*In vivo* recombination of co-transformed fragments, *i.e.* homologous recombination cloning (HR cloning) can be used to build the final expression vector within a cell. This method was not used in *Pichia pastoris* until recently, since the stability of plasmids in *P. pastoris* and homologous recombination in vivo was inefficient. It was recently reported that ARS plasmids offer an alternative strategy exploiting the yeast's recombination machinery for cloning. It was shown that HR cloning was successful by employing a panARS-based episomal vectors in *Pichia* pastoris (Camattari et al. 2016) and a split marker gene to avoid a high background from inefficient homologous recombination. Although this method and vectors allow the cloning of DNA fragments, the efficiency and flexibility in molecular design is quite limited.

In contrast, *S. cerevisiae* is known as an efficient host for cloning by in vivo recombination and HR cloning has been used in *Saccharomyces* to clone genes of interest into expression vectors (Oldenburg et al. 1997) and also for cloning of multiple fragments or assembly of whole vectors from multiple PCR products (van Leeuwen et al. 2015, Joska et al. 2014). Due to its high efficiency and easy application it has also been used for the creation of libraries ( Vina-Gonzalez et al. 2016) or creation of chimeric DNA (Arenhart et al. 2016).

**Summary**

Autonomously replicating sequences (ARS) with unprecedented technological features were identified. Surprisingly the identified ARS were found as parts of DNA elements which were tested as putative promoter sequences. Episomal plasmid vectors comprising said ARSs, a gene of interest operably linked to a promoter and a selection marker were more efficiently transformed compared to linear expression cassettes comprising said gene of interest, promoter and selection marker. If used under selective conditions, the plasmid vectors described herein resulted several-fold increased expression compared to genomic integration. Furthermore transformation rates were increased and the transformants showed more uniform expression and the employed ARS showed additional functionality as transcription terminators. Surprisingly a heterologous sequence from *Candida boidinii* showed the best technological features – better than any previously known endogenous ARS. The plasmid vectors described herein may be used in methods of protein expression as well as for screening large libraries required for protein or promoter engineering and discovery.

It was surprising that highly efficient library generation can be achieved by in vivo homologous recombination in *Pichia pastoris* using efficient CbARS based vector backbones (i.e. vector backbones comprising an ARS sequence of *Candida boidinii*) in combination with a KU70 *Pichia pastoris* deletion strain and relatively high amounts of DNA with a certain molar ratio (insert:vector backbone) without observing significant background when employing the *CbARS*-based vector backbone alone. Surprisingly no split selection marker was needed for such highly efficient approach for library construction with maximal flexibility in molecular design, allowing to design homologous recombination sites in any region of the vector backbone.

Specifically, the present disclosure provides for an episomal plasmid comprising a gene of interest (GOI) and an autonomously replicating sequence (ARS) which is not operably linked to the GOI, which ARS comprises or consists of a nucleotide sequence identified as any of SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:6-11, or a functionally active variant of any of the foregoing which is characterized by at least 60% sequence identity thereto, specifically at least any of 70%, 80%, 90%, or 95% sequence identity. The episomal plasmid is also referred to as plasmid which is defined by the structural features as claimed or further described herein, which is characterized by its function of being episomally replicable in a host cell, e.g. in Pichia.

Specifically, the ARS is not natively associated with the GOI.

Specifically, the ARS and/or the GOI is a nucleotide sequence which is heterologous to the plasmid. Specifically, both, the ARS and the GOI are heterologous sequences, more specifically from different sources, strains or species.

According to the present disclosure, the ARS and/or the GOI are also heterologous to a host cell, which has been engineered to synthesize the episomal plasmid in the host cell, or to introduce the episomal plasmid into the host cell, or which comprises the episomal plasmid and is cultivated in a host cell culture expressing or otherwise displaying the GOI.

The ARS consisting of the nucleotide sequence identified as SEQ ID NO:3 is a functionally active fragment of SEQ ID NO:2. Therefore, the functional variants of SEQ ID NO:2 shall also include fragments of SEQ ID NO:2 which are at least comprising SEQ ID NO:3.

The ARS consisting of the nucleotide sequence identified as any of SEQ ID NO:6-11 is an exemplary functionally active variant of SEQ ID NO:5. Further

Specifically, the promoter is a CAT1 or AOD promoter.

Specifically, the promoter comprises the nucleotide sequence of any of SEQ ID NO:4, or SEQ ID NO:5, or a functional variant characterized by a sequence identity of at least 60% of any of SEQ ID NO:4, or SEQ ID NO:5, specifically at least any of 70%, 80%, 90%, or 95% sequence identity.

Specifically, the episomal plasmid comprises a selection marker that is based on auxotrophy or chemical resistance, preferably wherein the selection marker is based on glycerol utilization, sucrose utilization, inulin utilization, cellobiose utilization, amino acid auxotrophy, thymidine auxotrophy, nitrogen source utilization, resistance to fluoracetamide, resistance to deoxyglucose, resistance to Zeocin or other antibiotics, resistance to a gene encoding a toxin. The selection marker provides for cultivating a host cell that comprises the episomal plasmid described herein, under selective pressure or under selective conditions, e.g., to select those host cells or plasmid comprising the GOI.

Specifically, the selection marker confers resistance to zeocin, geneticin or glycerol utilization.

According to one aspect, the present disclosure provides an episomal plasmid comprising an expression cassette, which expression cassette comprises a gene of interest (GOI) and regulatory sequences required for expressing said GOI, and an autonomously replicating sequence (ARS) which is not operably linked to the GOI, which ARS comprises or consists of a nucleotide sequence identified as any of SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:6-11, or a functionally active variant of any of the foregoing which is characterized by at least 60% sequence identity thereto.

According to a specific aspect, a eukaryotic host cell is provided which comprises the episomal plasmid described herein.

Specifically, the host cell is a yeast cell, preferably of the Pichia genus, preferably of *Pichia pastoris*, which is a wild-type strain or any mutated strain capable of being cultivated in a cell culture. Specifically, the mutated strain is a ku70 deletion strain of *P. pastoris*, wherein the ku70 gene (SEQ ID NO:87, identified in Figure 33) or an analogous ku70 gene is inactivated or inhibited. Specifically, mutated *P. pastoris* strains which are deficient in the ku70 gene expression are preferred for biosynthesis and production of episomal plasminds and libraries of episomal plasmids as described herein.

Specifically, the host cell is provided in a cell culture wherein the host cells are characterized by a genomic stability regarding the content of the episomal plasmid during at least 20 generations. Therefore, according to the present disclosure, the cell culture of the host cell described herein is provided in a fermenter and device, suitably used for batch, fed-batch, or continuous cultivation.

According to a further aspect, the present disclosure provides for a method of producing a protein of interest (POI) that is encoded by a GOI by cultivating the host cell described herein under conditions to express said GOI. Specifically, the host cell is

cultivated under selective conditions. Specifically, in the cell culture, protein expression is regulatable by the carbon source and/or the feed rate thereof.

Specifically, the POI expression is increased at least 1.5 fold, preferably at least 3 fold, more preferably at least 5 fold or at least 10 fold, compared to the POI expression employing genomic integration of the comparable expression cassette expressing said GOI, which does not contain a functional ARS, such as comprised in the episomal plasmid described herein.

Specifically, the transformation efficiency of the episomal plasmid is increased at least 20 fold or at least 50 fold, preferably at least 100 fold, at least 200 fold, or at least 300 fold, more preferably at least 500 fold compared to transformation employing genomic integration of the comparable expression cassette expressing said GOI, which does not contain a functional ARS, such as comprised in the episomal plasmid described herein.

According to a further aspect, the present disclosure provides for a library of episomal plasmids described herein, which comprises a repertoire of promoter variants and/or a repertoire of GOI variants, which are optionally coexpressed in a host cell culture.

Specifically, the promoter library may be used to discover variants of a parent promoter sequences with desired or improved properties.

Specifically a GOI library may be used to discover gene variants or a parent GOI sequence encoding variants of proteins or polypeptides with desired or improved properties.

The library may be provided as an in vitro library, or as a library of (ex vivo) host cells comprising the repertoire of episomal plasmids, which host cells may be of any eukaryotic or prokaryotic species.

According to a further aspect, the present disclosure provides for a method of selecting a host cell for a desired yield of a GOI expression, the method comprising:

i. contacting a plurality of host cells comprising the library of episomal plasmids described herein, wherein the library comprises a repertoire of promoter variants and wherein the expression level of said GOI is a function of said promoter variants;

ii. determining the expression level in individual host cells of said plurality; and

iii. selecting a host cell which is characterized by the desired expression level of said GOI.

Specifically, the method further comprises cultivating the selected host cell and producing a POI encoded by said GOI.

According to a further aspect, the present disclosure provides for a method of screening promoter variants from a repertoire of promoter variants, and selecting a promoter, the method comprising:

i. contacting a plurality of host cells comprising the library of episomal plasmids described herein, wherein the library comprises a repertoire of promoter variants and a reporter protein wherein any of the expression level of said reporter protein, transformation efficiency or clonal uniformity is a function of said promoter variants;

ii. determining changes in any of the expression level, transformation efficiency or clonal uniformity in individual host cells of said plurality;

iii. selecting a host cell which is characterized by the desired expression level, transformation efficiency or clonal uniformity; and

iv. identifying the promoter variant from the selected host cell.

According to a further aspect, the present disclosure provides for a method of screening variants of a POI encoded by a repertoire of GOI variants, the method comprising

i. contacting a plurality of host cells comprising the library of episomal plasmids described herein, wherein the library comprises a repertoire of GOI variants and wherein any of the GOI expression level, transformation efficiency or clonal uniformity is a function of GOI variants;

ii. determining changes in any of the expression level, transformation efficiency or clonal uniformity in individual host cells of said plurality;

iii. selecting a host cell which is characterized by the desired expression level, transformation efficiency or clonal uniformity; and

iv. identifying the GOI variant from the selected host cell.

According to a further aspect, the present disclosure provides for a method of biosynthesis of an episomal plasmid in a host cell which is of a KU70 deletion strain of *Pichia pastoris*, comprising

i. providing a linear vector backbone comprising recombination sites at its 5' and 3' ends, an ARS, wherein the ARS comprises or consists of a nucleotide sequence identified as any of SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:6-11, or a functionally active variant of any of the foregoing which is

characterized by at least 60% sequence identity thereto, and optionally further regulatory sequences;

ii. providing a vector insert comprising a GOI and 5' and 3' homologous sequences which are homologous to said recombination sites;

iii. introducing said linear vector backbone and said insert into said host cell and recombining said vector insert with said recombination sites by homologous recombination, thereby producing the episomal plasmid comprising said GOI.

The present disclosure provides for a method of biosynthesis of an episomal plasmid in a host cell which is of a KU70 deletion strain of *Pichia pastoris*, comprising

i. providing a linear vector backbone comprising recombination sites at its 5' and 3' ends, an ARS, and optionally further regulatory sequences;

ii. providing a vector insert comprising a GOI and 5' and 3' homologous sequences which are homologous to said recombination sites;

iii. introducing said linear vector backbone and said insert into said host cell and recombining said vector insert with said recombination sites by homologous recombination, thereby producing the episomal plasmid comprising said GOI.

Specifically, said linear vector backbone and said insert are introduced into said host cell at a molar ratio between 1: 1 and 1:10.

Specifically, the ARS comprises or consists of a nucleotide sequence identified as any of SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:6-11, or a functionally active variant of any of the foregoing which is characterized by at least 60% sequence identity thereto, specifically at least any of 70%, 80%, 90%, or 95% sequence identity.

Specifically, the vector backbone further comprises a selection marker, which is based on auxotrophy or chemical resistance, preferably wherein the selection marker is based on glycerol utilization, sucrose utilization, inulin utilization, cellobiose utilization, amino acid auxotrophy, thymidine auxotrophy, nitrogen source utilization, resistance to fluoracetamide, resistance to deoxyglucose, resistance to Zeocin or other antibiotics, resistance to a gene encoding a toxin.

Specifically, the 5' and 3' homologous sequences each comprise or consist of at least any one of 30, 50, 70, 100, 300 base pairs.

Specifically, the 5' and 3' homologous sequences are flanking the GOI or adjacent to the 5' and 3' end of the GOI.

Specifically, the 5' and 3' homologous sequences are located at both ends of the insert, or fused to the 5' and 3' end of the insert core sequence (which is without the 5' and 3' homologous sequences).

According to a further aspect, the present disclosure provides for the use of the biosynthesis method described herein for producing an episomal plasmid described herein.

According to a further aspect, the present disclosure provides for an episomal plasmid obtained by any one of the biosynthesis methods described herein.

According to a further aspect, the present disclosure provides for a method of producing a library of episomal plasmids using the biosynthesis method described herein, wherein the library comprises a repertoire of plasmid variants which differ in at least one point mutation in a predefined region of the GOI. The predefined region is typically a region which is varied by a suitable mutagenesis method (also referred to as region of mutagenesis).

Specifically, the variants are produced by mutagenesis of the vector inserts, in particular mutagenesis of the region of mutagenesis.

According to a further aspect, the present disclosure provides for a library of episomal plasmids obtained by the method described herein, which is characterized by a diversity of at least any of 10E2, 10E3, 10E4, 10E5, or 10E6 different variants. In certain embodiments, even a higher diversity can be produced, e.g. at least any of 10E7, 10E8, 10E9, 10E10, or 10E11 different variants.

Specifically, the library is incorporated into at least 10E6 transformed cells, preferably at least 10E7, 10E8, or 10E9 transformants. In certain embodiments, even a higher number of transformants can be produced, such as to cover the repertoire of episomal plasmid variants.

Specifically, the transformed cells are of the Pichia genus, preferably *P. pastoris*, or a KU70 deletion strain of *P. pastoris*.

According to the present disclosure, the GOI is encoding an antibody, an antibody fragment or antigen-binding sequence thereof, such as a scFv, Fab, VH and/or VL, a single CDR sequence or a set of CDR sequences, e.g. 3 CDR sequences. Specifically, GOI variants are produced by mutagenesis of one or more variable regions or antigen-binding sequences, such as to produce a repertoire of antigen-binding molecules which differ in at least one of epitope specificity or affinity of binding. Alternatively, GOI variants are produced by mutagenesis of one or more constant or framework sequences, such as to improve the stability or Fc function of the antigen-binding molecules.

Specifically, an antibody library is provided, which is characterized by the repertoire of GOI variants, each encoding an antibody or antigen-binding sequence thereof.

Specifically, the library is expressing a repertoire of POI variants, each representing an an antibody or antigen-binding sequence thereof. The library is suitably expressed by in vitro, ex vivo or in vivo expression systems, e.g. in a suitable host cell culture. Upon expression, the POI variants can be screened for selecting a POI variant with desired binding or other features.

**Claims**

1. An episomal plasmid comprising an expression cassette, which expression cassette comprises a gene of interest (GOI) and regulatory sequences required for expressing said GOI, and an autonomously replicating sequence (ARS) which is not operably linked to the GOI, which ARS comprises or consists of a nucleotide sequence identified as any of SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:6-11, or a functionally active variant of any of the foregoing which is characterized by at least 60% sequence identity thereto.

2. The episomal plasmid of claim 1, which further comprises a selection marker and optionally one or more regulatory sequences.

3. The episomal plasmid of any one of the claims 1 to 2, which comprises a promoter which is operably linked to the GOI.

4. The episomal plasmid of claim 3, wherein the promoter is a regulatable or constitutive promoter, preferably a promoter selected from the group consisting of AOX1, GAP, AOD, AOX2, DAS1, DAS2, ENO1, FLD1, FMD, GPM1, HSP82, ICL1, ILV5, KAR2, KEX2, PET9, PEX8, PGK1, PHO89/NSP, SSA4, TEF1, THI11, TPI1, YPT1, GTH1, GCW14, and GUT1, which promoter is preferably of *Pichia pastoris*, or a functional variant of any of the foregoing which is characterized by at least 60% sequence identity and functional as a promoter in a *P. pastoris* strain.

5. The episomal plasmid of any one of claims 3 or 4, wherein the promoter comprises the nucleotide sequence of any of SEQ ID NO:4, or SEQ ID NO:5, or a functional variant characterized by a sequence identity of at least 60% thereto.

6. The episomal plasmid of any one of claims 1 to 5, which comprises a selection marker that is based on auxotrophy or chemical resistance, preferably wherein the selection marker is based on glycerol utilization, sucrose utilization, inulin utilization, cellobiose utilization, amino acid auxotrophy, thymidine auxotrophy, nitrogen source utilization, resistance to fluoracetamide, resistance to deoxyglucose, resistance to Zeocin or other antibiotics, resistance to a gene encoding a toxin.

7. A eukaryotic host cell comprising the episomal plasmid of any one of claims 1 to 6.

8. The host cell of claim 7, wherein the host cell is a yeast cell, preferably of the Pichia genus, preferably of a *Pichia pastoris* strain, which is a wild-type strain or any mutated strain capable of being cultivated in a cell culture.

9. The host cell of claim 7 or 8 in a cell culture wherein the host cells are characterized by a genomic stability regarding the content of the episomal plasmid during at least 20 generations.

10. A method of producing a protein of interest (POI) that is encoded by a GOI by cultivating the host cell of any of claims 7 to 9 under conditions to express said GOI.

11. The method of claim 10, wherein the POI expression is increased at least 1.5 fold, preferably at least 3 fold, more preferably at least 5 fold or at least 10 fold, compared to the POI expression employing genomic integration of the comparable expression cassette expressing said GOI, which does not contain a functional ARS.

12. The method of claim 10 or 11, wherein transformation efficiency of the episomal plasmid is increased at least 20 fold or at least 50 fold, preferably at least 100 fold, at least 200 fold, or at least 300 fold, more preferably at least 500 fold compared to transformation employing genomic integration of the comparable expression cassette expressing said GOI, which does not contain a functional ARS.

13. A library of episomal plasmids of any one of claims 1 to 6, which comprises a repertoire of promoter variants and/or a repertoire of GOI variants, which are optionally coexpressed in a host cell culture.

14. A method of selecting a host cell for a desired yield of a GOI expression, the method comprising:

i. contacting a plurality of host cells comprising the library of episomal plasmids of claim 13, wherein the library comprises a repertoire of promoter variants and wherein the expression level of said GOI is a function of said promoter variants;
ii. determining the expression level in individual host cells of said plurality; and
iii. selecting a host cell which is characterized by the desired expression level of said GOI.

15. The method of claim 14, which further comprises cultivating the selected host cell and producing a POI encoded by said GOI.

16. A method of screening promoter variants from a repertoire of promoter variants, and selecting a promoter, the method comprising:

i. contacting a plurality of host cells comprising the library of episomal plasmids of claim 13, wherein the library comprises a repertoire of promoter variants and a reporter protein wherein any of the expression level of said reporter protein, transformation efficiency or clonal uniformity is a function of said promoter variants;

ii. determining changes in any of the expression level, transformation efficiency or clonal uniformity in individual host cells of said plurality;

iii. selecting a host cell which is characterized by the desired expression level, transformation efficiency or clonal uniformity; and

iv. identifying the promoter variant from the selected host cell.

17. A method of screening variants of a POI encoded by a repertoire of GOI variants, the method comprising

i. contacting a plurality of host cells comprising the library of episomal plasmids of claim 13, wherein the library comprises a repertoire of GOI variants and wherein any of the GOI expression level, transformation efficiency or clonal uniformity is a function of GOI variants;

ii. determining changes in any of the expression level, transformation efficiency or clonal uniformity in individual host cells of said plurality;

iii. selecting a host cell which is characterized by the desired expression level, transformation efficiency or clonal uniformity; and

iv. identifying the GOI variant from the selected host cell.

18. A method of biosynthesis of an episomal plasmid in a host cell which is of a KU70 deletion strain of *Pichia pastoris*, comprising

i. providing a linear vector backbone comprising recombination sites at its 5' and 3' ends, an ARS, wherein the ARS comprises or consists of a nucleotide sequence identified as any of SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:6-11, or a functionally active variant of any of the foregoing which is characterized by at least 60% sequence identity thereto, and optionally further regulatory sequences;

ii. providing a vector insert comprising a GOI and 5' and 3' homologous sequences which are homologous to said recombination sites;

iii. introducing said linear vector backbone and said insert into said host cell and recombining said vector insert with said recombination sites by homologous recombination, thereby producing the episomal plasmid comprising said GOI.

19. The method of claim 18, wherein said linear vector backbone and said insert are introduced into said host cell at a molar ratio between 1: 1 and 1:10.

20. The method of any one of claims 18 to 19, wherein the vector backbone further comprises a selection marker, which is based on auxotrophy or chemical resistance, preferably wherein the selection marker is based on glycerol utilization, sucrose utilization, inulin utilization, cellobiose utilization, amino acid auxotrophy, thymidine auxotrophy, nitrogen source utilization, resistance to fluoracetamide, resistance to deoxyglucose, resistance to Zeocin or other antibiotics, resistance to a gene encoding a toxin.

21. The method of any one of claims 18 to 20, wherein the 5' and 3' homologous sequences each comprise or consist of at least any one of 30, 50, 70, 100, 300 base pairs.

22. Use of the method of any of claim 18 to 21 for producing an episomal plasmid of any of claims 1 to 7.

23. An episomal plasmid obtained by any one of the methods of claim 18 to 21.

24. A method of producing a library of episomal plasmids using the method of any of claims 18 to 21, wherein the library comprises a repertoire of plasmid variants which differ in at least one point mutation in a predefined region of the GOI.

25. The method of claim 24, wherein the variants are produced by mutagenesis of the vector inserts.

26. A library of episomal plasmids obtained by the method of claim 24 or 25, which is characterized by a diversity of at least any of 10E2, 10E3, 10E4, 10E5, or 10E6 different variants.

27. The library of claim 26, which is incorporated into at least 10E6 transformed cells, preferably at least 10E7, 10E8, or 10E9 transformants.

28. The library of claim 27, wherein the transformed cells are of the Pichia genus, preferably *P. pastoris*, or a KU70 deletion strain of *P. pastoris*.

29. The library of any of claims 26 to 28, wherein the GOI is encoding an antibody, an antibody fragment, or antigen-binding sequence thereof.