Supplementary Note

Supplementary Data 1 | Sequence alignment of prokaryotic Argonautes. We clustered 487 pAgo proteins identified in Refseq by sequence similarity and selected a nonredundant representative set (261 pAgos and 8 selected eAgos). The MUSCLE program[109] was used for construction of multiple sequence alignment followed by minor manual curation based on the PSI-BLAST and HHpred alignments. The region corresponds to the most conserved MID and PIWI domains. Sequences in the alignment are shown in the order corresponding to the tree presented in **Supplementary Data 2**. Each sequence is labelled with GenBank Identifier (GI) number and full systematic name of the host. Names of pAgos discussed in this manuscript are colored red, while eAgos are colored blue. The motifs that include catalytic residues are highlighted in yellow.

Supplementary Data 2 | Complete phylogenetic tree of prokaryotic Argonautes. Maximum-likelihood phylogenetic unrooted trees was built using the FastTree program[43] using a multiple alignment of 145 confidently aligned positions of MID and PIWI domains (**Supplementary Data 1**). The same program was also used to compute bootstrap values indicated for all branches. Color code: dark blue, Eukaryota; orange, Archaea; dark green, Bacteria; Collapsed branches are shown as triangles of the corresponding color. pAgos discussed in this manuscript are colored red. Each protein is labelled with GenBank Identifier (GI) number (if available), full systematic host name and PDB accession code (if available). Two parts of the tree, long Agos and short pAgos, are shown on the left and right panel respectively. Genes that are present in the predicted operons (or fused) with pAgo are shown on the left side of the Tree. Abbreviations: Sir2 1 and Sir2 2 are two distinct families of the predicted Sir2 like nuclease, RE1 and RE2 are two distinct families of restriction endonuclease superfamily. Other domains: TIR – predicted nuclease of TIP family, Schlafen – predicted ATPase, APAZ – “analog of PAZ” domain, Cas4 – Cas4 subfamily of restriction endonuclease superfamily, PLD – predicted nuclease of phospholipase D superfamily.

Supplementary Data 3 | Sequence alignment of eukaryotic Argonautes. The MUSCLE program[109] was used for construction of multiple sequence alignment. The region corresponds to the conserved PIWI domain and was used for the tree reconstruction (**Supplementary Data 4**). Each sequence is labelled with GenBank Identifier (GI) number (if available), protein name (if available) and full systematic name of the host.

Supplementary Data 4 | Complete phylogenetic tree of eukaryotic Argonautes. Maximum-likelihood phylogenetic unrooted trees was built using the FastTree program[43] using a multiple alignment of 177 confidently aligned positions the PIWI domain (**Supplementary Data 3**). The same program was also used to compute bootstrap values indicated for all branches. Color code: grey, Eukaryota; orange, Archaea; dark green, Bacteria. Each protein is labelled with Genbank Identifier (GI) number (if available), protein name (if available), and full systematic host name.

References

109. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. **32**, 1792–1797 (2004).