

## SELECTION SYSTEM CONTAINING NON-ANTIBIOTIC RESISTANCE SELECTION MARKER

### Field of the invention

The present invention relates to a novel selection system, which is based on the use of an *araD* gene, a mutated form of an *araD* gene, a complementary sequence thereof, or a catalytically active fragment thereof as a selection marker and to the use of a bacterial strain deficient of the *araD* gene. The present invention further relates to novel vectors containing an *araD* gene, a mutated form of an *araD* gene, a complementary sequence thereof, or a catalytically active fragment thereof and to novel bacterial strains deficient of an *araD* gene. The present invention additionally relates to a method of selecting the cells transformed with a plasmid, which contains the gene of interest.

### Background of the invention

An essential requirement for effective genetic engineering of bacteria and other cells propagated in cell cultures is the capacity to select the cells with a specific genotypic alteration. The most common selection strategy in recombinant DNA technology is to include a selection marker in the cloning vector or plasmid. A selection marker can be a cloned gene or a DNA sequence, which allows the separation of the host cells containing the selection marker from those not containing it. The selection marker together with a suitable selection medium maintains the cloning vector in the cells. Otherwise, since the replication of plasmids is an energetic burden for the bacterial host, in a growing culture the bacteria, which have lost the plasmid, would have a growth advantage over the cells with the plasmid.

For most purposes, an antibiotic resistance gene is a commonly used selection marker. However, for the production of recombinant therapeutics, where the goal is to generate a product, such as a DNA vaccine, in high yield for administration in patients, the use of antibiotic resistance genes presents problems: the spread of antibiotic resistant pathogens is a serious worldwide problem [Levy, S. B., J. Antimicrob. Chemother. 49 (2002) 25-30]. Therefore the antibiotic resistance genes cannot have extensive use in the pharmaceutical industry, and for instance, according to the regulations of the U.S. Food and Drug Administration, no antibiotic resistance genes are allowed in experimental DNA vaccines entering the third phase.

Alternatively, antibiotic-free selection systems have been suggested. Such antibiotic-free selection systems include bacterial toxin-antitoxin

systems [Engelberg-Kulka, H. and Glaser, G., Annu Rev Microbiol 53 (1999) 43-70], genes responsible for resistance against heavy metals, such as tellurium [Silver, S. and Phung, L. T., Annu Rev Microbiol 50 (1996) 753-789], and systems, in which the plasmid encodes a gene complementing a host auxotrophy [Wang, M.D., *et al.*, J. Bacteriol. 169 (1987) 5610-5614].

US Patent Application 2000/0014476 A1 generally discloses, inter alia, the use of a non-antibiotic selection marker, which may be a gene whose product is necessary for the metabolism of the cell under certain culturing conditions, such as a catabolism gene, which makes it possible for the cell to assimilate a certain substance present in the culture medium (specific carbon or nitrogen source) etc. No specific examples of such suitable genes are given. This approach is not necessarily applicable for commercial production, since the deletion an essential component, such as an amino acid or a carbon source, from the growth medium reduces the yield, which is not desirable. Additionally, the manipulation of the growth medium in terms of omitting an essential nutrient may considerably increase the cost of the growth medium, since commercially available nutrient mixtures must be replaced by individual nutrients.

For commercial therapeutic purposes it would be of advantage to use a gene, which is not essential for the growth of the host but whose manipulation still affects the growth in selected circumstances. Additionally, in view of the therapeutic use, it would be of advantage to use a gene, whose deletion leads to accumulation of compounds, which are toxic to the host cell but not toxic to mammals, including humans. Also it would be of advantage to use smaller genes, which in turn would allow the construction of smaller plasmids for which the energy consumption for replication is smaller and thus the growth rate of bacterial culture and plasmid yield are improved.

### Short description of the invention

The object of the present invention is to provide a novel antibiotic-free selection system, which avoids the problems of previously disclosed selection systems for use in the production of recombinant therapeutic products.

Another object of the invention is to provide a novel antibiotic-free selection system, which can be safely used in the production of recombinant therapeutic products in terms of the environment and the patient safety.

A further object of the invention is to provide a novel antibiotic-free selection system, which can be cost-effectively used in the production of recombinant therapeutic products using standard growth mediums.

5 A still further object of the invention is to provide a novel antibiotic-free selection system, which provides an increased growth rate and improved yield.

Yet another object of the present invention is to provide a novel vector containing a selection marker, which is non-toxic to the environment and to humans and which is capable of a long-term maintenance in the host.

10 Yet another object of the present invention is to provide a novel host cell containing a gene defect, which is not hazardous to the environment.

Still another object of the present invention is to provide a method for selection of cells carrying a gene of interest for the production of recombinant therapeutic products.

15 It was surprisingly found that the objects of the present invention are met by the use of the *araD* gene, a mutated form of an *araD* gene, a complementary sequence thereof, or a catalytically active fragment thereof as a selection marker and the use of a specific bacterial host deficient of the *araD* gene.

Accordingly, the present invention provides a novel selection system  
20 comprising a bacterial cell deficient of an *araD* gene into which a vector carrying an *araD* gene, a complementary sequence thereof, or a catalytically active fragment thereof has been added as a selection marker. One embodiment of the present invention relates to a selection system wherein the *araD* gene is the *araD* gene or the L-ribulose-5-phosphate 4-epimerase (EC 5.1.3.4.). An-  
25 other embodiment of the present invention relates to a selection system wherein the *araD* gene is mutated.

The present invention further provides novel vectors, which contain an *araD* gene, a mutated form of an *araD* gene, a complementary sequence thereof, or a catalytically active fragment thereof as a selection marker.

30 The present invention further provides novel bacterial strains, which are deficient of the *araD* gene.

The present invention further provides a method of selecting the cells transformed with a plasmid, which contains 1) the *araD* gene, a mutated form of an *araD* gene, a complementary sequence thereof, or a catalytically  
35 active fragment thereof as a selection marker and 2) the gene of interest, the

method comprising inserting said plasmid into the *araD* deficient host cell and growing the cells in a growth medium containing arabinose.

### Drawings

Figure 1 shows the use of arabinose as a carbon source by the *E. coli* cells (Lin, 1987).

Figure 2 shows the map of S6wtd1EGFP. The coding sequences for the d1EGFP, E2 and kanamycin resistance marker aminoglycoside-3'-O-phosphotransferase (kana) are indicated by arrows. Additional features are indicated by solid boxes: 10E2BS – ten BPV E2 binding sites with high affinity; CMV-tk - human cytomegalovirus immediately early promoter and HSV Th gene leader sequence; intron – rabbit beta-globin gene intron with optimized SD and SA sites; tkpa – HSV Tk gene polyadenylation signal; RSV LTR – Rous sarcoma virus long terminal repeat; bgh pA – bovine growth hormone gene polyadenylation signal; pUCori – bacterial origin of replication derived from the pUC18 plasmid.

Figure 3 shows the map of S6wtd1EGFPkana/araD1. The coding sequences for the d1EGFP, E2, kanamycin resistance marker aminoglycoside-3'-O-phosphotransferase (kana) and L-ribulose-5-phosphate 4-epimerase (araD) are indicated by arrows. Additional features are indicated by solid boxes: 10E2BS – ten BPV E2 binding sites with high affinity; CMV-tk - human cytomegalovirus immediately early promoter and HSV Th gene leader sequence; intron – rabbit beta-globin gene intron with optimized SD and SA sites; tkpa – HSV Tk gene polyadenylation signal; RSV LTR – Rous sarcoma virus long terminal repeat; bgh pA – bovine growth hormone gene polyadenylation signal; pUCori – bacterial origin of replication derived from the pUC18 plasmid.

Figure 4 shows the map of S6wtd1EGFPkana/araD2. The coding sequences for the d1EGFP, E2, kanamycin resistance marker aminoglycoside-3'-O-phosphotransferase (kana) and L-ribulose-5-phosphate 4-epimerase (araD) are indicated by arrows. Additional features are indicated by solid boxes: 10E2BS – ten BPV E2 binding sites with high affinity; CMV-tk - human cytomegalovirus immediately early promoter and HSV Th gene leader sequence; intron – rabbit beta-globin gene intron with optimized SD and SA sites; tkpa – HSV Tk gene polyadenylation signal; RSV LTR – Rous sarcoma virus long terminal repeat; bgh pA – bovine growth hormone gene polyadenylation signal; pUCori – bacterial origin of replication derived from the pUC18 plasmid.

Figure 5 shows the map of S6wtd1EGFP/*araD*1. The coding sequences for the d1EGFP, E2 and L-ribulose-5-phosphate 4-epimerase (*araD*) are indicated by arrows. Additional features are indicated by solid boxes: 10E2BS – ten BPV E2 binding sites with high affinity; CMV-tk - human cytomegalovirus immediately early promoter and HSV Th gene leader sequence; 5 intron – rabbit beta-globin gene intron with optimized SD and SA sites; tkpa – HSV Tk gene polyadenylation signal; RSV LTR – Rous sarcoma virus long terminal repeat; bgh pA – bovine growth hormone gene polyadenylation signal; pUCori – bacterial origin of replication derived from the pUC18 plasmid.

10 Figure 6 shows the map of S6wtd1EGFP/*araD*2. The coding sequences for the d1EGFP, E2 and L-ribulose-5-phosphate 4-epimerase (*araD*) are indicated by arrows. Additional features are indicated by solid boxes: 10E2BS – ten BPV E2 binding sites with high affinity; CMV-tk - human cytomegalovirus immediately early promoter and HSV Th gene leader sequence; 15 intron – rabbit beta-globin gene intron with optimized SD and SA sites; tkpa – HSV Tk gene polyadenylation signal; RSV LTR – Rous sarcoma virus long terminal repeat; bgh pA – bovine growth hormone gene polyadenylation signal; pUCori – bacterial origin of replication derived from the pUC18 plasmid.

Figure 7A and 7B shows the electrophoretic analysis of the plasmid DNA of the S6wtd1EGFP/*araD*1 (7A) and S6wtd1EGFP/*araD*2 (7B) extracted from the *E. coli* strain AG1delta *araD* grown in different media.

Figure 8 shows the restriction pattern analysis of the plasmid DNA of the S6wtd1EGFP/*araD*1 and S6wtd1EGFP/*araD*2 extracted from the *E. coli* strain AG1delta *araD*

25 Figure 9 shows the electrophoretic analysis of the S6wtd1EGFP/*araD*2 in stability assay.

Figure 10A and 10B shows the restriction pattern analysis of the S6wtd1EGFP/*araD*2 in stability assay.

Figure 11 shows the growth parameters of fed-batch fermentation of AG1Δ*araD* S6wtd1EGFP/*araD*2 measured and registered during fermentation. 30 The abbreviations are as follows: sPump = feeding speed; pO<sub>2</sub> = the oxygen concentration; Temp = growth temperature; mys = desired growth rate; OD = optical density at 600nm.

Figure 12 shows the scheme of lysis and purification of AG1Δ*araD* S6wtd1EGFP/*araD*2. 35

Figure 13 shows the *araD* locus sequence of clone #13.

Figure 14 shows the map of plasmid p3hCG.

Figure 15 shows the map of plasmid *paraDMgB*.

Figure 16 shows the map of plasmid p3*araD*1hCG.

Figure 17 shows the map of plasmid p3*araD*2hCG.

5        Figure 18 shows the results of the analysis of L-arabinose sensitivity of *E. coli* strains with disrupted *araD*.

Figure 19 shows the results of the analysis of the L-arabinose sensitivity in M9 and yeast extract medium with different glucose and arabinose concentrations.

10        Figure 20 shows the map of plasmid p2 MG C #11.

Figure 21 shows the map of plasmid *paraD* MG C #145.

Figure 22 shows the *E. coli* genomic fragment containing the *sgbE* gene.

Figure 23 shows the *E. coli* genomic fragment containing *ulaF* gene.

## 15    **Detailed description of the invention**

The present invention is based on an effort to find an alternative, antibiotic-free selection system, which could be used in the production of recombinant therapeutic products to be administered *in vivo*, especially in the production of DNA vaccines. Surprisingly it was found that the *araD* gene involved  
20    in the pentose phosphate pathway of both prokaryotic and eukaryotic organisms, such as mammals including humans, can be successfully used as a selection marker in an auxotrophic host cell for the plasmid. The use of the auxotrophy has the advantage of not involving a use or generation of toxic substances that could later contaminate the plasmid preparation.

25        An efficient selection system has been constructed on the basis of *araD/araC* genes [Ariza, R. R., *et al.*, Carcinogenesis 14 (1993) 303-305]. However, this selection system has been used in the studies on the mechanisms of mutagenesis but not used before as a selection marker for plasmid maintenance. Ariza *et al.* used a strain where the *araC* gene contains a termination codon and the *araD* gene is inactivated. A product of the *supF* gene,  
30    which codes for a suppressor tRNA, was introduced on the plasmid. In the presence of active suppressor tRNA, enzymatically active product from *araC* was produced causing cell growth arrest (because *araD* was inactive). This system allows to study the suppression of mutations by *supF* tRNA: in case  
35    *supF* is inactivated by mutation, the cells can grow on arabinose. Therefore, this selection system is based on *araC* gene and not on *araD* gene. *araD* was

not introduced into a plasmid, nor was the system designed or characterized for plasmid production purposes.

The *araD* gene codes for an enzyme which is responsible for epimerization of ribulose-5-phosphate to xylulose-5-phosphate (Fig. 1) and therefore allows the use arabinose in the pentose phosphate pathway [Engelsberg, E., *et al.*, J. Bacteriol. 84: (1962) 137-146]. If *araD* is inactivated, ribulose-5-phosphate accumulates in the bacterial cell leading to growth arrest.

If the chromosomal copy of *araD* is inactivated in the host cell and an intact copy of the *araD* gene, a mutated form of the *araD* gene, a complementary sequence thereof, or a catalytically active fragment thereof is inserted into the plasmid, the growth advantage of the plasmid-containing cells in medium containing L-arabinose is achieved as a result from two effects. First, the plasmid-containing cells can use arabinose as a carbon source, and second, the toxic ribulose-5-phosphate does not accumulate. This allows the use of rich growth media supplemented with arabinose. In rich media the *E. coli* cells grow fast and the plasmid yield is high. Inexpensive standard components of the bacterial growth media, such as yeast extract, can be used as an amino acid source. The traces of ribulose-5-phosphate that theoretically could contaminate the plasmid preparation are not a problem, when the preparation is administered *in vivo*, as ribulose-5-phosphate can be efficiently metabolized by human cells and is not toxic.

The use of mutated form of the *araD* gene offers particular advantages. Selection systems of the invention comprising a bacterial cell deficient of an *araD* gene into which a vector carrying a mutated form of the *araD* gene as a selection marker produce an optimal concentration of the *araD* gene product L-ribulose-5-phosphate 4-epimerase to afford rapid uninhibited growth of the bacteria. Similar advantaged are obtained by the use selection systems containing a vector carrying an intact *araD* gene but comprising deletions or mutations elsewhere in the *araD* gene locus.

The selection system of the invention comprises 1) a vector carrying an *araD* gene, a mutated form of the *araD* gene, a complementary sequence thereof, or a catalytically active fragment thereof as a selection marker and 2) a specific bacterial strain deficient of the *araD* gene into which the vector has been added. When the specific host deficient of the *araD* gene is cultured in the presence of arabinose, the only surviving cells are those containing the

vector, which contains an *araD* gene, a mutated form of the *araD* gene, a complementary sequence thereof, or a catalytically active fragment thereof.

In the selection system of the invention any expression vector commonly used in the production of therapeutic products can be employed, whereby the *araD* gene, a mutated form of the *araD* gene, a complementary sequence thereof, or a catalytically active fragment thereof is inserted into the vector using methods generally known in the art. In the present context, the *araD* gene preferably comprises the sequence identified by SEQ ID NO. 1, by SEQ ID NO. 19, or a sequence hybridizable thereto. However, any applicable *araD* genes are also contemplated. In the present context, the term "a catalytically active fragment of the *araD* gene" is any gene fragment coding a polypeptide or a protein capable of epimerization of L-ribulose-5-phosphate to D-xylulose-5-phosphate. In a specific embodiment of the invention the *araD* gene, a complementary sequence thereof, or a catalytically active fragment thereof is inserted in the vector capable of a long-term maintenance and thereby capable of providing a stable expression of the desired antigen(s).

In another specific embodiment of the invention a mutated form of an *araD* gene, a complementary sequence thereof, or a catalytically active fragment thereof is inserted in the vector capable of a long-term maintenance and thereby capable of providing a stable expression of the desired antigen(s).

In a specifically preferred embodiment of the invention the vector used is an expression vector comprising:

(a) a DNA sequence encoding a nuclear-anchoring protein operatively linked to a heterologous promoter, said nuclear-anchoring protein comprising (i) a DNA binding domain which binds to a specific DNA sequence, and (ii) a functional domain that binds to a nuclear component, or a functional equivalent thereof; and

(b) a multimerized DNA sequence forming a binding site for the nuclear anchoring protein, wherein said vector lacks a papilloma virus origin of replication, and

(c) an *araD* gene, a mutated form of an *araD* gene, a complementary sequence thereof, or a catalytically active fragment thereof.

Such vectors have been described in detail in the international patent application WO02/090558, which is incorporated herein by reference.

Most preferably the vector used in the selection method of the present invention is an expression vector comprising:

- (a) the E2 protein of Bovine Papilloma Virus type 1 (BPV), and  
(b) multiple binding sites of the BPV E2 protein incorporated into the vector as a cluster, where the sites can be as head-to-tail structures or can be included into the vector by spaced positioning, wherein said vector lacks a papilloma virus origin of replication, and  
5 (c) the *araD* gene, a complementary sequence thereof, or a catalytically active fragment thereof.

In the selection system of the invention in principle any known host deficient of the *araD* gene and suitable for use in the production of therapeutic products could be employed. In the present connection the term "deficient" denotes a host, in which the *araD* gene is either totally deleted or inactivated by any known method.

In a preferred embodiment of the invention an *Escherichia coli* strain, preferably commercially available *E. coli* strains DH5alpha-T1, AG1 or  
15 JM109, from which the *araD* gene has been deleted with generally known methods, such as those described below in the Examples, is used. In another preferred embodiment of the invention an *E. coli* strain, preferably *E. coli* strain DH5alpha-T1, AG1 or JM109, into which combined deletions have been made for depletion of other genes encoding proteins with L-ribulose-5-phosphate 4-epimerase activity. Alternatively, commercially available *E. coli* strains, preferably *E. coli* strains DH5alpha-T1, AG1 or JM109, in which the *araD* gene and/or other genes encoding proteins with L-ribulose-5-phosphate 4-epimerase activity have been inactivated by any known method can be employed. In the method for selection of cells carrying a gene of interest for the  
20 production of recombinant therapeutic products, the gene of interest is inserted into host cells deficient of an *araD* and/or other genes encoding proteins with L-ribulose-5-phosphate 4-epimerase activity using method well known in the art and the cells are cultured in a growth medium containing arabinose under culturing medium and conditions suitable the host in question.

Any growth medium suitable for culturing *E. coli* cells can be used.  
30 For commercial production the growth medium will naturally be optimized in terms of the yield. Examples of suitable growth media are commercially available growth media, such as M9 and LB (available from several manufacturers, such as Fermentas, Lithuania). The amount of arabinose added in the growth medium is not critical but naturally arabinose should be present in an amount  
35 that is sufficient for the total culturing period. As low amount as 0.1% has been

found sufficient for the selection. Typically arabinose is added to the medium in an amount of about 0.1% to about 2.0%, preferably in an amount of about 0.2% to about 1.0%, most preferably 0.2% to about 0.5%. However the effect of L-arabinose is observed at concentrations as low as 0.01% and L-arabinose  
5 can be added up to 5% in the growth medium. In a special embodiment, where L-arabinose is used both as a selecting agent and as a limited carbon source, 0.2% of L-arabinose is a suitable amount to be added into the growth medium.

The selection system of the invention is suitable for use in any expression system. It is especially suitable for use in the expression of recombinant therapeutic products, such as DNA vaccines, intended for use *in vivo*,  
10 since the problems associated with the use of antibiotic resistance genes are avoided. Likewise the selection system of the invention is suitable for use in the production of recombinant proteins.

The possible contamination of arabinose in the final product resulting from the preparation process is inconsequential, since arabinose is edible sugar contained in foods naturally and as an additive and thus not toxic to mammals including humans.

Additionally, the *araD* gene is smaller in size than the commonly used antibiotic resistance genes against, for instance, ampicillin and tetracycline and of similar size to kanamycin and chloramphenicol resistance genes. This affords an additional advantage, since it allows the construction of small plasmids for which the energy consumption for replication is smaller than for large plasmids. Thereby both the growth rate of bacterial culture and plasmid yield are increased.

25 The present invention may be better understood by reference to the following non-limiting Examples, which are provided as exemplary of the invention. The following examples are presented in order to more fully illustrate the preferred embodiments of the invention. They should in no way be construed, however, as limiting the broad scope of the invention.

## 30 Example 1

### Cloning of *araD* selection plasmids

For cloning *araD* selection constructs plasmid S6wtd1EGFP (Figure 2) was used. It has pMB1 origin of replication and kanamycin resistance marker as functional elements of plasmid backbone. The kanamycin resistance

in this plasmid is conferred by gene that is derived from *E. coli* transposon Tn903.

The *araD* gene was amplified using polymerase chain reaction (PCR) from *E. coli* DH5 $\alpha$  chromosome according to standard procedure. The PCR product was cloned into selected plasmids in two different orientations with the primer pairs s6araDL1 + s6araDR1 or s6araDL1 + s6araDR1, generating products named *araD*1 and *araD*2, respectively:

s6araDL1:

CGCCATGGTTCTCATGTTTGACAGCTTATCATCGATAAGCTTTA  
10 ATGCGGTAGTTTAGCACGAAGGAGTCAACATG (SEQ ID NO. 2);

s6araDR1:

CGCCATGGACTAGTAAAAAAGCCCGCTCATTAGGCGGGCT  
GTCATTACTGCCCGTAATATGC (SEQ ID NO. 3);

s6araDL2:

CGCCATGGACTAGTTCTCATGTTTGACAGCTTATCATCGATAAG  
15 CTTTAATGCGGTAGTTTAGCACGAAGGAGTCAACATG (SEQ ID NO. 4);

s6araDR2:

CGCCATGGAAAAAAGCCCGCTCATTAGGCGGGCTGTCAT-  
TACTGCCCGTAATATGC (SEQ ID NO. 5);

20 The primers were designed so that P2 promoter from plasmid pBR322 (used for driving the tetracycline resistance gene in pBR322) and termination sequence from *trp* operon of *E. coli* were added during PCR to the upstream and downstream of *araD* coding sequence, respectively.

PCR products of 814 and 815 bp were cloned into pUC18 vector linearized with HincII (Fermentas, Lithuania) and correct sequences were verified by sequencing using universal sequencing primers

M13F22: GCCAGGGTTTTCCAGTCACGA (SEQ ID NO. 6) and

M13R24: GAGCGGATAACAATTTACACAGG (SEQ ID NO. 7) and

*araD* specific primers

30 *araD* F311: CCAACTCACCGGCTGCTCTATC (SEQ ID NO. 8),

*araD* F614: AATGCCGAAGATGCGGTGCATAAC (SEQ ID NO. 9),

*araD* R700: TAACTGCGGCGCTAACTGAC (SEQ ID NO. 10), and

*araD* R421: GGTTGCTGGAATCGACTGAC (SEQ ID NO. 11).

The mutations in amplified sequences were repaired by recombination of different clones.

For cloning *araD* into S6wtd1EGFP, the vector was linearized by partial digestion with restriction enzyme *PagI* (position 4761) (Fermentas, Lithuania) and the DNA 5'-termini were dephosphorylated with Calf Intestine Alkaline Phosphatase (CIAP; Fermentas, Lithuania). *araD1* and *araD2* fragments were cut out from pUC18 with *NcoI* (Fermentas, Lithuania) and ligated to S6wtd1EGFP/*PagI*.

Both ligation mixtures were transformed into *E. coli DH5α* competent cells and plated onto dishes containing LB medium containing 50 µg/ml kanamycin and incubated at 37°C over night. Colonies were first analysed with colony PCR, after which the DNA was isolated and digested with different restriction enzymes.

The cloning resulted in plasmids S6wtd1EGFP*kana/araD1*, S6wtd1EGFP*kana/araD2*, which are shown in Figures 3 and 4.

To remove the kanamycin resistance marker gene from the plasmids, S6wtd1EGFP*kana/araD1* and S6wtd1EGFP*kana/araD2* were digested with restriction endonuclease *BclI* (Fermentas, Lithuania) and a 6473 bp vector fragment was self-ligated.

The ligation mixtures were transformed into an *E. coli AG1 ΔaraD* strain (see Example 3) and plated onto dishes containing M9 media supplemented with 2% L-arabinose and incubated at 37°C for 36 hours. Colonies were first analyzed with colony PCR, after which the DNA was isolated and digested with different restriction enzymes. The cloning resulted in plasmids S6wtd1EGFP/*araD1*, S6wtd1EGFP/*araD2*, respectively, are shown in Figures 5 and 6.

The bacterial colonies containing S6wtd1EGFP/*araD1* and S6wtd1EGFP/*araD2* were grown in two different media: LB supplemented with 2.5% L-arabinose and M9 supplemented with 0.2% L-arabinose at 37°C with vigorous shaking. The cells were harvested and the plasmid DNA was extracted from the cell using QIAprep Spin Miniprep Kit (QIAGEN) and analysed by agarose gel electrophoresis (Figures 7A and 7B, respectively).

The plasmid DNA samples from cultures in LB and M9 media were analysed by agarose gel electrophoresis before and after digestion with restriction endonuclease *PagI* (Fermentas, Lithuania), (Figure 8). The predicted sizes of the fragments obtained in the *PagI* digestion were 3954 and 2519 bp for S6wtd1EGFP/*araD1* and 4315 and 2157 bp for S6wtd1EGFP/*araD2*. Lambda DNA digested with *Eco91I* (M15 in Figure 8C) and lambda DNA digested with

EcoRI/ HindIII (Fermentas, Lithuania) (M3 in Figure 8C) were used as molecular weight markers. All analyzed bacterial clones contained the correct plasmid in the restriction enzyme analysis, but the DNA yield was very low when the plasmids were grown in LB media. Two of the analyzed bacterial clones from  
5 four S6wtd1EGFP/*araD*2 clones (#13 and #14 in Figure 8B) had higher growth rate when grown in M9 media supplemented with 0,2% L-arabinose (Figures 7 and 8), which resulted in higher plasmid yield per culture.

Further analysis of these two clones with improved growth was performed. These two plasmids had the same structure as the other plasmids as  
10 judged by restriction analysis. The plasmids were extracted from the bacteria and further characterized by sequencing the *araD* gene locus. The *araD* locus sequence of clone #13 (SEQ ID NO. 18; SEQ ID NO. 19) indicated that *araD* gene coding sequence carries a STOP codon instead of a codon for Glutamine in position 8 of L-ribulose-5-phosphate 4-epimerase. This mutation resulted  
15 from the replacement of Cytidine in codon 8 of L-ribulose-5-phosphate 4-epimerase (*araD* coding sequence) (5'-CAG-3') with the Thymidine, resulting in a STOP codon (5'-TAG-3'). The plasmid carrying such a mutation in *araD* gene effectively provided the ability to grow in the selective medium in the presence of L-arabinose although the coding sequence contains the STOP codon. It has  
20 been demonstrated that the STOP codon UAG is effectively read through by the ribosomes of *Escherichia coli*, when such a STOP is in the beginning of the coding sequence [for reference, see review Murgola, E. J., Annu. Rev. Genet. 19 (1985) 57-80]. Without binding by the theory, we hypothesized that the high yield of the plasmid, which is an indication of rapid uninhibited growth of the  
25 bacteria, requires an optimal concentration of the *araD* gene product L-ribulose-5-phosphate 4-epimerase.

The analysis of clone #14 *araD* locus sequence indicated that the *araD* coding sequence is perfect as predicted. However, the sequence rearrangements near the *araD* promoter covering the E2 protein binding sites were  
30 observed (see Figure 13, SEQ ID NO. 18). These data suggested additionally that such rearrangements near the promoter might result in the down-regulation of the promoter activity, therefore the level of the *araD* product.

## Example 2

### Cloning of mutated *araD* selection plasmids

For cloning of mutated *araD* selection constructs plasmid p3hCG (Figure 14) carrying kanamycin resistance [transposon Tn5 derived kanamycin resistance marker (neo) gene] was cleaved with the restriction endonucleases BcuI and HindIII, the ends were filled in using Klenow Fragment (Fermentas, Lithuania) and the fragment with the size of 4647 bp was purified from the gel after agarose gel electrophoresis. The pMB1 origin of replication and the *araD* sequence carrying the C to T mutation, which results in a STOP codon in position 8 of the *araD* gene coding sequence, was excised from the plasmid paraDMgB (Figure 15) with the restriction endonucleases BcuI and Eco52I, the ends were filled in using Klenow Fragment (Fermentas, Lithuania), and the DNA 5'-termini were dephosphorylated with Calf Intestine Alkaline Phosphatase (CIAP; Fermentas, Lithuania). The fragment with the size of 1532 bp was purified from the gel after agarose gel electrophoresis and ligated with the 4647 bp fragment obtained above. *Escherichia coli* AG1 *araD* deficient strain was transformed with this ligation mixture and plated onto agar plates containing selective M9 medium with 0.5% yeast extract, 2% L-arabinose and 25 µg/ml of kanamycin. The colonies were inspected 24 hours after the plating and showed that the size of the colonies was uniform. The plasmids were extracted from the bacteria and further characterized by sequencing of the *araD* gene locus.

The cloning resulted in plasmids p3araD1hCG and p3araD2hCG, which are shown in Figures 16 and 17, respectively. According to the sequence analysis, the bacteria contained un-rearranged plasmids with the mutation C to T in codon 8 (p3araD1hCG; Figure 16; p3araD2hCG, Figure 17).

When this experiment was repeated with the wild type sequence and transformed plates were inspected 24 hours after the transformation, the result was different. Two types of colonies were observed: first, large size colonies, and small colonies, which had a retarded growth. The sequence analysis of these plasmids indicated that *araD* gene coding sequence carries a STOP codon instead of a codon for glutamine (plasmid #3A, *araD2*) or the mutation had occurred in the Shine-Dalgarno sequence in the ribosomal binding site (AGGAG was replaced with AGTAG) (plasmid #2A, *araD2*). Plasmid #7 (*araD1*) had the correct sequence in all *araD* gene locus regions, however, the

bacteria grew very slowly and resulted in a 10 times lower plasmid yield when were grown in liquid media.

### Example 3

#### Construction of arabinose sensitive $\Delta araD$ *Escherichia coli* strains.

5 Three *E. coli* strains, DH5alpha T1, AG1 and JM109, were used to construct  $\Delta araD$  mutants. The *araD* gene in *E. coli* genome was disrupted using the method described by Datsenko and Wanner [PNAS 97 (2000) 6640-6645]. This method exploits a phage  $\lambda$  Red recombination system. Briefly, the strategy of this system is to replace a chromosomal sequence with a select-  
10 able antibiotic resistance gene that is generated by PCR by using primers with homology extensions. This is accomplished by Red-mediated recombination in these flanking homologies.

For transformation of the pKD46 (Datsenko and Wanner, *supra*), which encodes the phage  $\lambda$  recombination system, *E. coli*, the cells were made  
15 chemically competent using RF1 and RF2 solutions:

**RF1 100ml**

|                                       |                            |
|---------------------------------------|----------------------------|
| RbCl 1                                | 1.2 g                      |
| MnCl <sub>2</sub> · 4H <sub>2</sub> O | 0.99 g                     |
| 1 M KAc pH 7.5                        | 3 ml                       |
| CaCl <sub>2</sub> · 2H <sub>2</sub> O | 0.15 g                     |
| Glycerol                              | 15 g                       |
| pH 5.8                                | (add CH <sub>3</sub> COOH) |

**RF2 100ml**

|                                       |            |
|---------------------------------------|------------|
| 0.5 M MOPS                            | 2 ml       |
| RbCl                                  | 0.12 g     |
| CaCl <sub>2</sub> · 2H <sub>2</sub> O | 1.1 g      |
| Glycerol                              | 15 g       |
| pH 6.8                                | (add NaOH) |

20

The cells were grown in 2 ml of LB medium to OD<sub>600</sub> 0.2-0.5. The culture was centrifuged and the pellet was resuspended in 1 ml of RF1. The mixture was kept on ice for 10 min and centrifuged. The pellet was suspended

in 100  $\mu$ l of RF2 and the suspension was kept on ice for 30-45 min. Approximately 50 ng of pKD43 was added and the cells were kept on ice for additional 30 min followed by heat shock of 5 min at 37°C. After incubation for 10 min on ice 900  $\mu$ l of SOB medium was added to the transformed cells and the mixture  
5 was incubated at 37°C for one hour. Cells were plated on LB medium containing ampicillin (100  $\mu$ g/ml). The colonies were picked from the transformation plates and grown in 2 ml of the same medium to OD<sub>600</sub> of approximately 1 and glycerol stocks were made (2 ml culture + 0.6 ml 50% glycerol). The stocks were stored at -80°C.

10 For disruption of the *araD* gene a linear PCR product which contains kanamycin resistance gene was generated. Plasmid pKD13 (Datsenko and Wanner, PNAS vol. 97, no 12, June 2000) was used as the PCR template. Primers used were *ara*(pr1) and *ara*(pr4):

*ara*(pr1)

15 5'-CTCAAACGCCCAGGTATTAGAAGCCAACCTGGCGCTGCC-  
AAAACACGTGTAG GCTGGAGCTGCTTC 3' (SEQ ID NO. 12)

*ara*(pr4)

5'-GGTTTGATCACAAAGACGCCGCGCTCGCGATCAACGGCGC-  
ATTCCGGGGAT CCGTCGACC 3' (SEQ ID NO. 13)

20 These primers have the complement sequences with pKD13 for annealing in PCR and with the *araD* gene for homologous recombination.

The PCR reaction mixture was as follows: PFU native buffer (5  $\mu$ l), 10 mM dNTP (5  $\mu$ l), primer *ara*(pr1) 10  $\mu$ M (1  $\mu$ l), primer *ara*(pr4) 10  $\mu$ M (1  $\mu$ l), pKD13 100 ng (2  $\mu$ l), DMSO (4  $\mu$ l), PFU 2.5 U (1  $\mu$ l), and mQ water up to 50  
25  $\mu$ l.

The PCR procedure was as follows: denaturation 45 s, 96°C, annealing 45 s, 50°C, synthesis 2 min 30 s, 72°C, 25 cycles. The PCR product obtained was 1.4 kb.

Five reactions were performed simultaneously; the DNA was purified from 2% agarose gel using Ultrapure purification Kit (MoBio Laboratories Inc.) and eluted with 60  $\mu$ l of water. The DNA was concentrated with ethanol precipitation and dissolved in 5  $\mu$ l of water. The final concentration was 0.6  $\mu$ g/ $\mu$ l. An aliquot of 1.5  $\mu$ l was used in one electroporation.

The PCR product was electroporated into DH5alpha T1 pKD46,  
35 AG1 pKD46 (Datsenko and Wanner, *supra*), and JM109 pKD46 *E. coli* cells. First, 200 ml of YENB medium containing 10 mM of L-arabinose for the induc-

tion of the recombination system and 100  $\mu$ g/ml ampicillin was inoculated with an overnight culture of DH5alpha T1 pKD46, AG1 pKD46, and JM109 pKD46 *E. coli* cells. The cultures were grown at 30°C to OD<sub>600</sub> 0.8 (DH5alpha T1 and JM109) and 0.6 (AG1). The bacteria was collected by centrifugation at 4,000 g for 10 min at 4°C, washed twice with 20 ml of sterile water and once with 20 ml of sterile water containing 10% glycerol. The cells were suspended in 300  $\mu$ l water containing 10% glycerol. 40  $\mu$ l of competent cells were used in one electroporation.

The electroporation was performed with BioRad *E. coli* Pulser using 0.2 cm cuvettes and 2.5 kV. The purified PCR product (1.5  $\mu$ l) was added to the competent cells, kept on ice for 1 min, and immediately after the electroporation, 2 ml of warm SOB medium was added to the cells and the mixture was incubated at 37°C for 1 hour. The cells were plated on LB medium containing kanamycin (25  $\mu$ g/ml). 100 pg of large kanamycin resistant plasmid (GTU-MultiHIV C-clade) was used as a positive control, no plasmid was added to the negative control. The transformation efficiency was 10<sup>6</sup> for AG1 and 10<sup>7</sup> for JM109 for positive control. There were no colonies on the negative control plate, 215 colonies were obtained on JM109+PCR product plate, 70 colonies on AG1+PCR product plate and 50 colonies on DH5alpha T1+PCR product plate.

#### Example 4

##### Testing of the *E. coli* DH5alpha T1 $\Delta$ araD, AG1 $\Delta$ araD and JM109 $\Delta$ araD strains

The colonies obtained from the electroporation as described in Example 2 were tested for the presence of kanamycin resistance gene by colony PCR using primers araVlisF (5' CGGCACGAAGGAGTCAACAT 3'; SEQ ID NO. 14) and araVlisR (5' TGATAGAGCAGCCGGTGAGT 3'; SEQ ID NO. 15) which contain annealing sites on the araD gene near the insertion site. A PCR product of 272 bp was expected from the *E. coli* DH5alpha T1, AG1 and JM109 strains without insertion in araD and a 1545 bp product, if the PCR product had been inserted in the araD gene. Three colonies of DH5alpha T1  $\Delta$ araD, nine colonies of AG1 $\Delta$ araD and 14 colonies of JM109 $\Delta$ araD out of 15 were checked and each gave the 1545 bp product. It was therefore concluded that these strains contained the kanamycin resistance gene insertion.

To confirm the insertion of kanamycin gene another colony PCR was performed using primers *kanaSF* (5'TCAGATCCTTGGCGGCAAGA3'; SEQ ID NO. 16) and *araVR* (5'TGTAATCGACGCCGGAAGGT3'; SEQ ID NO. 17). These primers produce a 435 bp product, if the kanamycin resistance gene has been inserted into the *araD* gene. Six colonies from AG1 $\Delta$ *araD* and JM109 $\Delta$ *araD* strains and three colonies of DH5alpha T1  $\Delta$ *araD* strains were tested and all gave the correct product.

Six colonies of AG1 $\Delta$ *araD* and JM109 $\Delta$ *araD*, and three colonies of DH5alpha T1  $\Delta$ *araD* were plated on LB medium containing 25  $\mu$ g/ml of kanamycin and incubated at 37°C overnight to eliminate the pKD46 plasmid, which has a temperature sensitive replication origin. The cells were tested for ampicillin sensitivity by replica plating on LB medium and LB medium containing ampicillin. None grew on the medium containing ampicillin and it was concluded that the bacteria does not contain the pKD46 plasmid any more.

The arabinose sensitivity was tested on the produced AG1 $\Delta$ *araD* and JM109 $\Delta$ *araD* strains. One colony of AG1 $\Delta$ *araD* and one colony of JM109 $\Delta$ *araD* were each inoculated into 2 ml LB. The cultures were grown for 8 hours, diluted 1:100 into M9 medium containing 0.2% glycerol, 25  $\mu$ g/ml kanamycin, 0.01% thiamine (0.05% proline for JM109 $\Delta$ *araD*) and different concentrations of L-arabinose were added in the growth medium. The cultures were grown overnight at 37°C in shaker incubator and OD<sub>600</sub> was measured (Table 1).

**Table 1. Testing of arabinose sensitivity.**

| L-arabinose % | AG1 $\Delta$ <i>araD</i> OD <sub>600</sub> | JM109 $\Delta$ <i>araD</i> OD <sub>600</sub> |
|---------------|--------------------------------------------|----------------------------------------------|
| 0             | 3.2                                        | 1.9                                          |
| 0.1           | 0.03                                       | 0.03                                         |
| 0.2           | 0.030                                      | 0.026                                        |
| 0.5           | 0.030                                      | 0.020                                        |
| 1             | 0.024                                      | 0.025                                        |
| 2             | 0.017                                      | 0.021                                        |

As can be seen from Table 1, as low amount as 0.1% of L-arabinose is enough to inhibit the growth of the  $\Delta$ *araD* strains of the invention.

The arabinose sensitivity was further tested on AG1 $\Delta$ *araD*, DH5alphaT1  $\Delta$ *araD* and JM109 $\Delta$ *araD* as above but using lower concentrations

of L-arabinose. The results are given in Figure 18. As can be seen in Figure 18, as low an amount as 0.0005% of L-arabinose is enough to inhibit the growth of the  $\Delta araD$  strains of the invention.

Additionally the L-arabinose sensitivity was tested in M9 and yeast  
5 extract medium with different glucose and arabinose concentrations (0.2% glucose, 0.2% arabinose, 2% arabinose). The cultures were incubated at 37°C in a shaker incubator overnight. Then the OD<sub>600</sub> was measured to quantitate the cell density. The results are given in Figure 19.

Both concentrations of arabinose (0.2% and 2%) inhibited the  
10 growth of the  $\Delta araD$  strains of the invention. However, the growth of strains with intact *araD* gene was not inhibited.

Additionally the plasmid DNA yield of the  $\Delta araD$  strains was tested. Plasmid S6wtd1EGFP $\Delta araD$ 2 prepared in Example 1 was transformed into AG1 $\Delta araD$  and JM109 $\Delta araD$  strains. Competent cells were prepared with  
15 RF1 and RF2 solutions as described in Example 3.

The colonies from the transformation plates were inoculated into 2 ml of M9 medium containing 0.5% yeast extract and 25 µg/ml kanamycin + 0.01% thiamine + L-arabinose (2% and 0.2%).

The cultures were incubated at 37°C for 17 hours. Then the OD<sub>600</sub>  
20 was measured to quantitate the cell density and the plasmid DNA was extracted with Qiagen Miniprep Kit. Coefficient 2.8 (OD<sub>600</sub>/ml) was used for mini-prep isolation to get comparable results. The results are shown in Table 2.

DNA concentration was measured with spectrophotometer as OD at 260 nm. For microscopic analysis a drop of bacterial culture was applied on  
25 glass slide and covered with cover slip. The culture was visually inspected at a 100xmagnification with an objective in oil immersion.

**Table 2. Plasmid DNA yield of  $\Delta araD$  strains**

| Strain              | L-arabinose (%) | OD <sub>600</sub> | Plasmid DNA conc. ( $\mu\text{g}/\mu\text{l}$ ) | Plasmid DNA yield ( $\mu\text{g}$ per ml of culture) | Appearance in microscope |
|---------------------|-----------------|-------------------|-------------------------------------------------|------------------------------------------------------|--------------------------|
| AG1 $\Delta araD$   | 2               | 7.6               | 0.039                                           | 5.3                                                  | no filaments             |
| AG1 $\Delta araD$   | 0.2             | 5.8               | 0.057                                           | 5.9                                                  | no filaments             |
| JM109 $\Delta araD$ |                 |                   |                                                 |                                                      | very few                 |
| D                   | 2               | 4.9               | 0.043                                           | 3.8                                                  | filaments                |
| JM109 $\Delta araD$ |                 |                   |                                                 |                                                      | very few                 |
| D                   | 0.2             | 4.3               | 0.038                                           | 2.9                                                  | filaments                |
| DH5 $\alpha$ T1     |                 |                   |                                                 |                                                      | no filaments             |
| $\Delta araD$       | 2               | 6.6               | 0.017                                           | 3.5                                                  |                          |
| DH5 $\alpha$ T1     |                 |                   |                                                 |                                                      | no filaments             |
| $\Delta araD$       | 0.2             | 6.4               | 0.016                                           | 3.4                                                  |                          |

According to these results 0.2% L-arabinose is sufficient for obtaining the plasmid copy number at the same level as with 2% arabinose.

For this plasmid AG1 $\Delta araD$  seems to be better, because the plasmid yield is somewhat higher and cell densities also.

#### Example 5.

**Generation of an *Escherichia coli* strain with additional mutations within the genes potentially encoding L-ribulose-5-phosphate 4-epimerase.**

*E. coli* chromosome contains two additional coding sequences for L-ribulose-5-phosphate 4-epimerases in different operons. The *ulaF* and *sgbE* genes from L-ascorbate degradation pathway encode the genes with epimerase activity (Wen Shan Yew, Jhon A. Gerlt, J. Bacteriol. 184 (2002) 302-306. In order to increase the stringency of the selection and to avoid or knock out the possible adaptation mechanisms of *E. coli* strains due to other genes with epimerase activity, the coding sequences of the *UlaF* and *SgbE* genes in *E. coli* genome were interrupted. Such adaptation mechanisms could occur in long-term plasmid production under suitable conditions.

The *UlaF* and *SgbE* genes in *E. coli* strains DH5alphaT1ΔaraD and AG1ΔaraD were disrupted using the phage λ Red recombination system as described in Example 3.

First, the kanamycin-resistant gene in *E. coli* AG1ΔaraD and  
5 DH5αT1ΔaraD strains was eliminated. FLP recombinase expression plasmid pKD20 (Datsenko and Wanner, *supra*) is ampicillin resistant and temperature-sensitive. Kanamycin-resistant mutants were transformed with pCP20 (kanamycin-resistant gene is FRT-flanked), and ampicillin-resistant transformants were selected at 30°C (48 hours), after which the same colonies were purified  
10 non-selectively at 42°C (24 hours twice). Then they were tested for loss of kanamycin and ampicillin resistances.

The inactivation of the chromosomal *ulaF* gene (SEQ ID NO. 20) by the phage λ Red recombination system was performed using the primers *ulaFylem* and *ulaFalum*:

15 *ulaFylem*  
CAGCAGGTATTTGAAGCCAACATGGAGCTGCCGCGCTACG-  
GGCTGGTGTAGGCTGGAGCTGCTTC (SEQ ID NO. 21)  
*ulaFalum*  
AAACGGCTGCGGAATTAGACCAGTTATCTCCCGAGGAAGGAAA  
20 TTAATTCCGGGGATCCGTCGACC (SEQ ID NO. 22)

A lot of colonies were observed on both transformation plates. Fifteen colonies obtained from the electroporation were tested for the presence of the kanamycin resistance gene by colony PCR using primers *ulaFvalisR* and *ulaFvalisF*:

25 *ulaFvalisR*  
AAACGGCTGCGGAATTAGACC (SEQ ID NO. 23)  
*ulaFvalisF*  
GCCGTACCTGATTGAGATGTGGAG (SEQ ID NO. 24)

These primers contain annealing sites on the *UlaF* gene near the  
30 insertion site. A PCR product of 864 bp was expected from the *E. coli* DH5alphaT1ΔaraD and AG1ΔaraD strains without insertion in *UlaF* and a 1527 bp product, if the PCR product had been inserted in the *UlaF* gene. To confirm the insertion of the kanamycin gene another colony PCR was performed using primers *ulaFvalisR* (SEQ ID NO 23) and *kanaSF* (SEQ ID NO 16).

35 These primers produce a 428 bp product, if the kanamycin resistance gene has been inserted into the *UlaF* gene. Four colonies from

AG1 $\Delta$ *araD* $\Delta$ *ulaF* and DH5 $\alpha$ phat1 $\Delta$ *araD* $\Delta$ *ulaF* strains were tested and all gave the correct product. One colony from each strain was used further.

The elimination of the kanamycin-resistant gene in *E. coli* AG1 $\Delta$ *araD* $\Delta$ *ulaF* and DH5 $\alpha$ phat1 $\Delta$ *araD* $\Delta$ *ulaF* strains was performed as described above. The inactivation of the chromosomal *sgbE* gene (SEQ ID NO. 25) by the phage  $\lambda$  Red recombination system was performed as described in Example 3. The primers used were *sgbE*alum and *sgbE*ylem:

*sgbE*alum  
CGTTACAGCAAGGAACATATCAATTCGTAGTGCCGGGGCGATG  
10 AAGAATTCCGGGGATCCGTCGACC (SEQ ID NO. 26)  
*sgbE*ylem  
GCAGGAGGCTGGATTTATATGTTAGAGCAACTGAAAGCCG-  
ACGTGGTGTAGGCTGGAGCTGCTTC (SEQ ID NO. 27)

A lot of colonies were observed on both transformation plates. Fifteen colonies obtained from the electroporation were tested for the presence of kanamycin resistance gene by the colony PCR using primers *sgbE*evalisR and *sgbE*evalisF:

*sgbE*evalisR  
CGGCGTTACAGCAAGGAACATATC (SEQ ID NO. 28)  
20 *sgbE*evalisF  
ATTGAAGCGCGTATGCAGGAGG (SEQ ID NO. 29)

A PCR product of 792 bp was expected from the *E. coli* DH5 $\alpha$ phat1 $\Delta$ *araD* $\Delta$ *ulaF* $\Delta$ *sgbE* and AG1 $\Delta$ *araD* $\Delta$ *ulaF* $\Delta$ *sgbE* strains without insertion in *SgbE* and a 1413 bp product, if the PCR product had been inserted in the *SgbE* gene. To confirm the insertion of kanamycin gene another colony PCR was performed using primers *sgbE*evalisR (SEQ ID NO. 28) and *kanaSF* (SEQ ID NO. 16):

Fifteen colonies from both strains were tested and four gave the correct product.

30 The arabinose sensitivity was tested on the *E. coli* DH5 $\alpha$ phat1 $\Delta$ *araD* $\Delta$ *ulaF* $\Delta$ *sgbE* and AG1 $\Delta$ *araD* $\Delta$ *ulaF* $\Delta$ *sgbE* strains produced and compared to those of *E. coli* DH5 $\alpha$ phat1 $\Delta$ *araD* and AG1 $\Delta$ *araD* strains. One colony of each strain was inoculated into 2 ml of M9 medium containing 0.5% yeast extract, 25  $\mu$ g/ml of kanamycin, 0.2% glucose only or 0.2% or 2% L-arabinose, respectively. The results are shown in Table 3.  
35

**Table 3. Testing of arabinose sensitivity**

| Strain                                                     | OD <sub>600</sub> Glc | OD <sub>600</sub> Glc+0,2%<br>L-arabinose | OD <sub>600</sub> Glc+2%<br>L-arabinose |
|------------------------------------------------------------|-----------------------|-------------------------------------------|-----------------------------------------|
| AG1 $\Delta$ araD                                          | 7.3                   | 0.82                                      | 0.26                                    |
| DH5alphaT1<br>$\Delta$ araD                                | 7.7                   | 0.95                                      | 0.35                                    |
| AG1 $\Delta$ araD<br>$\Delta$ ulaF $\Delta$ sgeE           | 8.3                   | 0.82                                      | 0.35                                    |
| DH5alphaT1<br>$\Delta$ araD $\Delta$ ulaF<br>$\Delta$ sgeE | 7.5                   | 0.75                                      | 0.28                                    |

As can be seen from Table 3, there were no essential differences in the arabinose sensitivity of the strains of the invention. Similarly, when the plasmid DNA yield of the  $\Delta$ araD and  $\Delta$ araD $\Delta$ ulaF $\Delta$ sgeE strains was tested as described in Example 3 (the results are not shown), no differences were found between *E.coli* AG1 $\Delta$ araD and AG1 $\Delta$ araD $\Delta$ ulaF $\Delta$ sgeE or DH5alphaT1 $\Delta$ araD and DH5alphaT1 $\Delta$ araD $\Delta$ ulaF $\Delta$ sgeE strains.

#### 10 Example 6

##### Stability of S6wtd1EGFP/araD2

An important feature of the vaccination vector is the stability during propagation in bacterial cells. To test the stability of S6wtd1EGFP/araD2 in bacteria the plasmid was transformed into the *E. coli* AG1 $\Delta$ araD and JM109 $\Delta$ araD strains prepared in Example 3 and the intactness of the vector was followed by the plasmid DNA analysis during four generations.

The plasmid S6wtd1EGFP/araD2 was mixed with competent *E. coli* AG1 $\Delta$ araD and JM109 $\Delta$ araD cells and incubated on ice for 30 minutes. Subsequently, the cell suspension was subjected to a heat-shock for 3 minutes at 37°C followed by a rapid cooling on ice. One milliliter of LB medium was added to the sample and the mixture was incubated for 45 minutes at 37°C with vigorous shaking. Finally, a portion of the cells was plated onto M9 medium dishes containing 0.5% yeast extract, 2% L-arabinose and 25  $\mu$ g/ml of kana-

mycin. On the next day, the cells from one colony were transferred onto the new dish containing the same medium. This procedure was repeated until four passages of bacteria had been grown. Two colonies from each passage of both bacterial strains were used to inoculate of 2 ml of M9 medium containing  
5 0.5% yeast extract, 2% L-arabinose and 25  $\mu$ g/ml of kanamycin incubated overnight at 37°C with vigorous shaking. The cells were harvested and the plasmid DNA was extracted from the bacteria using QIAprep Spin Miniprep Kit (QIAGEN). The plasmid DNA samples before (Figure 9) and after the digestion with restriction endonuclease HindIII (Figure 10) (Fermentas, Lithuania) were  
10 analyzed by agarose gel electrophoresis in comparison with the original S6wtd1EGFP/araD2 DNA used for transformation (as control in Figures 9 and 10). Lambda DNA digested with EcoRI/HindIII (Fermentas, Lithuania) was used as a molecular weight marker (M3 in Figure 10).

Samples were digested with HindIII as shown in Figure 10A for *E. coli* AG1 $\Delta$ araD and in Figure 10B for JM109 $\Delta$ araD strain, patterns identical to the original S6wtd1EGFP/araD2 plasmid DNA were observed. The predicted sizes of the fragments resulted by HindIII digestion are 3274, 1688 and 1510 bp. It can be concluded that the vaccination vector S6wtd1EGFP/araD2 is stable when propagated in *E. coli* AG1 $\Delta$ araD and JM109 $\Delta$ araD strains.

## 20 Example 7

### Comparison of an antibiotic selection system with the L-arabinose selection system of the invention

In the comparison of an antibiotic selection system with the L-arabinose selection system of the invention the following growth media were  
25 used.

For *E. coli* AG1 carrying plasmid p2 MG C #11:

Medium 1: M9 medium plus 0.5% yeast extract, 0.2% glucose and  
25  $\mu$ g/ml of kanamycin (selective medium);

Medium 2: M9 medium plus 0.5% yeast extract and 0.2% glucose  
30 (non-selective medium);

Medium 3:

M9 medium plus 0.5% yeast extract, 0.2% L-arabinose and 25  
 $\mu$ g/ml of kanamycin; (selective medium); and

Medium 4: M9 medium plus 0.5% yeast extract and 0.2% L-  
35 arabinose (non-selective medium).

For *E. coli* AG1 $\Delta$ araD carrying paraD MG C #145:

Medium 5:

M9 medium plus 0.5% yeast extract, 0.2% L-arabinose and 25  $\mu$ g/ml of kanamycin (selective medium); and

5 Medium 6: M9 medium plus 0.5% yeast extract, 0.2% glucose and 25  $\mu$ g/ml of kanamycin (non-selective medium).

The plasmids p2 MG C #11 (Figure 20) and paraD MG C #145 (Figure 21) were transformed into *E. coli* AG1 and into *E. coli* AG1 $\Delta$ araD carrying the mutation C to T in codon 8. The transformed bacterial colonies were grown at 37°C overnight in an incubator. Next morning the colonies were inoculated into the selective and non-selective liquid media as indicated above. The inoculated cultures were grown in a shaker in 2 ml of the respective medium until they reached the stationary phase, and the density of the cultures was measured at OD<sub>600</sub>. The plasmid was extracted from the cultures and the plasmid DNA yield was determined by the measurement of the plasmid DNA at 260 nm. The plasmid yield was calculated on the basis that 50  $\mu$ g yields to an optical density of 1 at 260 nm.

Then an aliquot of 20  $\mu$ l from the stationary cultures was inoculated into fresh medium (dilution 100 times), and the cultures were grown until stationary phase (8-12 hours). The density of the cultures was measured at OD<sub>600</sub>, the plasmid was extracted and the yield was determined, and again an aliquot was inoculated into 2  $\mu$ l of the liquid medium. This procedure was repeated 7 times (preparations 1 to 7). The results of the experiment are provided in Table 5 below.

**Table 5. Comparison of an antibiotic selection system with the L-arabinose selection system of the invention**

| Medium number/<br>preparation number | OD <sub>600</sub> | Amount of<br>plasmid DNA<br>per 1 ml culture |
|--------------------------------------|-------------------|----------------------------------------------|
| 1/1                                  | 6.215             | 6.35µg                                       |
| 1/7                                  | 3.278             | 2.3µg                                        |
| 2/1                                  | 6.652             | 6.15µg                                       |
| 2/7                                  | 5.133             | 0.65µg                                       |
| 3/1                                  | 7.317             | 10.9µg                                       |
| 3/7                                  | 3.046             | 1.6µg                                        |
| 4/1                                  | 6.874             | 6µg                                          |
| 4/7                                  | 4.634             | 0.75µg                                       |
| 5/1                                  | 7.271             | 6.45µg                                       |
| 5/7                                  | 7.014             | 5.15µg                                       |
| 6/1                                  | 6.131             | 5.3µg                                        |
| 6/7                                  | 6.031             | 4.4µg                                        |

5 It can be concluded from these data that a plasmid carrying the kanamycin resistance gene and conferring *E. coli* the resistance in the presence of kanamycin is lost in the consecutive dilution/growing steps of the culture under the non-selective as well as under selective conditions. The yield of the plasmid from 1 ml culture drops 3 times under the selective conditions and  
10 10 times under the non-selective conditions at the seventh round of dilution (preparations 1/1 vs. 1/7 and 2/1 vs. 2/7, respectively, in Table 5). The same basic result is obtained, when the carbon source for *E. coli* carrying a plasmid with kanamycin resistance is L-arabinose instead of glucose (preparations 3/1 vs. 3/7 and 4/1 vs. 4/7, respectively, in Table 5). However, when the *araD* selection system of the invention is used in the plasmid, the plasmid DNA yield is  
15 high under both selective (preparation 5/1 vs. 5/7 in Table 5) and non-selective (preparation 6/1 vs. 6/7 in Table 5) conditions. Both under selective and non-selective conditions the plasmid DNA yield dropped over 7 generations approximately 20%. This indicates clearly that the plasmids carrying *araD* selec-

tion system of the invention are much more stable and grow efficiently under the selective as well as non-selective conditions.

### Example 8

#### Fed-batch fermentation of AG1 $\Delta$ araD S6wtd1EGFP/araD2

5           The *araD* gene based selection system was also tested in fed-batch fermentation for the purpose of production of plasmid containing bacteria. A single colony was picked from AG1 $\Delta$ araD S6wtd1EGFP/araD2 plate and inoculated into 250 ml M9 medium containing 0.5% yeast extract, 0.2% L-arabinose and 25  $\mu$ g/ml of kanamycin and incubated overnight at 37°C with  
10 vigorous shaking. After 18 hours the OD<sub>600</sub> of inoculum was 6.4. 160 ml of inoculum was added to fermentor containing 5 l Fermenter Starting Medium (8 g/l KH<sub>2</sub>PO<sub>4</sub>; 10 g/l NaCl; 5 g/l NH<sub>4</sub>Cl; 5 g/l yeast extract; 2 g/l L-arabinose; 2 g/l MgSO<sub>4</sub>, 25 mg/l kanamycin and 0.1 g/l thiamine; pH 6.7 with NH<sub>4</sub>OH). After 5.5 hours of growth automatic feeding was started with given growth speed of 0.15  
15 h<sup>-1</sup> (allows carbon-source limited growth) with fermenter feeding medium (300 g/l L-arabinose; 150 g/l yeast extract; 50 mg/l kanamycin; 0.2 g/l thiamine). Feeding speed was controlled by computer according to formulae  $F(t)=\mu_y S^* S_{in}/S_f$  where  $\mu_y S^*$  is desired growth rate,  $S_{in}$  is the amount of carbon source added to the time point and  $S_f$  is carbon source concentration in feeding  
20 medium. The growth was followed by measuring OD<sub>600</sub> and samples for plasmid DNA were taken. The data registered during fermentation is represented in Figure 11. Fermentation was terminated when 1 l of feeding medium was consumed. Final OD<sub>600</sub> was 45. The bacterial mass was collected by centrifugation and washed once with 2 l STE buffer. Yield of bacterial biomass was 410 g wet  
25 weight. The data for plasmid DNA content is shown in Table 6.

**Table 6. Plasmid DNA yield during AG1ΔaraD S6wtd1EGFP/araD2 fermentation**

| Time     | OD <sub>600</sub> | Plasmid DNA conc.<br>(μg/μl) | Plasmid DNA yield (μg<br>per ml of culture) |
|----------|-------------------|------------------------------|---------------------------------------------|
| Inoculum | 6.4               | 0.04                         | 4.6                                         |
| 4 h      | 3.1               | 0.02                         | 1.1                                         |
| 21 h     | 28                | 0.1                          | 50                                          |
| 24 h     | 37                | 0.13                         | 87                                          |
| 29 h     | 45                | 0.14                         | 113                                         |

- 5           The data in Table 6 indicate that the L-arabinose selection system works very well at high cell densities. It is probably because more plasmid copies in bacterial cell gives an advantage in the conditions of L-arabinose limitation by enabling the bacterium to use sugar more rapidly.

### Example 9

#### 10 Purification of AG1ΔaraD S6wtd1EGFP/araD2

The purification of AG1ΔaraD S6wtd1EGFP/araD2 was performed as follows (Figure 12):

##### a) Feeding preparation

- 15           Clear lysate was prepared according to Qiagen's Plasmid Purification Handbook, except RNase was not used.

- 20           200g of *E. coli* cell paste was resuspended in 2000ml of Resuspension Buffer and later equal volumes of P2 and P3 for lysis and neutralization were used. The cell debris was removed by centrifugation at 6000g for 30 minutes at 4°C. Clear lysate was poured through the paper towel, 1/10 of 10% Triton X-114 (Sigma) was added and solution was left on ice for 1 hour. (Triton X-114 has been shown to effectively reduce the level of endotoxins in protein, Liu et al., Clinical Biochemistry, 1997) After one hour nucleic acids were precipitated with 0,6 volumes of cold isopropanol. Supernatant was decanted and precipitate was stored overnight at -20°C.

## b) Plasmid DNA purification

Plasmid DNA purification was performed according to Amersham Pharmacia's three step supercoiled plasmid purification process, where few modifications were adopted.

5           Step 1. Precipitate was redissolved in 1500ml TE (10mM Tris-Cl, 1mM EDTA; pH 8.0) and loaded for RNA removal and buffer exchange on Sepharose 6 FF (Amersham Pharmacia), previously equilibrated with Buffer A – 2M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 100mM Tris Cl, 10mM EDTA, pH 7.5.

10           Step 2. Void volume was directed to the PlasmidSelect (Amersham Pharmacia) column (equilibrated with Buffer A) and after washing and elution with Buffer B2 (1,6M NaCl, 2M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 100mM Tris Cl, 10mM EDTA, pH 7.5), supercoiled plasmid DNA was captured.

15           Step 3. Eluted plasmid was diluted with five volumes of distilled, de-ionized water and loaded to SOURCE 30Q (Amersham Pharmacia) equilibrated with buffer C1 (0,4M NaCl, 100mM Tris Cl, 10mM EDTA, pH 7.5) . After washing, purified plasmid was eluted with Buffer C2 (1M NaCl, 100mM Tris Cl, 10mM EDTA, pH 7.5) and elution peak was collected. Fraction size was 150ml and it contained 100mg of endotoxins-free (<10 EU/mg) S6wtd1EGFP/araD2 plasmid.

### Claims

1. A selection system comprising a bacterial cell deficient of an *araD* gene into which a vector carrying an *araD* gene has been added as a selection  
5 marker.
2. A selection system according to claim 1, wherein the *araD* gene is L-ribulose-5-phosphate 4-epimerase gene (EC 5.1.3.4.).
3. A selection system according to claim 1 or 2, wherein the *araD* gene of the vector is mutated said vector generating the *araD* gene product L-  
10 ribulose-5-phosphate 4-epimerase.
4. A selection system according to claim 3, wherein the mutation introduces a stop codon into position 8 of the *araD* gene.
5. A selection system according to any one of claims 1 to 4, wherein the vector comprises a mutation in the Shine-Dalgarno sequence of the *araD*  
15 gene.
6. A selection system according to any one of claims 1 to 5, wherein the bacterial cell is an *Escherichia coli* cell.
7. A selection system according to claim 6, wherein the *E. coli* is *E. coli* strain JM109.
8. A selection system according to claim 6, wherein the *E. coli* is *E. coli* strain DH5 alpha-T1.  
20
9. A selection system according to claim 6, wherein the *E. coli* is *E. coli* strain AG1.
10. A selection system according to any one of claims 7 to 9, wherein  
25 said *E. coli* strain is deficient of the *araD* gene and *ulaF* gene.
11. A selection system according to any one of claims 7 to 9, wherein said *E. coli* strain is deficient of the *araD* gene and *sgbE* gene.
12. A selection system according to any one of claims 7 to 9, wherein said *E. coli* strain is deficient of *araD* gene, *ulaF* gene, and *sgbE* gene.

13. A selection system according to any one of the preceding claims, wherein said vector is an expression vector further comprising:

5 (a) a DNA sequence encoding a nuclear-anchoring protein operatively linked to a heterologous promoter, said nuclear-anchoring protein comprising (i) a DNA binding domain which binds to a specific DNA sequence, and (ii) a functional domain that binds to a nuclear component, and

(b) a multimerized DNA sequence forming a binding site for the nuclear anchoring protein, wherein said vector lacks a papilloma virus origin of replication.

10 14. A selection system according to claim 13, wherein:

(a) the nuclear-anchoring protein is the E2 protein of Bovine Papilloma Virus type 1 (BPV), and

15 (b) the multimerized DNA sequence are multiple binding sites for the BPV E2 protein incorporated into the vector as a cluster, where the sites can be as head-to-tail structures or can be included into the vector by spaced positioning.

15. A selection system according to claim 14, wherein the vector additionally comprises a deletion in the multimerized DNA sequence.

20 16. A selection system according to claim 14, wherein the vector additionally comprises a mutation in the Shine-Dalgarno sequence of the *araD* gene.

17. A vector comprising an *araD* gene as a selection marker which *araD* gene is mutated generating a stop codon at position 8, wherein said vector generates the *araD* gene product L-ribulose-5-phosphate 4-epimerase.

25 18. A vector according to claim 17, wherein the vector is an expression vector further comprising:

30 (a) a DNA sequence encoding a nuclear-anchoring protein operatively linked to a heterologous promoter, said nuclear-anchoring protein comprising (i) a DNA binding domain which binds to a specific DNA sequence, and (ii) a functional domain that binds to a nuclear component; and

(b) a multimerized DNA sequence forming a binding site for the nuclear anchoring protein, wherein said vector lacks a papilloma virus origin of replication.

19. A vector according to claim 18, wherein:

(a) the nuclear-anchoring protein is the E2 protein of Bovine Papilloma Virus type 1 (BPV), and

(b) the multimerized DNA sequence are multiple binding sites for the BPV E2 protein incorporated into the vector as a cluster, where the sites can be as head-to-tail structures or can be included into the vector by spaced positioning.

20. A vector according to claim 19 additionally comprising a deletion in the multimerized DNA sequence.

21. A vector according to claim 19 additionally comprising a mutation in the Shine-Dalgarno sequence of the *araD* gene.

22. *E. coli* strain DH5alpha-T1 deficient of the *araD* gene and *ulaF* gene.

23. *E. coli* strain DH5alpha-T1 deficient of the *araD* gene and *sgbE* gene.

24. *E. coli* strain DH5alpha-T1 deficient of the *araD* gene, *ulaF* gene, and *sgbE* gene.

25. *E. coli* strain AG1 deficient of the *araD* gene and *ulaF* gene.

26. *E. coli* strain AG1 deficient of the *araD* gene and *sgbE* gene.

27. *E. coli* strain AG1 deficient of the *araD* gene, *ulaF* gene, and *sgbE* gene.

28. *E. coli* strain JM109 deficient of the *araD* gene and *ulaF* gene.

29. *E. coli* strain JM 109 deficient of the *araD* gene and the *sgbE* gene.

30. *E. coli* strain JM 109 deficient of the *araD* gene, *ulaF* gene, and *sgbE* gene.

31. A method of selecting the cells transformed with a plasmid containing an *araD* gene, and the gene of interest, the method comprising

inserting the plasmid into the *araD* deficient host cell and growing the cells in a growth medium containing arabinose.

32. A method according to claim 31, wherein the *araD* gene is L-ribulose-5-phosphate 4-epimerase gene (EC 5.1.3.4.).

5           33. A method according to claim 31 or 32, wherein the *araD* gene of the vector is mutated said vector generating the *araD* gene product L-ribulose-5-phosphate 4-epimerase.

34. A method according to claim 33, wherein the mutation introduces a stop codon into position 8 of the *araD* gene.

10           35. A selection system as claimed in claim 1 and substantially as herein described with reference to the accompanying Examples.

36. A vector according to claim 17 and substantially as herein described with reference to the accompanying Examples.

15           37. Method according to claim 31 and substantially as herein described with reference to the accompanying Examples.

2031002.ST25.txt  
SEQUENCE LISTING

&lt;110&gt; FIT Biotech Oyj Plc.

&lt;120&gt; Novel selection system

&lt;130&gt; 2031002

&lt;160&gt; 29

&lt;170&gt; PatentIn version 3.1

&lt;210&gt; 1

&lt;211&gt; 780

&lt;212&gt; DNA

&lt;213&gt; Escherichia coli

<400> 1  
gtttogtttg attggctgtg gttttataca gtcattaactg cccgtaatat gccttcgcgc 60  
catgcttaacg cagatagtggt ttatccagca gcgtttgctg catatccggt aactgcggcg 120  
ctaactgacg gcagaatatc cccatataag cgacctcttc cagcacgatg gcgttatgca 180  
ccgcatcttc gccatttttg ccccatgcaa acggggccgtg ggaatggacc agaacgcggg 240  
gcatttgccg tgcatacgata ccctgttttt oaaagggttc tacgatgacg ttaccgggtt 300  
cccactcata ttccgccgttg atttctgcgt cggtcatttt gcgggtgcag ggaatggtgc 360  
cgtagaaata gtcggcgttg gtggtgccgg ttgctggaat cgactgaccc gcctgcgccc 420  
agatggtggc gtggcgcgag tgcgtatgca caatgcgcgc aatggagggg aatgcctgat 480  
agagcagccg gtgagttggc gtgtcggagg agggcttttt cgtaccttca accacttcac 540  
cggtttcgat gctaaccacg accatatcgt cagcgggtcat gacgctgtaa tcgacgcggg 600  
aaggtttgat caciaagacg ccgcgcctgc gatcaacggc gctgacgttg ccccatgtga 660  
gcgtgaccag gttgtgtttt ggcagcgcca ggttggttc taatacctgg cgtttgagat 720  
cttctaacat gttgactcct tcgtgccgga tgcgctttgc ttatccggcc tacaaaatcg 780

&lt;210&gt; 2

2031002.ST25.txt

&lt;211&gt; 76

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 2

cgccatgggt ctcattgttg acagcttatt atcgataagc tttaatgcgg tagtttagca 60

cgaaggagtc aacatg 76

&lt;210&gt; 3

&lt;211&gt; 64

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 3

cgccatggac tagtaaaaaa aagcccgctc attaggcggg ctgtcattac tgcccgtaat 60

atgc 64

&lt;210&gt; 4

&lt;211&gt; 81

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 4

cgccatggac tagttctcat gtttgacagc ttatcatcga taagctttaa tgccgtagtt 60

tagcacgaag gactcaacat g 81

&lt;210&gt; 5

&lt;211&gt; 58

&lt;212&gt; DNA

2031002.ST25.txt

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 5

cgccatggaa aaaaaagccc gctcattagg cgggctgtca ttactgcccg taatatgc 58

&lt;210&gt; 6

&lt;211&gt; 22

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 6

gccagggttt tcccagtcac ga 22

&lt;210&gt; 7

&lt;211&gt; 24

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 7

gagcggataa caatttcaca cagg 24

&lt;210&gt; 8

&lt;211&gt; 22

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 8

ccaactcacc ggctgctcta tc 22

2031002.ST25.txt

<210> 9  
<211> 24  
<212> DNA  
<213> Artificial Sequence

<220>

<223> Primer

<400> 9  
aatgccgaag atgcggtgca taac

24

<210> 10  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>

<223> Primer

<400> 10  
taactgcggc gctaactgac

20

<210> 11  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>

<223> Primer

<400> 11  
ggttgctgga atcgactgac

20

<210> 12  
<211> 66  
<212> DNA  
<213> Artificial Sequence

2031002.ST25.txt

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 12

ctcaaacgcc caggtattag aagccaacct ggcgctgcc aaacacgtgt aggctggagc 60  
tgcttc 66

&lt;210&gt; 13

&lt;211&gt; 60

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 13

ggtttgatca caaagacgcc gcgctcgca tcaacggcgc attccgggga tccgtcgacc 60

&lt;210&gt; 14

&lt;211&gt; 20

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 14

cggcacgaag gagtcaacat 20

&lt;210&gt; 15

&lt;211&gt; 20

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer

&lt;400&gt; 15

tgatagagca gccggtgagt 20

2031002.ST25.txt

<210> 16  
 <211> 20  
 <212> DNA  
 <213> Artificial Sequence

<220>

<223> Primer

<400> 16  
 tcagatcctt ggcggaaga 20

<210> 17  
 <211> 20  
 <212> DNA  
 <213> Artificial Sequence

<220>

<223> Primer

<400> 17  
 tgtaatcgac gccggaaggt 20

<210> 18  
 <211> 1030  
 <212> DNA  
 <213> Artificial

<400> 18  
 ggatccgacc ggcaacggta cagatccgac cggcaacggc acagatccga ccggcaacgg 60  
 tcagatccga ccggcaacgg tacagatccg accggcaacg gtacagatcc gaccggcaac 120  
 ggtacagatc cgaccggcaa cggtagatcc cggaccggca acggtacaga tccgaccggc 180  
 aacggtacag atccgaccgg caacggtaca gatcccccta gogaattgac tagttctcat 240  
 gtttgacagc ttatcatcga taagctttaa tgcggtagtt tagcacgaag gagtcaacat 300  
 gttagaagat ctcaaacgcc aggtattaga agccaacctg gcgctgocaa aacacaaact 360  
 ggtaacgctc acatggggca acgtcagcgc cgttgatcgc gagcgcgcg tctttgtgat 420  
 caaaccttcc ggctcgatt acagcgtcat gaccgctgac gatatggcgc tgggttagcat 480

2031002.ST25.txt

```

cgaaaccggt gaagtgggtg aaggtacgaa aaagccctcc tccgacacgc caactcaccg      540
gctgctctat caggcattcc cctccattgg cggcattgtg catacgcact cgggccacgc      600
caccatctgg gcgcaggcgg gtcagtcgat tccagcaacc ggcaccaccc acgccgacta      660
tttctacggc accattccct gcaccgcgaa aatgaccgac gcagaaatca acggcgaata      720
tgagtgggaa accggtaacg tcctcgtaga aacctttgaa aaacagggtg tcgatgcagc      780
gcaaattgccc ggcgttcttg tccattccca cggcccgttt gcatggggca aaaatgccga      840
agatgcgggtg cataacgcca tcgtgctgga agaggtcgct tatatgggga tattctgccg      900
tcagttagcg ccgcagttac cggatatgca gcaaacgctg ctggataaac actatctgcg      960
taagcatggc gcgaaggcat attacgggca gtaatgacag cccgcctaata gagcggggctt    1020
ttttttccat                                     1030

```

&lt;210&gt; 19

&lt;211&gt; 696

&lt;212&gt; DNA

&lt;213&gt; Escherichia coli

&lt;400&gt; 19

```

atgttagaag atctcaaacg ccaggtatta gaagccaacc tggcgctgcc aaaacacaac      60
ctgggtcagc tcacatgggg caacgtcagc gccgttgatc gcgagcgagg cgtctttgtg    120
atcaaaccctt cgggcgtcga ttacagcgtc atgaccgctg acgatatggg cgtgggttagc    180
atcgaaaccg gtgaagtggg tgaaggtacg aaaaagccct cctccgacac gccaaactcac    240
cggctgctct atcaggcatt cccctccatt ggcggcattg tgcatacgca ctgcgcgccac    300
gccaccatct gggcgcgagg gggtcagtcg attccagcaa ccggcaccac ccacgcogac    360
tattttctacg gcaccattcc ctgcacccgc aaaatgaccg acgcagaaat caacggcgaa    420
tatgagtggg aaaccggtaa cgtcatogta gaaacctttg aaaaacaggg tatcgatgca    480
gcgcaaattgc cggcggttct ggtccattcc caccggccgt ttgcatgggg caaaaatgcc    540
gaagatgcgg tgcataacgc catcgtgctg gaagaggtcg cttatatggg gatattctgc    600
cgtcagttag cggcgagtt accggatatg cagcaaacgc tgctggataa acactatctg    660
cgtaagcatg gcgcgaaggc atattacggg cagtaa                                696

```

&lt;210&gt; 20

&lt;211&gt; 687

&lt;212&gt; DNA

2031002.ST25.txt

&lt;213&gt; Escherichia coli

<400> 20  
atgcaaaagc taaaacagca ggtatattgaa gccaacatgg agctgccgcg ctacgggctg 60  
gtgaccttta cctggggcaa cgtcagcgct atcgaccgcg aacgcgggct ggtggtgatc 120  
aagcccagcg gcgttgcta cgaaaccatg aaagcggcgc atatggtggg ggttgatatg 180  
agcggcaagg tgggtggaagg ggagtatcgc ccatcttccg aactgcgac gcatctcgaa 240  
ctctaccgtc gttaccgcgc gcttgggtggc attgtccata cccactccac tcatgccacc 300  
gcatgggcgc aggcggggct ggcgatcccg gcgttaggca ccacgcacgc cgactacttc 360  
tttggcgaca ttccgtgtac gcgcgggtta agcgaagaag aggtgcaggg cgagtatgaa 420  
ctgaacaccg gcaaagtgat tatcgaaacg ctgggcaacg ccgagccgct gcatacgccg 480  
ggaattgtgg tgtatcagca cgggccgttc gcctggggga aagatgctca cgatgcggtg 540  
cataacgcgg tgggtgatgga agaagtggcg aaaatggcgt ggattgcccg cggcattaac 600  
ccacaactca atcacatcga cagcttcttg atgaataaac acttcatgcg taaacacggt 660  
cctaacgctt attacgggca gaagtag 687

&lt;210&gt; 21

&lt;211&gt; 65

&lt;212&gt; DNA

&lt;213&gt; Artificial

<400> 21  
cagcaggtat ttgaagccaa catggagctg ccgcgctacg ggctgggtgta ggctggagct 60  
gcttc 65

&lt;210&gt; 22

&lt;211&gt; 66

&lt;212&gt; DNA

&lt;213&gt; Artificial

<400> 22  
aaacggctgc ggaattagac cagttatctc ccgaggaagg aaattaattc cggggatccg 60  
tcgacc 66

&lt;210&gt; 23

2031002.ST25.txt

&lt;211&gt; 21

&lt;212&gt; DNA

&lt;213&gt; Artificial

&lt;400&gt; 23

aaacggctgc ggaattagac c

21

&lt;210&gt; 24

&lt;211&gt; 24

&lt;212&gt; DNA

&lt;213&gt; Artificial

&lt;400&gt; 24

gccgtacctg attgagatgt ggag

24

&lt;210&gt; 25

&lt;211&gt; 696

&lt;212&gt; DNA

&lt;213&gt; Escherichia coli

&lt;400&gt; 25

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| atgttagagc aactgaaagc cgacgtgctg gggcggaatc tggcgcttcc cgctoaccat  | 60  |
| ctgggtgacgt tcacctgggg caatgtcagc gcggtagacg aaacgcggca atggatggta | 120 |
| atcaaacctt ccggcgctga gtacgacgtg atgaccgccg acgatatggg ggtgggtgag  | 180 |
| atagccagcg gtaagggtgg ggaaggcagc aaaaaacctt cttccgatac accaacgcac  | 240 |
| ctggcgctct accgtcgcta tgccgaaatt ggcggtattg tgcataccca ctgcgccac   | 300 |
| gccaccatct ggtcacaggc cgggctggat ctccccgcct ggggcaccac ccacgccgat  | 360 |
| tatTTTTTtac gtgccatccc ctgcacgcga cagatgaccg cagaggagat taacggcgaa | 420 |
| tatgaatatc agaccggcga agtgatcatt gaaaccttcg aagaacgtgg caggagtccg  | 480 |
| gcacaaatcc cggcggtgct ggtgcattct caggcccggt tcgcatgggg taaaaacgcc  | 540 |
| gccgatgccg tgcataacgc cgtagtactc gaagaatgcg cctatatggg tctattctcg  | 600 |
| cggcagcttg cggcgcagct ccctgcgatg caaacgaac tgctggataa gcactacctg   | 660 |
| cgtaagcatg gggccaatgc ctattacggg cagtaa                            | 696 |

&lt;210&gt; 26

2031002.ST25.txt

&lt;211&gt; 67

&lt;212&gt; DNA

&lt;213&gt; Artificial

&lt;400&gt; 26

cgttacagca aggaacatat caattcgtag tgccggggcg atgaagaatt ccgggggatcc 60

gtcgacc 67

&lt;210&gt; 27

&lt;211&gt; 65

&lt;212&gt; DNA

&lt;213&gt; Artificial

&lt;400&gt; 27

gcaggaggct ggatttatat gttagagcaa ctgaaagccg acgtggtgta ggctggagct 60

gcttc 65

&lt;210&gt; 28

&lt;211&gt; 24

&lt;212&gt; DNA

&lt;213&gt; Artificial

&lt;400&gt; 28

cggcggttaca gcaaggaaca tatc 24

&lt;210&gt; 29

&lt;211&gt; 22

&lt;212&gt; DNA

&lt;213&gt; Artificial

&lt;400&gt; 29

attgaagcgc gtatgcagga gg 22