origin of the sequence which is incorporated into the concatemer at that position. The present invention provides an improved method of generating concatemers for sequencing which is particularly useful in the context of multiplex detection assays such as PEA and PLA, whereby sequencing throughput is increased by concatenating reporter DNA molecules from multiple pools in a predefined order, such that the location of each reporter DNA sequence within the resultant concatemers is indicative of the pool from which it originates. Each pool may be generated, for instance, from a separate sample, or using a separate panels of proximity probes. The method is particularly advantageous when each pool of reporter DNA molecules is generated using probes carrying the same set of nucleic acid moieties. The ability to assign each reporter DNA sequence in a concatemer to a particular pool of origin means that identical reporter sequences present within multiple pools can be distinguished based on their locations within the concatemers.

The method provided herein thus has particular utility in the context of proximity assays (e.g. PEA and PLA assays), but its utility is not limited to these assays. The method of the invention can be used in any context where it is desired to analyse a pool of DNA molecules.

Summary

In a first aspect, disclosed and provided herein is a method of detecting DNA sequences from multiple pools, wherein each pool comprises multiple species of DNA molecule, the method comprising:

(i) combining the pools;

(ii) generating multiple linear DNA concatemers of a pre-defined length, wherein each concatemer is generated by joining together one random DNA molecule from each pool in a pre-determined order such that the position of each DNA molecule within the concatemer indicates the pool from which it is derived and each concatemer comprises a pre-determined number of DNA molecules; and

(iii) sequencing the concatemers, thereby detecting a DNA sequence from each pool in each concatemer, wherein the DNA sequence from each pool is assigned to that pool based upon its position within its concatemer.

In particular, the pools may comprise DNA molecules which are capable of being concatenated in a pre-defined and directed order. In other words, the DNA molecules in each pool are capable of being concatenated, or linked, only to molecules from a pre-designated, or selected, other pool. Accordingly, each pool is designated, or allocated, a predesignated place or position in the concatemer. The concatemer thus has a pre-determined “pool order” of monomer positions, and the identity of the pool from which each monomer in the concatemer derives may be determined from the position of the monomer in

the concatemer. In other words, the position of each DNA molecule within the concatemer correlates to the pool from which it is derived. To allow a concatemer of a predefined order of pools to be constructed, each DNA molecule (i.e. monomer) may be linked to only one (if it is a terminal monomer) or two other DNA molecules (that is to say, each DNA molecule (monomer) may be linked to DNA molecules from only one (if it is a terminal monomer) or two other pools.

Thus, the DNA molecules in a pool may be prepared for concatenation. In an example, the method comprises, prior to step (i), a step of preparing multiple pools of DNA molecules for concatenation, wherein said preparing comprises providing the DNA molecules within each pool with defined end sequences which may be joined in a concatenation step, the DNA molecules in the same pool having the same end sequences and the different pools having different end sequences, such that a DNA molecule from one pool may only be joined to a DNA molecule from one or two pre-determined different pools. A DNA molecule may have one or two end sequences, depending on its position in the conacatemer. Further, a DNA molecule in a terminal position in the concatemer may be provided with a second end sequence for linkage to another molecule (i.e. a molecule which is other than a DNA molecule from a pool), e.g. a sequencing or other adaptor.

In a second example, disclosed and provided herein is a kit comprising:

(i) multiple proximity probe pairs, wherein each proximity probe comprises a binding domain specific for an analyte and a nucleic acid domain, and each proximity probe pair is specific for a different analyte, such that on proximal binding of the pair of proximity probes to their respective analyte the nucleic acid domains of the proximity probe pair are capable of interacting to generate a reporter DNA molecule, and wherein in each pair the nucleic acid domain of one proximity probe comprises a first universal primer binding site and a barcode sequence 3’ thereof, and the nucleic acid domain of the other proximity probe comprises a second universal primer binding site and a barcode sequence 3’ thereof;

(ii) a first primer pair, wherein the primers are designed to bind the first and second universal primer binding sites;

(iii) a set of assembly primer pairs suitable for preparing DNA molecules for directed assembly by USER assembly or Gibson assembly into a linear concatemer, wherein each primer comprises, from 5’ to 3', an assembly site and a hybridisation site, and in each primer pair the hybridisation sites are designed to bind the first and second universal primer binding sites;

(iv) enzymes suitable for assembling DNA fragments by USER assembly or Gibson assembly, wherein the enzymes are suitable for use in the same means of DNA assembly as the assembly primer pairs; and

(v) a second primer pair, wherein each primer comprises a sequencing adaptor, a sequencing primer binding site, an index sequence and a hybridisation site, wherein the hybridisation sites are designed to bind the assembly sites of the assembly primers designed to form the ends of the linear concatemer;

and wherein the first primer in the pair comprises a first sequencing adaptor, a first sequencing primer site and a first index sequence, and the second primer in the pair comprises a second sequencing adaptor, a second sequencing primer site and a second index sequence.

In one aspect, the present disclosure refers to a kit comprising: (i) multiple proximity probe pairs, wherein each proximity probe comprises: (I) an analyte-binding domain specific for an analyte; and (II) a nucleic acid domain, wherein both probes within each pair comprise analyte-binding domains specific for the same analyte, and each probe pair is specific for a different analyte, and wherein each probe pair is designed such that on proximal binding of the pair of proximity probes to their respective analyte the nucleic acid domains of the proximity probes interact to generate a reporter DNA molecule, and wherein in each pair one proximity probe comprises a nucleic acid domain comprising a first universal primer binding site and a barcode sequence 3' thereof, and the other proximity probe comprises a nucleic acid domain comprising a second universal primer binding site and a barcode sequence 3' thereof; (ii) a first primer pair, wherein the primers are designed to bind the first and second universal primer binding sites; (iii) a set of assembly primer pairs suitable for preparing DNA molecules for directed assembly by USER assembly or Gibson assembly into a linear concatemer, wherein each primer comprises, from 5' to 3', an assembly site and a hybridisation site, and in each primer pair the hybridisation sites are designed to bind the first and second universal primer binding sites; (iv) enzymes suitable for assembling DNA fragments by USER assembly or Gibson assembly, wherein the enzymes are suitable for use in the same means of DNA assembly as the assembly primer pairs; and (v) a second primer pair, wherein each primer comprises a sequencing adaptor, a sequencing primer binding site, an index sequence and a hybridisation site, wherein the hybridisation sites are designed to bind the assembly sites of the assembly primers designed to form the ends of the linear concatemer; and wherein the first primer in the pair comprises a first sequencing adaptor, a first sequencing primer binding site and a first index sequence, and the second primer in the pair comprises a second sequencing adaptor, a second sequencing primer binding site and a second index sequence.

In an example, the proximity probes may be probes for a PEA. In such an example, the proximity probe pair may comprise nucleic acid domains that hybridise to one another and template an extension reaction. Thus, the nucleic acid domain of one proximity probe may prime an extension reaction templated by the nucleic domain of the other probe of the pair. In another example the proximity probes may be probes for a PLA. In such an example, the proximity probe pair comprise nucleic acid domains that hybridise to a common ligation template such that may be ligated together, or nucleic acid domains that template the ligation of one or more added oligonucleotides, and/or prime the amplification of the ligation product.

Detailed Description

As mentioned above, the first aspect provides a method of detecting DNA sequences from multiple pools. The DNA sequences are detected by DNA sequencing. A given DNA sequence is identified by sequencing and thus its presence in a pool is confirmed.

A "pool" as used herein is ~~simply a composition or~~ a mixture (e.g. a solution) containing multiple species of DNA molecules. A "species" of DNA molecule means herein a DNA molecule with a particular sequence. Each pool therefore comprises multiple, or in other words a plurality of, different DNA molecules (i.e. DNA molecules having different sequences). By "multiple" or "plurality" as used herein is meant at least two. A pool comprising a plurality of different DNA molecules may be prepared or generated in any convenient or desired way. Different nucleic acid molecules may occur naturally in a sample, and different samples may represent different pools, Alternatively, pools may be prepared by mixing nucleic acid molecules. A pool of nucleic acid molecules may be generated, for example a pool of reporter nucleic acid molecules may be generated by a multiplex assay detecting multiple different analytes in a sample, as discussed further below. Thus each pool comprises at least two species of DNA molecule, e.g. at least 10, at least 50 or at least 100 or more species of DNA molecule. Multiple copies of each species of DNA molecule may be~~, and generally are,~~ present in the respective pools. The DNA sequences from each pool detected in the method are the sequences of, or sequences comprised within, the various