long. Then the cells are forced to initiate sporulation. During the sporulation the recombination process takes place. The resulting spores and recombinant sequences can be differentiated by selection for the appropriate flanking markers.

The ability of yeast to efficiently recombine homologous DNA sequences can also be exploited to increase the diversity of a library. When two genes that share 89,9% homology were mutated by PCR and transformed into wild type yeast, a chimeric library of 10e7 was created through in vivo homologous recombination, showing several cross-over points throughout the two genes (Swers et al Nucleic Acids Research 32(3) e36 (2004)).

A method of mitotic homeologous recombination is described by Nicholson et al (Genetics 154: 133-146 (2000)). Effects of defined mismatches contained in short inverted repeats on recombination rates in wild-type or MMR-defective strains were investigated.

It is the object of the present invention to provide an improved method of preparing and assembling a diversity of gene mosaics, especially for recombining long DNA fragments. As a result it would be desirable to provide respective libraries of variants for the selection of improved recombinants.

The object is achieved by the provision of the embodiments of the present application.

Summary ~~of the invention~~

The present disclosure~~invention~~ provides a novel method for generating a gene mosaic by somatic in vivo recombination, comprising

a) in a single step procedure
   (i) transforming a cell with at least one gene A having a sequence homology of less than 99.5% to another gene to be recombined that is an integral part of the cell genome or presented in the framework of a genetic construct,
   (ii) recombining said genes,
   (iii) generating a gene mosaic of the genes at an integration site of a target genome, wherein said at least one gene A has a single flanking target sequence either at the 5' end or 3' end anchoring to the 5´or 3´end of said integration site, and
b) selecting clones comprising the gene mosaic.

Specifically the disclosure~~invention~~ relates to a method for generating a gene mosaic by somatic in vivo recombination, comprising

a) in a single step procedure

(i) transforming a cell with at least one gene A having a sequence homology of less than 99.5% to a different gene B which is an integral part of the cell genome or presented in the framework of a genetic construct or expression cassette,

(ii) recombining said genes,

(iii) generating a gene mosaic of genes A and B at an integration site of a target genome, wherein said at least one gene A is linked to a single flanking target sequence either at the 5' end or 3' end of the genetic construct anchoring to the 5´or 3´end of said integration site and

b) selecting clones comprising the gene mosaic.

In one aspect, there is provided a method for generating a gene mosaic by somatic in vivo recombination of homeologous DNA sequences, comprising

a) in a single step procedure

(i) transforming a DNA repair deficient cell with at least one gene A having a sequence homology of at least 30% and less than 99.5% to another gene to be recombined that is an integral part of the cell genome or presented in the framework of a genetic construct employing at least one gene B to be recombined,

(ii) recombining said genes,

(iii) generating a gene mosaic of the genes at an integration site of a target genome, wherein said at least one gene A has a single flanking target sequence either at the 5' end or 3' end anchoring to the 5' or 3' end of said integration site, wherein the single flanking target sequence of gene A is anchoring to the 5' end of the integration site while gene B has a single flanking target sequence anchoring to the 3' end, and

b) selecting clones comprising at least one intragenic gene mosaic.

It is specifically preferred that a selection marker is used in the gene mosaic and the clones are selected according to the presence of the selection marker. For example, the gene mosaic comprises a selection marker, e.g. where said gene A is linked to a selection marker. Alternatively, selection may also be made by the presence of any product resulting of recombinants, e.g. through determining the yield or functional characteristics. Specifically one or more different selection markers may be used to differentiate the type of gene mosaics.

Specifically the method according to the disclosure~~invention~~ employs said another gene that is part of the target genome, e.g. the genome of the cell. In a preferred embodiment said anther gene is gene B being part of the genome of the cell.

According to an alternatively preferred embodiment, said another gene is a genetic construct separate from the target genome, such as a linear polynucleotide, and optionally integrated into the target genome in the course of the recombination.

According to a specific embodiment of the disclosure~~invention~~ the cell is co-transformed with at least one gene A and at least one gene B, wherein said single flanking target sequence of gene A is anchoring to the 5´end of an integration site on said target genome, and wherein gene B is linked to a single flanking target sequence anchoring to the 3' end of the integration site.

Specifically, the cell can be co-transformed with at least one gene A with a selection marker and at least one gene B, wherein said single flanking target sequence

Using the inventive method, multiple genes coding for polypeptides of a biochemical pathway can be assembled and recombined.

In another specific embodiment, the inventive method provides for recombination and eventual assembly of genes resulting in a non-coding sequence, such as a promoter, untranslated region, ribosomal binding site, terminator, etc.

Any recombination competent eukaryotic or prokaryotic host cell can be used for generating a gene mosaic by somatic in vivo recombination according to the present disclosure~~invention~~. According to a preferred embodiment of the disclosure~~invention~~, the cell is a repair deficient cell, e.g. a nucleic acid repair deficient cell, such as with DNA repair deficiency, or an MMR deficient cell.

Sepcifically, the cell is a eukaryotic cell, preferably a fungal, mammalian or plant cell, or prokaryotic cell.

Preferably the cell is a an *Aspergillus sp* or a fungal cell, preferably, it can be selected from the group consisting of the genera Saccharomyces, Candida, Kluyveromyces, Hansenula, Schizosaccaromyces, Yarrowia, Pichia and Aspergillus.

Preferably haploid strains, such as haploid yeast strains are employed.

Alternatively, prokaryotes, such as E.coli, Bacillus, Streptomyces, or mammalian cells, like HeLa cells or Jurkat cells, or plant cells, like Arabidopsis, may be used.

According to a specific embodiment, the flanking target sequence is at least 5 bp, preferably at least 10 bp, more preferably at least 20 bp, 50 bp, 100 bp up to 5,000 bp length. Specifically the flanking target sequence is linked to said gene or is an integral, terminal part of said gene. It is preferred that said the flanking target sequence has homology in the range of 30% to 99.5%, preferably less than 95%, less than 90%, less than 80%, hybridising with the anchoring sequence of said integration site,

When at least two different flanking target sequences anchoring to the target integration site of the genome are used according to the disclosure~~invention~~, it is preferred that they do not recombine with each other, preferably they share less than 30% homology.

Selection markers useful for the inventive method can be selected from the group consisting of any of the known nutrition auxotrophic markers, antibiotics resistance markers, fluorescent markers, knock-in markers, activator/binding domain markers and dominant recessive markers and colorimetric markers. Preferred markers can be temporally inactivated or functionally knocked out, and may be re-established to regain its marking property. Further preferred markers are traceable genes, wherein the marker is a function of either of the gene sequences A and/or the other gene(s),

such as gene B, without separate sequences with a marker function, so that the expression of the gene mosaic can be directly determined through detection of the mosaic itself. In this case the gene mosaic is directly traceable.

According to a specific embodiment, said genes are comprised in a linear polynucleotide, a vector or a yeast artificial chromosome. Specifically, gene A and/or other genes to be recombined are in the form of linear polynucleotides, preferably of 300 to 20.000 bp. Specifically, there would be no need to construct or employ plasmids or megaplasmids. The gene(s) can thus be used as such, i.e. without carrier.

The genes used for recombination and integration can also be comprised in any genetic construct, e.g. to be used as vector for carrying said gene(s). Said genes can thus be comprised in a genetic construct, e.g. a linear polynucleotide, a vector or a yeast artificial chromosome. These preferably include linear polynucleotides, plasmids, PCR constructs, artificial chromosomes, like yeast artificial chromosomes, viral vectors or transposable elements.

According to a specific embodiment of the disclosure~~invention~~ the integration site of the target genome is located on either of the genes, e.g. within a linear polynucleotide, a plasmid or chromosome, including artificial chromosomes.

The method according to the disclosure~~invention~~ specifically provides for the selection of at least one clone having an intragenic gene mosaic. Specifically, at least one clone having a gene assembly and at least one intragenic gene mosaic is selected.

Using the method according to the disclosure~~invention~~ gene mosaics of at least 3, preferably at least 9, up to 20.000 base pairs can be obtained, as well as gene mosaics, e.g. comprising at least one intragenic mosaic, preferably with at least 3 cross-over events, preferably at least 4, 5, or 10 cross-over events per 700 base pairs, more preferably per 600 bp, per 500 bp or even below. Typically a high degree of cross-over events provides for a large diversity of recombined genes, which may be used to produce a library for selecting suitable library members. The degree of mosaics or cross-over events can be understood as a quality parameter of such a library.

The genes which are modified according to the method of the disclosure~~invention~~ can be any genes useful for scientific or industrial purposes. These genes can be for example non-coding sequences, e.g. those which may be used for recombinant expression systems, or variants of polypeptides, in whole or in part, including those partial sequences, which do not encode a polypeptide with biological activity, which

polypeptides are specifically selected from the group consisting of enzymes, antibodies or parts thereof, cytokines, vaccine antigens, growth factors or peptides. If genes are modified, which encode a non coding sequence or an amino acid sequence as part of a polypeptide having a biological activity, also called "partial genes", it may be preferred that an assembly of such partial genes has functional features, e.g. encodes a polypeptide having a biological activity. Preferably a number of different genes, e.g. different partial genes, at a size ranging from 3 bp to 20.000 bp, specifically at least 100 bp, preferably from 300 bp to 20.000 bp, specifically up to 10.000 bp, are recombined, which number of different genes of is at least 2, more specifically at least 3, 4, 5, 6, 7, 8, 9, or at least 10 to produce a recombined gene sequence that is non-coding or encoding a recombinant polypeptide, e.g. having a biological activity, which is advantageously modulated, e.g. having an increased biological activity. The term "biological activity" as used in this regard specifically refers to an enzymatic activity, such as an activity that converts a particular substrate into a particular product. Preferred genes as diversified according to the ~~invention~~disclosure are coding for multi-chain polypeptides.

According to a particular embodiment of the ~~invention~~disclosure there is provided a method of cell display of gene variants, comprising creating a variety of gene mosaics in cells using the method according to the ~~invention~~disclosure, and displaying said variety on the surface of said cells to obtain a library of mosaics.

The library obtainable by such preferred display specifically comprises a high percentage of gene mosaics within a functional open reading frame (ORF), preferably at least 80%.

A library according to the ~~invention~~disclosure specifically may be in any suitable form, specifically a biological library comprising a variety of organisms containing the gene variants. The biological library according to the ~~invention~~disclosure may be contained in and/or specifically expressed by a population of organisms to create a repertoire of organisms, wherein individual organisms include at least one library member.

According to a specific embodiment~~aspect~~ of the ~~invention~~disclosure there is further provided an organism that comprises a gene variant from such a library, e.g. an organism selected from a repertoire of organisms. The organism as provided according to the ~~invention~~disclosure may be used to express a gene expression product in a suitable expression system, e.g. as a production host cell.

Figures

Claims

1. A method for generating a gene mosaic by somatic in vivo recombination, comprising

c) in a single step procedure

(iv) transforming a cell with at least one gene A having a sequence homology of less than 99.5% to another gene to be recombined that is an integral part of the cell genome or presented in the framework of a genetic construct,

(v) recombining said genes,

(vi) generating a gene mosaic of the genes at an integration site of a target genome, wherein said at least one gene A has a single flanking target sequence either at the 5' end or 3' end anchoring to the 5´or 3´end of said integration site, and

d) selecting clones comprising the gene mosaic.

2. A method according to claim 1, wherein a selection marker is used in the gene mosaic and the clones are selected according to the presence of the selection marker.

3. A method according to claim 1 or 2, wherein said another gene is part of the genome of the cell.

4. A method according to any of claims 1 to 3, wherein the cell is co-transformed with at least one gene A and at least one gene B, wherein said single flanking target sequence of gene A is anchoring to the 5´end of an integration site on said target genome, and wherein gene B is linked to a single flanking target sequence anchoring to the 3' end of the integration site.

5. A method according to any one of claims 1 to 4, wherein the cell is co-transformed with at least two different genes A1 and A2 and optionally with at least two different genes B1 and B2.

6. A method according to any one of claims 1 to 5, wherein at least one further gene C is co-transformed, which has a sequence hybridizing with a sequence of gene A and/or said another gene to obtain assembly of said further gene C to gene A and/or said another gene.

7. A method according to any of claims 1 to 6, wherein said gene A and/or said another gene is coding for a polypeptide or part of a polypeptide having an activity.

8. A method according to claim 6 or 7, wherein multiple genes coding for polypeptides of a biochemical pathway are recombined and assembled.

9. A method according to any one of claims 1 to 8, wherein the cell is a repair deficient cell.

10. A method according to any one of claims 1 to 9, wherein the cell is a eukaryotic cell, preferably a fungal, mammalian or plant cell, or prokaryotic cell.

11. A method according to claim 10, wherein the cell is a fungal cell, preferably selected from the group consisting of Saccharomyces, Candida, Kluyveromyces, Hansenula, Schizosaccaromyces, Yarrowia, Pichia and Aspergillus.

12. A method according to any one of claims 1 to 11, wherein the flanking target sequence is at least 5 bp.

13. A method according to any one of claims 1 to 12, wherein the flanking target sequence has homology in the range of 30% to 99.5% with the anchoring sequence of said integration site

14. A method according to any one of claims 1 to 13, wherein a selection marker is used, which is selected from the group consisting of nutrition auxotrophic markers, antibiotics resistance markers, fluorescent markers, knock-in markers, activator/binding domain markers and dominant recessive markers and colorimetric markers.

15. A method according to any one of claims 1 to 14, wherein said genes are comprised in a linear polynucleotide, a vector or a yeast artificial chromosome.

16. A method according to any one of claims 1 to 15, wherein said genes are linear polynucleotides, preferably of 300 to 20.000 bp.

17. A method according to any one of claims 1 to 16, wherein at least one clone having an intragenic gene mosaic is selected.

18. A method according to any one of claims 1 to 17, wherein at least one clone having a gene assembly and at least one intragenic gene mosaic is selected.

19. A method according to any one of claims 1 to 18, wherein gene mosaics of at least 3 up to 20.000 base pairs, preferably with at least 3 cross-over events per 700 bp are obtained.

20. A method according to any one of claims 1 to 19, wherein the genes are non-coding sequences or encoding variants of a polypeptide selected from the group consisting of enzymes, antibodies or parts thereof, cytokines, growth factors, vaccine antigens and peptides.

21. A method of cell display of gene variants, comprising creating a variety of gene mosaics in cells using the method according to any one of claims 1 to 20, and displaying said variety on the surface of said cells to obtain a library of mosaics.

22. A library of gene mosaics obtainable by a method of any one of claims 1 to 20, wherein at least 80% of the gene mosaics are contained within a functional ORF.

23. A library according to claim 22, which comprises a variety of organisms containing the gene variants.

24. An organism that comprises a gene variant from a library according to any of claims 22 or 23.

CLAIMS

1. A method for generating a gene mosaic by somatic in vivo recombination of homeologous DNA sequences, comprising

a) in a single step procedure

(i) transforming a DNA repair deficient cell with at least one gene A having a sequence homology of at least 30% and less than 99.5% to another gene to be recombined that is an integral part of the cell genome or presented in the framework of a genetic construct employing at least one gene B to be recombined,
(ii) recombining said genes,
(iii) generating a gene mosaic of the genes at an integration site of a target genome,
wherein said at least one gene A has a single flanking target sequence either at the 5' end or 3' end anchoring to the 5' or 3' end of said integration site, wherein the single flanking target sequence of gene A is anchoring to the 5' end of the integration site while gene B has a single flanking target sequence anchoring to the 3' end, and

b) selecting clones comprising at least one intragenic gene mosaic.

2. A method according to claim 1, wherein a selection marker is used in the gene mosaic and the clones are selected according to the presence of the selection marker.

3. A method according to claim 1 or 2, wherein gene mosaics of a non-coding sequence, a partial gene or genes of at least 3 up to 20.000 base pairs are obtained.

4. A method according to any one of claims 1 to 3, wherein the cell is co-transformed with at least two different variants of gene A to be recombined, which are genes A1 and A2, and optionally with at least two different variants of gene B to be recombined, which are genes B1 and B2.

5. A method according to any one of claims 1 to 4, wherein at least one further gene C is co-transformed, which has a sequence hybridizing with a sequence of gene A and/or said another gene to obtain assembly of said further gene C to gene A and/or said another gene, wherein one or more of the assembled genes has an intragenic gene mosaic.

6. A method according to any of claims 1 to 5, wherein said gene A and/or said another gene are non-coding sequences or coding for a polypeptide or part of a polypeptide having an activity.

7. A method according to claim 5 or 6, wherein multiple genes coding for polypeptides of a biochemical pathway are recombined and assembled.

8. A method according to any one of claims 1 to 7, wherein the cell is a cell with a knock-out of a DNA repair gene or a mismatch repair deficient cell.

9. A method according to any one of claims 1 to 8, wherein the cell is a eukaryotic cell, preferably a fungal, mammalian or plant cell, or prokaryotic cell.

10. A method according to claim 9, wherein the cell is a fungal cell, preferably selected from the group consisting of Saccharomyces, Candida, Kluyveromyces, Hansenula, Schizosaccaromyces, Yarrowia, Pichia and Aspergillus.

11. A method according to any one of claims 1 to 10, wherein a selection marker is used, which is selected from the group consisting of nutrition auxotrophic markers, antibiotics resistance markers, fluorescent markers, knock-in markers, activator/binding domain markers and dominant recessive markers and colorimetric markers.

12. A method according to any one of claims 1 to 11, wherein said genes are comprised in a linear polynucleotide, a vector or a yeast artificial chromosome.

13. A method according to any one of claims 1 to 12, wherein said genes are linear polynucleotides, preferably of 300 to 20.000 bp.

14. A method according to any one of claims 1 to 13, wherein the genes are non-coding sequences or encoding variants of a polypeptide selected from the group consisting of enzymes, antibodies or parts thereof, cytokines, growth factors, vaccine antigens and peptides.