METHODS FOR GENOMIC MODIFICATION

1. CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims benefit of priority of U.S. Provisional Application No. 61/479,821, filed on April 27, 2011; U.S. Provisional Application No. 61/500,741, filed on June 24, 2011; and U.S. Provisional Application No. 61/539,389, filed on September 26, 2011, the contents of which are hereby incorporated by reference in their entireties.

2. FIELD OF THE INVENTION

[0002] The methods and compositions provided herein generally relate to the fields of molecular biology and genetic engineering.

3. BACKGROUND

[0003] Genetic engineering techniques to introduce and integrate exogenous nucleic acids into a host cell genome are needed in a variety of fields. For example, in the field of synthetic biology, the fabrication of a genetically modified strain requires the insertion of customized DNA sequences into a chromosome of the host cell, and commonly, industrial scale production requires the introduction of dozens of genes into the host organism. Optimized designs for the industrial strain are arrived at empirically, requiring construction and *in vivo* testing of many DNA assemblies, alone and/or in concert with other biosynthetic pathway components.

[0004] Genetic engineering is highly reliant on gene targeting, which utilizes an extrachromosomal fragment of donor template DNA and invokes a cell's homologous recombination (HR) machinery to exchange a chromosomal sequence with an exogenous donor sequence. *See, e.g.,* Capecchi, *Science* 244:1288-1292 (1989). Gene targeting is limited in its efficiency; in plant and mammalian cells, only ~1 in 10^6 cells provided with excess template sequences undergo the desired gene modification. Yeast demonstrates an increased capacity for homologous recombination. However, the successful incorporation of exogenous DNA into yeast genomes is still a comparatively rare event (~1 in 10^5), and requires the use of a selectable marker to screen for recombinant cells which usually comprise only a single genomic modification. In addition, since only a limited cache of selectable markers are available for use in yeast, selectable marker(s) must be removed from a recombinant strain to allow for additional genomic modifications using the same markers, and in some instances, prior to releasing the host cell in a manufacturing or natural environment. Thus, independent of the efficiency at which integration can be achieved at any single locus, the one-at-a-time

serial nature of genomic engineering requires that making changes at multiple loci requires as many engineering cycles as there are loci to be modified.

[0005] The efficiency of gene targeting can be improved when combined with a targeted genomic double-stranded break (DSB) introduced near the intended site of integration. *See e.g.*, Jasin, M., *Trends Genet* 12(6):224-228 (1996); and Urnov *et al.*, *Nature* 435(7042):646-651 (2005). So called "designer nucleases" are enzymes that can be tailored to bind to a specific "target" sequence of DNA *in vivo* and introduce a double-strand break thereto. Such targeted double-strand breaks can be effected, for instance, by transforming a host cell with a plasmid containing a gene that encodes the designer nuclease. The host cell repairs these double-strand breaks by either homology-directed DNA repair or non-homologous end joining. In the course of the repair, either mechanism may be utilized to incorporate an exogenous donor DNA at the target site. If the nuclease is introduced into the cell at the same time as the donor DNA is introduced, the cell can integrate the donor DNA at the target loci.

[0006] The advent of designer nucleases has enabled the introduction of transgenes into particular target loci in crops (Wright *et al.*, *Plant J* 44:693-705 (2005)), to improve mammalian cell culture lines expressing therapeutic antibodies (Malphettes *et al.*, *Biotechnol Bioeng* 106(5):774-783 (2010)), and even to edit the human genome to evoke resistance to HIV (Urnov *et al.*, *Nat Rev Genet* 11(9):636-646 (2010)). While impactful, DSB-mediated HR has yet to be exploited to reduce the multiple rounds of engineering needed to integrate multiple DNA assemblies, for example, towards the construction of functional metabolic pathways in industrial microbes.

[0007] Thus, there exists a need for methods and compositions that allow for the simultaneous integration of a plurality of exogenous nucleic acids into specific regions of a host cell genome.

4. SUMMARY

[0008] Provided herein are methods and compositions for integrating one or more exogenous nucleic acids into specified genomic loci of a host cell. In some examples, a plurality of exogenous nucleic acids are simultaneously integrated with a single transformation reaction. In some examples, the methods comprise the introduction of one or more nucleases and one or more donor DNA assemblies into the cell to facilitate integration of the donor DNA at specified locations in the genome. The methods and compositions utilize the native homologous recombination machinery of the host cell, which recombination is further

enhanced by inducing targeted double-strand breaks in the host cell's genome at the intended sites of integration.

[0008a] In one aspect, the present invention refers to a method for simultaneously integrating a plurality of (n) exogenous nucleic acids into a plurality of (n) target sites of a host cell genome, wherein n is at least two, the method comprising:

(a) simultaneously contacting a host cell with:

(i) said plurality of exogenous nucleic acids, wherein:

x is an integer that varies from 1 to n, and for each integer x, each exogenous nucleic acid $(ES)_x$ comprises a first homology region $(HR1)_x$ and a second homology region $(HR2)_x$, wherein $(HR1)_x$ and $(HR2)_x$ are capable of initiating host cell mediated homologous recombination of $(ES)_x$ at a target site $(TS)_x$ selected from said plurality of (n) target sites of said host cell genome; and

(ii) for each said target site $(TS)_x$, a nuclease $(N)_x$ capable of cleaving at $(TS)_x$, whereupon said cleaving results in homologous recombination of $(ES)_x$ at $(TS)_x$;

and

(b) recovering a host cell wherein each exogenous nucleic acid $(ES)_x$ has integrated at its selected target site $(TS)_x$.

[0008b] In another aspect, the present invention refers to a host cell comprising:

(a) a plurality of (n) genomic target sites, wherein n is at least two;

(b) a plurality of (n) exogenous nucleic acids, wherein x is an integer that varies from 1 to n, and for each integer x, each exogenous nucleic acid $(ES)_x$ comprises a first homology region $(HR1)_x$ and a second homology region $(HR2)_x$, wherein $(HR1)_x$ and $(HR2)_x$ are capable of initiating host cell mediated homologous recombination of $(ES)_x$ at a target site $(TS)_x$ selected from said plurality of (n) genomic target sites; and

(c) for each said target site $(TS)_x$, a nuclease $(N)_x$ capable of cleaving at $(TS)_x$, whereupon said cleaving results in homologous recombination of $(ES)_x$ at $(TS)_x$.

[0008c] In yet another aspect, the present invention refers to a method for simultaneously integrating a plurality of (n) exogenous nucleic acids into a plurality of (n) target sites of a host cell genome, wherein n is at least two, the method comprising:

(a) simultaneously contacting a host cell with:

(i) a plurality of libraries, wherein x is an integer that varies from 1 to n, and for each integer x, each library $(L)_x$ comprises a plurality of exogenous nucleic acids, wherein each exogenous nucleic acid comprises, in a 5' to 3' orientation, a first homology region $(HR1)_x$, any nucleic acid of interest selected from the group $(D)_x$, and a second homology region $(HR2)_x$, wherein $(HR1)_x$ and $(HR2)_x$ are capable of initiating host cell mediated homologous recombination of said selected exogenous nucleic acid at a target site $(TS)_x$ of said host cell genome; and

(ii) for each said target site $(TS)_x$, a nuclease $(N)_x$ capable of cleaving at $(TS)_x$, whereupon said cleaving results in homologous recombination of an exogenous nucleic acid from $(L)_x$ at $(TS)_x$;

and

(b) recovering a host cell wherein an exogenous nucleic acid from each library $(L)_x$ has integrated at each selected target sequence $(TS)_x$.

[0008d] In a further aspect, the present invention refers to a host cell comprising:

(a) a plurality of (n) genomic target sites, wherein n is at least two;

(b) a plurality of libraries, wherein x is an integer that varies from 1 to n, and for each integer x, each library $(L)_x$ comprises a plurality of exogenous nucleic acids, wherein each exogenous nucleic acid comprises, in a 5' to 3' orientation, a first homology region $(HR1)_x$, any nucleic acid of interest selected from the group $(D)_x$, and a second homology region $(HR2)_x$, wherein $(HR1)_x$ and $(HR2)_x$ are capable of initiating host cell mediated homologous recombination of said selected exogenous nucleic acid at a target site $(TS)_x$ of said host cell genome; and

for each said target site $(TS)_x$, a nuclease $(N)_x$ capable of cleaving at $(TS)_x$, whereupon said cleaving results in homologous recombination of an exogenous nucleic acid from $(L)_x$ at $(TS)_x$.

[0009] Thus, in one example, disclosed herein is a method for integrating a plurality of exogenous nucleic acids into a host cell genome, the method comprising:

(a) contacting a host cell with:

(i) a plurality of exogenous nucleic acids, wherein each exogenous nucleic acid $(ES)_x$ comprises a first homology region $(HR1)_x$ and a second homology region $(HR2)_x$, wherein $(HR1)_x$ and $(HR2)_x$ are capable of initiating host cell mediated homologous recombination of $(ES)_x$ at a target site $(TS)_x$ of said host cell genome; and

(ii) for each said target site $(TS)_x$, a nuclease $(N)_x$ capable of cleaving at $(TS)_x$, whereupon said cleaving results in homologous recombination of $(ES)_x$ at $(TS)_x$; and

(b) recovering a host cell wherein each selected exogenous nucleic acid $(ES)_x$ has integrated at each selected target sequence $(TS)_x$,

wherein x is any integer from 1 to n wherein n is at least 2.

[0010] In some examples, $(HR1)_x$ is homologous to a 5' region of $(TS)_x$, and $(HR2)_x$, is homologous to a 3' region of $(TS)_x$.

[0011] In some examples, $(N)_x$ is capable of cleaving at a region positioned between said 5' and 3' regions of $(TS)_x$.

[0012] In some examples, a single nuclease is capable of cleaving each $(TS)_x$.

[0013] In some examples, n= 3, 4, 5, 6, 7, 8, 9 or 10. In some examples, n>10.

[0014] In some examples, said recovering does not require integration of a selectable marker. In some examples, said recovering occurs at a higher frequency as compared to not contacting the host cell with a nuclease capable of cleaving at said target site. In some examples, said recovering occurs at a frequency of about one every 10, 9, 8, 7, 6, 5, 4, 3, or 2 contacted host cells, or clonal populations thereof, screened. In some examples, said recovering comprises identifying said integrations by at least one method selected from the group consisting of PCR, Southern blot, restriction mapping, and DNA sequencing.

[0015] In some examples, $(N)_x$ is capable of cleaving an endogenous host genomic sequence, *e.g.*, a native loci within $(TS)_x$. In some examples, $(N)_x$ is capable of cleaving an exogenous sequence, *e.g.*, an introduced loci within $(TS)_x$.

[0016] In some examples, $(ES)_x$ further comprises a nucleic acid of interest $(D)_x$ positioned 3' of $(HR1)_x$ and 5' of $(HR2)_x$. In some examples, $(D)_x$ is selected from the

group consisting of a promoter, a nucleic acid sequence encoding an epitope tag, a gene of interest, a reporter gene, and a nucleic acid sequence encoding a termination codon.

[0017] In some examples, $(ES)_x$ is linear. In some examples, $(N)_x$ is provided as an expression vector comprising the nucleic acid sequence encoding $(N)_x$. In some examples, $(N)_x$ is transformed into the host cell as a purified protein. In some examples, $(N)_x$ is transformed into the host cell as purified RNA.

[0018] In some examples, the host cell comprises one or more heterologous nucleotide sequences encoding one or more enzymes of a biosynthetic pathway. In some examples, the one or more heterologous nucleotide sequences encoding one or more enzymes of a biosynthetic pathway are genomically integrated. In some examples, each exogenous nucleic acid $(ES)_x$ comprises a nucleic acid of interest $(D)_x$ positioned 3' of $(HR1)_x$ and 5' of $(HR2)_x$, encoding an enzyme of a biosynthetic pathway. In some examples, $(D)_x$ is a member of a library $(L)_x$ comprising a plurality of nucleic acid molecules that encode variants of an enzyme of a biosynthetic pathway.

[0019] In some examples, the host cell comprises one or more heterologous nucleotide sequences encoding one or more enzymes of a mevalonate (MEV) pathway for making isopentenyl pyrophosphate. In some examples, the one or more enzymes of the mevaloante pathway are selected from acetyl-CoA thiolase, HMG-CoA synthase, HMG-CoA reductase, mevalonate kinase, phosphomevalonate kinase and mevalonate pyrophosphate decarboxylase. In some examples, the host cell comprises a plurality of heterologous nucleic acids encoding all of the enzymes of a MEV pathway. In other words, the plurality of heterologous nucleic acids, taken together, encodes at least one enzyme of each class of enzymes of the MEV pathway listed above. In some examples, each exogenous nucleic acid $(ES)_x$ comprises a nucleic acid of interest $(D)_x$ positioned 3' of $(HR1)_x$ and 5' of $(HR2)_x$, encoding a terpene synthase. In some examples, the terpene synthase is selected from the group consisting of a monoterpene synthase, a diterpene synthase, a sesquiterpene synthase, a sesterterpene synthase, a triterpene synthase, a tetraterpene synthase, and a polyterpene synthase.

[0020] In some examples, $(N)_x$ is selected from the group consisting of an endonuclease, *e.g.,* a meganuclease, a zinc finger nuclease, a TAL-effector DNA binding domain-nuclease fusion protein (TALEN), a transposase, and a site-specific recombinase, wherein x is 1 or any integer from 1 to n. In some examples, the zinc finger nuclease is a fusion protein comprising the cleavage domain of a TypeIIS restriction endonuclease fused to an engineered zinc finger binding domain. In some examples, the TypeIIS restriction

endonuclease is selected from the group consisting of HO endonuclease and Fok I endonuclease. In some examples, the zinc finger binding domain comprises 3, 5 or 6 zinc fingers. In some examples, the endonuclease is a homing endonuclease selected from the group consisting of: an LAGLIDADG homing endonuclease, an HNH homing endonuclease, a His-Cys box homing endonuclease, a GIY-YIG homing endonuclease, and a cyanobacterial homing endonuclease. In some examples, the endonuclease is selected from the group consisting of: H-DreI, I-SceI, I-SceII, I-SceIII, I-SceIV, I-SceV, I-SceVI, I-SceVII, I-CeuI, I-CeuAIIP, I-CreI, I-CrepsbIP, I-CrepsbIIP, I-CrepsbIIIP, I-CrepsbIVP, I-TliI, I-PpoI, Pi-PspI, F-SceI, F-SceII, F-SuvI, F-CphI, F-TevI, F-TevII, I-AmaI, I-AniI, I-ChuI, I-CmoeI, I-CpaI, I-CpaII, I-CsmI, I-CvuI, I-CvuAIP, I-DdiI, I-DdiII, I-DirI, I-DmoI, I-HmuI, I-HmuII, I-HsNIP, I-LlaI, I-MsoI, I-NaaI, I-NanI, I-NclIP, I-NgrIP, I-NitI, I-NjaI, I-Nsp236IP, I-PakI, I-PboIP, I-PcuIP, I-PcuAI, I-PcuVI, I-PgrIP, I-PobIP, I-PorI, I-PorIIP, I-PbpIP, I-SpBetaIP, I-ScaI, I-SexIP, I-SneIP, I-SpomI, I-SpomCP, I-SpomIP, I-SpomIIP, I-SquIP, I-Ssp68031, I-SthPhiJP, I-SthPhiST3P, I-SthPhiSTe3bP, I-TdeIP, I-TevI, I-TevII, I-TevIII, i-UarAP, i-UarHGPAIP, I-UarHGPA13P, I-VinIP, I-ZbiIP, PI-MgaI, PI-MtuI, PI-MtuHIP PI-MtuHIIP, PI-PfuI, PI-PfuII, PI-PkoI, PI-PkoII, PI-Rma43812IP, PI-SpBetaIP, PI-SceI, PI-TfuI, PI-TfuII, PI-ThyI, PI-TliI, or PI-TliII. In particular examples, the endonuclease is Fcph-I.

[0021] In some examples, the endonuclease is modified to specifically bind an endogenous host cell genomic sequence, wherein the modified endonuclease no longer binds to its wild type endonuclease recognition sequence. In some examples, the modified endonuclease is derived from a homing endonuclease selected from the group consisting of: an LAGLIDADG homing endonuclease, an HNH homing endonuclease, a His-Cys box homing endonuclease, a GIY-YIG homing endonuclease, and a cyanobacterial homing endonuclease. In some examples, the modified endonuclease is derived from an endonuclease selected from the group consisting of: H-DreI, I-SceI, I-SceII, I-SceIII, I-SceIV, I-SceV, I-SceVI, I-SceVII, I-CeuI, I-CeuAIIP, I-CreI, I-CrepsbIP, I-CrepsbIIP, I-CrepsbIIIP, I-CrepsbIVP, I-TliI, I-PpoI, Pi-PspI, F-SceI, F-SceII, F-SuvI, F-CphI, F-TevI, F-TevII, I-AmaI, I-AniI, I-ChuI, I-CmoeI, I-CpaI, I-CpaII, I-CsmI, I-CvuI, I-CvuAIP, I-DdiI, I-DdiII, I-DirI, I-DmoI, I-HmuI, I-HmuII, I-HsNIP, I-LlaI, I-MsoI, I-NaaI, I-NanI, I-NclIP, I-NgrIP, I-NitI, I-NjaI, I-Nsp236IP, I-PakI, I-PboIP, I-PcuIP, I-PcuAI, I-PcuVI, I-PgrIP, I-PobIP, I-PorI, I-PorIIP, I-PbpIP, I-SpBetaIP, I-ScaI, I-SexIP, I-SneIP, I-SpomI, I-SpomCP, I-SpomIP, I-SpomIIP, I-SquIP, I-Ssp68031, I-SthPhiJP, I-SthPhiST3P, I-SthPhiSTe3bP, I-TdeIP, I-TevI, I-TevII, I-TevIII, i-UarAP, i-UarHGPAIP, I-UarHGPA13P, I-VinIP, I-ZbiIP, PI-MgaI, PI-MtuI, PI-MtuHIP PI-

MtuHIIP, PI-PfuI, PI-PfuII, PI-PkoI, PI-PkoII, PI-Rma43812IP, PI-SpBetaIP, PI-SceI, PI-TfuI, PI-TfuII, PI-ThyI, PI-TliI, or PI-TliII.

[0022] In some examples, the host cell is a fungal cell, a bacterial cell, a plant cell, an animal cell, or a human cell. In particular examples, the host cell is a yeast cell. In some examples, the yeast cell is a haploid yeast cell. In some examples, the yeast cell is a *Saccharomyces cerevisiae* cell. In some examples, the *Saccharomyces cerevisiae* cell is of the Baker's yeast, Mauri, Santa Fe, IZ-1904, TA, BG-1, CR-1, SA-1, M-26, Y-904, PE-2, PE-5, VR-1, BR-1, BR-2, ME-2, VR-2, MA-3, MA-4, CAT-1, CB-1, NR-1, BT-1 or AL-1 strain.

[0023] In another example, disclosed herein is a method for markerless integration of an exogenous nucleic acid into a target site of a yeast cell genome, the method comprising:

(a) contacting a host yeast cell with:

(i) an exogenous nucleic acid (ES) comprising a first homology region (HR1) and a second homology region (HR2), wherein (HR1) and (HR2) are capable of initiating host cell mediated homologous recombination at said target site (TS); and

(ii) a nuclease (N) capable of cleaving at (TS), whereupon said cleaving results in homologous recombination of (ES) at (TS);

and

(b) recovering a host cell having (ES) integrated at (TS), wherein said recovering does not require integration of a selectable marker.

[0024] In another example, disclosed herein is a modified host cell generated by any of the methods of genomically integrating one or more exogenous nucleic acids described herein. In some examples, the modified host cell comprises:

(a) a plurality of exogenous nucleic acids, wherein each exogenous nucleic acid $(ES)_x$ comprises a first homology region $(HR1)_x$ and a second homology region $(HR2)_x$, wherein $(HR1)_x$ and $(HR2)_x$ are capable of initiating host cell mediated homologous recombination of $(ES)_x$ at a target site $(TS)_x$ of said host cell genome; and

(b) for each said target site $(TS)_x$, a nuclease $(N)_x$ capable of cleaving at $(TS)_x$, whereupon said cleaving results in homologous recombination of $(ES)_x$ at $(TS)_x$;

wherein x is any integer from 1 to n wherein n is at least 2.

[0025] In some examples, the modified host cell is a yeast cell and comprises:

(a) an exogenous nucleic acid (ES) comprising a first homology region (HR1) and a second homology region (HR2), wherein (HR1) and (HR2) are capable of initiating host cell mediated homologous recombination at a target site (TS) of the host cell genome; and

(b) a nuclease (N) capable of cleaving at (TS), whereupon said cleaving results in homologous recombination of (ES) at (TS);

wherein (ES) does not comprise a selectable marker.

[0026] In another example, disclosed herein is a composition comprising:

(a) a yeast cell;

(b) a plurality of exogenous nucleic acids, wherein each exogenous nucleic acid $(ES)_x$ comprises:

(i) a first homology region $(HR1)_x$ and a second homology region $(HR2)_x$, wherein $(HR1)_x$ and $(HR2)_x$ are capable of initiating host cell mediated homologous recombination of $(ES)_x$ at a selected target site $(TS)_x$ of a yeast cell genome; and

(ii) a nucleic acid of interest $(D)_x$ positioned 3' of $(HR1)_x$ and 5' of $(HR2)_x$;

(c) a plurality of nucleases, wherein each nuclease $(N)_x$ capable of cleaving at $(TS)_x$, whereupon said cleaving results in homologous recombination of $(ES)_x$ at $(TS)_x$;

wherein x is any integer from 1 to n wherein n is at least 2.

[0027] In another example, disclosed herein is a kit useful for performing the methods for genomically integrating one or more exogenous nucleic acids described herein. In some examples, the kit comprises:

(a) a plurality of exogenous nucleic acids, wherein each exogenous nucleic acid $(ES)_x$ comprises:

(i) a first homology region $(HR1)_x$ and a second homology region $(HR2)_x$, wherein $(HR1)_x$ and $(HR2)_x$ are capable of initiating host cell mediated homologous recombination of $(ES)_x$ at a selected target site $(TS)_x$ of a yeast cell genome; and

(ii) a nucleic acid of interest $(D)_x$ positioned 3' of $(HR1)_x$ and 5' of $(HR2)_x$;

(b) a plurality of nucleases, wherein each nuclease $(N)_x$ capable of cleaving at $(TS)_x$, whereupon said cleaving results in homologous recombination of $(ES)_x$ at $(TS)_x$;

wherein x is any integer from 1 to n wherein n is at least 2.

[0028] In some examples, $(D)_x$ is selected from the group consisting of a selectable marker, a promoter, a nucleic acid sequence encoding an epitope tag, a gene of interest, a reporter gene, and a nucleic acid sequence encoding a termination codon. In some

examples, the kit further comprises a plurality of primer pairs $(P)_x$, wherein each primer pair is capable of identifying integration of $(ES)_x$ at $(TS)_x$ by PCR. In some examples, $(ES)_x$ is linear. In some examples, $(ES)_x$ is circular.

[0029] In a particular example, the kit enables site-specific integration of an exogenous nucleic acid at a unique target site within any of the approximately 6000 genetic loci of the yeast genome. In these examples, $n \geq 6000$, wherein each $(TS)_x$ is unique to a specific locus of the yeast cell genome.

5. BRIEF DESCRIPTION OF THE FIGURES

[0030] FIG. 1 provides an exemplary embodiment of markerless genomic integration of an exogenous nucleic acid using a site-specific nuclease.

[0031] FIG. 2 provides an exemplary embodiment of simultaneous genomic integration of a plurality of exogenous nucleic acids using a plurality of site-specific nucleases. HR1 – upstream homology region; HR2 – downstream homology region; TS – target site; N – site-specific nuclease; D – nucleic acid of interest.

[0032] FIG. 3 provides a schematic representation of the MEV pathway for isoprenoid production.

[0033] FIG. 4 provides an exemplary embodiment of the methods of generating combinatorial integration libraries provided herein. The hatch marks represent individual exogenous nucleic acid members of each library $(L)_x$.

[0034] FIG. 5 provides results of colony PCR of 96 colonies of yeast cells transformed with empty vector DNA and linear "donor" DNA encoding functional EmGFP. The yeast cells comprised copies of "target" nucleic acid encoding a truncated, non-functional EmGFP genomically integrated at each of the HO, YGR250c, and NDT80 loci. Separate PCR reactions were performed to probe the HO, YGR250c, and NDT80 loci with primers specific to nucleic acid encoding functional EmGFP. No PCR products were observed, indicating that no replacements of the target nucleic acid encoding non-functional EmGFP with donor nucleic acid encoding functional EmGFP occurred.

[0035] FIG. 6 provides results of colony PCR of 96 colonies of yeast cells transformed with pZFN.gfp DNA and linear "donor" DNA encoding functional EmGFP. The yeast cells comprised copies of "target" nucleic acid encoding a truncated, non-functional EmGFP genomically integrated at each of the HO, YGR250c, and NDT80 loci. pZFN.gfp encodes a zinc finger nuclease which recognizes and cleaves a nucleic acid sequence specific to the non-functional EmGFP coding sequence. Separate PCR reactions were performed to probe the HO,

YGR250c, and NDT80 loci with primers specific to nucleic acid encoding functional EmGFP. Numerous PCR products were observed, indicating successful replacement of the non-functional EmGFP integrations with DNA expressing functional EmGFP. 23 colonies have all 3 loci replaced.

[0036] FIG. 7 provides the sequiterpene titers of Strain B, a parental farnesene-producing yeast strain comprising enzymes of the mevalonate pathway and a plasmid encoding farnesene synthase (FS); Strain D, a derivative strain of Strain B in which 4 copies of amorphadiene synthase (ADS) have been genomically integrated; and Strain E, a derivative strain of Strain D in which the plasmid encoding FS has been lost. Nearly 100% of the sesquiterpene capacity of parental Strain B is maintained in Strains D and E with only the addition of multiple copies of ADS.

[0037] FIG. 8, provides results for cells co-transformed with linear donor DNAs for the SFC1 (GFP donor DNA) and YJR030c (ADE2 donor DNA) loci, the YJR030c endonuclease plasmid (pCUT006) and SFC1 endonuclease plasmid (pCUT058). 80% of colonies selected on URA dropout + Kan agar plates were GFP positive. Of these colonies, 91% were positive for ADE2 integration. In total, 72.8% of colonies had successfully integrated the markerless donor DNA at both loci.

6. DETAILED DESCRIPTION OF THE EMBODIMENTS

6.1 Definitions

[0038] As used herein, the terms "cleaves," "cleavage" and/or "cleaving" with respect to a nuclease, *e.g.* a homing endonuclease, zinc-finger nuclease or TAL-effector nuclease, refer to the act of creating a double-stranded break (DSB) in a particular nucleic acid. The DSB can leave a blunt end or sticky end (*i.e.,* 5' or 3' overhang), as understood by those of skill in the art.

[0039] As used herein, the term "engineered host cell" refers to a host cell that is generated by genetically modifying a parent cell using genetic engineering techniques (*i.e.*, recombinant technology). The engineered host cell may comprise additions, deletions, and/or modifications of nucleotide sequences to the genome of the parent cell.

[0040] As used herein, the term "heterologous" refers to what is not normally found in nature. The term "heterologous nucleotide sequence" refers to a nucleotide sequence not normally found in a given cell in nature. As such, a heterologous nucleotide sequence may be: (a) foreign to its host cell (*i.e.,* is "exogenous" to the cell); (b) naturally found in the host cell (*i.e.,* "endogenous") but present at an unnatural quantity in the cell (*i.e.,* greater or lesser

quantity than naturally found in the host cell); or (c) be naturally found in the host cell but positioned outside of its natural locus.

[0041] As used herein, the term "homology" refers to the identity between two or more nucleic acid sequences, or two or more amino acid sequences. Sequence identity can be measured in terms of percentage identity (or similarity or homology); the higher the percentage, the more near to identical the sequences are to each other. Homologs or orthologs of nucleic acid or amino acid sequences possess a relatively high degree of sequence identity when aligned using standard methods. Methods of alignment of sequences for comparison are well known in the art. Various programs and alignment algorithms are described in: Smith & Waterman, *Adv. Appl. Math.* 2:482, 1981; Needleman & Wunsch, *J. Mol. Biol.* 48:443, 1970; Pearson & Lipman, *Proc. Natl. Acad. Sci.* USA 85:2444, 1988; Higgins & Sharp, *Gene*, 73:237-44, 1988; Higgins & Sharp, *CABIOS* 5:151-3, 1989; Corpet *et al.*, *Nuc. Acids Res.* 16:10881-90, 1988; Huang *et al. Computer Appls. Biosc.* 8, 155-65, 1992; and Pearson *et al.*, *Meth. Mol. Bio.* 24:307-31, 1994. Altschul *et al.*, *J. Mol. Biol.* 215:403-10, 1990, presents a detailed consideration of sequence alignment methods and homology calculations. The NCBI Basic Local Alignment Search Tool (BLAST) (Altschul *et al.*, *J. Mol. Biol.* 215:403-10, 1990) is available from several sources, including the National Center for Biological Information (NCBI, National Library of Medicine, Building 38A, Room 8N805, Bethesda, Md. 20894) and on the Internet, for use in connection with the sequence analysis programs blastp, blastn, blastx, tblastn and tblastx. Additional information can be found at the NCBI web site.

[0042] As used herein, the term "markerless" refers to integration of a donor DNA into a target site within a host cell genome without accompanying integration of a selectable marker. In some embodiments, the term also refers to the recovery of such a host cell without utilizing a selection scheme that relies on integration of selectable marker into the host cell genome. For example, in certain embodiments, a selection marker that is episomal or extrachromasomal may be utilized to select for cells comprising a plasmid encoding a nuclease capable of cleaving a genomic target site. Such use would be considered "markerless" so long as the selectable marker is not integrated into the host cell genome.

[0043] As used herein, the term "polynucleotide" refers to a polymer composed of nucleotide units as would be understood by one of skill in the art. Preferred nucleotide units include but are not limited to those comprising adenine (A), guanine (G), cytosine (C), thymine (T), and uracil (U). Useful modified nucleotide units include but are not limited to those comprising 4-acetylcytidine, 5-(carboxyhydroxylmethyl)uridine, 2-O-methylcytidine, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylamino-methyluridine,

dihydrouridine, 2-O-methylpseudouridine, 2-O-methylguanosine, inosine, N6-isopentyladenosine, 1-methyladenosine, 1-methylpseudouridine, 1-methylguanosine, 1-methylinosine, 2,2-dimethylguanosine, 2-methyladenosine, 2-methylguanosine, 3-methylcytidine, 5-methylcytidine, N6-methyladenosine, 7-methylguanosine, 5-methylaminomethyluridine, 5-methoxyaminomethyl-2-thiouridine, 5-methoxyuridine, 5-methoxycarbonylmethyl-2-thiouridine, 5-methoxycarbonylmethyluridine, 2-methylthio-N6-isopentyladenosine, uridine-5-oxyacetic acid-methylester, uridine-5-oxyacetic acid, wybutoxosine, wybutosine, pseudouridine, queuosine, 2-thiocytidine, 5-methyl-2-thiouridine, 2-thiouridine, 4-thiouridine, 5-methyluridine, 2-O-methyl-5-methyluridine, 2-O-methyluridine, and the like. Polynucleotides include naturally occurring nucleic acids, such as deoxyribonucleic acid ("DNA") and ribonucleic acid ("RNA"), as well as nucleic acid analogs. Nucleic acid analogs include those that include non-naturally occurring bases, nucleotides that engage in linkages with other nucleotides other than the naturally occurring phosphodiester bond or that include bases attached through linkages other than phosphodiester bonds. Thus, nucleotide analogs include, for example and without limitation, phosphorothioates, phosphorodithioates, phosphorotriesters, phosphoramidates, boranophosphates, methylphosphonates, chiral-methyl phosphonates, 2-O-methyl ribonucleotides, peptide-nucleic acids (PNAs), and the like.

[0044] Conventional notation is used herein to describe polynucleotide sequences: the left-hand end of a single-stranded polynucleotide sequence is the 5'-end; the left-hand direction of a double-stranded polynucleotide sequence is referred to as the 5'-direction.

[0045] As used herein, the term "simultaneous," when used with respect to multiple integration, encompasses a period of time beginning at the point at which a host cell is co-transformed with a nuclease, *e.g.* a plasmid encoding a nuclease, and more than one donor DNA to be integrated into the host cell genome, and ending at the point at which the transformed host cell, or clonal populations thereof, is screened for successful integration of the donor DNAs at their respective target loci. In some embodiments, the period of time encompassed by "simultaneous" is at least the amount of time required for the nuclease to bind and cleave its target sequence within the host cell's chromosome(s). In some embodiments, the period of time encompassed by "simultaneous" is at least 6, 12, 24, 36, 48, 60, 72, 96 or more than 96 hours, beginning at the point at which the a host cell is co-transformed with a nuclease, *e.g.* a plasmid encoding a nuclease, and more than one donor DNA.

6.2 Methods of Integrating Exogenous Nucleic Acids

[0046] Provided herein are methods of integrating one or more exogenous nucleic acids into one or more selected target sites of a host cell genome. In certain embodiments, the methods comprise contacting the host cell with one or more integration polynucleotides, *i.e.*, donor DNAs, comprising an exogenous nucleic acid to be integrated into the genomic target site, and one or more nucleases capable of causing a double-strand break near or within the genomic target site. Cleavage near or within the genomic target site greatly increases the frequency of homologous recombination at or near the cleavage site.

[0047] In a particular aspect, provided herein is a method for markerless integration of an exogenous nucleic acid into a target site of a host cell genome, the method comprising:

(a) contacting a host cell with:

(i) an exogenous nucleic acid (ES) comprising a first homology region (HR1) and a second homology region (HR2), wherein (HR1) and (HR2) are capable of initiating host cell mediated homologous recombination at said target site (TS); and

(ii) a nuclease (N) capable of cleaving at (TS), whereupon said cleaving results in homologous recombination of (ES) at (TS);

and

(b) recovering a host cell having (ES) integrated at (TS), wherein said recovering does not require integration of a selectable marker.

[0048] FIG. 1 provides an exemplary embodiment of markerless genomic integration of an exogenous nucleic acid using a site-specific nuclease. A donor polynucleotide is introduced to a host cell, wherein the polynucleotide comprises a nucleic acid of interest (D) flanked by a first homology region (HR1) and a second homology region (HR2). HR1 and HR2 share homology with 5' and 3' regions, respectively, of a genomic target site (TS). A site-specific nuclease (N) is also introduced to the host cell, wherein the nuclease is capable of recognizing and cleaving a unique sequence within the target site. Upon induction of a double-stranded break within the target site by the site-specific nuclease, endogenous homologous recombination machinery integrates the nucleic acid of interest at the cleaved target site at a higher frequency as compared to a target site not comprising a double-stranded break. This increased frequency of integration obviates the need to co-integrate a selectable marker in order to select transformants having undergone a recombination event. By eliminating the need for selectable markers, for example, during construction of an engineered microbe, the time needed to build a strain comprising a complete and functional biosynthetic pathway is greatly reduced. In addition, engineering strategies are no longer limited by the need to recycle

selectable markers due to there being a limited cache of markers available for a given host organism.

[0049] In some embodiments, markerless recovery of a transformed cell comprising a successfully integrated exogenous nucleic acid occurs within a frequency of about one every 1000, 900, 800, 700, 600, 500, 400, 300, 200 or 100 contacted host cells, or clonal populations thereof, screened. In particular embodiments, markerless recovery of a transformed cell comprising a successfully integrated exogenous nucleic acid occurs within a frequency of about one every 90, 80, 70, 60, 50, 40, 30, 20, or 10 contacted host cells, or clonal populations thereof, screened. In more particular embodiments, markerless recovery of a transformed cell comprising a successfully integrated exogenous nucleic acid occurs within a frequency of about one every 9, 8, 7, 6, 5, 4, 3, or 2 contacted host cells, or clonal populations thereof, screened. In more particular embodiments, the host cell is a yeast cell, and the increased frequency of integration derives from yeast's increased capacity for homologous recombination relative to other host cell types.

[0050] A variety of methods are available to identify those cells having an altered genome at or near the target site without the use of a selectable marker. In some embodiments, such methods seek to detect any change in the target site, and include but are not limited to PCR methods, sequencing methods, nuclease digestion, *e.g.,* restriction mapping, Southern blots, and any combination thereof.

[0051] In another aspect, provided herein is a method for integrating a plurality of exogenous nucleic acids into a host cell genome, the method comprising:

(a) contacting a host cell with:

(i) a plurality of exogenous nucleic acids, wherein each exogenous nucleic acid $(ES)_x$ comprises a first homology region $(HR1)_x$ and a second homology region $(HR2)_x$, wherein $(HR1)_x$ and $(HR2)_x$ are capable of initiating host cell mediated homologous recombination of $(ES)_x$ at a target site $(TS)_x$ of said host cell genome; and

(ii) for each said target site $(TS)_x$, a nuclease $(N)_x$ capable of cleaving at $(TS)_x$, whereupon said cleaving results in homologous recombination of $(ES)_x$ at $(TS)_x$; and

(b) recovering a host cell wherein each selected exogenous nucleic acid $(ES)_x$ has integrated at each selected target sequence $(TS)_x$,

wherein x is any integer from 1 to n wherein n is at least 2.

[0052] FIG. 2 provides an exemplary embodiment of simultaneous genomic integration of a plurality of exogenous nucleic acids using a plurality of site-specific nucleases. In this example, three polynucleotides are introduced to a host cell, wherein each polynucleotide comprises an exogenous nucleic acid $(ES)_x$ comprising a nucleic acid of interest $(D)_x$, wherein x=1, 2 or 3. Each $(D)_x$ is flanked by a first homology region $(HR1)_x$ and a second homology region $(HR2)_x$. $(HR1)_x$ and $(HR2)_x$ share homology with 5' and 3' regions, respectively, of a selected target site $(TS)_x$, of three total unique target sites in the genome. A plurality of site-specific nucleases $(N)_x$ is also introduced to the host cell, wherein each $(N)_x$ is capable of recognizing and cleaving a unique sequence within its corresponding target site, $(TS)_x$. Upon cleavage of a target site $(TS)_x$ by its corresponding site-specific nuclease $(N)_x$, endogenous homologous recombination machinery facilitates integration of the corresponding nucleic acid interest $(D)_x$ at $(TS)_x$.

[0053] In particular embodiments, each exogenous nucleic acid $(ES)_x$, optionally comprising a nucleic acid of interest $(D)_x$, is integrated into its respective genomic target site $(TS)_x$ simultaneously, *i.e.*, with a single transformation of the host cell with the plurality of integration polynucleotides and plurality of nucleases. In some embodiments, the methods are useful to simultaneously integrate any plurality of exogenous nucleic acids $(ES)_x$, that is, where x is any integer from 1 to n wherein n is at least 2, in accordance with the variables recited for the above described method. In some embodiments, the method of simultaneous integration provided herein is useful to simultaneously integrate up to 10 exogenous nucleic acids $(ES)_x$ into 10 selected target sites $(TS)_x$, that is, where x is any integer from 1 to n wherein n= 2, 3, 4, 5, 6, 7, 8, 9, or 10. In some embodiments, the method of simultaneous integration provided herein is useful to simultaneously integrate up to 20 exogenous nucleic acids $(ES)_x$ into 20 selected target sites $(TS)_x$, that is, where x is any integer from 1 to n wherein n= 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20. In some embodiments, n =2. In some embodiments, n = 3. In some embodiments, n = 4. In some embodiments, n = 5. In some embodiments, n = 6. In some embodiments, n = 7. In some embodiments, n = 8. In some embodiments, n = 9. In some embodiments, n = 10. In some embodiments, n = 11. In some embodiments, n = 12. In some embodiments, n = 13. In some embodiments, n = 14. In some embodiments, n = 15. In some embodiments, n = 16. In some embodiments, n = 17. In some embodiments, n = 18. In some embodiments, n = 19. In some embodiments, n = 20. In some embodiments, the method of simultaneous integration provided herein is useful to simultaneously integrate more than 20 exogenous nucleic acids.

[0054] As with integration of a single exogenous nucleic acid at a single target site, the simultaneous multiple integration of a plurality of exogenous nucleic acids occurs at a substantially higher frequency as compared to not contacting the target sites with a nuclease capable of inducing a double-stranded break. In some embodiments, during the simultaneous integration of a plurality of exogenous nucleic acids at multiple loci, *i.e.,* in the presence of multiple nucleases, the frequency of integration at any single loci is substantially higher compared to the frequency of integration at the same locus during a single integration event, *i.e.,* in the presence of a single nuclease. Such an advantage is demonstrated in Example 6 (Section 7.5.2) below. Without being bound by theory, it is believed that the presence and activity of multiple nucleases, creating double-strand breaks (DSBs) at a plurality of target sites, enriches for transformants that successfully repair the DSBs by integrating donor DNA(s) at the cut site, and/or selects against transformants unable to repair the DSBs. Since DSBs are toxic to cells, it is believed that an increased number of nucleases leads to more DSBs, and correspondingly, an enrichment for cells able to repair the DSBs through HR-mediated integration of donor DNA(s).

[0055] In some embodiments, this increased frequency of integration obviates the requirement for co-integration of one or more selectable markers for the identification of the plurality of recombination events. In some embodiments, markerless recovery of a transformed cell comprising a plurality of successfully integrated exogenous nucleic acid occurs within a frequency of about one every 1000, 900, 800, 700, 600, 500, 400, 300, 200 or 100 contacted host cells, or clonal populations thereof, screened. In particular embodiments, markerless recovery occurs within a frequency of about one every 90, 80, 70, 60, 50, 40, 30, 20, or 10 contacted host cells, or clonal populations thereof, screened. In more particular embodiments, markerless recovery occurs within a frequency of about one every 9, 8, 7, 6, 5, 4, 3, or 2 contacted host cells, or clonal populations thereof, screened. In more particular embodiments, the host cell is a yeast cell, and the increased frequency of integration derives from yeast's increased capacity for homologous recombination relative to other host cell types.

6.2.1. Methods for Metabolic Pathway Engineering

[0056] The methods and compositions described herein provide particular advantages for constructing recombinant organisms comprising optimized biosynthetic pathways, for example, towards the conversion of biomass into biofuels, pharmaceuticals or biomaterials. Functional non-native biological pathways have been successfully constructed in microbial hosts for the production of precursors to the antimalarial drug artemisinin (*see, e.g.,* Martin *et*

al., *Nat Biotechnol* 21:796-802 (2003); fatty acid derives fuels and chemicals (*e.g.*, fatty esters, fatty alcohols and waxes; *see, e.g.*, Steen *et al.*, *Nature* 463:559-562 (2010); methyl halide-derived fuels and chemicals (*see, e.g.*, Bayer *et al.*, *J Am Chem Soc* 131:6508-6515 (2009); polyketide synthases that make cholesterol lowering drugs (*see, e.g.*, Ma *et al.*, *Science* 326:589-592 (2009); and polyketides (*see, e.g.*, Kodumal, *Proc Natl Acad Sci* USA 101:15573-15578 (2004).

[0057] Traditionally, metabolic engineering, and in particular, the construction of biosynthetic pathways, has proceeded in a one-at-a-time serial fashion whereby pathway components have been introduced, *i.e.*, integrated into the host cell genome at a single loci at a time. The methods of integration provided herein can be utilized to reduce the time typically required to engineer a host cell, for example, a microbial cell, to comprise one or more heterologous nucleotide sequences encoding enzymes of a new metabolic pathway, *i.e.*, a metabolic pathway that produces a metabolite that is not endogenously produced by the host cell. In other particular embodiments, the methods of integration provided herein can be used to efficiently engineer a host cell to comprise one or more heterologous nucleotide sequences encoding enzymes of a metabolic pathway that is endogenous to the host cell, *i.e.*, a metabolic pathway that produces a metabolite that is endogenously produced by the host cell. In one example, a design strategy may seek to replace three native genes of a host cell with a complementary exogenous pathway. Modifying these three endogenous loci using the current state of the art requires three separate transformations. By contrast, the methods of simultaneous multiple integration provided herein enables all three integrations to be performed in a single transformation, thus reducing the rounds of engineering needed by three-fold. Moreover, the methods enable the porting of DNA assemblies, comprising optimized pathway components integrated at multiple sites in one host cell chassis, to analogous sites in a second host cell chassis. By reducing the number of rounds needed to engineer a desired genotype, the pace of construction of metabolic pathways is substantially increased.

6.2.1.1 Isoprenoid Pathway Engineering

[0058] In some embodiments, the methods provided herein can be utilized to simultaneously introduce or replace one or more components of a biosynthetic pathway to modify the product profile of an engineered host cell. In some embodiments, the biosynthetic pathway is the isoprenoid pathway.

[0059] Terpenes are a large class of hydrocarbons that are produced in many organisms. When terpenes are chemically modified *(e.g.*, via oxidation or rearrangement of the carbon

skeleton) the resulting compounds are generally referred to as terpenoids, which are also known as isoprenoids. Isoprenoids play many important biological roles, for example, as quinones in electron transport chains, as components of membranes, in subcellular targeting and regulation via protein prenylation, as photosynthetic pigments including carotenoids, chlorophyll, as hormones and cofactors, and as plant defense compounds with various monoterpenes, sesquiterpenes, and diterpenes. They are industrially useful as antibiotics, hormones, anticancer drugs, insecticides, and chemicals.

[0060] Terpenes are derived by linking units of isoprene (C_5H_8), and are classified by the number of isoprene units present. Hemiterpenes consist of a single isoprene unit. Isoprene itself is considered the only hemiterpene. Monoterpenes are made of two isoprene units, and have the molecular formula $C_{10}H_{16}$. Examples of monoterpenes are geraniol, limonene, and terpineol. Sesquiterpenes are composed of three isoprene units, and have the molecular formula $C_{15}H_{24}$. Examples of sesquiterpenes are farnesenes and farnesol. Diterpenes are made of four isoprene units, and have the molecular formula $C_{20}H_{32}$. Examples of diterpenes are cafestol, kahweol, cembrene, and taxadiene. Sesterterpenes are made of five isoprene units, and have the molecular formula $C_{25}H_{40}$. An example of a sesterterpenes is geranylfarnesol. Triterpenes consist of six isoprene units, and have the molecular formula $C_{30}H_{48}$. Tetraterpenes contain eight isoprene units, and have the molecular formula $C_{40}H_{64}$. Biologically important tetraterpenes include the acyclic lycopene, the monocyclic gamma-carotene, and the bicyclic alpha- and beta-carotenes. Polyterpenes consist of long chains of many isoprene units. Natural rubber consists of polyisoprene in which the double bonds are cis.

[0061] Terpenes are biosynthesized through condensations of isopentenyl pyrophosphate (isopentenyl diphosphate or IPP) and its isomer dimethylallyl pyrophosphate (dimethylallyl diphosphate or DMAPP). Two pathways are known to generate IPP and DMAPP, namely the mevalonate-dependent (MEV) pathway of eukaryotes (FIG. 3), and the mevalonate-independent or deoxyxylulose-5-phosphate (DXP) pathway of prokaryotes. Plants use both the MEV pathway and the DXP pathway. IPP and DMAPP in turn are condensed to polyprenyl diphosphates (*e.g.*, geranyl disphosphate or GPP, farnesyl diphosphate or FPP, and geranylgeranyl diphosphate or GGPP) through the action of prenyl disphosphate synthases (*e.g.*, GPP synthase, FPP synthase, and GGPP synthase, respectively). The polyprenyl diphosphate intermediates are converted to more complex isoprenoid structures by terpene synthases.

[0062] Terpene synthases are organized into large gene families that form multiple products. Examples of terpene synthases include monoterpene synthases, which convert GPP

into monoterpenes; diterpene synthases, which convert GGPP into diterpenes; and sesquiterpene synthases, which convert FPP into sesquiterpenes. An example of a sesquiterpene synthase is farnesene synthase, which converts FPP to farnesene. Terpene synthases are important in the regulation of pathway flux to an isoprenoid because they operate at metabolic branch points and dictate the type of isoprenoid produced by the cell. Moreover, the terpene synthases hold the key to high yield production of such terpenes. As such, one strategy to improve pathway flux in hosts engineered for heterologous isoprenoid production is to introduce multiple copies of nucleic acids encoding terpene synthases. For example, in engineered microbes comprising the MEV pathway where the production of sesquiterpenes such as farnesene is desired, a sesquiterpene synthase, *e.g.,* a farnesene synthase is utilized as the terminal enzyme of the pathway, and multiple copies of farnesene synthase genes may be introduced into the host cell towards the generation of a strain optimized for farnesene production.

[0063] Because the biosynthesis of any isoprenoid relies on the same pathway components upstream of the prenyl disphosphate synthase and terpene synthase, these pathway components, once engineered into a host "platform" strain, can be utilized towards the production of any sesquiterpene, and the identity of the sesquiterpene can be dictated by the particular sesquiterpene synthase introduced into the host cell. Moreover, where production of terpenes having different isoprene units is desired, for example a monoterpene instead of a sesquiterpene, both the prenyl diphosphate synthase and the terpene synthase can be replaced to produce the different terpene while still utilizing the upstream components of the pathway.

[0064] Accordingly, the methods and compositions provided herein can be utilized to efficiently modify a host cell comprising an isoprenoid producing pathway, *e.g.,* the MEV pathway to produce a desired isoprenoid. In some embodiments, the host cell comprises the MEV pathway, and the methods of simultaneous multiple integration provided herein can be utilized to simultaneously introduce multiple copies of a prenyl diphosphate synthase and /or a terpene synthase to define the terpene product profile of the host cell. In some embodiments, the prenyl diphosphate synthase is GPP synthase and the terpene synthase is a monoterpene synthase. In some embodiments, the prenyl diphosphate synthase is FPP synthase and the terpene synthase is a sesquiterpene synthase. In some embodiments, the prenyl diphosphate synthase is GGPP synthase and the terpene synthase is a diterpene synthase. In other embodiments, the host cell comprises the MEV pathway and a prenyl diphosphate synthase and/or a terpene synthase for the production of a first type of terpene, for example, farnesene, and the methods of simultaneous multiple integration provided herein can be utilized to

simultaneously replace one or more copies of the prenyl diphosphate synthase and/or a terpene synthase to produce a second type of terpene, for example, amorphadiene. These embodiments are exemplified in Examples 3 and 4 below. The methods provided herein can be similarly utilized towards the construction and/or modification of any biosynthetic pathway which utilizes multiple copies of pathway components, and are particularly useful for engineering host cells whose product profile can be readily modified with the addition or exchange of multiple copies of a single pathway component.

6.2.1.2 Methods of Generating Combinatorial Integration Libraries

[0065] Once biosynthetic pathways are constructed, the expression levels of all the components need to be orchestrated to optimize metabolic flux and achieve high product titers. Common approaches for optimizing flux include varying the identity of the pathway component gene, the codon optimization of the gene, the use of solubility tags, the use of truncations or known mutations, and the expression context of the gene (*i.e.* promoter and terminator choice). To sample this variability in the course of building a strain using traditional methods requires generating and archiving an impractically large number of strains. For example, if a strain engineer plans to integrate constructs at three loci, and has devised 10 variants for each locus, 1,000 strains would need to be generated to fully sample the combinatorial diversity. Since pathway genes work in concert, and not all metabolite intermediates can easily be screened for, it is often impossible to evaluate the individual contribution of the pathway genes after each integration cycle. Thus, strain engineers routinely make choices that severely limit the design space that they sample when constructing a novel metabolic pathway.

[0066] To better identify the optimal pathway design, the methods of genomic modification provided herein can be utilized to generate strains comprising combinatorial libraries of rationally designed integration constructs. The methods rely on the introduction of one or more nucleases and one or more donor DNA assemblies into the cell to facilitate multiple simultaneous integration of donor DNA at specified locations in the genome. However, to generate a diversity of engineered strains, the methods comprise co-transforming a library of donor DNAs, *i.e.,* a mixture of integration constructs for each targeted locus, such that combinatorial integration libraries of host strains can be generated (FIG. 4). The high frequency of multiple integrations achieved means that the resultant strains can reasonably be screened directly for product without extensive genomic quality control, and the identity of top strains can be determined after screening, for example, by sequencing. This method removes

the burden of individual strain generation, quality control and archiving, and allows the engineer to generate diverse integration combinations in a single tube, and sort out the best performing strains by screening, *e.g.,* for the terminal product of the pathway.

[0067] Thus, in some embodiments, the methods for integrating a plurality of exogenous nucleic acids into a host cell genome provided herein comprise:

(d) contacting a host cell with:

(i) a plurality of libraries, wherein each library (L)x comprises a plurality of exogenous nucleic acids, wherein a selected exogenous nucleic acid comprises, in a 5' to 3' orientation, a first homology region $(HR1)_x$, any nucleic acid of interest selected from the group $(D)_x$, and a second homology region $(HR2)_x$, wherein $(HR1)_x$ and $(HR2)_x$ are capable of initiating host cell mediated homologous recombination of said selected exogenous nucleic acid at a target site $(TS)_x$ of said host cell genome; and

(ii) for each said target site $(TS)_x$, a nuclease $(N)_x$ capable of cleaving at $(TS)_x$, whereupon said cleaving results in homologous recombination of said selected exogenous nucleic acid at $(TS)_x$;

and

(e) recovering a host cell wherein an exogenous nucleic acid from each library $(L)_x$ has integrated at each selected target sequence $(TS)_x$,

wherein x is any integer from 1 to n wherein n is at least 2.

[0068] A schematic representation of this method is provided in FIG. 4.

[0069] Also provided herein is a host cell comprising:

(a) a plurality of libraries, wherein each library (L)x comprises a plurality of exogenous nucleic acids, wherein a selected exogenous nucleic acid comprises, in a 5' to 3' orientation, a first homology region $(HR1)_x$, any nucleic acid of interest selected from the group $(D)_x$, and a second homology region $(HR2)_x$, wherein $(HR1)_x$ and $(HR2)_x$ are capable of initiating host cell mediated homologous recombination of said selected exogenous nucleic acid at a target site $(TS)_x$ of said host cell genome; and

(b) for each said target site $(TS)_x$, a nuclease $(N)_x$ capable of cleaving at $(TS)_x$, whereupon said cleaving results in homologous recombination of said selected exogenous nucleic acid at $(TS)_x$,

wherein x is any integer from 1 to n wherein n is at least 2.

[0070] In some embodiments, each library $(L)_x$ comprises exogenous nucleic acids encoding enzymes of a common biosynthetic pathway. In some embodiments, the group $(D)_x$

comprises at least 10^1, 10^2, 10^3, 10^4, 10^5, 10^6, or more than 10^6 unique nucleic acids of interest. In some embodiments, each library $(L)_x$ comprises a plurality of exogenous nucleic acids encoding variants of an enzyme of a biosynthetic pathway. As used herein, the term "variant" refers to an enzyme of a biosynthetic pathway that compared to a selected enzyme has a different nucleotide or amino acid sequence. For example, in some embodiments, a library $(L)_x$ comprises sesquiterpene synthase variants, and compared to the wild-type version of the selected sesquiterpene synthase, the sesquiterpene synthase variant may comprise nucleotide additions, deletions, and/or substitutions that may or may not result in changes to the corresponding amino acid sequence. In other embodiments, the enzyme variant comprises amino acid additions, deletions and/or substitutions relative to a reference enzyme, *e.g.*, the wild-type version.

[0071] In some embodiments, the host cell comprises one or more heterologous nucleotide sequences encoding one or more enzymes of a biosynthetic pathway prior to said contacting. In some embodiments, the one or more heterologous nucleotide sequences encoding one or more enzymes of a biosynthetic pathway are genomically integrated.

6.3 Integration Polynucleotides

[0072] Advantageously, an integration polynucleotide, *i.e.*, donor DNA, facilitates integration of one or more exogenous nucleic acid constructs into a selected target site of a host cell genome. In preferred embodiments, an integration polynucleotide comprises an exogenous nucleic acid $(ES)_x$ comprising a first homology region $(HR1)_x$ and a second homology region $(HR2)_x$, and optionally a nucleic acid of interest positioned between $(HR1)_x$ and $(HR2)_x$. In some embodiments, the integration polynucleotide is a linear DNA molecule. In other embodiments, the integration polynucleotide is a circular DNA molecule.

[0073] The integration polynucleotide can be generated by any technique apparent to one skilled in the art. In certain embodiments, the integration polynucleotide is generated using polymerase chain reaction (PCR) and molecular cloning techniques well known in the art. *See, e.g.*, *PCR Technology: Principles and Applications for DNA Amplification*, ed. HA Erlich, Stockton Press, New York, N.Y. (1989); Sambrook *et al.*, 2001, *Molecular Cloning – A Laboratory Manual*, 3rd edition, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY; *PCR Technology: Principles and Applications for DNA Amplification*, ed. HA Erlich, Stockton Press, New York, N.Y. (1989); U.S. Patent No. 8,110,360.

6.3.1. Genomic Integration Sequences

[0074] In preferred embodiments, an integration polynucleotide comprises an exogenous nucleic acid $(ES)_x$ comprising a first homology region $(HR1)_x$ and a second homology region $(HR2)_x$, wherein $(HR1)_x$ and $(HR2)_x$ are capable of initiating host cell mediated homologous recombination at a selected target site $(TS)_x$ within the host cell genome. To integrate an exogenous nucleic acid into the genome by homologous recombination, the integration polynucleotide preferably comprises $(HR1)_x$ at one terminus and $(HR2)_x$ at the other terminus. In some embodiments, $(HR1)_x$ is homologous to a 5' region of the selected genomic target site $(TS)_x$, and $(HR2)_x$, is homologous to a 3' region of the selected target site $(TS)_x$. In some embodiments, $(HR1)_x$ is about 70%, 75%, 80%, 85%, 90%, 95% or 100% homologous to a 5' region of the selected genomic target site $(TS)_x$. In some embodiments, $(HR2)_x$, is about 70%, 75%, 80%, 85%, 90%, 95% or 100% homologous to a 3' region of the selected target site $(TS)_x$.

[0075] In certain embodiments, $(HR1)_x$ is positioned 5' to a nucleic acid of interest $(D)_x$. In some embodiments, $(HR1)_x$ is positioned immediately adjacent to the 5' end of $(D)_x$. In some embodiments, $(HR1)_x$ is positioned upstream to the 5' of $(D)_x$. In certain embodiments, $(HR2)_x$ is positioned 3' to a nucleic acid of interest $(D)_x$. In some embodiments, $(HR2)_x$ is positioned immediately adjacent to the 3' end of $(D)_x$. In some embodiments, $(HR2)_x$ is positioned downstream to the 3' of $(D)_x$.

[0076] Properties that may affect the integration of an integration polynucleotide at a particular genomic locus include but are not limited to: the lengths of the genomic integration sequences, the overall length of the excisable nucleic acid construct, and the nucleotide sequence or location of the genomic integration locus. For instance, effective heteroduplex formation between one strand of a genomic integration sequence and one strand of a particular locus in a host cell genome may depend on the length of the genomic integration sequence. An effective range for the length of a genomic integration sequence is 50 to 5,000 nucleotides. For a discussion of effective lengths of homology between genomic integration sequences and genomic loci. *See*, Hasty *et al.*, *Mol Cell Biol* 11:5586-91 (1991).

[0077] In some embodiments, $(HR1)_x$ and $(HR2)_x$ can comprise any nucleotide sequence of sufficient length and sequence identity that allows for genomic integration of the exogenous nucleic acid $(ES)_x$ at any yeast genomic locus. In certain embodiments, each of $(HR1)_x$ and $(HR2)_x$ independently consists of about 50 to 5,000 nucleotides. In certain embodiments, each of $(HR1)_x$ and $(HR2)_x$ independently consists of about 100 to 2,500 nucleotides. In certain embodiments, each of $(HR1)_x$ and $(HR2)_x$ independently consists of about 100 to 1,000 nucleotides. In certain embodiments, each of $(HR1)_x$ and $(HR2)_x$ independently consists of

about 250 to 750 nucleotides. In certain embodiments, each of $(HR1)_x$ and $(HR2)_x$ independently consists of about 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000, 3100, 3200, 3300, 3400, 3500, 3600, 3700, 3800, 3900, 4000, 4100, 4200, 4300, 4400, 4500, 4600, 4700, 4800, 4900 or 5,000 nucleotides. In some embodiments, each of $(HR1)_x$ and $(HR2)_x$ independently consists of about 500 nucleotides.

6.3.2. Nucleic Acids of Interest

[0078] In some embodiments, the integration polynucleotide further comprises a nucleic acid of interest $(D)_x$. The nucleic acid of interest can be any DNA segment deemed useful by one of skill in the art. For example, the DNA segment may comprise a gene of interest that can be "knocked in" to a host genome. In other embodiments, the DNA segment functions as a "knockout" construct that is capable of specifically disrupting a target gene upon integration of the construct into the target site of the host cell genome, thereby rendering the disrupted gene non-functional. Useful examples of a nucleic acid of interest $(D)_x$ include but are not limited to: a protein-coding sequence, reporter gene, fluorescent marker coding sequence, promoter, enhancer, terminator, transcriptional activator, transcriptional repressor, transcriptional activator binding site, transcriptional repressor binding site, intron, exon, poly-A tail, multiple cloning site, nuclear localization signal, mRNA stabilization signal, integration loci, epitope tag coding sequence, degradation signal, or any other naturally occurring or synthetic DNA molecule. In some embodiments, $(D)_x$ can be of natural origin. Alternatively, $(D)_x$ can be completely of synthetic origin, produced *in vitro*. Furthermore, $(D)_x$ can comprise any combination of isolated naturally occurring DNA molecules, or any combination of an isolated naturally occurring DNA molecule and a synthetic DNA molecule. For example, $(D)_x$ may comprise a heterologous promoter operably linked to a protein coding sequence, a protein coding sequence linked to a poly-A tail, a protein coding sequence linked in-frame with a epitope tag coding sequence, and the like. The nucleic acid of interest $(D)_x$ may be obtained by standard procedures known in the art from cloned DNA (*e.g.*, a DNA "library"), by chemical synthesis, by cDNA cloning, or by the cloning of genomic DNA, or fragments thereof, purified from the desired cell, or by PCR amplification and cloning. *See*, for example, Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual*, 3d. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (2001); Glover, D.M. (ed.), *DNA Cloning: A Practical Approach*, 2d. ed., MRL Press, Ltd., Oxford, U.K. (1995).

[0079] In particular embodiments, the nucleic acid of interest $(D)_x$ does not comprise nucleic acid encoding a selectable marker. In these embodiments, the high efficiency of integration provided by the methods described herein allows for the screening and identification of integration events without the requirement for growth of transformed cells on selection media. However, in other embodiments where growth on selective media is nonetheless desired, the nucleic acid of interest $(D)_x$ can comprise a selectable marker that may be used to select for the integration of the exogenous nucleic acid into a host genome.

[0080] A wide variety of selectable markers are known in the art (see, for example, Kaufinan, *Meth. Enzymol.*, 185:487 (1990); Kaufman, *Meth. Enzymol.*, 185:537 (1990); Srivastava and Schlessinger, *Gene*, 103:53 (1991); Romanos *et al.*, in *DNA Cloning 2: Expression Systems*, 2nd Edition, pages 123-167 (IRL Press 1995); Markie, *Methods Mol. Biol.*, 54:359 (1996); Pfeifer *et al.*, *Gene*, 188:183 (1997); Tucker and Burke, *Gene*, 199:25 (1997); Hashida-Okado *et al.*, FEBS Letters, 425:117 (1998)). In some embodiments, the selectable marker is a drug resistant marker. A drug resistant marker enables cells to detoxify an exogenous drug that would otherwise kill the cell. Illustrative examples of drug resistant markers include but are not limited to those which confer resistance to antibiotics such as ampicillin, tetracycline, kanamycin, bleomycin, streptomycin, hygromycin, neomycin, Zeocin™, and the like. In other embodiments, the selectable marker is an auxotrophic marker. An auxotrophic marker allows cells to synthesize an essential component (usually an amino acid) while grown in media that lacks that essential component. Selectable auxotrophic gene sequences include, for example, hisD, which allows growth in histidine free media in the presence of histidinol. Other selectable markers include a bleomycin-resistance gene, a metallothionein gene, a hygromycin B-phosphotransferase gene, the AURI gene, an adenosine deaminase gene, an aminoglycoside phosphotransferase gene, a dihydrofolate reductase gene, a thymidine kinase gene, a xanthine-guanine phosphoribosyltransferase gene, and the like. In other embodiments, the selectable marker is a marker other than one which rescues an auxotophic mutation. For example, the host cell strain can comprise mutations other than auxotrophic mutations, for example, mutations that are not lethal to the host and that also do not cause adverse effects on the intended use of the strain, *e.g.,* industrial fermentation, so long as the mutations can be identified by a known selection method.

[0081] Host cell transformants comprising a chromosomally integrated polynucleotide can also be identified by selecting host cell transformants exhibiting other traits encoded by individual DNA segments or by combinations of DNA segments, *e.g.*, expression of peptides that emit light, or by molecular analysis of individual host cell colonies, *e.g.*, by restriction

enzyme mapping, PCR amplification, or sequence analysis of isolated assembled polynucleotides or chromosomal integration sites.

6.4 Nucleases

[0082] In some embodiments of the methods described herein, a host cell genome is contacted with one or more nucleases capable of cleaving, *i.e.,* causing a double-stranded break at a designated region within a selected target site. In some embodiments, a double-strand break inducing agent is any agent that recognizes and/or binds to a specific polynucleotide recognition sequence to produce a break at or near the recognition sequence. Examples of double-strand break inducing agents include, but are not limited to, endonucleases, site-specific recombinases, transposases, topoisomerases, and zinc finger nucleases, and include modified derivatives, variants, and fragments thereof.

[0083] In some embodiments, each of the one or more nucleases is capable of causing a double-strand break at a designated region within a selected target site $(TS)_x$. In some embodiments, the nuclease is capable of causing a double-strand break at a region positioned between the 5' and 3' regions of $(TS)_x$ with which $(HR1)_x$ and $(HR2)_x$ share homology, respectively. In other embodiments, the nuclease is capable of causing a double-strand break at a region positioned upstream or downstream of the 5' and 3' regions of $(TS)_x$.

[0084] A recognition sequence is any polynucleotide sequence that is specifically recognized and/or bound by a double-strand break inducing agent. The length of the recognition site sequence can vary, and includes, for example, sequences that are at least 10, 12, 14, 16, 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, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70 or more nucleotides in length.

[0085] In some embodiments, the recognition sequence is palindromic, that is, the sequence on one strand reads the same in the opposite direction on the complementary strand. In some embodiments, the nick/cleavage site is within the recognition sequence. In other embodiments, the nick/cleavage site is outside of the recognition sequence. In some embodiments, cleavage produces blunt end termini. In other embodiments, cleavage produces single-stranded overhangs, *i.e.,* "sticky ends," which can be either 5' overhangs, or 3' overhangs.

[0086] In some embodiments, the recognition sequence within the selected target site can be endogenous or exogenous to the host cell genome. When the recognition site is an endogenous sequence, it may be a recognition sequence recognized by a naturally-occurring, or

native double-strand break inducing agent. Alternatively, an endogenous recognition site could be recognized and/or bound by a modified or engineered double-strand break inducing agent designed or selected to specifically recognize the endogenous recognition sequence to produce a double-strand break. In some embodiments, the modified double-strand break inducing agent is derived from a native, naturally-occurring double-strand break inducing agent. In other embodiments, the modified double-strand break inducing agent is artificially created or synthesized. Methods for selecting such modified or engineered double-strand break inducing agents are known in the art. For example, amino acid sequence variants of the protein(s) can be prepared by mutations in the DNA. Methods for mutagenesis and nucleotide sequence alterations include, for example, Kunkel, (1985) *Proc Natl Acad Sci* USA 82:488-92; Kunkel, *et al.*, (1987) *Meth Enzymol* 154:367-82; U.S. Pat. No. 4,873,192; Walker and Gaastra, eds. (1983) *Techniques in Molecular Biology* (MacMillan Publishing Company, New York) and the references cited therein. Guidance regarding amino acid substitutions not likely to affect biological activity of the protein is found, for example, in the model of Dayhoff, *et al.*, (1978) *Atlas of Protein Sequence and Structure* (Natl Biomed Res Found, Washington, D.C.). Conservative substitutions, such as exchanging one amino acid with another having similar properties, may be preferable. Conservative deletions, insertions, and amino acid substitutions are not expected to produce radical changes in the characteristics of the protein, and the effect of any substitution, deletion, insertion, or combination thereof can be evaluated by routine screening assays. Assays for double strand break inducing activity are known and generally measure the overall activity and specificity of the agent on DNA substrates containing recognition sites.

[0087] In some embodiments of the methods provided herein, one or more of the nucleases is an endonuclease. Endonucleases are enzymes that cleave the phosphodiester bond within a polynucleotide chain, and include restriction endonucleases that cleave DNA as specific sites without damaging the bases. Restriction endonucleases include Type I, Type II, Type III, and Type IV endonucleases, which further include subtypes. Restriction endonucleases are further described and classified, for example in the REBASE database (webpage at rebase.neb.com; Roberts, *et al.*, (2003) *Nucleic Acids Res* 31:418-20), Roberts, *et al.*, (2003) *Nucleic Acids Res* 31:1805-12, and Belfort, *et al.*, (2002) in *Mobile DNA II*, pp. 761-783, Eds. Craigie, *et al.*, ASM Press, Washington, D.C.

[0088] As used herein, endonucleases also include homing endonucleases, which like restriction endonucleases, bind and cut at a specific recognition sequence. However the recognition sites for homing endonucleases are typically longer, for example, about 18 bp or

more. Homing endonucleases, also known as meganucleases, have been classified into the following families based on conserved sequence motifs: an LAGLIDADG (SEQ ID NO: 50) homing endonuclease, an HNH homing endonuclease, a His-Cys box homing endonuclease, a GIY-YIG (SEQ ID NO: 51) homing endonuclease, and a cyanobacterial homing endonuclease. *See, e.g.*, Stoddard, *Quarterly Review of Biophysics* 38(1): 49-95 (2006). These families differ greatly in their conserved nuclease active-site core motifs and catalytic mechanisms, biological and genomic distributions, and wider relationship to non-homing nuclease systems. *See,* for example, Guhan and Muniyappa (2003) *Crit Rev Biochem Mol Biol* 38:199-248; Lucas, *et al.*, (2001) *Nucleic Acids Res* 29:960-9; Jurica and Stoddard, (1999) *Cell Mol Life Sci* 55:1304-26; Stoddard, (2006) *Q Rev Biophys* 38:49-95; and Moure, *et al.*, (2002) *Nat Struct Biol* 9:764. Examples of useful specific homing endonucleases from these families include, but are not limited to: I-CreI (*see*, Rochaix *et al.*, *Nucleic Acids Res.* 13: 975-984 (1985), I-MsoI (*see*, Lucas *et al.*, *Nucleic Acids Res.* 29: 960-969 (2001), I-SceI (*see*, Foury *et al.*, *FEBS Lett.* 440: 325-331 (1998), I-SceIV (see, Moran *et al.*, *Nucleic Acids Res.* 20: 4069-4076 (1992), H-DreI (*see*, Chevalier *et al.*, *Mol. Cell* 10: 895-905 (2002), I-HmuI (*see*, Goodrich-Blair *et al.*, *Cell* 63: 417-424 (1990); Goodrich-Blair *et al.*, *Cell* 84: 211-221 (1996), I-PpoI (*see*, Muscarella *et al.*, *Mol. Cell. Biol.* 10: 3386-3396 (1990), I-DirI (*see*, Johansen *et al.*, *Cell* 76: 725-734 (1994); Johansen, *Nucleic Acids Res.* 21: 4405 (1993), I-NjaI (*see*, Elde *et al.*, *Eur. J. Biochem.* 259: 281-288 (1999); De Jonckheere *et al.*, *J. Eukaryot. Microbiol.* 41: 457-463 (1994), I-NanI (*see*, Elde *et al.*, S. *Eur. J. Biochem.* 259: 281-288 (1999); De Jonckheere *et al.*, *J. Eukaryot. Microbiol.* 41: 457-463 (1994)), I-NitI (*see*, De Jonckheere *et al.*, *J. Eukaryot. Microbiol.* 41: 457-463 (1994); Elde *et al.*, *Eur. J. Biochem.* 259: 281-288 (1999), I-TevI (*see*, Chu *et al.*, *Cell* 45: 157-166 (1986), I-TevII (*see*, Tomaschewski *et al.*, *Nucleic Acids Res.* 15: 3632-3633 (1987), I-TevIII (*see*, Eddy *et al.*, *Genes Dev.* 5: 1032-1041 (1991), F-TevI (*see*, Fujisawa *et al.*, *Nucleic Acids Res.* 13: 7473-7481 (1985), F-TevII (*see*, Kadyrov *et al.*, *Dokl. Biochem.* 339: 145-147 (1994); Kaliman, *Nucleic Acids Res.* 18: 4277 (1990), F-CphI (*see*, Zeng *et al.*, *Curr. Biol.* 19: 218-222 (2009), PI-MgaI (*see*, Saves *et al.*, *Nucleic Acids Res.* 29:4310-4318 (2001), I-CsmI (*see*, Colleaux *et al.*, *Mol. Gen. Genet.* 223:288-296 (1990), I-CeuI (*see*, Turmel *et al.*, *J .Mol. Biol.* 218: 293-311 (1991) and PI-SceI (*see*, Hirata *et al.*,,*J. Biol. Chem.* 265: 6726-6733 (1990).

[0089] In some embodiments of the methods described herein, a naturally occurring variant, and/or engineered derivative of a homing endonuclease is used. Methods for modifying the kinetics, cofactor interactions, expression, optimal conditions, and/or recognition site specificity, and screening for activity are known. *See*, for example, Epinat, *et*

al., (2003) *Nucleic Acids Res* 31:2952-62; Chevalier, *et al.*, (2002) *Mol Cell* 10:895-905; Gimble, *et al.*, (2003) *Mol Biol* 334:993-1008; Seligman, *et al.*, (2002) *Nucleic Acids Res* 30:3870-9; Sussman, *et al.*, (2004) *J Mol Biol* 342:31-41; Rosen, *et al.*, (2006) *Nucleic Acids Res* 34:4791-800; Chames, *et al.*, (2005) *Nucleic Acids Res* 33:e178; Smith, *et al.*, (2006) *Nucleic Acids Res* 34:e149; Gruen, *et al.*, (2002) *Nucleic Acids Res* 30:e29; Chen and Zhao, (2005) *Nucleic Acids Res* 33:e154; WO2005105989; WO2003078619; WO2006097854; WO2006097853; WO2006097784; and WO2004031346. Useful homing endonucleases also include those described in WO04/067736; WO04/067753; WO06/097784; WO06/097853; WO06/097854; WO07/034262; WO07/049095; WO07/049156; WO07/057781; WO07/060495; WO08/152524; WO09/001159; WO09/095742; WO09/095793; WO10/001189; WO10/015899; and WO10/046786.

[0090] Any homing endonuclease can be used as a double-strand break inducing agent including, but not limited to: H-DreI, I-SceI, I-SceII, I-SceIII, I-SceIV, I-SceV, I-SceVI, I-SceVII, I-CeuI, I-CeuAIIP, I-CreI, I-CrepsbIP, I-CrepsbIIP, I-CrepsbIIIP, I-CrepsbIVP, I-TliI, I-PpoI, Pi-PspI, F-SceI, F-SceII, F-SuvI, F-CphI, F-TevI, F-TevII, I-AmaI, I-AniI, I-ChuI, I-CmoeI, I-CpaI, I-CpaII, I-CsmI, I-CvuI, I-CvuAIP, I-DdiI, I-DdiII, I-DirI, I-DmoI, I-HmuI, I-HmuII, I-HsNIP, I-LlaI, I-MsoI, I-NaaI, I-NanI, I-NclIP, I-NgrIP, I-NitI, I-NjaI, I-Nsp236IP, I-PakI, I-PboIP, I-PcuIP, I-PcuAI, I-PcuVI, I-PgrIP, I-PobIP, I-PorI, I-PorIIP, I-PbpIP, I-SpBetaIP, I-ScaI, I-SexIP, I-SneIP, I-SpomI, I-SpomCP, I-SpomIP, I-SpomIIP, I-SquIP, I-Ssp68031, I-SthPhiJP, I-SthPhiST3P, I-SthPhiSTe3bP, I-TdeIP, I-TevI, I-TevII, I-TevIII, I-UarAP, I-UarHGPAIP, I-UarHGPA13P, I-VinIP, I-ZbiIP, PI-MgaI, PI-MtuI, PI-MtuHIP PI-MtuHIIP, PI-PfuI, PI-PfuII, PI-PkoI, PI-PkoII, PI-Rma43812IP, PI-SpBetaIP, PI-SceI, PI-TfuI, PI-TfuII, PI-ThyI, PI-TliI, or PI-TliII, or any variant or derivative thereof.

[0091] In some embodiments, the endonuclease binds a native or endogenous recognition sequence. In other embodiments, the endonuclease is a modified endonuclease that binds a non-native or exogenous recognition sequence and does not bind a native or endogenous recognition sequence.

[0092] In some embodiments of the methods provided herein, one or more of the nucleases is a TAL-effector DNA binding domain-nuclease fusion protein (TALEN). TAL effectors of plant pathogenic bacteria in the genus *Xanthomonas* play important roles in disease, or trigger defense, by binding host DNA and activating effector-specific host genes. *see, e.g.,* Gu *et al.* (2005) *Nature* 435:1122-5; Yang *et al.*, (2006) *Proc. Natl. Acad. Sci.* USA 103:10503-8; Kay *et al.*, (2007) *Science* 318:648-51; Sugio *et al.*, (2007) *Proc. Natl. Acad. Sci.* USA 104:10720-5; Romer *et al.*, (2007) *Science* 318:645-8; Boch *et al.*, (2009) *Science* 326(5959):1509-12; and

Moscou and Bogdanove, (2009) 326(5959):1501. A TAL effector comprises a DNA binding domain that interacts with DNA in a sequence-specific manner through one or more tandem repeat domains. The repeated sequence typically comprises 34 amino acids, and the repeats are typically 91-100% homologous with each other. Polymorphism of the repeats is usually located at positions 12 and 13, and there appears to be a one-to-one correspondence between the identity of repeat variable-diresidues at positions 12 and 13 with the identity of the contiguous nucleotides in the TAL-effector's target sequence.

[0093] The TAL-effector DNA binding domain may be engineered to bind to a desired target sequence, and fused to a nuclease domain, *e.g.*, from a type II restriction endonuclease, typically a nonspecific cleavage domain from a type II restriction endonuclease such as FokI (*see e.g.*,Kim *et al.* (1996) *Proc. Natl. Acad. Sci.* USA 93:1156-1160). Other useful endonucleases may include, for example, HhaI, HindIII, Nod, BbvCI, EcoRI, BglI, and AlwI. Thus, in preferred embodiments, the TALEN comprises a TAL effector domain comprising a plurality of TAL effector repeat sequences that, in combination, bind to a specific nucleotide sequence in the target DNA sequence, such that the TALEN cleaves the target DNA within or adjacent to the specific nucleotide sequence. TALENS useful for the methods provided herein include those described in WO10/079430 and U.S. Patent Application Publication No. 2011/0145940.

[0094] In some embodiments, the TAL effector domain that binds to a specific nucleotide sequence within the target DNA can comprise 10 or more DNA binding repeats, and preferably 15 or more DNA binding repeats. In some embodiments, each DNA binding repeat comprises a repeat variable-diresidue (RVD) that determines recognition of a base pair in the target DNA sequence, wherein each DNA binding repeat is responsible for recognizing one base pair in the target DNA sequence, and wherein the RVD comprises one or more of: HD for recognizing C; NG for recognizing T; NI for recognizing A; NN for recognizing G or A; NS for recognizing A or C or G or T; N* for recognizing C or T, where * represents a gap in the second position of the RVD; HG for recognizing T; H* for recognizing T, where * represents a gap in the second position of the RVD; IG for recognizing T; NK for recognizing G; HA for recognizing C; ND for recognizing C; HI for recognizing C; HN for recognizing G; NA for recognizing G; SN for recognizing G or A; and YG for recognizing T.

[0095] In some embodiments of the methods provided herein, one or more of the nucleases is a site-specific recombinase. A site-specific recombinase, also referred to as a recombinase, is a polypeptide that catalyzes conservative site-specific recombination between its compatible recombination sites, and includes native polypeptides as well as derivatives, variants and/or

fragments that retain activity, and native polynucleotides, derivatives, variants, and/or fragments that encode a recombinase that retains activity. For reviews of site-specific recombinases and their recognition sites, see, Sauer (1994) *Curr Op Biotechnol* 5:521-7; and Sadowski, (1993) *FASEB* 7:760-7. In some embodiments, the recombinase is a serine recombinase or a tyrosine recombinase. In some embodiments, the recombinase is from the Integrase or Resolvase families. In some embodiments, the recombinase is an integrase selected from the group consisting of FLP, Cre, lambda integrase, and R. For other members of the Integrase family, see for example, Esposito, *et al.*, (1997) *Nucleic Acids Res* 25:3605-14 and Abremski, *et al.*, (1992) *Protein Eng* 5:87-91. Methods for modifying the kinetics, cofactor interaction and requirements, expression, optimal conditions, and/or recognition site specificity, and screening for activity of recombinases and variants are known, see for example Miller, *et al.*, (1980) *Cell* 20:721-9; Lange-Gustafson and Nash, (1984) *J Biol Chem* 259:12724-32; Christ, *et al.*, (1998) *J Mol Biol* 288:825-36; Lorbach, *et al.*, (2000) *J Mol Biol* 296:1175-81; Vergunst, *et al.*, (2000) *Science* 290:979-82; Dorgai, *et al.*, (1995) *J Mol Biol* 252:178-88; Dorgai, *et al.*, (1998) *J Mol Biol* 277:1059-70; Yagu, *et al.*, (1995) *J Mol Biol* 252:163-7; Sclimente, *et al.*, (2001) *Nucleic Acids Res* 29:5044-51; Santoro and Schultze, (2002) *Proc Natl Acad Sci* USA 99:4185-90; Buchholz and Stewart, (2001) *Nat Biotechnol* 19:1047-52; Voziyanov, *et al.*, (2002) *Nucleic Acids Res* 30:1656-63; Voziyanov, *et al.*, (2003) *J Mol Biol* 326:65-76; Klippel, *et al.*, (1988) *EMBO J* 7:3983-9; Arnold, *et al.*, (1999) *EMBO J* 18:1407-14; WO03/08045; WO99/25840; and WO99/25841. The recognition sites range from about 30 nucleotide minimal sites to a few hundred nucleotides. Any recognition site for a recombinase can be used, including naturally occurring sites, and variants. Variant recognition sites are known, see for example Hoess, *et al.*, (1986) *Nucleic Acids Res* 14:2287-300; Albert, *et al.*, (1995) *Plant J* 7:649-59; Thomson, *et al.*, (2003) *Genesis* 36:162-7; Huang, *et al.*, (1991) *Nucleic Acids Res* 19:443-8; Siebler and Bode, (1997) *Biochemistry* 36:1740-7; Schlake and Bode, (1994) *Biochemistry* 33:12746-51; Thygarajan, *et al.*, (2001) *Mol Cell Biol* 21:3926-34; Umlauf and Cox, (1988) *EMBO J* 7:1845-52; Lee and Saito, (1998) *Gene* 216:55-65; WO01/23545; WO99/25821; WO99/25851; WO01/11058; WO01/07572 and U.S. Pat. No. 5,888,732.

[0096] In some embodiments of the methods provided herein, one or more of the nucleases is a transposase. Transposases are polypeptides that mediate transposition of a transposon from one location in the genome to another. Transposases typically induce double strand breaks to excise the transposon, recognize subterminal repeats, and bring together the ends of the excised transposon, in some systems other proteins are also required to bring together the

ends during transposition. Examples of transposons and transposases include, but are not limited to, the Ac/Ds, Dt/rdt, Mu-M1/Mn, and Spm(En)/dSpm elements from maize, the Tam elements from snapdragon, the Mu transposon from bacteriophage, bacterial transposons (Tn) and insertion sequences (IS), Ty elements of yeast (retrotransposon), Ta1 elements from Arabidopsis (retrotransposon), the P element transposon from Drosophila (Gloor, *et al.*, (1991) *Science* 253:1110-1117), the Copia, Mariner and Minos elements from Drosophila, the Hermes elements from the housefly, the PiggyBack elements from Trichplusia ni, Tc1 elements from C. elegans, and IAP elements from mice (retrotransposon).

[0097] In some embodiments of the methods provided herein, one or more of the nucleases is a zinc-finger nuclease (ZFN). ZFNs are engineered double-strand break inducing agents comprised of a zinc finger DNA binding domain and a double strand break inducing agent domain. Engineered ZFNs consist of two zinc finger arrays (ZFAs), each of which is fused to a single subunit of a non-specific endonuclease, such as the nuclease domain from the *Fok*I enzyme, which becomes active upon dimerization. Typically, a single ZFA consists of 3 or 4 zinc finger domains, each of which is designed to recognize a specific nucleotide triplet (GGC, GAT, *etc.*). Thus, ZFNs composed of two "3-finger" ZFAs are capable of recognizing an 18 base pair target site; an 18 base pair recognition sequence is generally unique, even within large genomes such as those of humans and plants. By directing the co-localization and dimerization of two FokI nuclease monomers, ZFNs generate a functional site-specific endonuclease that creates a double-stranded break (DSB) in DNA at the targeted locus.

[0098] Useful zinc-finger nucleases include those that are known and those that are engineered to have specificity for one or more target sites (TS) described herein. Zinc finger domains are amenable for designing polypeptides which specifically bind a selected polynucleotide recognition sequence, for example, within the target site of the host cell genome. ZFNs consist of an engineered DNA-binding zinc finger domain linked to a non-specific endonuclease domain, for example nuclease domain from a Type IIs endonuclease such as HO or FokI. Alternatively, engineered zinc finger DNA binding domains can be fused to other double-strand break inducing agents or derivatives thereof that retain DNA nicking/cleaving activity. For example, this type of fusion can be used to direct the double-strand break inducing agent to a different target site, to alter the location of the nick or cleavage site, to direct the inducing agent to a shorter target site, or to direct the inducing agent to a longer target site. In some examples a zinc finger DNA binding domain is fused to a site-specific recombinase, transposase, or a derivative thereof that retains DNA nicking and/or cleaving activity. Additional functionalities can be fused to the zinc-finger binding domain,

including transcriptional activator domains, transcription repressor domains, and methylases. In some embodiments, dimerization of nuclease domain is required for cleavage activity.

[0099] Each zinc finger recognizes three consecutive base pairs in the target DNA. For example, a 3 finger domain recognized a sequence of 9 contiguous nucleotides, with a dimerization requirement of the nuclease, two sets of zinc finger triplets are used to bind a 18 nucleotide recognition sequence. Useful designer zinc finger modules include those that recognize various GNN and ANN triplets (Dreier, *et al.*, (2001) *J Biol Chem* 276:29466-78; Dreier, *et al.*, (2000) *J Mol Biol* 303:489-502; Liu, *et al.*, (2002) *J Biol Chem* 277:3850-6), as well as those that recognize various CNN or TNN triplets (Dreier, *et al.*, (2005) *J Biol Chem* 280:35588-97; Jamieson, *et al.*, (2003) *Nature Rev Drug Discov* 2:361-8). *See* also, Durai, *et al.*, (2005) *Nucleic Acids Res* 33:5978-90; Segal, (2002) *Methods* 26:76-83; Porteus and Carroll, (2005) *Nat Biotechnol* 23:967-73; Pabo, *et al.*, (2001) *Ann Rev Biochem* 70:313-40; Wolfe, *et al.*, (2000) *Ann Rev Biophys Biomol Struct* 29:183-212; Segal and Barbas, (2001) *Curr Opin Biotechnol* 12:632-7; Segal, *et al.*, (2003) *Biochemistry* 42:2137-48; Beerli and Barbas, (2002) *Nat Biotechnol* 20:135-41; Carroll, *et al.*, (2006) *Nature Protocols* 1:1329; Ordiz, *et al.*, (2002) *Proc Natl Acad Sci* USA 99:13290-5; Guan, *et al.*, (2002) *Proc Natl Acad Sci* USA 99:13296-301; WO2002099084; WO00/42219; WO02/42459; WO2003062455; US20030059767; US Patent Application Publication Number 2003/0108880; U.S. Pat. Nos. 6,140,466, 6,511,808 and 6,453,242. Useful zinc-finger nucleases also include those described in WO03/080809; WO05/014791; WO05/084190; WO08/021207; WO09/042186; WO09/054985; and WO10/065123.

6.5 Genomic Target Sites

[00100] In the methods provided herein, a nuclease is introduced to the host cell that is capable of causing a double-strand break near or within a genomic target site, which greatly increases the frequency of homologous recombination at or near the cleavage site. In preferred embodiments, the recognition sequence for the nuclease is present in the host cell genome only at the target site, thereby minimizing any off-target genomic binding and cleavage by the nuclease.

[00101] In some embodiments, the genomic target site is endogenous to the host cell, such as a native locus. In some embodiments, the native genomic target site is selected according to the type of nuclease to be utilized in the methods of integration provided herein. If the nuclease to be utilized is a zinc finger nuclease, optimal target sites may be selected using a number of publicly available online resources. *See, e.g.,* Reyon *et al., BMC Genomics* 12:83

(2011), which is hereby incorporated by reference in its entirety. For example, Oligomerized Pool Engineering (OPEN) is a highly robust and publicly available protocol for engineering zinc finger arrays with high specificity and *in vivo* functionality, and has been successfully used to generate ZFNs that function efficiently in plants, zebrafish, and human somatic and pluripotent stem cells. OPEN is a selection-based method in which a pre-constructed randomized pool of candidate ZFAs is screened to identify those with high affinity and specificity for a desired target sequence. ZFNGenome is a GBrowse-based tool for identifying and visualizing potential target sites for OPEN-generated ZFNs. ZFNGenome provides a compendium of potential ZFN target sites in sequenced and annotated genomes of model organisms. ZFNGenome currently includes a total of more than 11.6 million potential ZFN target sites, mapped within the fully sequenced genomes of seven model organisms; *S. cerevisiae, C. reinhardtii, A. thaliana, D. melanogaster, D. rerio, C. elegans*, and *H. sapiens.* Additional model organisms, including three plant species; *Glycine max* (soybean), *Oryza sativa* (rice), *Zea mays* (maize), and three animal species *Tribolium castaneum* (red flour beetle), *Mus musculus* (mouse), *Rattus norvegicus* (brown rat) will be added in the near future. ZFNGenome provides information about each potential ZFN target site, including its chromosomal location and position relative to transcription initiation site(s). Users can query ZFNGenome using several different criteria (*e.g.*, gene ID, transcript ID, target site sequence).

[00102] If the nuclease to be utilized is a TAL-effector nuclease, in some embodiments, optimal target sites may be selected in accordance with the methods described by Sanjana *et al., Nature Protocols,* 7:171-192 (2012), which is hereby incorporated by reference in its entirety. In brief, TALENs function as dimers, and a pair of TALENs, referred to as the left and right TALENs, target sequences on opposite strands of DNA. TALENs are engineered as a fusion of the TALE DNA-binding domain and a monomeric FokI catalytic domain. To facilitate FokI dimerization, the left and right TALEN target sites are chosen with a spacing of approximately 14–20 bases. Therefore, for a pair of TALENs, each targeting 20-bp sequences, an optimal target site should have the form 5′-$TN^{19}N^{14\text{-}20}N^{19}$A-3′, where the left TALEN targets 5′-TN^{19}-3′ and the right TALEN targets the antisense strand of 5′-N^{19}A-3′ (N = A, G, T or C).

[00103] In other embodiments of the methods provided herein, the genomic target site is exogenous to the host cell. For example, one or more genomic target sites can be engineered into the host cell genome using traditional methods, *e.g.,* gene targeting, prior to performing the methods of integration described herein. In some embodiments, multiple copies of the same target sequence are engineered into the host cell genome at different loci, thereby

facilitating simultaneous multiple integration events with the use of only a single nuclease that specifically recognizes the target sequence. In other embodiments, a plurality of different target sequences is engineered into the host cell genome at different loci. In some embodiments, the engineered target site comprises a target sequence that is not otherwise represented in the native genome of the host cell. For example, homing endonucleases target large recognition sites (12-40 bp) that are usually embedded in introns or inteins, and as such, their recognition sites are extremely rare, with none or only a few of these sites present in a mammalian-sized genome. Thus, in some embodiments, the exogenous genomic target site is a recognition sequence for a homing endonuclease. In some embodiments, the homing nuclease is selected from the group consisting of : H-DreI, I-SceI, I-SceII, I-SceIII, I-SceIV, I-SceV, I-SceVI, I-SceVII, I-CeuI, I-CeuAIIP, I-CreI, I-CrepsbIP, I-CrepsbIIP, I-CrepsbIIIP, I-CrepsbIVP, I-TliI, I-PpoI, Pi-PspI, F-SceI, F-SceII, F-SuvI, F-CphI, F-TevI, F-TevII, I-AmaI, I-AniI, I-ChuI, I-CmoeI, I-CpaI, I-CpaII, I-CsmI, I-CvuI, I-CvuAIP, I-DdiI, I-DdiII, I-DirI, I-DmoI, I-HmuI, I-HmuII, I-HsNIP, I-LlaI, I-MsoI, I-NaaI, I-NanI, I-NclIP, I-NgrIP, I-NitI, I-NjaI, I-Nsp236IP, I-PakI, I-PboIP, I-PcuIP, I-PcuAI, I-PcuVI, I-PgrIP, I-PobIP, I-PorI, I-PorIIP, I-PbpIP, I-SpBetaIP, I-ScaI, I-SexIP, I-SneIP, I-SpomI, I-SpomCP, I-SpomIP, I-SpomIIP, I-SquIP, I-Ssp68031, I-SthPhiJP, I-SthPhiST3P, I-SthPhiSTe3bP, I-TdeIP, I-TevI, I-TevII, I-TevIII, I-UarAP, I-UarHGPAIP, I-UarHGPA13P, I-VinIP, I-ZbiIP, PI-MgaI, PI-MtuI, PI-MtuHIP PI-MtuHIIP, PI-PfuI, PI-PfuII, PI-PkoI, PI-PkoII, PI-Rma43812IP, PI-SpBetaIP, PI-SceI, PI-TfuI, PI-TfuII, PI-ThyI, PI-TliI, or PI-TliII, or any variant or derivative thereof. In particular embodiments, the exogenous genomic target site is the recognition sequence for I-SceI, VDE (PI-SceI), F-CphI, PI-MgaI or PI-MtuII, each of which are provided below.

Table 1: Recognition and cleavage sites for select homing endonucleases.

Nuclease	**Recognition sequence**
I-SceI	TAGGGATAACAGGGTAAT (SEQ ID NO:52)
VDE (PI-SceI)	TATGTCGGGTGCGGAGAAAGAGGTAATGAAA (SEQ ID NO:53)
F-CphI	GATGCACGAGCGCAACGCTCACAA (SEQ ID NO:54)
PI-MgaI	GCGTAGCTGCCCAGTATGAGTCAG (SEQ ID NO:55)
PI-MtuII	ACGTGCACTACGTAGAGGGTCGCACCGCACCGATCTACAA (SEQ ID NO:56)

6.6 Delivery

[00104] In some embodiments, the one or more nucleases useful for the methods described herein are provided, *e.g.,* delivered into the host cell as a purified protein. In other embodiments, the one or more nucleases are provided via polynucleotide(s) comprising a nucleic acid encoding the nuclease. In other embodiments, the one or more nucleases are introduced into the host cell as purified RNA which can be directly translated in the host cell nucleus.

[00105] In certain embodiments, an integration polynucletide, a polynucleotide encoding a nuclease, or a purified nuclease protein as described above, or any combination thereof, may be introduced into a host cell using any conventional technique to introduce exogenous protein and/or nucleic acids into a cell known in the art. Such methods include, but are not limited to, direct uptake of the molecule by a cell from solution, or facilitated uptake through lipofection using, *e.g.,* liposomes or immunoliposomes; particle-mediated transfection; *etc.* See, *e.g.,* U.S. Patent No. 5,272,065; Goeddel *et al.*, eds, 1990, *Methods in Enzymology*, vol. 185, Academic Press, Inc., CA; Krieger, 1990, *Gene Transfer and Expression -- A Laboratory Manual*, Stockton Press, NY; Sambrook *et al.*, 1989, *Molecular Cloning -- A Laboratory Manual*, Cold Spring Harbor Laboratory, NY; and Ausubel *et al.*, eds., Current Edition, *Current Protocols in Molecular Biology*, Greene Publishing Associates and Wiley Interscience, NY. Particular methods for transforming cells are well known in the art. *See* Hinnen *et al., Proc. Natl. Acad. Sci.* USA 75:1292-3 (1978); Cregg *et al., Mol. Cell. Biol.* 5:3376-3385 (1985). Exemplary techniques include but are not limited to, spheroplasting, electroporation, PEG 1000 mediated transformation, and lithium acetate or lithium chloride mediated transformation.

[00106] In some embodiments, biolistics are utilized to introduce an integration polynucletide, a polynucleotide encoding a nuclease, a purified nuclease protein, or any combination thereof into the host cell, in particular, host cells that are otherwise difficult to transform/transfect using conventional techniques, such as plants. Biolistics work by binding the transformation reaction to microscopic gold particles, and then propelling the particles using compressed gas at the target cells.

[00107] In some embodiments, the polynucleotide comprising nucleic acid encoding the nuclease is an expression vector that allows for the expression of a nuclease within a host cell. Suitable expression vectors include but are not limited to those known for use in expressing genes in *Escherichia coli*, yeast, or mammalian cells. Examples of *Escherichia coli* expression vectors include but are not limited to pSCM525, pDIC73, pSCM351, and pSCM353.

Examples of yeast expression vectors include but are not limited to pPEX7 and pPEX408. Other examples of suitable expression vectors include the yeast-*Escherichia coli* pRS series of shuttle vectors comprising CEN.ARS sequences and yeast selectable markers; and 2μ plasmids. In some embodiments, a polynucleotide encoding a nuclease can be modified to substitute codons having a higher frequency of usage in the host cell, as compared to the naturally occurring polynucleotide sequence. For example the polynucleotide encoding the nuclease can be modified to substitute codons having a higher frequency of usage in *S. cerevisiae*, as compared to the naturally occurring polynucleotide sequence.

[00108] In some embodiments where the nuclease functions as a heterodimer requiring the separate expression of each monomer, as is the case for zinc finger nucleases and TAL-effector nucleases, each monomer of the heterodimer may be expressed from the same expression plasmid, or from different plasmids. In embodiments where multiple nucleases are introduced to the cell to effect double-strand breaks at different target sites, the nucleases may be encoded on a single plasmid or on separate plasmids.

[00109] In certain embodiments, the nuclease expression vector further comprises a selectable marker that allows for selection of host cells comprising the expression vector. Such selection can be helpful to retain the vector in the host cell for a period of time necessary for expression of sufficient amounts of nuclease to occur, for example, for a period of 12, 24, 36, 48, 60, 72, 84, 96, or more than 96 hours, after which the host cells may be grown under conditions under which the expression vector is no longer retained. In certain embodiments, the selectable marker is selected from the group consisting of: URA3, hygromycin B phosphotransferase, aminoglycoside phosphotransferase, zeocin resistance, and phosphinothricin N-acetyltransferase. In some embodiments, the nuclease expression vector vector may comprise a counter-selectable marker that allows for selection of host cells that do not contain the expression vector subsequent to integration of the one or more donor nucleic acid molecules. The nuclease expression vector used may also be a transient vector that has no selection marker, or is one that is not selected for. In particular embodiments, the progeny of a host cell comprising a transient nuclease expression vector loses the vector over time.

[00110] In certain embodiments, the expression vector further comprises a transcription termination sequence and a promoter operatively linked to the nucleotide sequence encoding the nuclease. In some embodiments, the promoter is a constitutive promoter. In some embodiments, the promoter is an inducible promoter. Illustrative examples of promoters suitable for use in yeast cells include, but are not limited to the promoter of the TEF1 gene of *K. lactis*, the promoter of the PGK1 gene of *Saccharomyces cerevisiae*, the promoter of the

TDH3 gene of *Saccharomyces cerevisiae*, repressible promoters, *e.g.,* the promoter of the CTR3 gene of *Saccharomyces cerevisiae*, and inducible promoters, *e.g.,* galactose inducible promoters of *Saccharomyces cerevisiae* (*e.g.,* promoters of the GAL1, GAL7, and GAL10 genes).

[00111] In some embodiments, an additional nucleotide sequence comprising a nuclear localization sequence (NLS) is linked to the 5' of the nucleotide sequence encoding the nuclease. The NLS can facilitate nuclear localization of larger nucleases (>25 kD). In some embodiments, the nuclear localization sequence is an SV40 nuclear localization sequence. In some embodiments, the nuclear localization sequence is a yeast nuclear localization sequence.

[00112] A nuclease expression vector can be made by any technique apparent to one skilled in the art. In certain embodiments, the vector is made using polymerase chain reaction (PCR) and molecular cloning techniques well known in the art. *See, e.g., PCR Technology: Principles and Applications for DNA Amplification*, ed. HA Erlich, Stockton Press, New York, N.Y. (1989); Sambrook *et al.*, 2001, *Molecular Cloning – A Laboratory Manual*, 3rd edition, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

6.7 Host Cells

[00113] In another aspect, provided herein is a modified host cell generated by any of the methods of genomically integrating one or more exogenous nucleic acids described herein. Suitable host cells include any cell in which integration of a nucleic acid or "donor DNA" of interest into a chromosomal or episomal locus is desired. In some embodiments, the cell is a cell of an organism having the ability to perform homologous recombination. Although several of the illustrative embodiments are demonstrated in yeast (*S. cerevisiae*), it is believed that the methods of genomic modification provided herein can be practiced on all biological organisms having a functional recombination system, even where the recombination system is not as proficient as in yeast. Other cells or cell types that have a functional homologous recombination systems include bacteria such as *Bacillus subtilis* and *E. coli* (which is RecE RecT recombination proficient; Muyrers *et al.*, *EMBO rep.* 1: 239-243, 2000); protozoa (*e.g.*, *Plasmodium*, *Toxoplasma*); other yeast (*e.g.*, *Schizosaccharomyces pombe*); filamentous fungi (*e.g., Ashbya gossypii*); plants, for instance the moss *Physcomitrella patens* (Schaefer and Zryd, *Plant J.* 11: 1195-1206, 1997); and animal cells, such as mammalian cells and chicken DT40 cells (Dieken *et al.*, *Nat. Genet.* 12:174-182, 1996).

[00114] In some embodiments, the host cell is a prokaryotic cell. In some embodiments, the host cell is a eukaryotic cell. In some embodiments, the cell is a fungal cell (for instance, a

yeast cell), a bacteria cell, a plant cell, or an animal cell (for instance, a chicken cell). In some embodiments, the host cell is a mammalian cell. In some embodiments, the host cell is a Chinese hamster ovary (CHO) cell, a COS-7 cell, a mouse fibroblast cell, a mouse embryonic carcinoma cell, or a mouse embryonic stem cell. In some embodiments, the host cell is an insect cell. In some embodiments, the host cell is a S2 cell, a Schneider cell, a S12 cell, a 5B1-4 cell, a Tn5 cell, or a Sf9 cell. In some embodiments, the host cell is a unicellular eukaryotic organism cell.

[00115] In particular embodiments, the host cell is a yeast cell. Useful yeast host cells include yeast cells that have been deposited with microorganism depositories (*e.g.* IFO, ATCC, *etc.*) and belong to the genera *Aciculoconidium, Ambrosiozyma, Arthroascus, Arxiozyma, Ashbya, Babjevia, Bensingtonia, Botryoascus, Botryozyma, Brettanomyces, Bullera, Bulleromyces, Candida, Citeromyces, Clavispora, Cryptococcus, Cystofilobasidium, Debaryomyces, Dekkara, Dipodascopsis, Dipodascus, Eeniella, Endomycopsella, Eremascus, Eremothecium, Erythrobasidium, Fellomyces, Filobasidium, Galactomyces, Geotrichum, Guilliermondella, Hanseniaspora, Hansenula, Hasegawaea, Holtermannia, Hormoascus, Hyphopichia, Issatchenkia, Kloeckera, Kloeckeraspora, Kluyveromyces, Kondoa, Kuraishia, Kurtzmanomyces, Leucosporidium, Lipomyces, Lodderomyces, Malassezia, Metschnikowia, Mrakia, Myxozyma, Nadsonia, Nakazawaea, Nematospora, Ogataea, Oosporidium, Pachysolen, Phachytichospora, Phaffia, Pichia, Rhodosporidium, Rhodotorula, Saccharomyces, Saccharomycodes, Saccharomycopsis, Saitoella, Sakaguchia, Saturnospora, Schizoblastosporion, Schizosaccharomyces, Schwanniomyces, Sporidiobolus, Sporobolomyces, Sporopachydermia, Stephanoascus, Sterigmatomyces, Sterigmatosporidium, Symbiotaphrina, Sympodiomyces, Sympodiomycopsis, Torulaspora, Trichosporiella, Trichosporon, Trigonopsis, Tsuchiyaea, Udeniomyces, Waltomyces, Wickerhamia, Wickerhamiella, Williopsis, Yamadazyma, Yarrowia, Zygoascus, Zygosaccharomyces, Zygowilliopsis*, and *Zygozyma*, among others.

[00116] In some embodiments, the yeast host cell is a *Saccharomyces cerevisiae* cell, a *Pichia pastoris* cell, a *Schizosaccharomyces pombe* cell, a *Dekkera bruxellensis* cell, a *Kluyveromyces lactis* cell, a *Arxula adeninivorans* cell, or a *Hansenula polymorpha* (now known as *Pichia angusta*) cell. In a particular embodiment, the yeast host cell is a *Saccharomyces cerevisiae* cell. In some embodiments, the yeast host cell is a *Saccharomyces fragilis* cell or a *Kluyveromyces lactis* (previously called *Saccharomyces lactis*) cell. In some embodiments, the yeast host cell is a cell belonging to the genus *Candida*, such as *Candida*

lipolytica, *Candida guilliermondii*, *Candida krusei*, *Candida pseudotropicalis*, or *Candida utilis*. In another particular embodiment, the yeast host cell is a *Kluveromyces marxianus* cell.

[00117] In particular embodiments, the yeast host cell is a *Saccharomyces cerevisiae* cell selected from the group consisting of a Baker's yeast cell, a CBS 7959 cell, a CBS 7960 cell, a CBS 7961 cell, a CBS 7962 cell, a CBS 7963 cell, a CBS 7964 cell, a IZ-1904 cell, a TA cell, a BG-1 cell, a CR-1 cell, a SA-1 cell, a M-26 cell, a Y-904 cell, a PE-2 cell, a PE-5 cell, a VR-1 cell, a BR-1 cell, a BR-2 cell, a ME-2 cell, a VR-2 cell, a MA-3 cell, a MA-4 cell, a CAT-1 cell, a CB-1 cell, a NR-1 cell, a BT-1 cell, and a AL-1 cell. In some embodiments, the host cell is a *Saccharomyces cerevisiae* cell selected from the group consisting of a PE-2 cell, a CAT-1 cell, a VR-1 cell, a BG-1 cell, a CR-1 cell, and a SA-1 cell. In a particular embodiment, the *Saccharomyces cerevisiae* host cell is a PE-2 cell. In another particular embodiment, the *Saccharomyces cerevisiae* host cell is a CAT-1 cell. In another particular embodiment, the *Saccharomyces cerevisiae* host cell is a BG-1 cell.

[00118] In some embodiments, the yeast host cell is a cell that is suitable for industrial fermentation, *e.g.*, bioethanol fermentation. In particular embodiments, the cell is conditioned to subsist under high solvent concentration, high temperature, expanded substrate utilization, nutrient limitation, osmotic stress due, acidity, sulfite and bacterial contamination, or combinations thereof, which are recognized stress conditions of the industrial fermentation environment.

6.8 Kits

[00119] In another aspect, provided herein is a kit useful for performing the methods for genomically integrating one or more exogenous nucleic acids described herein. In some embodiments, the kit comprises:

(a) a plurality of exogenous nucleic acids, wherein each exogenous nucleic acid $(ES)_x$ comprises:

(i) a first homology region $(HR1)_x$ and a second homology region $(HR2)_x$, wherein $(HR1)_x$ and $(HR2)_x$ are capable of initiating host cell mediated homologous recombination of $(ES)_x$ at a selected target site $(TS)_x$ of a host cell genome; and

(ii) a nucleic acid of interest $(D)_x$ positioned 3' of $(HR1)_x$ and 5' of $(HR2)_x$;

(b) a plurality of nucleases, wherein each nuclease $(N)_x$ capable of cleaving at $(TS)_x$, whereupon said cleaving results in homologous recombination of $(ES)_x$ at $(TS)_x$;

wherein x is any integer from 1 to n wherein n is at least 2.

[00120] In some embodiments, $(D)_x$ is selected from the group consisting of a selectable marker, a promoter, a nucleic acid sequence encoding an epitope tag, a gene of interest, a reporter gene, and a nucleic acid sequence encoding a termination codon. In some embodiments, the kit further comprises a plurality of primer pairs $(P)_x$, wherein each primer pair is capable of identifying integration of $(ES)_x$ at $(TS)_x$ by PCR. In some embodiments, $(ES)_x$ is linear. In some embodiments, $(ES)_x$ is circular.

[00121] In a particular embodiment, the kit enables site-specific integration of an exogenous nucleic acid at a unique target site within any of the approximately 6000 genetic loci of the yeast genome. In these embodiments, n= ≥6000, wherein each $(TS)_x$ is unique to a single locus of the yeast cell genome.

[00122] In some embodiments, the kit further comprises instructions for use that describe methods for integrating one or more exogenous nucleic acids into any genetic locus of a host yeast cell.

7. EXAMPLES

7.1 Example 1: Simultaneous multiple integration of a plurality of exogenous nucleic acids

[00123] The methods and compositions described herein are implemented to create a modified yeast cell comprising two exogenous nucleic acids integrated at two loci of the yeast cell genome in a single transformation step, wherein recovery of the modified yeast cell does not require the use of selectable marker(s).

[00124] A host strain is provided comprising: (a) a previously introduced recognition site for the F-CphI endonuclease positioned within the NDT80 locus; and (b): a previously introduced recognition site for the I-SceI endonuclease positioned within the HO locus. The host cell is simultaneously transformed with: (1) an expression plasmid encoding F-CphI; (2) an expression plasmid encoding I-SceI; (3) a linear DNA comprising an expression cassette encoding green fluorescent protein (GFP), flanked by two stretches of >500 bp sequence corresponding to the 5' and 3' regions of the NDT80 locus; and (4) a linear DNA comprising an expression cassette encoding lacZ, flanked by two stretches of >500 bp sequence corresponding to the 5' and 3' regions of the HO locus. As an alternative to inclusion of the expression plamids encoding F-CphI and I-SceI, respectively, purified F-CphI and I-SceI protein are included in the transformation reaction. A non-double strand break control is performed by transforming host cells with the linear integration constructs (3) and (4) only, in the absence of F-CphI and I-SceI expression plasmid or purified protein.

[00125] Experimental and control transformants are plated on selection-free media, and colonies from each plate are visualized for expression of GFP and lacZ, respectively. Colony PCR is independently performed with a primer pair which anneals upstream and downstream of the junction between the integrated integration construct (3) or (4), respectively, and their respective target sequences, to confirm fidelity and frequency of integration.

7.2 Example 2: Simultaneous multiple integration of a plurality of exogenous nucleic acids

[00126] This Example provides results which demonstrate simultaneous integration of three exogenous nucleic acids at three different loci of a *S. cerevisiae* host following the induction of targeted double-stranded breaks in the host cell genome. In brief, an exogenous "target" nucleic acid sequence encoding a truncated, non-functional copy of Emerald Green Fluorescent Protein (emgfpΔ) was integrated into the HO, YGR250c and NDT80 loci, respectively, of host yeast cells. Recombinant cells were transformed with linear "donor" DNA encoding an intact, functional copy of Emerald Green Fluorescent Protein (EmGFP) and either: (1) empty vector; or (2) an expression vector, pZFN.gfp, encoding a zinc-finger nuclease (ZFN.gfp) that specifically recognizes and cleaves a sequence within the emgfpΔ coding sequence. Transformed colonies were screened by colony PCR (cPCR) for the replacement of one, two or three genomically integrated copies of the target emgfpΔ coding sequence with the donor EmGFP coding sequence.

7.2.1. Construction and Integration of Target DNA

[00127] To generate exogenous genomic target sites for nuclease-mediated double-strand breaks, target DNAs encoding emgfpΔ were constructed using RYSE-mediated assembly, as described in U.S. Patent No. 8,110,360, the contents of which are hereby incorporated by reference in their entirety. Nucleotides 450 to 462 of the wild-type EmGFP coding sequence (SEQ ID NO:1) were replaced with the following sequence: 5'-CGTCTAAATCATG-3' (SEQ ID NO:2), resulting in the introduction of: (1) a premature stop codon at position 152 of EmGFP (emgfpΔ); and (2) the recognition/cleavage sequence for ZFN.gfp.

[00128] For the targeted integration of the emgfpΔ coding sequence into each of the HO, YGR250c and NDT80 loci, the emgfpΔ coding sequence was flanked with ~200-500 bp of upstream and downstream homologous sequences for each loci (SEQ ID NOS:3-8). A unique selectable marker was also incorporated into each construct, positioned 5' to the emgfpΔ coding sequence, for selection of colonies having successful integration events. The HO integration construct included KanR, the YGR250c integration construct included URA3, and

the NDT80 integration construct included NatR. Each integration construct was transformed sequentially into a naïve CEN.PK2 haploid yeast strain (strain A), and the strain was confirmed to have three integrated copies of the emgfpΔ coding sequence.

7.2.2. Construction of ZFN Yeast Expression Plasmid

[00129] Zinc finger nucleases consist of two functional domains: a DNA-binding domain comprised of a chain of zinc finger proteins and a DNA-cleaving domain comprised of the nuclease domain of *FokI*. The endonuclease domain of *FokI* functions as an obligate heterodimer in order to cleave DNA, and thus, a pair of ZFNs is required to bind and cut its target sequence. The target sequence of ZFN.gfp (CompoZr® Zinc Finger Nuclease, Sigma-Aldrich, St. Louis, MO) is:
5'-ACAACTACAACAGCCACAACgtctatATCATGGCCGACAAGCA-3' (SEQ ID NO: 9), with the recognition sequence indicated in uppercase and the cleavage sequence indicated in lowercase.

[00130] A high-copy ZFN.gfp yeast expression plasmid, pZFN.gfp, was constructed as follows. The genes ZFN.gfp.1 and ZFN.gfp.2, each encoding one member of the ZFN.gfp obligate heterodimer, were PCR-amplified from a mammalian expression plasmid and fused to the divergent $P_{GAL1,10}$ promoter and ADH1 and CYC1 terminators, respectively. Individual PCR products of P_{GAL10}>ZFN.gfp.1-T_{ADH1} and P_{GAL1}>ZFN.gfp.2-T_{CYC1}, along with a linearized vector backbone comprising a LEU2 selectable marker, were co-transformed into a naïve yeast strain for *in vivo* assembly via homologous recombination of overlapping ends. The PCR products recombined at the pGAL1,10 promoter sequence and assembled into the vector backbone via homologous sequences added by the terminal primers. Transformants were selected on minimal media lacking leucine, isolated, and grown in liquid media. The plasmids from multiple clones were extracted from yeast using the Zymoprep Yeast Plasmid Miniprep I kit (Zymo Research). The eluent from the extraction protocol was then transformed into *E. coli* XL-1 blue chemically competent cells. Plasmids were propagated overnight in *E. coli* and miniprepped (Qiagen, Valencia, CA). Correct clones were identified by restriction mapping.

7.2.3. Transformation with Donor DNA and Induction of Double-Strand Breaks

[00131] A standard lithium acetate/SSDNA/PEG protocol (Gietz and Woods, *Methods Enzymol.* 350:87-96 (2002)) was used to co-transform strain A with linear "donor" DNA encoding EmGFP and either: (1) empty vector; or (2) the pZFN.gfp expression vector. The

EmGFP coding sequence differs from the emgfpΔ coding sequence at positions within the recognition/cleavage site for ZFN.gfp, namely positions 450 (C→G), 456 (A→T), 461 (T→C) and 462 (G→C). Thus, ZFN.gfp is expected to recognize and cleave within the emgfpΔ sequence but not within the EmGFP sequence.

[00132] One microgram of the appropriate plasmid DNA was co-transformed with 70 ul of linear EmGFP DNA (~300 ng/ul). All transformations were recovered overnight in YP + 2% galactose to induce ZFN expression. Various dilutions were plated onto minimal media agar plates lacking leucine to select for colonies transformed with plasmid DNA. Plates were incubated for 3 days at 30°C.

7.2.4. Confirmation of Multiple Simultaneous Integration

[00133] Colony PCR was performed to determine the frequency of replacement of the emgfpΔ coding sequence with the EmGFP coding sequence at each target locus. DNA was prepped from 96 colonies from each transformation and probed with primer pairs specific for EmGFP and HO, EmGFP and NDT80, and EmGFP and YGR250c, respectively, such that successful integration of the EmGFP coding sequence at each locus was expected to produce an amplicon of a predicted size, while non-integration was expected to produce no amplicon.

Table 2. Primer sequences for cPCR verification of multiple integration of the EmGFP coding sequence

Primer Name	Description	Sequence	SEQ ID NO
KMH749-Fixed GFP-fwd	Forward primer specific to Em.GFP	CAACTACAACAGCCACAAGGTCTATATCACC	SEQ ID NO: 10
CR813	Reverse primer for HO locus	CTCTAACGCTGTTGGTAGATTG	SEQ ID NO: 11
KMH773-NDT80-Ar	Reverse primer for NDT80 locus	ACCATGTGATAATACACTACTAATGTGACTACTAGTTGA	SEQ ID NO: 12
KMH679-YGR250c 3' rev	Reverse primer for YGR250c locus	TCAGACGCGTTCGGAGGAGAGTGCATTCAC	SEQ ID NO: 13

[00134] As indicated in FIG. 5, of the 96 colonies transformed with linear EmGFP donor DNA (SEQ ID NO:1) and empty vector control, no amplicons were produced during PCR, indicating that there were no successful integration events, *i.e.,* replacements at any of the three loci comprising the target emgfpΔ coding sequence in the absence of a double-strand break.

By contrast, of the 96 colonies transformed with linear EmGFP DNA and pZFN.gfp, 2 colonies had one locus replaced, 4 colonies had two loci replaced, and 23 colonies had all three loci replaced with the EmGFP coding sequence (FIG. 6). Colony PCR results were corroborated by visualizing the fluorescence of transformed colonies on plates (data not shown). None of the colonies transformed with EmGFP DNA and empty vector appeared green, indicating that none of the target emgfpΔ coding sequences were replaced with functional EmGFP coding sequences. By contrast, ~20% of colonies transformed with EmGFP DNA and pZFN.gfp appeared green, roughly correlating with the frequency of integration events observed by cPCR.

[00135] These results demonstrate that induction of multiple targeted double-strand breaks in the genome of a host cell can facilitate simultaneous multiple targeted integration of exogenous donor nucleic acids.

7.3 Example 3: Simultaneous multiple integration of terpene synthase genes to facilitate conversion of a farnesene producing strain to an amorphadiene producing strain

[00136] This Example provides results which demonstrate simultaneous integration of three sesquiterpene synthase genes at three different engineered loci of a *S. cerevisiae* host engineered for high mevalonate pathway flux. As a result, a parental strain producing farnesene and comprising a plasmid-based copy of the farnesene synthase gene was converted into an amorphadiene producing strain comprising multiple genomically integrated copies of amorphadiene synthase. In brief, URA3, NatR and KanR marker cassettes flanked by F-CphI sites were integrated at the Gal80, HXT3 and Matα locus, respectively, of the host strain. The host was then co-transformed with a plasmid encoding the F-CphI endonuclease as well as three linear "donor" DNA constructs containing distinct codon optimizations of the amorphadiene synthase (ADS) gene expressed from the Gal1 promoter and terminated by the CYC1 terminator (ADS cassette), each flanked by homology regions for their respective target locus. Transformed colonies were screened by colony PCR (cPCR) for the replacement of one, two or three genomically integrated target marker loci with the ADS cassettes. A triply-integrated strain was identified and further engineered by integrating a fourth ADS cassette, and the resulting strain was cultured under conditions allowing for loss of the plasmid encoding farnesene synthase, such that its product profile was fully converted from farnesene to amorphadiene.

7.3.1. Construction of a Parental Farnesene Producing Strain

[00137] A farnesene-producing yeast strain, Y3639, useful for the multiple simultaneous integration of exogenous donor DNAs encoding amorphadiene synthase, was prepared as follows.

[00138] Strains Y93 (MAT A) and Y94 (MAT alpha) were generated by replacing the promoter of the ERG9 gene of yeast strains Y002 and Y003 (CEN.PK2 background MAT A or MAT alpha, respectively; ura3-52; trp1-289; leu2-3,112; his3Δ1; MAL2-8C; SUC2; van Dijken *et al.* (2000) *Enzyme Microb. Technol.* 26:706-714), respectively, with the promoter of the MET3 gene of *Saccharomyces cerevisiae*. To this end, exponentially growing Y002 and Y003 cells were transformed with integration construct i8 (SEQ ID NO: 14), which comprised the kanamycin resistance marker (KanMX) flanked by the promoter and terminator of the Tef1 gene of *Kluyveromyces lactis,* the ERG9 coding sequence, a truncated segment of the ERG9 promoter (trunc. PERG9), and the MET3 promoter (PMET3), flanked by ERG9 upstream and downstream sequences. Host cell transformants were selected on medium comprising 0.5 μg/mL Geneticin (Invitrogen Corp., Carlsbad, CA), and selected clones were confirmed by diagnostic PCR, yielding strains Y93 and Y94.

[00139] Strains Y176 (MAT A) and Y177 (MAT alpha) were generated by replacing the coding sequence of the ADE1 gene in strains Y93 and Y94, respectively, with the coding sequence of the LEU2 gene of *Candida glabrata* (*Cg*LEU2). To this end, the 3.5 kb *Cg*LEU2 genomic locus was PCR amplified from *Candida glabrata* genomic DNA (ATCC, Manassas, VA) using primers 61-67-CPK066-G (SEQ ID NO: 15) and 61-67-CPK067-G (SEQ ID NO: 16), and transforming the PCR product into exponentially growing Y93 and Y94 cells. Host cell transformants were selected on CSM-L, and selected clones were confirmed by diagnostic PCR, yielding strains Y176 and Y177.

[00140] Strain Y188 was generated by introducing into strain Y176 an additional copy of the coding sequences of the ERG13, ERG10, and ERG12 genes of *Saccharomyces cerevisiae*, and a truncated coding sequence of the HMG1 gene of *Saccharomyces cerevisiae*, each under regulatory control of a galactose inducible promoter of the GAL1 or GAL10 gene of *Saccharomyces cerevisiae*. To this end, exponentially growing Y176 cells were transformed with 2 μg of expression plasmids pAM491 and pAM495 digested with PmeI restriction endonuclease (New England Biolabs, Beverly, MA). Host cell transformants were selected on CSM lacking uracil and histidine (CSM-U-H), and selected clones were confirmed by diagnostic PCR, yielding strain Y188.

[00141] Strain Y189 was generated by introducing into strain Y177 an additional copy of the coding sequences of the ERG20, ERG8, and ERG19 genes of *Saccharomyces cerevisiae*, and a truncated coding sequence of the HMG1 gene of *Saccharomyces cerevisiae*, each under regulatory control of a galactose inducible promoter of the GAL1 or GAL10 gene of *Saccharomyces cerevisiae*. To this end, exponentially growing Y188 cells were transformed with 2 μg of expression plasmids pAM489 and pAM497 digested with PmeI restriction endonuclease. Host cell transformants were selected on CSM lacking tryptophan and histidine (CSM-T-H), and selected clones were confirmed by diagnostic PCR, yielding strain Y189.

[00142] Strain Y238 was generated by mating strains Y188 and Y189, and by introducing an additional copy of the coding sequence of the IDI1 gene of *Saccharomyces cerevisiae* and a truncated coding sequence of the HMG1 gene of *Saccharomyces cerevisiae*, each under regulatory control of a galactose inducible promoter of the GAL1 or GAL10 gene of *Saccharomyces cerevisiae*. To this end, approximately 1 x 10^7 cells of strains Y188 and Y189 were mixed on a YPD medium plate for 6 hours at room temperature, diploid cells were selected on CSM-H-U-T, and exponentially growing diploids were transformed with 2 μg of expression plasmid pAM493 digested with PmeI restriction endonuclease. Host cell transformants were selected on CSM lacking adenine (CSM-A), and selected clones were confirmed by diagnostic PCR, yielding strain Y238.

[00143] Strains Y210 (MAT A) and Y211 (MAT alpha) were generated by sporulating strain Y238. The diploid cells were sporulated in 2% potassium acetate and 0.02% raffinose liquid medium, and approximlately 200 genetic tetrads were isolated using a Singer Instruments MSM300 series micromanipulator (Singer Instrument Co, LTD. Somerset, UK). Spores were selected on CSM-A-H-U-T, and selected clones were confirmed by diagnostic PCR, yielding strains Y210 (MAT A) and Y211 (MAT alpha).

[00144] Strain Y221 was generated by transforming exponentially growing Y211 cells with vector pAM178. Host cell transformants were selected on CSM-L.

[00145] Strain Y290 was generated by deleting the coding sequence of the GAL80 gene of strain Y221. To this end, exponentially growing Y221 cells were transformed with integration construct i32 (SEQ ID NO: 17), which comprised the hygromycin B resistance marker (hph) flanked by the promoter and terminator of the Tef1 gene of *Kluyveromyces lactis* flanked by GAL80 upstream and downstream sequences. Host cell transformants were selected on medium comprising hygromycin B, and selected clones were confirmed by diagnostic PCR, yielding strain Y290.

[00146] Strain Y318 was generated by removing the pAM178 vector from strain Y290 by serial propagation in leucine-rich media, and testing individual colonies for their inability to grow on CSM-L, yielding strain Y318.

[00147] Strain Y409 was generated by introducing a heterologous nucleotide sequence encoding a β-farnesene synthase into strain Y318. To this end, exponentially growing Y318 cells were transformed with expression plasmid pAM404. Host cell transformants were selected on CSM-L, yielding strain Y409.

[00148] Strain Y419 was generated by rendering the GAL promoters of strain Y409 constitutively active. To this end, exponentially growing Y409 cells were transformed with integration construct i33 (SEQ ID NO: 18), which comprised the nourseothricin resistance marker of *Streptomyces noursei* (NatR) flanked by the promoter and terminator of the Tef1 gene of *Kluyveromyces lactis*, and the coding sequence of the GAL4 gene of *Saccharomyces cerevisiae* under regulatory control of an "operative constitutive" version of its native promoter (PGAL4oc; Griggs & Johnston (1991) *PNAS* **88**(19):8597-8601) and the GAL4 terminator (TGAL4), flanked by upstream and downstream sequences of the modified ERG9 promoter and coding sequences. Host cell transformants were selected on medium comprising nourseothricin, and selected clones were confirmed by diagnostic PCR, yielding strain Y419.

[00149] Strain Y677 was generated by introducing at the modified GAL80 locus of strain Y419 an additional copy of the coding region of the ERG12 gene of *Saccharomyces cerevisiae* under regulatory control of the promoter of the GAL1 gene of *Saccharomyces cerevisiae*. To this end, exponentially growing Y677 cells were transformed with integration construct i37 (SEQ ID NO: 19), which comprised the kanamycin resistance marker of *Streptomyces noursei* (KanR) flanked by the promoter and terminator of the Tef1 gene of *Kluyveromyces lactis*, and the coding and terminator sequences of the ERG12 gene of *Saccharomyces cerevisiae* flanked by the GAL1 promoter (PGAL1) and the ERG12 terminator (TERG12). Host cell transformants were selected on medium comprising kanamycin, and selected clones were confirmed by diagnostic PCR, yielding strain Y677.

[00150] Strain Y1551 was generated from strain Y677 by chemical mutagenesis. Mutated strains were screened for increased production of β–farnesene, yielding strain Y1551.

[00151] Strain Y1778 was generated from strain Y1551 by chemical mutagenesis. Mutated strains were screened for increased production of β–farnesene, yielding strain Y1778.

[00152] Strain Y1816 was generated by replacing the HXT3 coding sequence of strain Y1778 with two copies of an acetoacetyl-CoA thiolase coding sequence, one being derived

from *Saccharomyces cerevisiae* and the other from *C. butylicum*, and one copy of the coding sequence of the HMGS gene of *B. juncea*. To this end, exponentially growing Y1778 cells were transformed with integration construct i301 (SEQ ID NO: 20), which comprised the hygromycin B resistance marker (hyg) flanked by the promoter and terminator of the Tef1 gene of *Kluyveromyces lactis*, the coding sequence of the ERG10 gene of *Saccharomyces cerevisiae* flanked by a truncated TDH3 promoter (tPTDH3) and the AHP1 terminator (TAHP1), the coding sequence of the acetoacetyl-CoA thiolase gene of *C. butylicum* (thiolase) flanked by the YPD1 promoter (PYPD1) and CCW12 terminator (TCCW12), and the coding sequence of the HMGS gene of *B. juncea* (HMGS) preceded by the TUB2 promoter (PTUB2), flanked by upstream and downstream sequences of the HXT3 gene of *Saccharomyces cerevisiae*. Host cell transformants were selected on medium comprising hygromycin B, and selected clones were confirmed by diagnostic PCR, yielding strain Y1816.

[00153] Strain Y2055 was generated from strain Y1778 by chemical mutagenesis. Mutant strains were screened for increased production of β–farnesene, yielding strain Y2055.

[00154] Strain Y2295 was generated from strain Y2055 by chemical mutagenesis. Mutant strains were screened for increased production of β–farnesene, yielding strain Y2295.

[00155] Strain Y3111 was generated by switching the mating type of strain Y2295 from MAT A to MAT alpha. To this end, exponentially growing Y2295 cells were transformed with integration construct i476 (SEQ ID NO: 21), which comprised the MAT alpha mating locus and the hygromycin B resistance marker (hygA). Host cell transformants were selected on medium comprising hygromycin B, and selected clones were confirmed by diagnostic PCR, yielding strain Y3111.

[00156] Strain Y2168 was generated from strain Y1816 by chemical mutagenesis. Mutant strains were screened for increased production of β–farnesene, yielding strain Y2168.

[00157] Strain Y2446 was generated from strain Y2168 by chemical mutagenesis. Mutant strains were screened for increased production of β–farnesene, yielding strain Y2446.

[00158] Strain Y3118 was generated by inserting into the native URA3 locus of strain Y2446 the coding sequence, promoter, and terminator of the GAL80 gene of *Saccharomyces cerevisiae*. To this end, exponentially growing Y2446 cells were transformed with integration construct i477 (SEQ ID NO: 22), which comprised the promoter, terminator, and coding sequence of the GAL80 gene of *Saccharomyces cerevisiae* (GAL80) flanked by overlapping URA3 sequences (which enable loop-out excision of the GAL80 gene by homologous

recombination and restoration of the original URA3 sequence). Host cell transformants were selected on medium comprising 5-FOA, yielding strain Y3118.

[00159] Strain Y3215 was generated by mating strains Y3111 and Y3118. Approximately 1 x 10^7 cells of strains Y3111 and Y3118 were mixed on a YPD medium plate for 6 hours at room temperature to allow for mating, followed by plating on YPD agar plate to isolate single colonies. Diploids were identified by screening by colony PCR for the presence of both the hphA-marked MAT alpha locus and the wild-type MAT A locus.

[00160] Strain Y3000 was generated by sporulating strain Y3215 and looping out the GAL80 coding sequence. The diploid cells were sporulated in 2% potassium acetate and 0.02% raffinose liquid medium. Random spores were isolated, plated on YPD agar, grown for 3 days, and then replica-plated to CSM-U to permit growth only of cells lacking GAL80 (*i.e.*, having a functional URA3 gene). Spores were then tested for β–farnesene production, the best producer was identified, and the presence of integration construct i301 was confirmed by diagnostic PCR, yielding strain Y3000.

[00161] Strain Y3284 was generated by removing the URA3 marker from strain Y3000. To this end, exponentially growing Y3000 cells were transformed with integration construct i94 (SEQ ID NO: 23), which comprised the hisG coding sequence of *Salmonella,* and the coding sequence of the ERG13 gene and a truncated coding sequence of the HMG1 gene of *Saccharomyces cerevisiae* under control of a galactose inducible promoter of the GAL1 or GAL10 gene of *Saccharomyces cerevisiae*, flanked by upstream and downstream sequences of the URA3 gene of *Saccharomyces cerevisiae.* Host cell transformants were selected on medium comprising 5-FOA, and selected clones were confirmed by diagnostic PCR, yielding strain Y3284.

[00162] Strain Y3385 was generated by replacing the NDT80 coding sequence of strain Y3284 with an additional copy of the coding sequence of an acetyl-CoA synthetase gene of *Saccharomyces cerevisiae* and the coding sequence of the PDC gene of *Z. mobilis*. To this end, exponentially growing Y3385 cells were transformed with integration construct i467 (SEQ ID NO: 24), which comprised the URA3 marker, the coding sequence of the ACS2 gene of *Saccharomyces cerevisiae* (ACS2) flanked by the HXT3 promoter (PHXT3) and PGK1 terminator (TPGK1), and the coding sequence of the PDC gene of *Z. mobilis* (zmPDC) flanked by the GAL7 promoter (PGAL7) and the TDH3 terminator (TTDH3), flanked by upstream and downstream NDT80 sequences. Host cell transformants were selected on CSM-U, and selected clones were confirmed by diagnostic PCR, yielding strain Y3385.

[00163] Strain Y3547 was generated from strain Y3385 by chemical mutagenesis. Mutated strains were screened for increased production of β–farnesene, yielding strain Y3547.

[00164] Strain Y3639 was generated from strain Y3547 by chemical mutagenesis. Mutated strains were screened for increased production of β–farnesene, yielding strain Y3639.

7.3.2. Construction and Integration of Target DNA

[00165] Exogenous genomic target sites for FcphI endonuclease-mediated double-strand breaks were integrated into three different loci of strain Y3639. Three target site cassettes were constructed using PCR assembly of overlapping fragments, each comprising the recognition sequence for the FcphI endonuclease and the coding sequence for: (1) URA3 (flanked by homology regions for the modified Gal80 locus) (SEQ ID NO: 25); (2) NatR (flanked by homology regions for the modified HXT3 locus) (SEQ ID NO: 26); and (3) KanR (flanked by homology regions for the modified Matα locus) (SEQ ID NO: 27), respectively. Each target site cassette was serially transformed into Y3639, and the strain was confirmed by colony PCR to have three integrated copies of the F-CphI-flanked marker cassettes at the correct loci ("strain B").

7.3.3. Construction of F-CphI Yeast Expression Plasmid

[00166] The F-CphI yeast expression plasmid pAM1799, comprising a HygR selectable marker, has been described previously in U.S. Patent No. 7,919,605, which is hereby incorporated by reference in its entirety.

7.3.4. Transformation with Donor DNA and Induction of Double-Strand Breaks

[00167] The standard lithium acetate/SSDNA/PEG protocol (Gietz and Woods, *Methods Enzymol.* 2002;350:87-96) was modified to include a 30 minute, 30 degree incubation of the cells prior to the 42 degree heat shock. This method was used to co-transform strain B with pAM1799, encoding FcphI endonuclease, and three linear "donor" DNAs, each comprising a codon optimized coding sequence for amorphadiene synthase (ADS) of *Artemisia annua*, flanked by homology regions to the modified Gal80 (SEQ ID NO: 28), HXT3 (SEQ ID NO: 29) and Matα loci (SEQ ID NO: 30), respectively, of strain B.

[00168] One microgram of pAM1799 was co-transformed with ~100 ng of each of the ADS donor DNAs. All transformations were recovered overnight in YP + 2% galactose to induce F-CphI expression. Various dilutions were plated onto YPD agar plates containing hygromycin to select for colonies transformed with plasmid DNA. Plates were incubated for 3 days at 30°C.

7.3.5. Confirmation of Multiple Simultaneous Integration

[00169] Colony PCR (cPCR) was performed to determine the frequency of replacement of the F-CphI-flanked marker cassette coding sequences with the ADS cassette coding sequence. DNA was prepped from 20 colonies probed with primer pairs specific for ADS and the Gal80 locus, ADS and the HXT3 locus, and ADS and the Matα locus, respectively, such that successful integration of the ADS cassette coding sequence at each locus was expected to produce an amplicon of a predicted size, while non-integration was expected to produce no amplicon. PCR reactions to produce amplicons from the 5' and 3' ends of each locus were attempted in multiplex. In some cases, only the 5' or the 3' amplicon was successfully detected, but proper integration of the ADS cassette was confirmed at these loci by sequencing larger PCR fragments.

[00170] Table 3. Primer sequences for cPCR verification of multiple integration of the ADS cassette coding sequence

Primer Name	Description	Sequence	SEQ ID NO
CUT24	Gal80 locus US FOR	GTTTCTTTTGGATTGCGCTTGCC	SEQ ID NO: 31
ART12	ADS codon v2 5' REV	TACTGACAACCACATGTTAC	SEQ ID NO: 32
ART45	ADS ORF 5' REV	TACTGCTTCGGTAGTAGTTTCACC CTTCA	SEQ ID NO: 33
ART210	Gal80 locus DS REV	GGGAAGTCCAATTCAATAGT	SEQ ID NO: 34
HJ207	HXT3 locus US FOR	CATCTTCTCGAGATAACACCTGG AG	SEQ ID NO: 35
KB349	CYC1T FOR	ACGCGTGTACGCATGTAAC	SEQ ID NO: 36
HJ602	HXT3 locus DS REV	CAATTGGGGTTCTGGCAGTC	SEQ ID NO: 37
CUT76	Matα locus US FOR	GAAGCCTGCTTTCAAAATTAAGA ACAAAGC	SEQ ID NO: 38

HJ632	Matα locus DS REV	GAATTTACCTGTTCTTAGCTTGTA CCAGAG	SEQ ID NO: 39

[00171] Of the 20 colonies screened by cPCR, 14 had ADS integrated at the Gal80 locus, 17 had ADS integrated at the HXT3 locus, and four had ADS integrated at the Matα locus. The low rate of integration at the Matα locus can be explained by self-closure at this locus mediated by a direct repeat sequence flanking the F-CphI sites. In total, 6 clones had ADS integrated at a single site, 10 clones had ADS integrated at two sites, and three clones had ADS integrated at all three loci ("strains C"). The triply integrated strains were further confirmed by sequencing longer PCR products encompassing both flanks.

1.1.5 Completion of the integrated ADS strain and sesquiterpene assay

[00172] The triply integrated ADS strains were further engineered by integrating a final copy of ADS marked with a URA cassette (SEQ ID NO: 40) at the His3 locus using a standard protocol, and a resulting strain was confirmed for this fourth copy ("strain D"). Finally, strain D cells were passaged in non-selective media to lose the Leu+ marked high copy farnesene synthase plasmid (pAM404) ("strain E").

[00173] Several isolates of strain E were assayed for sesquiterpene production alongside strain D and the original parent strain B. In brief, isolates of strains B, D and E were incubated in separate wells of a 96-well plate containing 360 μL of Bird Seed Medium (BSM) with 2% sucrose per well (preculture). After 3 days of incubation at 33.5°C with 999 rpm agitation, 14.4 μL of each well was inoculated into a well of a new 96-well plate containing 360 μL of fresh BSM with 4% sucrose (production culture). After another 2 days of incubation at 33.5°C with 999 rpm agitation, samples were taken and analyzed for sesquiterpene production by gas chromatography (GC) analysis. Samples were extracted with methanol-heptane (1:1 v/v), and the mixtures were centrifuged to remove cellular material. An aliquot of the methanol-heptane extract was diluted into heptane, and then injected onto a methyl silicone stationary phase using a pulsed split injection. Farnesene and amorphadiene were separated by boiling point using GC with flame ionization detection (FID). Trans-β-caryophyllene was used as a retention time marker to monitor successful injection and elution during the specified GC oven profile.

[00174] As shown in FIG. 7, total sesquiterpene production remained nearly identical (3-3.5 g/L) for all strains, but the product profile was successfully converted from Farnesene (strain B) to mixed product (strain D) to amorphadiene (strain E).

[00175] These results demonstrate that induction of multiple targeted double-strand breaks in the genome of a host cell can facilitate simultaneous multiple integrations of a functional gene cassette, in this case facilitating conversion of a farnesene-producing strain into an amorphadiene-producing strain in a single transformation.

7.4 Example 4: Simultaneous replacement of multiple integrated copies of farnesene synthase with amorphadiene synthase

[00176] This Example provides results which demonstrate the simultaneous replacement of four genomically integrated terpene synthase genes, facilitated by designer nuclease-induced double-strand breaks within the synthase coding regions. In brief, an existing farnesene production strain, derived from strain Y3639 (described in Example 3) but comprising four integrated rather than extrachromasomal copies of the farnesene synthase (FS) gene, was co-transformed with a plasmid encoding a designer TAL-effector nuclease (TALEN) and four linear donor DNAs encoding new terpene synthase genes. The designer TALEN is capable of binding to and cleaving a sequence unique to the integrated farnesene synthase genes. Transformed colonies were screened by colony PCR (cPCR) and strains with one, two or three or four genomically integrated target marker loci were identified.

7.4.1. Construction and Integration of Target DNA

[00177] Four donor cassettes, each comprising a terpene synthase coding sequence flanked by homology regions (~500 bp) to its respective target loci, were assembled by overlap PCR. Three of the donor DNAs comprised ADS coding sequences and no selectable marker (SEQ ID NOs: 41-43), while the final donor DNA was a cassette comprising a novel codon optimization of the farnesene synthase (FS) fused to a URA3 marker cassette (SEQ ID NO: 44). None of the donor DNAs contained the target site recognized by the FS-specific TALEN (5'-TAGTGGAGGAATTAAAAGAGGAAGTTAAGAAGGAATTGATAACTATCAA-3' (SEQ ID NO:45)).

[00178] For the replacement of the four integrated FS cassettes in the strain (Strain F), the hyg+ marked TALEN plasmid was co-transformed into the host strain along with ~500 ng of each linear donor DNA using the protocol described in Example 3. Various dilutions were plated onto CSM-URA + Hyg plates and incubated at 30 degrees for 3 day.

7.4.2. Confirmation of Multiple Simultaneous Integration

[00179] After selection for the TALEN plasmid and integration of the URA3 marked codon-FS cassette on CSM-URA + Hyg plates, colony PCR was performed to determine the

frequency of replacement of the integrated FS cassettes with the unmarked ADS cassettes at three loci. DNA was prepped from 20 colonies and probed with primer pairs specific for integration of the ADS cassette at the NDT80, DIT1 and ERG10 loci, such that successful integration of the ADS cassette coding sequence at each locus was expected to produce an amplicon of a predicted size, while non-integration was expected to produce no amplicon.

[00180] Table 4. Primer sequences for cPCR verification of replacement of multiple farnesene synthase cassettes with amorphadiene synthase cassettes.

Primer Name	Description	Sequence	SEQ ID NO
HJ272	NDT80 5' FOR	ATAACAATATTATAAAAAGCGCT TAA	SEQ ID NO: 46
ART45	ADS ORF 5' REV	TACTGCTTCGGTAGTAGTTTCACC CTTCA	SEQ ID NO: 47
HJ643	DIT1 5' FOR	AAAATCCTTATATTATTGGCCC	SEQ ID NO: 48
HJ799	ERG10 5' FOR	GTAGCCTAAAACAAGCGCC	SEQ ID NO: 49

[00181] Three out of 48 clones examined had integrated a single ADS cassette in addition to the URA3-marked FS, one clone had integrated two ADS cassettes, and one clone had integrated all three ADS cassettes. Multiple integration results were further confirmed by sequencing longer PCR products encompassing both flanks.

[00182] These results demonstrate that expression of a site-specific designer nuclease in a host cell comprising a biosynthetic pathway can facilitate the simultaneous replacement of multiple integrated copies of a pathway gene with new pathway genes in a single transformation step.

7.5 Example 5: Simultaneous multiple integration of markerless DNA constructs into two loci cut with distinct designer nucleases

[00183] This Example provides results which demonstrate the simultaneous integration of two markerless DNA constructs at two native target sites, each site being cut with a distinct designer nuclease. In brief, an ADE$^-$ host strain was co-transformed with: (1) a linear DNA fragment comprising a GFP cassette (flanked by upstream and downstream regions homologous to the SFC1 locus); (2) a linear DNA fragment comprising an ADE2 cassette (flanked with upstream and downstream regions homologous to the YJR030c locus); and (3) plasmid(s) encoding designer nucleases that target sequences in the native SFC1 and YJR030c open reading frames, respectively. After selection for the plasmid(s), transformed colonies were screened visually for GFP fluorescence and for white color, indicating complementation of the ADE$^-$ phenotype. Colony PCR (cPCR) was also performed to confirm replacement of

both loci. Interestingly, a significant improvement in the rate of integration at both target loci was observed when the designer endonucleases were used in combination compared to the rate of integration when only a single designer nuclease was used.

7.5.1. Construction of donor DNA cassettes

[00184] Two donor DNAs were generated using PCR assembly of overlapping fragments: (1) a linear DNA fragment comprising a GFP cassette flanked by ~500 bp of upstream and downstream regions homologous to the SFC1 locus (SEQ ID NO: 58); and (2) a linear DNA fragment comprising an ADE2 cassette flanked by ~500 bp of upstream and downstream regions homologous to the YJR030c locus (SEQ ID NO: 59).

7.5.2. Construction of heterodimeric ZFN Expression Plasmids

[00185] A plasmid encoding the YJR030c-specific zinc finger nuclease (ZFN) was constructed in two ways. In the first version, the two ORFs of a heterodimeric ZFN under expression of a divergent Gal1-10 promoter and terminated by the Adh1 and CYC1 terminators were cloned into a Kan marked CEN-ARS vector by a three part gap repair in yeast (pCUT006). A second version was also constructed wherein both ORFs of the heterodimeric ZFN were expressed from the Gal10 promoter as a single ORF with the monomers separated by a DNA sequence encoding a cleavable peptide linker. This second plasmid was constructed by a three-part ligation using linkers produced by type IIS restriction enzyme digest of PCR fragments into a Kan marked CEN-ARS vector backbone (pCUT016). A plasmid encoding the SFC1-specific ZFN was also constructed as a single ORF using the same linker strategy, marker and backbone (pCUT015). The marker was then changed to URA by means of a gap repair reaction in yeast (pCUT058). To construct a single plasmid for expression of both the YJR030c and SFC1-specific nucleases, the single ORFs from pCUT16 and pCUT15 were subcloned into a new CEN-ARS Kan+ vector backbone, and expressed from the Gal1-10 divergent promoter with Cyc1and Adh1 terminators (pCUT032).

7.5.1. Transformation with Donor DNA and Induction of Double-Strand Breaks

[00186] One microgram of each designer nuclease plasmid DNA, or the plasmid containing both designer endonucleases on a single plasmid, was co-transformed with ~1 microgram of each of the donor DNAs. All transformations were recovered overnight in YP + 2% galactose to induce nuclease expression. Various dilutions were plated onto URA dropout + Kan agar plates (for the dual plasmids) or YPD + Kan to select for colonies transformed with plasmid DNA. Plates were incubated for 3-4 days at 30°C.

7.5.2. Confirmation of Multiple Simultaneous Integration

[00187] Marker-less integration at the SFC1 locus was scored by observation of GFP fluorescence under UV light using appropriate filters. Marker-less integration of ADE2 was scored by observation of a white colony color, indicating complementation of the ADE2 deletion phenotype (pink colonies) in the host strain. In a typical experiment, 50-150 colonies were assayed. The visual scoring strategy was confirmed in a subset of colonies by colony PCR using primers 5' of the integration construct and an internal reverse primer. Integration at each locus was expected to produce an amplicon of a predicted size, while non-integration was expected to produce no amplicon. The cPCR results confirmed the accuracy of the visual scoring method.

[00188] Table 4. Primer sequences for successful cPCR verification of multiple integration of the ADS cassette coding sequence

Primer Name	Description	Sequence	SEQ ID NO
CUT351	SFC1 5' cPCR	GCGAATGAGCCATGAATTATTAACCGC	SEQ ID NO: 63
CUT350	YJR030c 5' cPCR	AGATGAAACGAATTACTAGCATTTTATCCGTTC	SEQ ID NO: 64
CUT371	ADE2 cassette REV	TAACTACCATTACTCAGTGTACTTGATTGTTTTGTCCGATTTTCTTG	SEQ ID NO: 65
HJ788	GFP cassette REV	GCCGGGTGACAGAGAAATATTG	SEQ ID NO: 66

[00189] As indicated in FIG. 8, in cells co-transformed with linear donor DNAs for the SFC1 and YJR030c loci, and the YJR030c endonuclease plasmid (pCUT006) and SFC1 endonuclease plasmid (pCUT058), 80% of colonies selected on URA dropout + Kan agar plates were GFP positive. Of these colonies, 91% were positive for ADE2 integration. In total, 72.8% of colonies had integrated the donor DNA at both loci.

[00190] In cells co-transformed with linear donor DNA for the SFC1 locus and the designer nuclease plasmid targeting SFC1 (pCUT015), 50% of the cells were positive for GFP. When cells were co-transformed with linear donor DNA for the YJR030c locus and the designer nuclease plasmid targeting the YJR030c locus (pCUT016), only 5% of the cells were positive for ADE2 integration. When the host cells were co-transformed with linear DNAs for the SFC1 and YJR030c loci, and the SFC1/YJR030c designer nuclease plasmid (pCUT032), 76% of the cells were GFP positive, and 63% were ADE2 positive. This result is notable in that it

demonstrates an unexpectedly significant improvement in integration efficiency when multiple sites are targeted by designer endonucleases.

[00191] These results demonstrate that induction of multiple targeted double-strand breaks at native loci in the genome of a host cell can facilitate simultaneous, multiple, marker-less integrations of functional gene cassettes.

[00192] All publications, patents and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference. Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be readily apparent to those of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

WHAT IS CLAIMED:

1. A method for simultaneously integrating a plurality of (n) exogenous nucleic acids into a plurality of (n) target sites of a host cell genome, wherein n is at least two, the method comprising:

(a) simultaneously contacting a host cell with:

(i) said plurality of exogenous nucleic acids, wherein:

x is an integer that varies from 1 to n, and for each integer x, each exogenous nucleic acid $(ES)_x$ comprises a first homology region $(HR1)_x$ and a second homology region $(HR2)_x$, wherein $(HR1)_x$ and $(HR2)_x$ are capable of initiating host cell mediated homologous recombination of $(ES)_x$ at a target site $(TS)_x$ selected from said plurality of (n) target sites of said host cell genome; and

(ii) for each said target site $(TS)_x$, a nuclease $(N)_x$ capable of cleaving at $(TS)_x$, whereupon said cleaving results in homologous recombination of $(ES)_x$ at $(TS)_x$;

and

(b) recovering a host cell wherein each exogenous nucleic acid $(ES)_x$ has integrated at its selected target site $(TS)_x$.

2. A host cell comprising:

(a) a plurality of (n) genomic target sites, wherein n is at least two;

(b) a plurality of (n) exogenous nucleic acids, wherein x is an integer that varies from 1 to n, and for each integer x, each exogenous nucleic acid $(ES)_x$ comprises a first homology region $(HR1)_x$ and a second homology region $(HR2)_x$, wherein $(HR1)_x$ and $(HR2)_x$ are capable of initiating host cell mediated homologous recombination of $(ES)_x$ at a target site $(TS)_x$ selected from said plurality of (n) genomic target sites; and

(c) for each said target site $(TS)_x$, a nuclease $(N)_x$ capable of cleaving at $(TS)_x$, whereupon said cleaving results in homologous recombination of $(ES)_x$ at $(TS)_x$.

3. The method of claim 1, or the host cell of claim 2, wherein $(HR1)_x$ is homologous to a 5' region of $(TS)_x$, and $(HR2)_x$, is homologous to a 3' region of $(TS)_x$.

4. The method of claim 1, or the host cell of claim 2, wherein $(N)_x$ is capable of cleaving at a region positioned between said 5' and 3' regions of $(TS)_x$.

5. The method of claim 1, or the host cell of claim 2, wherein a single nuclease is capable of cleaving each $(TS)_x$.

6. The method or host cell of any one of claims 1 to 5, wherein n= 3, 4, 5, 6, 7, 8, 9 or 10.

7. The method of any one of claims 1 and 3 to 6, wherein said recovering does not require integration of a selectable marker.

8. The method of any one of claims 1 and 3 to 6, wherein said recovering occurs at a higher frequency as compared to not contacting the host cell with a nuclease capable of cleaving at said target site.

9. The method of any one of claims 1 and 3 to 6, wherein said recovering occurs at a frequency of about one every 10, 9, 8, 7, 6, 5, 4, 3, or 2 contacted host cells, or clonal populations thereof, screened.

10. The method or host cell of any one of claims 1, 2, and 3 to 9, wherein $(N)_x$ is capable of cleaving an endogenous genomic sequence within $(TS)_x$.

11. The method or host cell of any one of claims 1, 2, and 3 to 9, wherein $(N)_x$ is capable of cleaving an exogenous sequence within $(TS)_x$.

12. The method or host cell of claim 11, wherein the exogenous sequence is a recognition sequence for a homing endonuclease.

13. The method or host cell of claim 12, wherein the homing endonuclease is F-cphI.

14. The method or host cell of any one of claims 1 to 13, wherein $(ES)_x$ further comprises a nucleic acid of interest $(D)_x$ positioned 3' of $(HR1)_x$ and 5' of $(HR2)_x$.

15. The method or host cell of claim 14, wherein $(D)_x$ is selected from the group consisting of a selectable marker, a promoter, a nucleic acid sequence encoding an epitope tag, a gene of interest, a reporter gene, and a nucleic acid sequence encoding a termination codon.

16. The method or host cell of any one of claims 1 to 15, wherein $(ES)_x$ is linear.

17. The method or host cell of any one of claims 1, 2, and 3 to 16, wherein $(N)_x$ is provided as an expression vector comprising a nucleic acid sequence encoding $(N)_x$.

18. The method or host cell of any one of claims 1, 2, and 3 to 17, wherein $(N)_x$ is selected from the group consisting of an endonuclease, a zinc finger nuclease, a TAL-effector DNA binding domain-nuclease fusion protein (TALEN), a transposase, and a site-specific recombinase.

19. The method or host cell of claim 18, wherein the endonuclease is modified to specifically bind an endogenous genomic sequence, wherein the modified endonuclease no longer binds to its wild type endonuclease recognition sequence.

20. The method or host cell of claim 18 or 19, wherein the endonuclease is a homing endonuclease selected from the group consisting of: an LAGLIDADG homing endonuclease, an HNH homing endonuclease, a His-Cys box homing endonuclease, a GIY-YIG homing endonuclease, and a cyanobacterial homing endonuclease.

21. The method of or host cell of claim 18 or 19, wherein the endonuclease is derived from an endonuclease selected from the group consisting of: H-DreI, I-SceI, I-SceII, I-SceIII, I-SceIV, I-SceV, I-SceVI, I-SceVII, I-CeuI, I-CeuAIIP, I-CreI, I-CrepsbIP, I-CrepsbIIP, I-CrepsbIIIP, I-CrepsbIVP, I-TliI, I-PpoI, Pi-PspI, F-SceI, F-SceII, F-SuvI, F-CphI, F-TevI, F-TevII, I-AmaI, I-AniI, I-ChuI, I-CmoeI, I-CpaI, I-CpaII, I-CsmI, I-CvuI, I-CvuAIP, I-DdiI, I-DdiII, I-DirI, I-DmoI, I-HmuI, I-HmuII, I-HsNIP, I-LlaI, I-MsoI, I-NaaI, I-NanI, I-NclIP, I-NgrIP, I-NitI, I-NjaI, I-Nsp236IP, I-PakI, I-PboIP, I-PcuIP, I-PcuAI, I-

PcuVI, I-PgrIP, I-PobIP, I-PorI, I-PorIIP, I-PbpIP, I-SpBetaIP, I-ScaI, I-SexIP, I-SneIP, I-SpomI, I-SpomCP, I-SpomIP, I-SpomIIP, I-SquIP, I-Ssp68031, I-SthPhiJP, I-SthPhiST3P, I-SthPhiSTe3bP, I-TdeIP, I-TevI, I-TevII, I-TevIII, i-UarAP, i-UarHGPAIP, I-UarHGPA13P, I-VinIP, I-ZbiIP, PI-MgaI, PI-MtuI, PI-MtuHIP PI-MtuHIIP, PI-PfuI, PI-PfuII, PI-PkoI, PI-PkoII, PI-Rma43812IP, PI-SpBetaIP, PI-SceI, PI-TfuI, PI-TfuII, PI-ThyI, PI-TliI, or PI-TliII.

22. The method or host cell of any one of claims 1 to 21, wherein the host cell is selected from the group consisting of a fungal cell, a bacterial cell, a plant cell, and an animal cell.

23. The method or host cell of any one of claims 1 to 22, wherein the host cell is a yeast cell.

24. The method or host cell of claim 23, wherein the yeast cell is a *Saccharomyces cerevisiae* cell.

25. A method for simultaneously integrating a plurality of (n) exogenous nucleic acids into a plurality of (n) target sites of a host cell genome, wherein n is at least two, the method comprising:

(a) simultaneously contacting a host cell with:

(i) a plurality of libraries, wherein x is an integer that varies from 1 to n, and for each integer x, each library $(L)_x$ comprises a plurality of exogenous nucleic acids, wherein each exogenous nucleic acid comprises, in a 5' to 3' orientation, a first homology region $(HR1)_x$, any nucleic acid of interest selected from the group $(D)_x$, and a second homology region $(HR2)_x$, wherein $(HR1)_x$ and $(HR2)_x$ are capable of initiating host cell mediated homologous recombination of said selected exogenous nucleic acid at a target site $(TS)_x$ of said host cell genome; and

(ii) for each said target site $(TS)_x$, a nuclease $(N)_x$ capable of cleaving at $(TS)_x$, whereupon said cleaving results in homologous recombination of an exogenous nucleic acid from $(L)_x$ at $(TS)_x$;

and

(b) recovering a host cell wherein an exogenous nucleic acid from each library $(L)_x$ has integrated at each selected target sequence $(TS)_x$.

26. A host cell comprising:

(a) a plurality of (n) genomic target sites, wherein n is at least two;

(b) a plurality of libraries, wherein x is an integer that varies from 1 to n, and for each integer x, each library $(L)_x$ comprises a plurality of exogenous nucleic acids, wherein each exogenous nucleic acid comprises, in a 5' to 3' orientation, a first homology region $(HR1)_x$, any nucleic acid of interest selected from the group $(D)_x$, and a second homology region $(HR2)_x$, wherein $(HR1)_x$ and $(HR2)_x$ are capable of initiating host cell mediated homologous recombination of said selected exogenous nucleic acid at a target site $(TS)_x$ of said host cell genome; and

for each said target site $(TS)_x$, a nuclease $(N)_x$ capable of cleaving at $(TS)_x$, whereupon said cleaving results in homologous recombination of an exogenous nucleic acid from $(L)_x$ at $(TS)_x$.