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 ~~embodiments~~examples, a plurality of exogenous nucleic acids are simultaneously integrated with a single transformation reaction. In some ~~embodiments~~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 ~~aspect~~example, ~~provided~~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 ~~embodiments~~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 ~~embodiments~~examples, $(N)_x$ is capable of cleaving at a region positioned between said 5' and 3' regions of $(TS)_x$.

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

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

[0014] In some ~~embodiments~~examples, said recovering does not require integration of a selectable marker. In some ~~embodiments~~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 ~~embodiments~~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 ~~embodiments~~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 ~~embodiments~~examples, $(N)_x$ is capable of cleaving an endogenous host genomic sequence, *e.g.*, a native loci within $(TS)_x$. In some ~~embodiments~~examples, $(N)_x$ is capable of cleaving an exogenous sequence, *e.g.*, an introduced loci within $(TS)_x$.

[0016] In some ~~embodiments~~examples, $(ES)_x$ further comprises a nucleic acid of interest $(D)_x$ positioned 3' of $(HR1)_x$ and 5' of $(HR2)_x$. In some ~~embodiments~~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 ~~embodiments~~examples, $(ES)_x$ is linear. In some ~~embodiments~~examples, $(N)_x$ is provided as an expression vector comprising the nucleic acid sequence encoding $(N)_x$. In some ~~embodiments~~examples, $(N)_x$ is transformed into the host cell as a purified protein. In some ~~embodiments~~examples, $(N)_x$ is transformed into the host cell as purified RNA.

[0018] In some ~~embodiments~~examples, the host cell comprises one or more heterologous nucleotide sequences encoding one or more enzymes of a biosynthetic pathway. In some ~~embodiments~~examples, the one or more heterologous nucleotide sequences encoding one or more enzymes of a biosynthetic pathway are genomically integrated. In some ~~embodiments~~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 ~~embodiments~~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 ~~embodiments~~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 ~~embodiments~~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 ~~embodiments~~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 ~~embodiments~~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 ~~embodiments~~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 ~~embodiments~~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 ~~embodiments~~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 ~~embodiments~~examples, the TypeIIS restriction

endonuclease is selected from the group consisting of HO endonuclease and Fok I endonuclease. In some ~~embodiments~~examples, the zinc finger binding domain comprises 3, 5 or 6 zinc fingers. In some ~~embodiments~~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 ~~embodiments~~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 ~~embodiments~~examples, the endonuclease is Fcph-I.

[0021] In some ~~embodiments~~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 ~~embodiments~~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 ~~embodiments~~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 ~~embodiments~~examples, the host cell is a fungal cell, a bacterial cell, a plant cell, an animal cell, or a human cell. In particular ~~embodiments~~examples, the host cell is a yeast cell. In some ~~embodiments~~examples, the yeast cell is a haploid yeast cell. In some ~~embodiments~~examples, the yeast cell is a *Saccharomyces cerevisiae* cell. In some ~~embodiments~~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 ~~aspect~~example, ~~provided~~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 ~~aspect~~example, ~~provided~~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 ~~embodiments~~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 ~~embodiments~~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 ~~aspect~~example, ~~provided~~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 ~~aspect~~example, ~~provided~~disclosed herein is a kit useful for performing the methods for genomically integrating one or more exogenous nucleic acids described herein. In some ~~embodiments~~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 ~~embodiments~~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

~~embodiments~~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 ~~embodiments~~examples, $(ES)_x$ is linear. In some ~~embodiments~~examples, $(ES)_x$ is circular.

[0029] In a particular ~~embodiment~~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 ~~embodiments~~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,