ENDORIBONUCLEASE COMPOSITIONS AND METHODS OF USE THEREOF

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Patent Application No. 61/333,163, filed May 10, 2010, U.S. Provisional Patent Application No. 61/365,627, filed July 19, 2010, and U.S. Provisional Patent Application No. 61/413,287, filed November 12, 2010, each of which applications is incorporated herein by reference in its entirety.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

[0002] This invention was made with government support under Grant No. T32 GM07232 awarded by the National Institutes of Health and Grant No. MCB-0950971 awarded by the National Science Foundation. The government has certain rights in the invention.

BACKGROUND

[0003] DNA restriction enzymes transformed molecular biology in the 1970s by making it possible to cleave specific DNA sequences at will. Sequencing of RNA molecules currently entails copying the RNA into a DNA strand that is then sequenced by conventional methods. This approach, also known as RNASeq, is robust and can yield many millions of sequence reads. However, the necessity of generating cDNA introduces inherent bias due to sequence-dependent efficiencies of individual steps.

Literature

[0004] Carte et al. (2008) *Genes Dev.* 22:3489; U.S. Patent Publication No. 2010/0093026.

SUMMARY

[0005] The present disclosure provides variant Csy4 endoribonucleases, nucleic acids encoding the variant Csy4 endoribonucleases, and host cells genetically modified with the nucleic acids. The variant Csy4 endoribonucleases find use in a variety of applications, which are also provided. The present disclosure also provides methods of

detecting a specific sequence in a target polyribonucleotide; and methods of regulating production of a target RNA in a eukaryotic cell.

[0005a] In one aspect, there is provided a variant Csy4 endoribonuclease comprising an amino acid sequence having at least about 95% amino acid sequence identity to the amino acid sequence set forth in SEQ ID NO: 8, wherein the variant Csy4 endoribonuclease comprises an amino acid substitution at His-29, wherein the variant Csy4 endoribonuclease is enzymatically inactive in the absence of imidazole, and wherein the variant Csy4 endoribonuclease is activatable in the presence of imidazole.

[0005b] In another aspect, there is provided a nucleic acid comprising a nucleotide sequence encoding the variant Csy4 endoribonuclease as disclosed herein.

[0005c] In another aspect, there is provided a recombinant expression vector comprising a nucleotide sequence encoding the variant Csy4 endoribonuclease as disclosed herein.

[0005d] In another aspect, there is provided an *in vitro* genetically modified host cell comprising the recombinant expression vector as disclosed herein.

[0005e] In another aspect, there is provided a kit for purifying a target RNA present in a mixed population of nucleic acids, the kit comprising: the variant Csy4 endoribonuclease as disclosed herein.

[0005f] In another aspect, there is provided a method of isolating a target RNA present in a mixed population of nucleic acids, the method comprising:

a) contacting a mixed population of nucleic acids with the variant Csy4 endoribonuclease as disclosed herein, where the variant Csy4 endoribonuclease is immobilized on an insoluble support, wherein the mixed population of nucleic acids comprises a tagged target RNA comprising a recognition nucleotide sequence that is specifically bound by the immobilized variant Csy4 endoribonuclease, forming a tagged target RNA-immobilized variant Csy4 endoribonuclease complex, wherein said contacting is carried out in a binding solution lacking imidazole;

b) adding imidazole to the binding solution to a final concentration of from about 100 mM to about 500 mM, thereby forming a reactivation solution that enzymatically reactivates the immobilized variant Csy4 endoribonuclease, wherein the reactivated immobilized variant Csy4 endoribonuclease cleaves the target RNA from the tag; and

c) collecting the released target RNA.

Brief Description Of The Drawings

[0006] Figures 1A-C depict specific recognition of a pre-crRNA substrate by Pa14Csy4. The nucleotide sequence depicted is 5'-GUUCACUGCCGUAUAGGCAGCUAAGAAA-3' (SEQ ID NO:1).

[0007] Figures 2A-C depict crystal structures of Csy4 bound to RNA substrate.

[0008] Figures 3A and 3B depict: a detailed view of the catalytic center of Csy4 (Figure 3A); and cleavage activity of Csy4 wild-type (WT) and mutants (Figure 3B).

[0009] Figure 4 depicts invariant amino acids among 12 Csy4 sequences. Pa (SEQ ID NO:8); Yp (SEQ ID NO:34); Ec89 (SEQ ID NO:39); Dn (SEQ ID NO:79); Ab (SEQ ID NO:84); MP1 (SEQ ID NO:2); MP01 (SEQ ID NO:3); SW (SEQ ID NO:4); Pm (SEQ ID NO:85); Pw (SEQ ID NO:13); and Dd (SEQ ID NO:10).

[0010] Figures 5A-5BD present an amino acid sequence alignment of various Csy4 polypeptides, as well as the nucleotide sequences of RNA sequences recognized by each Csy4 polypeptide.

[0011] Figure 6 depicts examples of amino acid sequences of enzymatically inactive, sequence-specific endoribonucleases.

[0012] Figure 7 depicts an example of a method for detecting a specific sequence in a target polyribonucleotide.

[0013] Figure 8 depicts the effect of imidazole on activation of various enzymatically inactive Csy4 variants.

[0014] Figure 9 depicts an exemplary method of isolating a target RNA. A Csy4 target stem-loop (SEQ ID NO:103) is shown.

[0015] Figure 10 depicts an exemplary method of regulating expression of a target RNA in a eukaryotic cell. A Csy4 RNA substrate sequence (SEQ ID NO:103) is shown.

Definitions

[0016] As used herein, "polyribonucleotide" refers to a polymeric form of ribonucleotides, and includes RNA, RNA containing deoxyribonucleotide(s), and DNA containing ribonucleotide(s). A polyribonucleotide can in some cases include one or