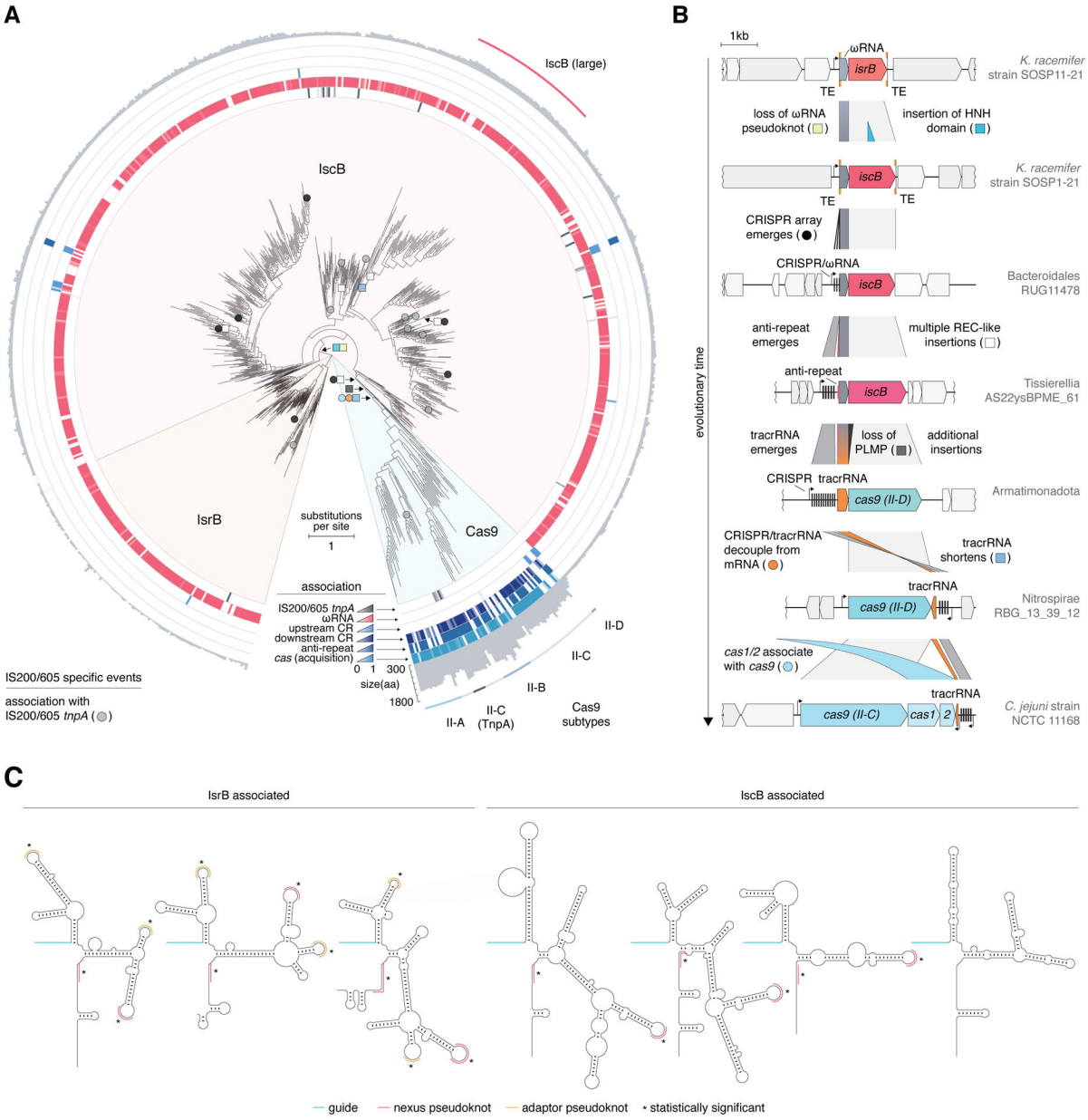


**Fig. 4. Diversity and evolution of IscB**

**(A)** Phylogenetic tree of IsrB, IscB, and Cas9. Associations with IS200/605 TnpA, ωRNA, CRISPR arrays, anti-repeats (where applicable), and *Cas* acquisition genes. ORF size of cluster representative is shown on the second outermost ring. Notable groups are shown as colored arcs on the outermost ring. First occurrences of evolutionary events in each clade are marked by colored circles/squares, as described in (B). CR: CRISPR array.

**(B)** Parsimonious evolutionary timeline linking IsrB to Cas9 with exemplifying loci. Colors of protein of interest indicate distinct stages in the evolution of IsrB to Cas9.

**(C)** Structural diversity and evolution of ωRNAs in IsrB and IscB systems.


**Fig. 5. Exploration of the diversity of IS200/605 superfamily nucleases**

**(A)** Evolution between IS200/605 transposon superfamily-encoded nucleases and associated RNAs. Dashed lines reflect tentative/unknown relationships.

**(B)** Locations of IscB loci and fragments in the *I. tetrasporus* genome. Intact locus is labeled as “ChlorIscB.”

**(C)** Small RNA-seq of *I. tetrasporus*.

**(D)** Weblogo of ChlorIscB cleavage TAM using a reprogrammed guide in an IVTT TAM screen.

**(E)** Weblogo of OgeuIscB TAM using a reprogrammed guide in an IVTT TAM screen.

**(F)** Targeted OgeuIscB-mediated indel formation at the *VEGFA* locus in HEK293FT cells ordered by abundance, with indel size on the left.

**(G)** OgeuIscB-mediated indel formation at multiple sites in HEK293T cells, $*P < 0.05$).
**(H)** Small RNA-seq of ωRNA from IsrB locus in *K. racemifer* strain SOSP1-21.
**(I)** Weblogo of *Desulfovigula thermocuniculi* (DthIsrB) TAM using a reprogrammed guide in an IVTT TAM screen.
**(J)** DthIsrB mediates ωRNA-guided non-target strand nicking in a TAM- and target-dependent manner in an IVTT cleavage assay using 5' strand-specific labeled targets.
**(K)** Small RNA-seq of ωRNA from TnpB locus in *K. racemifer* strain SOSP1-21.
**(L)** Comparison of ωRNAs from *K. racemifer* IscB and TnpB loci.
**(M)** Secondary structure prediction of KraTnpB-associated ωRNA.
**(N)** Weblogo of *A. macrosporangiidus* TnpB (AmaTnpB) TAM using a reprogrammed guide in an IVTT TAM screen.
**(O)** *In vitro*-reconstituted AmaTnpB cleavage of dsDNA substrates in the presence or absence of ωRNA, target, and/or TAM.
**(P)** AmaTnpB performs ωRNA-guided TAM-independent target-dependent cleavage of 3' Cy5.5-labeled ssDNA substrates.
**(Q)** AmaTnpB cleaves a 3' Cy5.5-labeled collateral ssDNA substrate in the presence of TAM- and target-containing dsDNA or target-containing ssDNA substrates. Contig accession and position information for all displayed loci are listed in Table S6.

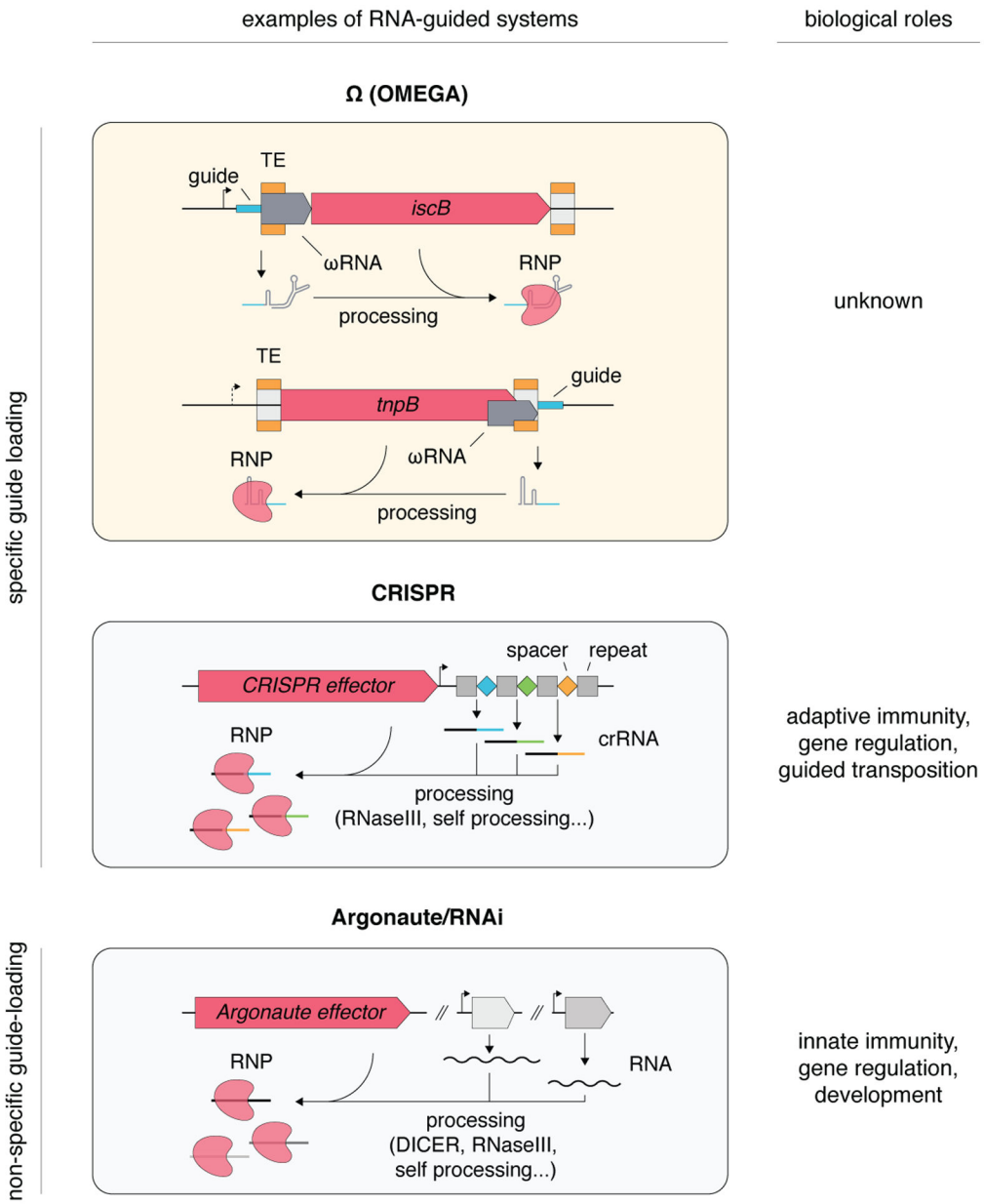


**Fig. 6. Naturally occurring RNA-guided DNA-targeting systems**

Comparison of Ω (OMEGA) systems with other known RNA-guided systems. In contrast to CRISPR systems, which capture spacer sequences and store them in the locus within the CRISPR array, Ω systems may transpose their loci (or *trans*-acting loci) into target sequences, converting targets into ωRNA guides.