

557.e516, (2018)) in yeast. Despite tremendous promise, these applications suffered from lower-than-expected efficiency and context-restriction, which likely stem from elements in the endogenous form of the retron. These include (1) a branched structure with a phosphodiester bond linking the 5' end of the ssDNA to a 2' hydroxyl of the msr RNA, (2) invariant flanking regions that may be required for retron reverse transcription, but are not part of the repair template, (3) limited total length, and (4) a native poly T stretch that functions as a terminator for Pol III transcription.

SUMMARY

Engineered retrons, modified to enhance production of multicopy single-stranded DNA (msDNA), are provided that solve many of the existing problems relating to efficiency and low copy numbers. Also described herein are vector systems encoding such engineered retrons and methods of using engineered retrons and vector systems in various applications such as CRISPR/Cas-mediated genome editing, recombineering, cellular barcoding, and molecular recording.

In one aspect, the present disclosure provides an engineered retron comprising:

- a) a pre-msr sequence;
- b) an *msr* gene encoding multicopy single-stranded RNA (msRNA);
- c) an *msd* gene encoding multicopy single-stranded DNA (msDNA);
- d) a post-msd sequence comprising a self-complementary region having sequence complementarity to the pre-msr sequence, wherein the self-complementary region has a length 1 to 50 nucleotides longer than a wild-type complementary region such that the engineered retron is capable of enhanced production of the msDNA; and
- e) a *ret* gene encoding a reverse transcriptase.

In one ~~aspect~~example, an engineered retron is provided, the engineered retron comprising: a) a pre-msr sequence; b) an *msr* gene encoding multicopy single-stranded RNA (msRNA); c) an *msd* gene encoding multicopy single-stranded DNA (msDNA); d) a post-msd sequence comprising a self-complementary region having sequence complementarity to the pre-msr sequence, wherein the self-complementary region has a length of at least 1 to 50 nucleotides longer than a wild-type complementary region such that the engineered retron is capable of enhanced production of the msDNA; and e) a *ret* gene encoding a reverse transcriptase.

The self-complementary region is formed by hydrogen bonding between the 3' and 5' ends of the ncRNA. In certain ~~embodiment~~examples, the complementary region has a length that is at least 1, at least 2, at least 4, at least 6, at least 8, at least 10, at least 12, at least 14, at least 16, at least 18, at least 20, at least 30, at least 40, or at least 50 nucleotides longer than the wild-type complementary region. For example, the self-complementary region may have a length ranging from 1 to 50 nucleotides longer than the wild-type complementary region, including any length within this range, such as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 nucleotides longer. In certain ~~embodiment~~examples, the self-complementary region

has a length ranging from 1 to 16 nucleotides longer than the wild-type complementary region.

In certain [embodimentexamples](#), the *msr* gene and the *msd* gene are provided in a trans arrangement or a cis arrangement. In some [embodimentexamples](#), the *ret* gene is provided in a trans arrangement with respect to the *msr* gene and/or the *msd* gene.

In certain [embodimentexamples](#), the *msr* gene, *msd* gene, and *ret* gene are derived from a bacterial retron including, without limitation, a myxobacteria retron (e.g., Mx65, Mx162), an *Escherichia coli* retron (e.g., E67, Ec73, EC83, EC86, EC107), a *Salmonella enterica* retron (e.g., msDNA-St85) and a *Vibrio cholerae* retron (e.g., Vc81, Vc95, Vc137).

In certain [embodimentexamples](#), the engineered retron further comprises a heterologous sequence of interest. The heterologous sequence may be inserted, for example, into the *msr* gene or the *msd* gene. For example, the heterologous sequence can be inserted into the loop of the *msd* stem loop. In some [embodimentexamples](#), the heterologous sequence encodes a polypeptide or peptide. In other [embodimentexamples](#), the heterologous sequence encodes a donor polynucleotide comprising a 5' homology arm that hybridizes to a 5' genomic target sequence and a 3' homology arm that hybridizes to a 3' genomic target sequence flanking a nucleotide sequence comprising an intended edit to be integrated at a genomic target locus by homology directed repair (HDR) or recombineering. In yet other [embodimentexamples](#), the heterologous sequence comprises a CRISPR protospacer DNA sequence. In one [embodimentexample](#), the CRISPR protospacer DNA sequence comprises a modified "AAG" protospacer adjacent motif (PAM).

In certain [embodimentexamples](#), the engineered retron further comprises a barcode sequence. The barcode sequence may be located, for example, in a hairpin loop of the msDNA.

In one aspect, the present disclosure provides a vector system comprising one or more vectors comprising the engineered retron disclosed herein.

In another [aspectexample](#), a vector system is provided, the vector system comprising one or more vectors comprising an engineered retron described herein. In certain [embodimentexamples](#), the *msr* gene and the *msd* gene are provided by the same vector or different vectors. In some [embodimentexamples](#), the *msr* gene, the *msd* gene, and the *ret* gene are provided by the same vector, wherein the vector comprises a promoter operably linked to the *msr* gene and the *msd* gene. In some [embodimentexamples](#), the promoter is further operably linked to the *ret* gene. In other [embodimentexamples](#), the vector further

comprises a second promoter operably linked to the *ret* gene. In certain embodimentexamples, the *msr* gene, the *msd* gene, and the *ret* gene are provided by different vectors.

In certain embodimentexamples, one or more of the vectors of the vector system are viral vectors or nonviral vectors (e.g., plasmids).

In certain embodimentexamples, the vector system comprises an engineered retron comprising a heterologous sequence encoding a donor polynucleotide comprising a 5' homology arm that hybridizes to a 5' genomic target sequence and a 3' homology arm that hybridizes to a 3' genomic target sequence flanking the donor polynucleotide sequence. The donor polynucleotide sequence can replace or edit a genomic target locus, for example, by homology directed repair (HDR) or recombineering.

In certain embodimentexamples, the vector system further comprises a vector encoding an RNA-guided nuclease. Exemplary RNA-guided nucleases include, without limitation, Cas nucleases (e.g., Cas9, Cpf1) and engineered RNA-guided FokI-nuclease.

In certain embodimentexamples, the vector system further comprises a vector encoding bacteriophage recombination proteins for recombineering. In some embodimentexamples, the vector is a replication-defective prophage encoding the bacteriophage recombination proteins.

In certain embodimentexamples, the vector system comprises an engineered retron comprising a heterologous sequence encoding a CRISPR protospacer DNA sequence. In some embodimentexamples, the vector system further comprises a vector encoding a Cas1 or Cas2 protein. In some embodimentexamples, the vector system further comprises a vector comprising a CRISPR array sequence.

In one aspect, the present disclosure provides an isolated host cell comprising the engineered retron disclosed herein or the vector system disclosed herein.

In another aspectexample, an isolated host cell is provided, the host cell comprising an engineered retron or a vector system described herein.

In certain embodimentexample, the host cell is a prokaryotic, archeon, or eukaryotic host cell. For example, the host cell may be a bacterial, protist, fungal, animal, or plant host cell. In some embodimentexamples, the host cell is a mammalian host cell. The host cell may be a human or nonhuman mammalian host cell. In other embodimentexamples, the host cell is an artificial cell or a genetically modified cell.

In one aspect, the present disclosure provides a kit comprising the engineered retron disclosed herein, the vector system disclosed herein, or the host cell disclosed herein.

In another aspectexample, a kit comprising an engineered retron, described herein, or a vector system or a host cell comprising such an engineered retron is provided. In

some embodiment examples, the kit further comprises instructions on methods of using the engineered retron.

In one aspect, the present disclosure provides a method of genetically modifying a cell comprising:

- a) transfecting a cell with the engineered retron disclosed herein;
- b) introducing an RNA-guided nuclease and guide RNA into the cell, wherein the RNA-guided nuclease forms a complex with the guide RNA, said guide RNAs directing the complex to the genomic target locus, wherein the RNA-guided nuclease creates a double-stranded break in the genomic DNA at the genomic target locus, and the donor polynucleotide generated by the engineered retron is integrated at the genomic target locus recognized by its 5' homology arm and 3' homology arm by homology directed repair (HDR) to produce a genetically modified cell.

5 In another aspect example, a method of genetically modifying a cell is provided. In some cases, the method includes transfecting a cell with an engineered retron. For example, the method can include: a) transfecting a cell with an engineered retron comprising a heterologous sequence encoding a donor polynucleotide comprising a 5' homology arm that hybridizes to a 5' genomic target sequence and a 3' homology arm that hybridizes to a 3' genomic target sequence flanking a nucleotide sequence comprising an intended edit to be integrated at a genomic target locus by
10 homology directed repair (HDR); and b) introducing an RNA-guided nuclease and guide RNA into the cell, wherein the RNA-guided nuclease forms a complex with the guide RNA, said guide RNAs directing the complex to the genomic target locus, wherein the RNA-guided nuclease creates a double-stranded break in the genomic DNA at the genomic target locus, and the donor polynucleotide generated by the
15 engineered retron is integrated at the genomic target locus recognized by its 5' homology arm and 3' homology arm by homology directed repair (HDR). HDR with an engineered retron encoding a donor polynucleotide can be used, for example, to create a gene replacement, gene knockout, deletion, insertion, inversion, or point mutation. In some cases, HDR with an engineered retron encoding a donor
20 polynucleotide can be used, for example, to repair a gene, gene knockout, deletion, insertion, inversion, or point mutation. Such methods can thereby create a genetically modified cell. In some embodiment examples, the method further comprises phenotyping the genetically modified cell or sequencing the genome of the genetically modified cell.

In one aspect, the present disclosure provides a method of genetically modifying a cell by recombineering, the method comprising:

- a) transfecting the cell with the engineered retron disclosed herein; and
- b) introducing bacteriophage recombination proteins into the cell, wherein the bacteriophage recombination proteins mediate homologous recombination at a target locus such that the donor polynucleotide generated by the engineered retron is integrated at the target locus recognized by its 5' homology arm and 3' homology arm to produce a genetically modified cell.

25 In another aspect example, a method of genetically modifying a cell by recombineering is provided, the method comprising: a) transfecting the cell with an engineered retron comprising a heterologous sequence encoding a donor polynucleotide comprising a 5' homology arm that hybridizes to a 5' genomic target sequence and a 3' homology arm that hybridizes to a 3' genomic target sequence
30 flanking a nucleotide sequence comprising an intended edit to be integrated at a genomic target locus by recombineering; and b) introducing bacteriophage recombination proteins into the cell, wherein the bacteriophage recombination proteins mediate homologous recombination at the target locus such that the donor polynucleotide generated by the engineered retron is integrated at the target locus
35 recognized by its 5' homology arm

and 3' homology arm to produce a genetically modified cell. Recombineering with an engineered retron encoding a donor polynucleotide can be used, for example, to create a gene replacement, gene knockout, deletion, insertion, inversion, or point mutation. In certain embodimentexamples, the donor polynucleotide is used to modify a plasmid, bacterial artificial chromosome (BAC), or a bacterial chromosome in a bacterial cell by recombineering. In some embodimentexamples, the method further comprises phenotyping the genetically modified cell or sequencing the genome of the genetically modified cell.

In certain embodimentexamples, the bacteriophage recombination proteins are introduced into a bacterial cell by insertion of a replication-defective λ prophage into the bacterial genome. In one embodimentexample, the bacteriophage comprises *exo*, *bet*, and *gam* genes.

In one aspect, the present disclosure provides a method of barcoding a cell comprising transfecting a cell with the engineered retron disclosed herein.

In another aspectexample, a method of barcoding a cell is provided, the method comprising transfecting a cell with an engineered retron comprising a barcode, as described herein.

In one aspect, the present disclosure provides a method of producing an *in vivo* molecular recording system comprising:

- a) introducing a Cas1 protein or a Cas2 protein of a CRISPR adaptation system into a host cell;
- b) introducing a CRISPR array nucleic acid sequence comprising a leader sequence and at least one repeat sequence into the host cell, wherein the CRISPR array nucleic acid sequence is integrated into genomic DNA or a vector in the host cell; and
- c) introducing a plurality of engineered retrons disclosed herein into the host cell, wherein each retron comprises a different protospacer DNA sequence that can be processed and inserted into the CRISPR array nucleic acid sequence.

In another aspectexample, a method of producing an *in vivo* molecular recording system is provided, the method comprising: a) introducing a Cas1 protein or a Cas2 protein of a CRISPR adaptation system into a host cell; b) introducing a CRISPR array nucleic acid sequence comprising a leader sequence and at least one repeat sequence into the host cell, wherein the CRISPR array nucleic acid sequence is integrated into genomic DNA or into a vector in the host cell; and c) introducing a plurality of engineered retrons comprising CRISPR protospacer DNA sequences into the host cell, wherein each retron comprises a different protospacer DNA sequence that can be processed and inserted into the CRISPR array nucleic acid sequence. In certain embodimentexamples, the Cas1 protein or the Cas2 protein are provided by a vector. In certain embodimentexamples, the engineered retron is provided by a vector. In certain embodimentexamples, the plurality of engineered retrons comprises at least three different protospacer DNA sequences.

In one aspect, the present disclosure provides an engineered cell comprising an *in vivo* molecular recording system comprising:

- a) a Cas1 protein or a Cas2 protein of a CRISPR adaptation system;
- b) a CRISPR array nucleic acid sequence comprising a leader sequence and at least one repeat sequence into the host cell, wherein the CRISPR array nucleic acid sequence is integrated into genomic DNA or a vector in the engineered cell; and
- c) a plurality of engineered retrons disclosed herein, wherein each retron comprises a different protospacer DNA sequence that can be processed and inserted into the CRISPR array nucleic acid sequence.

In another aspectexample, an engineered cell comprising an *in vivo* molecular recording system is provided, the engineered cell comprising: a) a Cas1 protein or a Cas2 protein of a CRISPR adaptation system; b) a CRISPR array nucleic acid sequence comprising a leader sequence and at least one repeat sequence into the host cell, wherein the CRISPR array nucleic acid sequence is integrated into genomic

DNA or a vector in the engineered cell; and c) a plurality of engineered retons, each comprising CRISPR protospacer DNA sequences, wherein each retron comprises a different protospacer DNA sequence that can be processed and inserted into the CRISPR array nucleic acid sequence. In certain embodimentexamples, the Cas1 protein or the Cas2 protein are provided by a vector. In certain embodimentexamples, the engineered retron is provided by a vector. In certain embodimentexamples, the plurality of engineered retons comprises at least three different protospacer DNA sequences.

In one aspect, the present disclosure provides a kit comprising the engineered cell disclosed herein and instructions for *in vivo* molecular recording.

In another aspectexample, a kit comprising an engineered cell comprising an *in vivo* molecular recording system, as described herein, is provided. In some embodimentexamples, the kit further comprises instructions for *in vivo* molecular recording.

In one aspect, the present disclosure provides a method of producing recombinant msDNA comprising:

- a) transfecting a host cell with the engineered retron disclosed herein or the vector system disclosed herein; and
- b) culturing the host cell under suitable conditions, wherein the msDNA is produced.

In another aspectexample, a method of producing recombinant msDNA is provided, the method comprising: a) transfecting a host cell with an engineered retron or vector system described herein; and b) culturing the host cell under suitable conditions, wherein the msDNA is produced.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGS. 1A-1D show schematics of retron operons and the potential uses of retons. **FIG. 1A** shows a schematic of a retron operon that encodes a msr, msd, and reverse transcriptase, where the reverse transcriptase can synthesize a DNA copy of a portion of the *msd* gene encoding multicopy single-stranded DNA. **FIG. 1B** illustrates that recombineering is a potential use of retons, where Beta can protect the ssDNA and can promote annealing of the ssDNA to a complementary ssDNA target, for example, a DNA target in a cell. **FIG. 1C** illustrates that CRISPR/Cas9 gene editing is a potential use of retons, where the retron can provide a ssDNA template that can repair a variant or mutant target site. **FIG. 1D** illustrates that molecular recording is a potential use of retons (e.g., as provided in WO2018191525 A1, which is specifically incorporated by reference herein in its entirety).

FIG. 2A-2B shows retron elements and their assembly. **FIG. 2A** shows retron elements: (1) the msr and the 5' end of the reverse-transcribed msd are covalently bonded to the priming guanosine via a 2'-5' linkage, this branched structure impede use in genome engineering, (2) invariant flanking regions that may be required for retron reverse transcription, and so they cannot easily be part of a repair template, (3) a stem that currently is thought to have a limited total length. Another issue for