Bescheid/Protokoll (Anlage)   **Communication/Minutes** (Annex)   **Notification/Procès-verbal** (Annexe)

Datum
Date   17.08.2004
Date

Blatt
Sheet   1
Feuille

Anmelde-Nr.:
Application No.: 02 798 716.3
Demande n°:

The examination is being carried out on the **following application documents**:

Description, Pages

1-61   as originally filed

Claims, Numbers

1-46   as originally filed

Drawings, Sheets

1/21-21/21   as originally filed

The following documents (D) are considered relevant to this application; the numbering will be adhered to in the rest of the procedure:

D1: US-B-6 238 9241 (DUJON BERNARD ET AL) 29 May 2001 (2001-05-29)

D2: WO 96/14408 A (PASTEUR INSTITUT ;UNIV PARIS CURIE (FR)) 17 May 1996 (1996-05-17)

D3: COHEN-TANNOUDJI MICHEL ET AL: "I-Scel-induced gene replacement at a natural locus in embryonic stem cells." MOLECULAR AND CELLULAR BIOLOGY, vol. 18, no. 3, March 1998 (1998-03), pages 1444-1448, XP002238353 ISSN: 0270-7306

D4: VAN DIJK R ET AL: "Tagging Hansenula polymorpha genes by random integration of linear DNA fragments (RALF)." MGG MOLECULAR GENETICS AND GENOMICS, vol. 266, no. 4, December 2001 (2001-12), pages 646-656, XP002238354 ISSN: 1617-4615

D5: THERMES VIOLETTE ET AL: "I-Scel meganuclease mediates highly efficient transgenesis in fish." MECHANISMS OF DEVELOPMENT, vol. 118, no. 1-2, 20 October 2002 (2002-10-20), pages 91-98, XP002238356 ISSN: 0925-4773

Bescheid/Protokoll (Anlage)   **Communication/Minutes** (Annex)   **Notification/Procès-verbal** (Annexe)

Datum
Date   17.08.2004
Date

Blatt
Sheet   2
Feuille

Anmelde-Nr.:
Application No.: 02 798 716.3
Demande n°:

Patentability - Article 52(4) EPC, Rule 23d

The present application does not meet the requirements of Article 52(4) EPC, because the subject-matter of claim 30 refers to methods of treatment of the human or animal body by therapy which are excluded from patentability as being not susceptible of industrial application.

Claims 1, 17-20 need to be restricted to "in vitro" methods and "non-human" in order to not violate the provisions of Rule 23d.

Novelty - Article 54 EPC

The present application does not meet the requirements of Article 52(1) EPC, because the subject-matter of claims 1-4, 6-46 is not new in the sense of Article 54(1) and (2) EPC.

D1 (or D2) discloses a method directed to introducing a double strand break in a host genome containing an (engineered) I-SceI site. A nucleic acid can be introduced at this site by co-transforming the cell with a plasmid containing a gene segment flanked by 2 I-SceI sites and a plasmid expressing I-SceI. The method is unwittingly identical to that of the present application which is based on the finding that integration of the linearized DNA occurs in the absence of host I-SceI sites.

With regard to D1, D2 novelty could be restored by specifying either of the following:

Option 1: That the target genome contains no I-SceI sites (as specified in claim 5).
Option 2: That the target genome has not been engineered to contain further I-SceI sites.

Out of the 2 options, the examiner is in favour of no 1 for the following reasons:

The method of the present application concerns random integration. If the genome contains I-SceI sites integration at those sites will be favoured relative to random integration (see D2 pages 51-53 which notices a low frequency of random integration relative to site specific integration by homologous recombination at the double strand break). Thus, it could even be considered an essential feature of the method that no I-SceI sites are present in the target genome.

The above objection also applies to composition claim no 34.

Product by process claims are only available if the products themselves fulfill the criteria

Bescheid/Protokoll (Anlage)   **Communication/Minutes** (Annex)   **Notification/Procès-verbal** (Annexe)

Datum
Date   17.08.2004

Blatt
Sheet   3
Feuille

Anmelde-Nr.:
Application No.: 02 798 716.3
Demande n°:

of patentability, i.e that they are novel and inventive, see Guidelines C-III 4.7b. This is not the case with claims 26-28, as at least claims 26, 27 are anticipated by D3, D4.

Inventive step - Article 56 EPC

The closest state of the art D4 discloses random integration of a linear fragment in the yeast species H. polymorpha. A slight (1.8) fold increase of transformants is obtained by the addition of a restriction endonuclease. The present application differs from D4 in that the restriction endonuclease is provided in vivo (i.e. circular DNA is introduced to the host cell). Furthermore, in D4 the addition of the restriction enzyme is associated with an increase of integration at restriction sites. Thus, with regard to random integration D4 teaches away from the solutions proposed in the present application. Thus, in summary the application provides an alternative method for random integration of a DNA construct. The method appears to be associated with the technical advantage of higher frequency of integration (see D5). In hindsight, the success of the method can be rationalized by I-Sce I protecting the ends of the linearized DNA from the action of cellular exonucleases/ligases. As there are no suggestions in the prior art that would prompt a person skilled in the art to develop the methods of the present application, an inventive step is acknowledged provided that novelty is restored.

Conclusion

To meet the requirements of Rule 27(1)(b) EPC, the document D1, D2, D4 should be identified in the description and the relevant background art disclosed therein should be briefly discussed. Any passages in the description relating to non-patentable subject-matter (see above) should be deleted or redrafted. The applicant is minded to delete phrases such as "incorporated by reference" and "spirit and scope" of the application if present.

The Applicant is requested to file new claims which take account of the above comments. When filing amended claims the Applicant should at the same time bring the description into conformity with the amended claims.

The attention of the applicant is drawn to the fact that the application may not be amended in such a way that it contains subject-matter which extends beyond the content of the application as filed (Article 123(2) EPC).

RANDOM INTEGRATION OF A POLYNUCLEOTIDE BY *IN VIVO* LINEARIZATION

BACKGROUND

Field of the invention

The invention concerns a process of *in vivo* liberation a linear DNA fragment from a vector in order to integrate this fragment into the cellular genome. The invention further concerns the use of this process and the results of this use.

Brief description of the prior art

The first breakthrough of reverse genetics 20 years ago was Transgenesis. The transgenesis is the technique which allows to introduce a exogenous DNA sequence into a host cell. For example, DNA micro-injection in an egg fertilized by mammal leads, in a certain number of cases, to the integration of micro-injected DNA molecule into the genome of the fertilized egg. Transgenesis implies that a foreign DNA fragment is introduced into the genome of a multicellular organism and transmitted to progeny. Therefore, the foreign DNA must be present in a stable form in the embryo at an early stage of development in order to be transmitted into progeny.

DNA transfection in mammal cells by means of precipitated calcium phosphate precipitation can also lead, in a certain number of cases, to the integration of the exogneous DNA to the genome of the cell host which is called stable transfection. The exogenous DNA can be introduced into cells under two forms, either linear or circular. When the DNA is introduced under linear form, the linear fragment is prepared *in vitro* before its intoduction into the host cell, generally by excision of the desired fragment with restriction enzyme from a plasmid for example. As regards the DNA under circular form, the introduced DNA is generally a supercoiled plasmid.

All cells have systems of maintenance and repair of their DNA. One of the particular signals of stress which activates the intervention of these systems is the generation of DNA free ends in the cell. The cell has then two solutions to resolve this type of problem:

-       The first solution is the degradation of the DNA presenting the free ends (for example the case where the cell should eliminate a DNA being able to have a viral activity). This solution is based on the presence in these systems of maintenance and repair of exonucleases, which degrade the DNA by digesting it from the free ends.

2

- An other solution is the recombination of the free ends with the cellular genome. This recombination can take different forms, notably the integration of the exogenous molecule into the genome of the cell.

As a consequence, the works aiming at integrating of the exogenous DNA by
5 introduction of naked double-stranded DNA under circular or linear form meet three types of major problems as well as certain number of collateral problems.

- The first one of these problems is the efficiency of the integration of the exogenous molecule in the DNA of the cell host. This lack of efficiency implies the injection of very hight amount of exogenous DNA for a very few number of integration. In
10 some species such as fish, plants or insects, the exonucleases are so efficient that the exogenous DNA is never integrated in the chromosome and it stays episomal. For example, if the circular form is used for the exogenous DNA, a nick is necessary for having integration. Indeed, the integration process needs the presence of a free end. If the linear form is used, most of the exogenous DNA is degraded by the exonucleases due to
15 the presence of the free ends.

- The second problem is the integrity of the integrated DNA. The linear fragment of exogenous DNA exposes its free ends to the cellular exonucleases before its arrival to the nucleus. This prolonged exhibition of its ends in the cell limits in a significant way the chances of the exogenous DNA to be wholly integrated into the cellular genome. A way of
20 increasing the chances to be wholly integrated into the genome consists in adding some cohesive single-stranded DNA overhangs at each ends of the exogenous DNA, for example made by digestion with the same restriction enzyme. These cohesive extremities allow the DNA fragments to associate in a multimer and to prevent the complete degradation of the DNA before its integration. Another way of increasing the chances is to
25 surround the DNA fragment to integrate with long and neutral DNA sequences. In the case of the exogenous DNA is included in a supercoiled plasmid, the intregrated fragment comprises not only the exogenous DNA but also the whole vector.

- The third problem is the control of the number of integrated copies. The multimerization, deliberated or not, of the exogenous DNA presents the inconvenience to
30 favor the insertion of several copies of the exogenous DNA fragment. Similarly, when the exogenous DNA has a circular form, the plasmid is integrated as a concatemer. This multiple insertion has for consequence to introduce a chromosome instability and to result in problems of regulation of the expression due to the presence of the same gene in multiple copies.

35   Transgenesis is, more than ever, an essentiel tool for biologists. The study of human diseases relies to a great extent on the use of animal models. Vast gene sequence informations are available for the pharmaceutical industry. These informations would

provide new targets for drug. However, the function of the majority of the genes and related proteins in an organism remains unknown and the efficiency of target validation does not appear to have significantly changed.

Genetic studies and structural genomics have shown that biochemical pathways and physiology are highly conserved throughout the animal world. Animal models can therefore be used to investigate processes relevant to human diseases. Animal model can be used to efficiently identify and validate optimal screening targets. Screening target selection can be improved and will lead to fewer failures and a more efficient drug development process.

10   Mouse is a well known mammalian model which is abundantly used. The most widely used method for the production of transgenic animals is the microinjection of DNA into the pronuclei of fertilized embryos. This method is rather efficient for the production of transgenic mice but is much less efficient for the production of large transgenic mammals such as cows and sheep. Moreover, the transgenic animals from the available

15   transgenesis method are often mosaic for the transgene resulting in the lack of transmission of the transgene to the progeny. Some animals present a high resistance to transgenesis such as fish or bird. Among them, the problem of fish transgenesis is more detailed below.

Tank fish have risen to high popularity as vertebrate models in developmental

20   biology and genetics. Indeed, the zebrafish (*Danio rerio*) is a popular model system for vertebrate developmental studies because it offers the opportunity to combine classical genetic analysis with an easily accessible and manupilable embryo. Genetic studies of the zebrafish benefit from the 2-3 month generation time, the ability of females to routinely lay hundreds of egg, and the small size of the adults. Embryological studies benefit from the

25   large, transparent embryos.

However, the usefulness of tank fish is still limited by the lack of some methodological tools, above all a simple and efficient technology for transgenesis, which has become a major technique in fundamental research and has varied applications in agronomy and biomedecine. In particular, attempts to establish embryonic stem (ES) cells

30   in fish as cellular vectors for transgenesis, have so far been unsuccessful.

Therefore, the method of choice to generate transgenic fish remains the injection of high concentrations of DNA (approximately 10^6 plasmid copies) in the cytoplasm of one cell-stage embryos. Plasmids have been microinjected in linearized and circular form, and in both types of experiments transgenesis has been achieved. This technology is fast and

35   easy, due to the transparency and great size of most fish eggs, but unfortunately, it is also rather inefficient, with a frequency of genomic integration and germline transmission which usually lies in a range of a few percents [Stuart et al., 1988 *Development* **103**, 403-412;

Stuart et al., 1990, *Development* **109**, 577-584; Culp et al., 1991, *Proc Natl Acad Sci USA*, **88**, 7953-57; Lin et al., 1994 *Dvelopmental biology* **161**, 77-83; Collas et al., 1998 *Transgenic Research*, **7**, 303-309]. Recent studies have shown that this is likely to be due to late and mosaic integrations: DNA persists in an unintegrated form in the egg cytoplasm and is inherited only by a subset of blastomeres. After injection, the plasmid sequences are transiently amplified and form long concatemers consisting of many unit-length copies of the plasmid arranged in tandem (Stuart et al. 1988, *supra*). The foreign DNA usually inserts into one site in the host genome but usually consists of tandem arrays of the original injected construct (Culp et al., 1991, *Supra*).

It is generally accepted that increasing the frequency of transgenic fish generated by plasmid microinjection is difficult. Injecting higher amounts of DNA is toxic to the embryos so one must attempt to improve the efficiency of integration. Several attempts to improve the rate of integration and transmission of transgenes have been performed. For example, the use of DNA-NLS complexes has been reported, although most authors found no improvements with this technology [Liang et al., 2000 *Mol Reprod Dev* **55**, 8-13]. In principle, technologies using flanking repeats of adeno-associated virus [Fu, 1998 *Nature Biotech*, **16**, 253-257] or of transposons [Izsvak et al., 2000 *J Mol Biol* **302**, 93-102] may also increase transgene integration. However, to the best knowledge of the inventors, positive results have not been not reported with these techniques, which also suffer from the potentially deleterious presence of repeats in the plasmids. Moreover, these vectors are limited for the size of the DNA sequences that can be engineered into them.

Therefore, numbers of methods have emerged to improve and develop new ways to increase the transgenesis efficiency. Nevertheless, a very limited amount of methods allows better control and numbers of species remain resistant or very inefficient for this technology. Main problems encountered are :

- Transgenesis control and efficiency

- Early integration events (germ line transmission)

- The control of integrated copy number.

SUMMARY

The present invention exactly aims at offering a method allowing to improve the efficiency of DNA's random integration during the transgenesis and to allow a better control of the integrity of the integrated DNA, as well as to reduce the number of copies of integrated DNA.

The method for random integrating an exogenous DNA into the genomic DNA according to the invention consists in linearization of the DNA polynucleotide to be integrated having no 5' and 3' free ends from a vector into the host cell.

The method according to the invention consists in a method for randomly integrating a polynucleotide into a host cell genome by (*in vivo* or *in ovo*) preparation into the host cell of said linear polynucleotide having free 5' and 3' ends, said method comprising:

a) Introducing into said host cell a vector having no free 5' and 3' ends and comprising the polynucleotide sequence to be linearized, said vector comprising at least one cleavage site and said cleavage site being found in the host cell genome at less than 5 copies, preferably 2 copies, and more preferably said cleavage site is not found in the host cell genome; and,

b) causing cleavage of said site(s) in said host cell, thereby creating or releasing said polynucleotide in a linear form having free 5' and 3' ends from said vector into said host cell; and,

c) maintaining the host cell under conditions and for a period of time sufficient to cause the random integration of said linearized polynucleotide into said host cell genome.

Optionally, said method further comprises, prior to step (b), an aditional step of introducing into said host cell a cleaving agent or a vector comprising a nucleic acid encoding said cleaving agent. Preferably, said method further comprises, prior to step (b), an aditional step of introducing into said host cell said cleaving agent. Optionally, said vector comprising a nucleic acid encoding said cleaving agent is an expression vector or a mRNA. Optionally, said host cell is a transgenic cell expressing said cleaving agent. Said polynucleotide sequence to be integrated according to the present invention is unable to undergo homologous recombination with the host cell genome. Optionally, said polynucleotide to be integrated have less than 70 % indentity with a host cell genomic sequence, preferably less than 60 or 50 %, more preferably less than 40, 30 or 20 % identity. Optionally, the 5' and 3' sequences of said polynucleotide to be integrated have no homology with a host cell genomic sequence, preferably less than 90 % indentity, more preferably less than 80 or 70 %, still more preferably less than 50, 40, 30 or 20 % identity, wherein said 5' and 3' sequences are 5 kb long, preferably between 3 kb and 1.5 kb long, more preferably 1 kb, 500 bp or 100 bp long. Preferably, the cleaved sites do not generate cohesive ends. Preferably, said polynucleotide sequence to be excised does not comprise any cleavage site. Preferably, said vector comprising said polynucleotide sequence to be linearized further comprises said nucleic acid encoding said cleaving agent. Preferably, said polynucleotide sequence is flanked by at least one cleavage site. Preferably, said polynucleotide sequence to be linearized is flanked by two cleavage sites. Preferably, said

6

cleavage site is an endonuclease site and said cleaving agent is the corresponding endonuclease. More preferably, said endonuclease has a recognition site of at least 12 nucleotides. Still more preferably, said endonuclease is a meganuclease, notably one meganuclease of Figure 2. Optionally, said meganuclease is selected from the group
5   consisting of I-Ceu I, I-Cre I, I-Chu I, I-Csm I, I-Dmo I, I-Pan I, I-Sce I, I-Sce II, I-Sce III, I-Sce IV, F-Sce I, F-Sce II, PI-Aae I, PI-Ape I, PI-Ceu I, PI-Cir I, PI-Ctr I, PI-Dra I, PI-Mav I, PI-Mfl I, PI-Mgo I, PI-Mja I, PI-Mka I, PI-Mle I, PI-Mtu I, PI-MtuH I, PI-Pab III, PI-Pfu I, PI-Pho I, PI-Pko I, PI-Psp I, PI-Rma I, PI-Sce I, PI-Ssp I, PI-Tfu I, PI-Tfu II, PI-Tli I, PI-Tli II, PI-Tsp I, PI-Tsp II, PI-Bsp I, PI-Mch I, PI-Mfa I, PI-Mga I, PI-Mga II, PI-Min I, PI-Mma I, PI-
10 Msh I, PI-Msm II, PI-Mth I, PI-Tag I, PI-Thy II, I-Ncr I, I-Ncr II, I-Pan II, I-Tev I, I-Ppo I, I-Dir I, I-Hmu I, I-Hmu II, I-Tev II, I-Tev III, F-Sce I, F-Sce II (HO), F-Suv I, F-Tev I, and F-Tev II. Preferably, said meganuclease is selected from the group consisting of I-Ceu I, I-Cre I, and I-Sce I. Optionally, said endonuclease is synthetic. Preferably, said vector is a double-stranded DNA vector. Optionally, said vector is a plasmid or a viral vector. Preferably, said
15 vector is a plasmid. Preferably, said polynucleotide sequence is a sequence encoding a polypeptide or an antisense, a regulatory sequence, or a recognition sequence for a molecule. Preferably, said host cell is selected from the group consisting of a stem cell, a somatic cell, a gamete, a blastomer and an egg. More preferably, said host cell is selected from the group consisting of a stem cell, a blastomer and an egg. Optionally, said method
20 for randomly integrating a polynucleotide into a host cell genome is used for stable transfection or transgenesis.

In one embodiment of the present invention, the method consists in a method for randomly integrating a polynucleotide into the host cell genome by (*in vivo* or *in ovo*) preparation into the host cell of said linear polynucleotide having free 5' and 3' ends from
25 a vector, said method comprising the following steps:

a) Introducing in said host cell a vector having no free 5' and 3' ends and comprising said polynucleotide sequence to be linearized or excised, said vector comprising at least one endonuclease site and said endonuclease site being found in the host cell genome at less than 5 copies, preferably 2 copies, and more preferably said
30 endonuclease site is not found in the host cell genome;

b) Optionally, introducing in said host cell either the endonuclease which cleaves said endonuclease site present in said vector or an expression vector comprising a nucleic acid encoding said endonuclease; and,

c) causing cleavage of said site in said host cell, thereby creating or releasing said
35 polynucleotide in a linear form having free 5' and 3' ends from said vector into said host cell; and,

7

d) maintaining the host cell under conditions and for a period of time sufficient to cause the random integration of said linearized or excised polynucleotide into said host cell genome.

Preferably, step b) consists in introducing in said host cell the endonuclease which 5 cleaves said endonuclease site. Preferably, steps a) and b) are simultaneous. Optionally, said vector comprising a nucleic acid encoding said endonuclease is an expression vector or a mRNA. Optionally, said host cell is a transgenic cell expressing said endonuclease. Said polynucleotide sequence to be integrated according to the present invention is unable to make significantly an homologous recombination with the host cell genome. 10 Optionally, said polynucleotide to be integrated have less than 70 % indentity with a host cell genomic sequence, preferably less than 60 or 50 %, more preferably less than 40, 30 or 20 % identity. Optionally, the 5' and 3' sequences of said polynucleotide to be integrated have no homology with a host cell genomic sequence, preferably less than 90 % indentity, more preferably less than 80 or 70 %, still more preferably less than 50, 40, 15 30 or 20 % identity, wherein said 5' and 3' sequences are 5 kb long, preferably between 3 kb and 1.5 kb long, more preferably 1 kb, 500 bp or 100 bp long. Preferably, the cleaved sites do not generate cohesive ends. Optionally said vector further comprises a nucleic acid sequence encoding the endonuclease. Optionally, said polynucleotide to be linearized or excised and said nucleic acid encoding said endonuclease are each 20 comprised by a distinct vector. Preferably, said polynucleotide to be linearized or excised does not comprise any endonuclease site. Preferably, said polynucleotide sequence is flanked by at least one endonuclease site. Preferably, said polynucleotide sequence to be linearized is flanked by two endonuclease sites. Preferably, said endonuclease has a recognition site of at least 12 nucleotides. Still more preferably, said endonuclease is a 25 meganuclease, notably one meganuclease of Figure 2. Optionally, said meganuclease is selected from the group consisting of I-Ceu I, I-Cre I, I-Chu I, I-Csm I, I-Dmo I, I-Pan I, I-Sce I, I-Sce II, I-Sce III, I-Sce IV, F-Sce I, F-Sce II, PI-Aae I, PI-Ape I, PI-Ceu I, PI-Cir I, PI-Ctr I, PI-Dra I, PI-Mav I, PI-Mfl I, PI-Mgo I, PI-Mja I, PI-Mka I, PI-Mle I, PI-Mtu I, PI-MtuH I, PI-Pab III, PI-Pfu I, PI-Pho I, PI-Pko I, PI-Psp I, PI-Rma I, PI-Sce I, PI-Ssp I, PI-30 Tfu I, PI-Tfu II, PI-Tli I, PI-Tli II, PI-Tsp I, PI-Tsp II, PI-Bsp I, PI-Mch I, PI-Mfa I, PI-Mga I, PI-Mga II, PI-Min I, PI-Mma I, PI-Msh I, PI-Msm II, PI-Mth I, PI-Tag I, PI-Thy II, I-Ncr I, I-Ncr II, I-Pan II, I-Tev I, I-Ppo I, I-Dir I, I-Hmu I, I-Hmu II, I-Tev II, I-Tev III, F-Sce I, F-Sce II (HO), F-Suv I, F-Tev I, and F-Tev II. Preferably, said meganuclease is selected from the group consisting of I-Ceu I, I-Cre I, and I-Sce I. Optionally, said endonuclease is synthetic. 35 Preferably, said vector is double-stranded. Optionally, said polynucleotide can comprise a sequence encoding a polypeptide or an antisense, a regulatory sequence such as promoter and enhancer, and/or a recognition sequence for a molecule. Preferably, said

8

host cell is selected from the group consisting of a stem cell, a somatic cell, a gamete, a blastomer and an egg. More preferably, said host cell is selected from the group consisting of a stem cell, a blastomer and an egg. Optionally, said method for randomly integrating a polynucleotide into a host cell genome is used for stable transfection or

5   transgenesis.

The present invention relates to a composition for transgenesis or for stable transfection comprising :

1) a vector having no 5' and 3' free ends and comprising a transgene to be randomly integrated, said vector comprising at least one cleavage site which is found in

10 the host cell genome at less than 5 copies, preferably 2 copies, and more preferably said cleavage site is not found in the host cell genome; and,

2) a cleaving agent or a vector comprising a nucleic acid encoding said cleaving agent.

Preferably, said composition is used for transgenesis. Preferably, said composition

15 comprises the cleavage agent. Optionally, said vector comprising a nucleic acid encoding said cleaving agent is an expression vector or a mRNA. Said transgene to be integrated according to the present invention is unable to make significantly an homologous recombination with the host cell genome. Optionally, said transgene to be integrated have less than 70 % indentity with a host cell genomic sequence, preferably less than 60 or 50

20 %, more preferably less than 40, 30 or 20 % identity. Optionally, the 5' and 3' sequences of said transgene to be integrated have no homology with a host cell genomic sequence, preferably less than 90 % indentity, more preferably less than 80 or 70 %, still more preferably less than 50, 40, 30 or 20 % identity, wherein said 5' and 3' sequences are 5 kb long, preferably between 3 kb and 1.5 kb long, more preferably 1 kb, 500 bp or 100 bp

25 long. Preferably, the cleaved sites do not generate cohesive ends. Preferably, said transgene does not comprise any cleavage site. Optionally said vector comprising said transgene further comprises a nucleic acid sequence encoding the cleaving agent. Optionally, said transgene and said nucleic acid encoding said cleaving agent are each comprised by a distinct vector. Preferably, said transgene is flanked by at least one

30 cleavage site. Preferably, said transgene is flanked by two cleavage sites. Preferably, said cleavage site is an endonuclease site and said cleaving agent is the corresponding endonuclease. Preferably, said endonuclease has a recognition site of at least 12 nucleotides. Still more preferably, said endonuclease is a meganuclease, notably one meganuclease of Figure 2. Optionally, said meganuclease is selected from the group

35 consisting of I-Ceu I, I-Cre I, I-Chu I, I-Csm I, I-Dmo I, I-Pan I, I-Sce I, I-Sce II, I-Sce III, I-Sce IV, F-Sce I, F-Sce II, PI-Aae I, PI-Ape I, PI-Ceu I, PI-Cir I, PI-Ctr I, PI-Dra I, PI-Mav I, PI-Mfl I, PI-Mgo I, PI-Mja I, PI-Mka I, PI-Mle I, PI-Mtu I, PI-MtuH I, PI-Pab III, PI-Pfu I, PI-

Pho I, PI-Pko I, PI-Psp I, PI-Rma I, PI-Sce I, PI-Ssp I, PI-Tfu I, PI-Tfu II, PI-Tli I, PI-Tli II, PI-Tsp I, PI-Tsp II, PI-Bsp I, PI-Mch I, PI-Mfa I, PI-Mga I, PI-Mga II, PI-Min I, PI-Mma I, PI-Msh I, PI-Msm II, PI-Mth I, PI-Tag I, PI-Thy II, I-Ncr I, I-Ncr II, I-Pan II, I-Tev I, I-Ppo I, I-Dir I, I-Hmu I, I-Hmu II, I-Tev II, I-Tev III, F-Sce I, F-Sce II (HO), F-Suv I, F-Tev I, and F-Tev II.

5   Preferably, said meganuclease is selected from the group consisting of I-Ceu I, I-Cre I, and I-Sce I. Optionally, said endonuclease is synthetic. Preferably, said vector is double-stranded. Preferably, said vector is a plasmid. Optionally, said transgene can comprise a sequence encoding a polypeptide or an antisense, a regulatory sequence such as promoter and enhancer, and/or a recognition sequence for a molecule.

10   The invention relates to the use of the composition according to the present invention for producing transgenic cells, non-human animals or plants. Preferably, said non-human animals are selected from non human mammals, birds, reptiles, amphibians and fishes. For example, the invention contemplates cattle (cows), goats, rabbit, rodents, marmots, monkeys, insects (spider, butterflies, fly), fishes, calmar, amphoxius, xenopes,

15   birds, chickens, ascidies and ovine races (sheeps). More particularly, the invention contemplates fishes such as sticklebass, astyanax, medaka and zebrafish, birds like chickens and rodents such as mice.

The present invention also relates to the resulting cells from any method of *in vivo* or *in ovo* linearization of a polynucleotide and of random polynucleotide integration

20   according to the present invention and, their uses, for example for production of proteins or other genes, biomolecules, biomaterials, transgenic plants, vaccines, transgenic plants and animals or for treatment or prophylaxis of a condition or disorder in an individual. More particularly, the invention concerns any non-human transgenic animal and any transgenic plant comprising a cell resulting from any method of *in vivo* linearization of a

25   polynucleotide and of random polynucleotide integration according to the present invention. The invention also relates to any use of a cell, of a non-human transgenic animal or of a transgenic plant according to the present invention for production of protein, antisense, biomolecule, biomaterial, or vaccine. The invention further relates to the use of any resulting cell for treatment or prophylaxis of a condition or disorder in an individual.

30   Moreover, the invention concerns a method of treating or prophylaxis of a genetic disease in an individual in need thereof by random integration of a polynucleotide comprising the steps of :

a) Introducing in said individual cell a vector having no free 5' and 3' ends and comprising said polynucleotide, said vector comprising at least one cleavage site which is

35   found in the individual cell genome at less than 5 copies, preferably 2 copies, and more preferably said cleavage site is not found in the host cell genome;

b) causing cleavage of said site(s) in said individual cell, thereby creating or releasing said polynucleotide in a linear form having free 5' and 3' ends from said vector into said individual cell; and,

c) maintaining the host cell under conditions and for a period of time sufficient to
5 cause the random integration of said linearized or excised polynucleotide into said individual cell genome; said random integration of said polynucleotide compensates the genetic defect causing said genetic disease.

Optionally, said method further comprises, prior to step (b), an aditional step of introducing into said individual cell a cleaving agent or a vector comprising a nucleic acid
10 encoding said cleaving agent. Said polynucleotide sequence to be integrated according to the present invention is unable to make significantly an homologous recombination with the host cell genome. Optionally, said polynucleotide to be integrated have less than 70 % indentity with a host cell genomic sequence, preferably less than 60 or 50 %, more preferably less than 40, 30 or 20 % identity. Optionally, the 5' and 3' sequences of said
15 polynucleotide to be integrated have no homology with a host cell genomic sequence, preferably less than 90 % indentity, more preferably less than 80 or 70 %, still more preferably less than 50, 40, 30 or 20 % identity, wherein said 5' and 3' sequences are 5 kb long, preferably between 3 kb and 1.5 kb long, more preferably 1 kb, 500 bp or 100 bp long. Preferably, the cleaved sites do not generate cohesive ends. Preferably, said
20 polynucleotide sequence to be excised does not comprise any cleavage site. Preferably, said vector comprising said polynucleotide sequence to be linearized or excised further comprises said nucleic acid encoding said cleaving agent. Preferably, said polynucleotide sequence to be linearized or excised is flanked by at least one cleavage site. Preferably, said polynucleotide sequence to be linearized or excised is flanked by two cleavage sites.
25 Preferably, said cleavage site is an endonuclease site and said cleaving agent is the corresponding endonuclease. More preferably, said endonuclease has a recognition site of at least 12 nucleotides. Still more preferably, said endonuclease is a meganuclease, notably one meganuclease of Figure 2. Optionally, said meganuclease is selected from the group consisting of I-Ceu I, I-Cre I, I-Chu I, I-Csm I, I-Dmo I, I-Pan I, I-Sce I, I-Sce II,
30 I-Sce III, I-Sce IV, F-Sce I, F-Sce II, PI-Aae I, PI-Ape I, PI-Ceu I, PI-Cir I, PI-Ctr I, PI-Dra I, PI-Mav I, PI-Mfl I, PI-Mgo I, PI-Mja I, PI-Mka I, PI-Mle I, PI-Mtu I, PI-MtuH I, PI-Pab III, PI-Pfu I, PI-Pho I, PI-Pko I, PI-Psp I, PI-Rma I, PI-Sce I, PI-Ssp I, PI-Tfu I, PI-Tfu II, PI-Tli I, PI-Tli II, PI-Tsp I, PI-Tsp II, PI-Bsp I, PI-Mch I, PI-Mfa I, PI-Mga I, PI-Mga II, PI-Min I, PI-Mma I, PI-Msh I, PI-Msm II, PI-Mth I, PI-Tag I, PI-Thy II, I-Ncr I, I-Ncr II, I-Pan II, I-Tev I, I-
35 Ppo I, I-Dir I, I-Hmu I, I-Hmu II, I-Tev II, I-Tev III, F-Sce I, F-Sce II (HO), F-Suv I, F-Tev I, and F-Tev II. Preferably, said meganuclease is selected from the group consisting of I-Ceu I, I-Cre I, and I-Sce I. Optionally, said endonuclease is synthetic. Preferably, said

vector is a double-stranded DNA vector. Optionally, said vector is a plasmid or a viral vector. Preferably, said vector is a plasmid. Preferably, said polynucleotide sequence is a sequence encoding a polypeptide or an antisense, a regulatory sequence, or a recognition sequence for a molecule. Preferably, said individual cell is a stem cell or a somatic cell.

5   The invention relates to a method for producing a transgenic animal. More particularly, the invention concerns a method for producing a non-human transgenic animal, wherein embryonic stem cells are tranfected by the method according to the present invention and screened for random integration event, the cells are injected into embryos at a stage at which they are capable of integrating the transfected cells, for
10 example at the blastocyst stage, the embryos are then reimplanted in a surrogate mother, and the chimeric individuals obtained at the end of gestation, and in which colonization by embryonic stem cells of the germline is observed are mated to obtain transgenic animals. Otherwise, the invention concerns a method for producing a non-human transgenic animal, wherein fertilized egg are tranfected by the method according to the present
15 invention, and either the eggs are reimplanted in a surrogate mother and the transgenic individuals obtained at the end of gestation or the eggs are incubated in condition appropriate for the developping of the transgenic animal.

BRIEF DESCRIPTION OF THE DRAWINGS

20   **Figure 1** is a schematic diagram representing one embodiment of the method for integrating a transgene, said transgene being *in vivo* released from a supercoiled plasmid vector. The transgene is flanked by two cleavage sites. The cleavage agent is added and makes *in* vivo a double-stranded break at the two cleavage sites. The linear transgene is released and one copy of the transgene is integrated into the genome.

25   **Figure 2** is a table disclosing the known meganuclease. Th means theorical, Exp experimental demonstration, and Pot potential.

        Figure 3 is a schematic diagram representing some ways for the preparation of linear vector according to the present invention.

        Figure 4 is a schematic diagram of the DNA constructs described in the Example
30 1. "I-Sce I" refers to I-Sce I endonuclease recognition and cleavage site. "IR" refers to invert repeat. "pA$_{SV40}$" refers to the polyadenylation signal of SV40.

        Figure 5 is a schematic diagram disclosing the validation system used in Example 1. "I-Sce I", "IR" , "pA$_{SV40}$" respectively refer to I-Sce I endonuclease recognition and cleavage site, invert repeat, and the polyadenylation signal of SV40.

35   **Figure 6** is a half-tome reproduction of an electrophoresis gel from the southern analysis after a *EcoR* V digestion of the transfected cell genome with both pVirtkU3Rβgeo

and pCMV-I Scel and a diagram describing the obtained restriction fragment. +IN means that the cells were also transfected with a plasmid encoding an integrase. If one copy of the fragment $\Delta U_3 R\beta geopA_{SV40}$ is integrated in the cell genome, the Southern analysis shows one band at 1.2 kb and another one of at least 2.5 kb. If the fragment $\Delta U_3 R\beta geopA_{SV40}$ is integrated in the cell genome with at least two consecutive copies, the Southern analysis shows at least one band at 1.2 kb and one band at 2.5 kb.

Figure 7 shows the results of the stable transfection on COS cells with the in vivo linearization method.

Figure 7A is a diagram illustrating the structure of the ppSNIE construct. HindIII and I-Sce I cleavage sites (I-ScelCS) are indicated as well as the location of the NEO probe (Probe).

Figure 7B is a half-tone reproduction of a Southern Blot analysis. The left blot corresponds to a co-transfection with the plasmids ppSNIE and pCMV-I-Scel (-). The right blot corresponds to a co-transfection with the plamsids ppSNIE and pCMV-I-Scel (+).

Figure 8 shows the transient expression of GFP in α-actin –GFP injected embryos.

Figure 8A is a diagram illustrating the improvement of GFP transient expression when the meganuclease is added to the injection buffer. Right panels: dechorionated embryos expressing GFP at different levels. For routine observations, chorions are not removed. S: strong; M: moderate; W: weak.

Figure 8B is a table summarising the frequencies of the different patterns of GFP muscular expression grouped as shown in Figure 7A. N (negative), W (weak), M (moderate) and strong (S).

Figure 9 shows the efficiency of transgenesis in medaka.

Figure 9A is a table showing the efficiency of transgenesis in medaka following injection of p□act-GFPM2. Lane 1: linear fragment corresponding to the α-actin promoter and the GFP reporter gene, obtained by in vitro I-Scel digestion and agarose gel purification.

Figure 9B is a half-tone reproduction of a Southern blot analysis of insert structures in p□act-GFPM2 lines injected with (lines 1 to 4) or without (lane C) the meganuclease. P: control p□act-GFPM2 digested with BamHI. □-actin promoter: grey box. EGFP, polyA: white box. The restriction enzyme used in the represented filter is BamHI. The probe used was obtained with XhoI and EcoRV digests and corresponds to the actin promoter and the GFP reporter gene. On Southerns, two fragments of 1 and 2kb corresponding respectively to the GFP/pA and to the downstream region of the actin promoter (3'□p) are observed in lane 1 to 4. 5'□p+pBluSK: 4,8kb band, diagnostic of the upstream region of the promoter linked with the plasmid sequence.

Figure 10 is a schematic diagram of the plasmid constructs used for the meganuclease mediated transgenesis by in ovo linearization of the transgene by an I-SceI expressing vector.

Figure 10A is a diagram of the Megafluo plasmid construct. The arrows I-Sce I
5 indicate the location of the I-Sce I cleavage site. A strong promoter (promoter of the mouse Gas 5 gene) is directing the transcription of a fluorescent reporter (the DSRed1-E5 from Clontech). Two I-Sce I recognition/cleavage sites in the same orientation are flanking the transgene subcloned in a derivate of pUC vector.

Figure 10B is a diagram of the I-Sce I expressing vector (pI-SceI/EGFP). The I-Sce
10 I protein coding sequence was subcloned into the MCS of pIRES2-EGFP (Clontech). The CMV promoter is driving the transcription from a single RNA of both I-Sce I protein and the EGFP fluorescent reporter as a bicistronic expression is produced.

Figure 11 shows the results of the meganuclease mediated transgenesis by in ovo linearization of the transgene by an I-SceI expressing vector.

15   Figure 11A is a half-tone reproduction of an agarose gel showing the genomic DNA extraction. Lines 1, 25, 2, 3, 4 are genomic DNA extraction of three-weeks old mice. Lines A and B are genomic DNA extraction of stillborn mice. About 500 ng of DBA were loaded on a 1 % TAE agarose gel. Genomic DNA from stillborn animals (A and B) are degraded as expected from dead animal but DNA can still be used for a PCR analysis.

20   Figure 11B is a half-tone reproduction of an agarose gel showing the PCR amplification of a mouse β-globin gene fragment and of the reporter sequence of Megafluo (Red5). Left Panel: all genomic DNA samples are suitable for PCR amplification as the 494bp PCR product expected for the ß-globin gene is detected. Right Panel : animals 1 and 25 and the stillborn aimal A are transgenics as the 484bp fragment of the
25 Megafluo reporter gene (DSRedE1-5) are detected. Lines 1, 25, 2, 3, 4 are genomic DNA extraction of three-weeks old mice. Lines A and B are genomic DNA extraction of stillborn mice. Negative control (-) corresponds to the pI-SceI/EGFP plasmid. Positive control (+) corresponds to megafluo plasmid. Wt corresponds to genomic DNA extracted from B6SJL uninjected mice. C corresponds PCR contamination control performed without DNA.

30   Figure 11C is a half-tone reproduction of an agarose gel showing the PCR amplification of a sub-fragment of the pI-Sce I/EGFP. Positive control (+) corresponds to pI-SceI/EGFP plasmid. Wt corresponds to genomic DNA extracted from B6SJL uninjected mice. C corresponds PCR contamination control performed without DNA..

Figure 12 shows the results of the "classical transgenesis". Figure 12 is a half-tone
35 reproduction of an agarose gel showing the PCR amplification of the reporter sequence of Megaflu (Red5). Lines 14 to 21 correspond to the 8 newborn.

14

Figure **13** is a schematic diagram of the plasmid constructs used for the meganuclease mediated transgenesis by in ovo linearization of the transgene by an PI-SceI expressing vector. The Gas5 promoter drives the expression of the reporter gene LagoZ (a nearly CPG-free LacZ sequence). These two sequences are flanked by two PI-
5 SceI recognition/cleaving sites on the same orientation separating the backbone sequences of the vector. A single recognition/cleaving I-SceI site is present in between the promoter and the reporter gene.

Figure **14** shows the results of the meganuclease mediated transgenesis by in ovo linearization of the transgene by I-Sce I or PI-Sce I protein with the PIFF-Lago construct.
10 Figure 14 is a half-tone reproduction of an agarose gel showing the PCR results of a fragment of the reporter gene (LagoZ). Positive control corresponds to PIFFLago plasmid. Wt coresponds to genomic DNA from B6SJL uninjected mouse. C corresponds to PCR contamination control performed without DNA.

Left Panel : PCR genotyping on the genomic DNAs of the nine 12dpi-embryos
15 injected with PIFFLago plasmid plus I-Scel protein is performed. The 370bp PCR product of the Lago reporter is detected in four out nine embryos (number 101,102,103,105).

Right panel : PCR genotyping on the genomic DNAs of the nine 14dpi-embryos injected with PIFFLago plasmid plus PI-Scel protein is performed. The 370bp PCR product of the Lago reporter is detected in three out ten embryos (number 111,112,114).

20   Figure **15** shows the results of the meganuclease mediated transgenesis by in ovo linearization of the transgene by I-Sce I protein with the PIFF-Lago construct. Figure 14 is a half-tone reproduction of a Southern blot analysis.

Genotyping of the nine 12dpi-embryos injected with PIFFLago plasmid plus I-Scel protein is performed by Southern Blot experiments. Genomic DNAs were digested with
25 EcoRI restriction enzyme. Two different digoxigenine-labelled probes were used ; one hybridizing to the Gas5 sequences and the second to the Lago reporter gene. Detection is performed with the chemioluminescent CDP-Star (Roche) substrate on hyperfilm ECL films (Amersham). A 3.1kbp fragment is detected with the Gas5 probe and corresponds to the endogenous mouse Gas5 gene. Animals 105, 103 and 101 are transgenics as an
30 additional band is detected. This fragment is as well detected when the blot is deshybridized and reprobed with the Lago probe (right panel).

DETAILED DESCRIPTION

Definitions

35   By « transgenesis » is intended the introduction of new DNA sequences into the genome, preferably resulting in the production of transgenic animals or plants.

As used interchangeably herein, the terms "nucleic acid" "oligonucleotide", and "polynucleotide" include RNA, DNA, or RNA/DNA hybrid sequences of more than one nucleotide in either single chain or duplex form. The term "polynucleotide" refers to a polymer of units comprising a purine or pyrimidine, a ribose or deoxyribose sugar moiety,

5   and a phosphate group, or phosphodiester linkage. "polynucleotides" also refers to polynucleotide comprising "modified nucleotides" which comprise at least one of the following modifications (a) an alternative linking group, (b) an analogous form of purine, (c) an analogous form of pyrimidine, or (d) an analogous sugar.

By « in vivo » is intended in a cell, said cell being isolated or comprised in an

10 organism. The term « in vivo » encompasses « in ovo ». By « in ovo » is intended in an egg.

Flanked : A polynucleotide to be linearized or excised is flanked by a cleavage site if such a site is present at or near either or both ends of the polynucleotide. There can be one cleavage site present or near one end of the polynucleotide to be linearized or

15 excised or there can be two cleavage sites, one at or near each end of the polynucleotide to be linearized or excised. By "near" is preferably intended in the present invention that the cleavage site is located at less than 1 kb, preferably less than 500 bp, more preferably less than 200, or 100 bp, of the end of the polynucleotide to be integrated.

By cleavage, cleaving is intended in the present invention the formation of a DNA

20 double-stranded break. By "cleaving agent" is intended the agent able to cleave the "cleavage site".

By "endonuclease" is intended an enzyme which makes a double-stranded break in the DNA molecule at highly specific locations. This endonuclease can be natural. Preferably the enzyme is a homing endonuclease or a meganuclease. This endonuclease

25 can also be synthetic.

By "biomolecule" is intended in the present invention any organic molecule that is an essential part of a living organism such as polypeptide, protein, DNA or RNA polynucleotide.

By "free ends" is intended blunt ends and 5' or 3' overhangs available for

30 exonuclease degradation. Therefore, in the case of a linear molecule, any modification of the 5' and 3' abolishing or significatively decreasing the exonuclease degradation will not be considered as free ends. For example, a linear polynucleotide comprising secondary structures or modified nucleotides at its extremities confering a exonuclease resistance is not considered as having free ends.

35   "Cells," or "host cells", are terms used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding

generations due to either mutation or environmental influences, such progeny may not be identical to the parent cell, but are still included within the scope of the term as used herein. The cell can be a stem cell (preferably an embryonic stem cell), a somatic cell, a gamete, a blastomer and an egg (preferably a fertilized egg).

5   By « exogenous polynucleotide » is intended a polynucleotide without any similarity with the host cell chromosome. By « no similarity » is intended less than 50 % identity, preferably 40 or 30 % identity, more preferably less than 20 % identity with one sequence of the host cell chromosome. The polynucleotide similarity is so low that the polynucleotide is unable to make homologous recombination with the host cell 10 chromosome.

"Identity" refers to sequence identity between two nucleic acid molecules. Identity can be determined by comparing a position in each sequence which may be aligned for purposes of comparison. When a position in the compared sequence is occupied by the same base, then the molecules are identical at that position. A degree of similarity or 15 identity between nucleic acid sequences is a function of the number of identical or matching nucleotides at positions shared by the nucleic acid sequences. Various alignment algorithms and/or programs may be used to calculate the identity between two sequences, including FASTA, or BLAST which are available as a part of the GCG sequence analysis package (University of Wisconsin, Madison, Wis.), and can be used 20 with, e.g., default settings.

As used herein, the term "transgene" means a nucleic acid sequence (or an antisense transcript thereto) which has been introduced into a cell genome. A transgene could be partly or entirely heterologous, i.e., foreign, to the transgenic animal/plant or cell into which it is introduced; or, is homologous to an endogenous gene of the transgenic 25 animal or cell into which it is introduced, but it is inserted into the animal's genome at a location which differs from that of the natural gene. A transgene can include one or more transcriptional regulatory sequences and any other nucleic acid, such as introns, that may be necessary for optimal expression of a selected nucleic acid. Throughout the present invention, the polynucleotide to be linearized or excised and integrated is a transgene.

30   The "non-human animals" or "transgenic animals" of the invention include in, but not restrictive to, mammalians such as rodents, non-human primates, sheep, dog, cow, chickens, amphibians, reptiles, fishes, ascidies. Preferred non-human animals are selected from the rodent family including rat and mouse, most preferably mouse, though transgenic amphibians, such as members of the Xenopus genus, and transgenic 35 chickens, cow, sheep, and fishes can also provide important tools.

A "transgenic animal" or "transgenic plant" refers to any animal or plant, in which one or more of the cells of the animal or plant contain a transgene introduced by way of

human intervention, such as by transgenic techniques well known in the art. The transgene is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include

5 classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. This molecule is integrated within a chromosome. Moreover, "transgenic animal" or "transgenic plant" also includes those recombinant animals or plants in which gene disruption of one or more genes is caused by human intervention, including antisense technique. By transgenic animal is also intended

10 transgenic embryo.

The term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of preferred vector is an episome, i.e., a nucleic acid capable of extra-chromosomal replication. Preferred vectors are those capable of autonomous replication and/or expression of nucleic acids to which they are

15 linked. Vectors capable of directing the expression of genes to which they are operatively linked are referred to herein as "expression vectors". A vector according to the present invention comprises, but is not limited to, a YAC (yeast artificial chromosome), a BAC (bacterial artificial), a baculovirus vector, a phage, a phagemid, a cosmid, a viral vector, a plasmid, a RNA vector or a linear or circular DNA or RNA molecule which may consist of a

20 chromosomal, non chromosomal, semi-synthetic or synthetic DNA. In general, expression vectors of utility in recombinant DNA techniques are often in the form of "plasmids" which refer generally to circular double stranded DNA loops which, in their vector form are not bound to the chromosome. Large numbers of suitable vectors are known to those of skill in the art and commercially available, such as the following bacterial vectors: pQE70,

25 pQE60. pQE-9 (Qiagen), pbs, pDIO, phagescript, psiXI74. pbluescript SK. pbsks. pNH8A. pNH16A, pNH18A, pNH46A (Stratagene) ; ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia); pWLNEO. pSV2CAT, pOG44, pXT1, pSG (Stratagene); pSVK3, pBPV, pMSG, pSVL (Pharmacia); pQE-30 (QIAexpress).

Viral vectors include retrovirus, adenovirus, parvovirus (e. g., adenoassociated

30 viruses), coronavirus, negative strand RNA viruses such as orthomyxovirus (e. g., influenza virus), rhabdovirus (e. g., rabies and vesicular stomatitis virus), paramyxovirus (e. g. measles and Sendai), positive strand RNA viruses such as picornavirus and alphavirus, and double stranded DNA viruses including adenovirus, herpesvirus (e. g., Herpes Simplex virus types 1 and 2, Epstein-Barr virus, cytomegalovirus), and poxvirus

35 (e. g., vaccinia, fowlpox and canarypox). Other viruses include Norwalk virus, togavirus, flavivirus, reoviruses, papovavirus, hepadnavirus, and hepatitis virus, for example. Examples of retroviruses include: avian leukosis-sarcoma, mammalian C-type, B-type

viruses, Dtype viruses, HTLV-BLV group, lentivirus, spumavirus (Coffin, J. M., Retroviridae: The viruses and their replication, In Fundamental Virology, Third Edition, B. N. Fields, et al., Eds., Lippincott-Raven Publishers, Philadelphia, 1996). Other examples include murine leukemia viruses, murine sarcoma viruses, mouse mammary tumor virus,
5   bovine leukemia virus, feline leukemia virus, feline sarcoma virus, avian leukemia virus, human T-cell leukemia virus, baboon endogenous virus, Gibbon ape leukemia virus, Mason Pfizer monkey virus, simian immunodeficiency virus, simian sarcoma virus, Rous sarcoma virus and lentiviruses. Other examples of vectors are described, for example, in McVey et al., U. S5,801,030, the teachings of which are incorporated herein by reference.
10   Vectors according the present invention can comprise transposon (Ivicz et al. 1997, *Cell*, **91** , 501-510 ; Raz et al, 1998, *Current Biology*, **8**, 82-88 ; the disclosure of which is incorporated herein by reference).

Vectors can comprise selectable markers (for example, neomycin phosphotransferase, histidinol dehydrogenase, dihydrofolate reductase, hygromycin
15 phosphotransferase, herpes simplex virus thymidine kinase, adenosine deaminase, glutamine synthetase, and hypoxanthine-guanine phosphoribosyl transferase for eukaryotic cell culture ; TRP1 for S. cerevisiae; tetracycline, rifampicin or ampicillin resistance in E. coli; etc...). However, the invention is intended to include such other forms of expression vectors which serve equivalent functions and which become known in
20 the art subsequently hereto.

General overview of the invention

The present invention exactly aims at offering a method allowing to improve the efficiency of DNA's random integration during the transgenesis or the stable transfection
25 and to allow a better control of the integrity of the integrated DNA, as well as to reduce the number of copies of integrated DNA. The more interesting application of the present invention is the transgenesis, more particularly in species in which the transgenesis process are unefficient.

The method according to the invention consists in (*in vivo*) linearization or excision
30 into the host cell of the DNA polynucleotide to be integrated in the host cell genome (see Figure 1). The invention concerns a method for randomly integrating an exogenous DNA into the genomic DNA of the host cell wherein the polynucleotide to be integrated is linearized or excised *in vivo* from a vector. The *in vivo* linearization or excision of the polynucleotide to be integrated allows to generate a linear fragment with free ends available for recombining with the genomic DNA of the host cell. According to this method,
35 said vector has no free ends allowing the protection of the transgene against the

degradation. The method according to the present invention does not use the homologous recombination. The polynucleotide integration is random and not targeted.

Therefore, the invention concerns a method for *in vivo* generation of a linear polynucleotide with 5' and 3' free ends, corresponding essentially to the transgene, from
5   polynucleotide having no free end, said linear polynucleotide being randomly integrated into the host cell genome. In a particular embodiment of this method, said polynucleotide is comprised by the vector having a cleavage site, wherein said cleavage site is found in the host cell genome at less than 5 copies, preferably 2 copies, and more preferably said cleavage site is not found in the host cell genome. In a preferred embodiment of this
10 method, said polynucleotide is comprised by the vector and is flanked by a cleavage site, wherein said cleavage site is found in the host cell genome at less than 5 copies, preferably 2 copies, and more preferably said cleavage site is not found in the host cell genome. Preferably, the cleaved site does not generate cohesive ends. Preferably, said polynucleotide to be *in vivo* linearized or excised does not comprise any cleavage site.
15 Optionally said vector further comprises a nucleic acid sequence encoding the cleaving agent. Preferably, said cleavage site is a endonuclease site and said cleaving agent is the corresponding endonuclease. Optionally, the nucleic acid encoding the cleaving agent is comprised by an expression vector. Optionally, the nucleic acid encoding the cleaving agent is a mRNA. Optionally, said polynucleotide can comprise a sequence encoding a
20 polypeptide or an antisense, a regulatory sequence such as promoter and enhancer, and/or a recognition sequence for a molecule.

The method according to the invention consists in a method for randomly integrating a polynucleotide into the host cell genome by *in vivo* preparation said linear
25 polynucleotide having free 5' and 3' ends from a vector, said method comprising the following steps:

a) Introducing in said host cell a vector having no free 5' and 3' ends and comprising said polynucleotide, said vector comprising at least one cleavage site which is found in the host cell genome at less than 5 copies, preferably 2 copies, and more
30 preferably said cleavage site is not found in the host cell genome;

b) causing cleavage of said site(s) in said host cell, thereby creating or releasing said polynucleotide in a linear form from said vector having free 5' and 3' ends into said host cell; and,

c) maintaining the host cell under conditions and for a period of time sufficient to
35 cause the random integration of said linearized or excised polynucleotide into said host cell genome.

20

Preferably, the method according to the invention consists in a method for randomly integrating a polynucleotide into the host cell genome by *in vivo* preparation said linear polynucleotide having free 5' and 3' ends from a vector, said method comprising the following steps:

5   a) Introducing in said host cell a vector having no free 5' and 3' ends and comprising said polynucleotide, said polynucleotide being flanked at least one cleavage site which is found in the host cell genome at less than 5 copies, preferably 2 copies, and more preferably said cleavage site is not found in the host cell genome;

b) causing cleavage of said site(s) in said host cell, thereby creating or releasing 10 said polynucleotide in a linear form from said vector having free 5' and 3' ends into said host cell; and,

c) maintaining the host cell under conditions and for a period of time sufficient to cause the random integration of said linearized or excised polynucleotide into said host cell genome.

15   In one embodiment of the prensent invention, the method for randomly integrating a polynucleotide into the host cell genome by *in vivo* preparation of said linear polynucleotide having free 5' and 3' ends from a vector comprises the following steps:

a) Introducing in said host cell a vector having no free 5' and 3' ends and comprising said polynucleotide, said polynucleotide being flanked by at least one 20 cleavage site which is found in the host cell genome at less than 5 copies;

b) Introducing the cleaving agent or a vector encoding said cleaving agent; and,

c) causing cleavage of said site(s) in said host cell, thereby creating or releasing said polynucleotide in a linear form from said vector having free 5' and 3' ends into said host cell; and,

25   d) maintaining the host cell under conditions and for a period of time sufficient to cause the random integration of said linearzed or excised polynucleotide into said host cell genome.

In a further embodiment of the present invention, the method for randomly integrating a polynucleotide into the host cell genome by *in vivo* preparation of said linear 30 polynucleotide having free 5' and 3' ends from a vector comprises the following steps:

a) Introducing in said host cell :

- a vector having no free 5' and 3' ends and comprising said polynucleotide, said polynucleotide being flanked by at least one cleavage site which is found in the host cell genome at less than 5 copies; and,

35   - the cleaving agent;

b) causing cleavage of said site(s) in said host cell, thereby creating or releasing said polynucleotide in a linear form from said vector having free 5' and 3' ends into said host cell; and,

c) maintaining the host cell under conditions and for a period of time sufficient to
5   cause the random integration of said linearzed or excised polynucleotide into said host cell genome.

The process of the invention is remarkable because the release of the polynucleotide to be integrated can only be done *in vivo* after the crossing of the plasmidic membrane of the host cell, presumably within the nucleus. After the crossing of the
10 plasmidic membrane, the release of the ends of the polynucleotide, preferably by the endonuclease, can be achieved at any moment or in any compartment of the host cell.

Surprisingly, the integrity of the polynucleotide to be integrated is maintained, the efficiency of the transgenesis is significantly increased compared to either circular DNA or *in vitro* linearized DNA and the polynucleotide is integrated at a very low copy number. For
15 mice, the rate of transgenesis is increased by 3 to 5 times with the method according to the present invention. Moreover, the integrity of the transgene is excellent and the transgene is often integrated in the cell genome at less than three copies, more frequently only one copy. By conventional methods, at least 10 to 20 copies are integrated as a concantemer and, if only one copy is integrated, which is very rare event, the integrity of
20 the transgene is not maintained.

More particularly, the present disclosure shows that meganuclease-mediated transgenesis is indeed a very simple technique that is spectacularly more efficient than the other methods currently reported in fish. The method according to the present invention allows to efficiently integrate DNA in the fish genome, more particularly the
25 medakafish *(Oryzias latipes)* or zebrafish (*Danio rerio*) genome (see Example 2). The expression of the transgene is spectacularly improved. Even more striking is the dramatic increase in germline transmission. Whereas in classical egg injection experiments, it does not, in most cases, exceed a few percents, due to late chimeric integrations, the frequency of germline transmission in fish co-injected with the meganuclease was boosted to 50%,
30 suggesting that a single insertion occurred at one-cell stage of the founder fish. Moreover, single low copy integrations occur in most cases. Therefore, meganuclease-induced *in ovo* linearisation, by limiting degradation of preintegrative DNA free ends, is a simple and efficient process to improve transgenesis by egg injection. Moreover, the generation of polynucleotide free-ends in cell may created recombination foci which could improve the
35 transgene integration.

Therefore, the method accroding to the present invention presents several advantages among which the increase of the efficiency of the transgenesis or of the stable transfection and the decrease of the copy number of the integrated polynucleotide.

The method according to the present invention concerns the random integration of 5 the transgene. Therefore, the transgene is not designed to undergo homologous recombination with the host cell genome. The 5' and 3' sequence of the polynucleotide to be integrated into the genome do not have significant homologies with one locus of the host genome. Optionally, said polynucleotide to be integrated have less than 70 % indentity with a host cell genomic sequence, preferably less than 60 or 50 %, more 10 preferably less than 40, 30 or 20 % identity. Optionally, the 5' and 3' sequences of said polynucleotide to be integrated have less than 90 % indentity with a host cell genomic sequence, preferably less than 80 or 70 %, more preferably less than 50, 40, 30 or 20 % identity, wherein said 5' and 3' sequences are less than 5 kb long, preferably less than 3 kb or 1.5 kb long, more preferably less than 1 kb, 500 bp or 100 bp long.

15 For introducing a transgene in a mouse egg by homologous recombination, the man skilled in the art usually flanks the transgene with about 5 kb of homologous sequence, never less than 3 kb. The homologous recombination in egg is a extremely rare event. Indeed, with 5 kb of homologous sequence, the homologous recombination occurs in only 1 clone among 10 millions of fertilized eggs. In embryonic fish cells, homologous 20 recombination never occurs, more particularly in fish fertilized egg.

For transgenesis, the cleaving agent is preferably co-injected with the vector comprising the transgene. Indeed, a co-injection of the cleaving agent avoids the delay due to the expression of the cleaving agent. An early integration of the transgene is important for decreasing the mosaïcism.

25 In one embodiment of the present invention, the method consists in a method for randomly integrating a polynucleotide into the host cell genome by *in vivo* preparation said linear polynucleotide having free 5' and 3' ends from a vector, said method comprising the following steps:

a) Introducing in said host cell a vector having no free 5' and 3' ends and 30 comprising said polynucleotide, said polynucleotide being flanked by at least one endonuclease site which is found in the host cell genome at less than 5 copies, preferably less than 2, more preferably never found in the host cell genome;

b) Optionally, introducing in said host cell either the endonuclease which cleaves said endonuclease site present in said vector or a vector comprising a nucleic acid 35 encoding said restriction endonuclease; and,

23

c) causing cleavage of said site(s) in said host cell, thereby creating or releasing said polynucleotide in a linear form having free 5' and 3' ends from said vector into said host cell; and,

d) maintaining the host cell under conditions and for a period of time sufficient to
5 cause the random integration of said excised polynucleotide into said host cell genome.

In a particular embodiment of this method, said polynucleotide to be linearized or excised is flanked by two endonuclease sites. Preferably, the step b) consists in introducing in said host cell the endonuclease which cleaves said endonuclease site present in said vector. Preferably, steps a) and b) are simultaneous. Preferably, the
10 cleaved sites do not generate cohesive ends. Optionally said vector further comprises a nucleic acid sequence encoding the endonuclease. Optionally, said polynucleotide to be linearized or excised and said nucleic acid encoding said endonuclease are each comprised by a distinct vector. Preferably, said polynucleotide to be linearized or excised does not comprise any cleavage site. Preferably, said vector is double-stranded.
15 Optionally, said polynucleotide can comprise a sequence encoding a polypeptide or an antisense, a regulatory sequence such as promoter and enhancer, and/or a recognition sequence for a molecule. Preferably, said host cell is selected from the group consisting of a stem cell, a somatic cell, a gamete, a blastomer and an egg.

The present invention relates to a composition for transgenesis comprising :
20   1) a vector having no 5' and 3' free ends and comprising a transgene to be integrated, said transgene being flanked by at least one cleavage site which is found in the host cell genome at less than 5 copies, preferably 2 copies, and more preferably said cleavage site is not found in the host cell genome; and,

2) a cleaving agent or a vector comprising a nucleic acid encoding said cleaving
25 agent.

Preferably, said composition comprises the cleaving agent. Optionally, said vector comprising a nucleic acid encoding said cleaving agent is an expression vector or a mRNA. Said transgene to be integrated according to the present invention is unable to make significantly an homologous recombination with the host cell genome. Optionally,
30 said transgene to be integrated have less than 70 % indentity with a host cell genomic sequence, preferably less than 60 or 50 %, more preferably less than 40, 30 or 20 % identity. Optionally, the 5' and 3' sequences of said transgene to be integrated have no homology with a host cell genomic sequence, preferably less than 90 % indentity, more preferably less than 80 or 70 %, still more preferably less than 50, 40, 30 or 20 % identity,
35 wherein said 5' and 3' sequences are 5 kb long, preferably between 3 kb and 1.5 kb long, more preferably 1 kb, 500 bp or 100 bp long. Preferably, the cleaved sites do not generate cohesive ends. Preferably, said transgene does not comprise any cleavage site.

24

Optionally said vector comprising said transgene further comprises a nucleic acid sequence encoding the cleavage agent. Optionally, said transgene and said nucleic acid encoding said cleavage agent are each comprised by a distinct vector. Preferably, said transgene is flanked by two cleavage sites. Preferably, said cleavage site is an

5 endonuclease site and said cleaving agent is the corresponding endonuclease. Preferably, said endonuclease has a recognition site of at least 12 nucleotides. Still more preferably, said endonuclease is a meganuclease, notably one meganuclease of Figure 2. Optionally, said meganuclease is selected from the group consisting of I-Ceu I, I-Cre I, I-Chu I, I-Csm I, I-Dmo I, I-Pan I, I-Sce I, I-Sce II, I-Sce III, I-Sce IV, F-Sce I, F-Sce II, PI-

10 Aae I, PI-Ape I, PI-Ceu I, PI-Cir I, PI-Ctr I, PI-Dra I, PI-Mav I, PI-Mfl I, PI-Mgo I, PI-Mja I, PI-Mka I, PI-Mle I, PI-Mtu I, PI-MtuH I, PI-Pab III, PI-Pfu I, PI-Pho I, PI-Pko I, PI-Psp I, PI-Rma I, PI-Sce I, PI-Ssp I, PI-Tfu I, PI-Tfu II, PI-Tli I, PI-Tli II, PI-Tsp I, PI-Tsp II, PI-Bsp I, PI-Mch I, PI-Mfa I, PI-Mga I, PI-Mga II, PI-Min I, PI-Mma I, PI-Msh I, PI-Msm II, PI-Mth I, PI-Tag I, PI-Thy II, I-Ncr I, I-Ncr II, I-Pan II, I-Tev I, I-Ppo I, I-Dir I, I-Hmu I, I-Hmu II, I-Tev

15 II, I-Tev III, F-Sce I, F-Sce II (HO), F-Suv I, F-Tev I, and F-Tev II. Preferably, said meganuclease is selected from the group consisting of I-Ceu I, I-Cre I, and I-Sce I. Optionally, said endonuclease is synthetic. Preferably, said vector is double-stranded. Preferably, said vector is a plasmid. Optionally, said transgene can comprise a sequence encoding a polypeptide or an antisense, a regulatory sequence such as promoter and

20 enhancer, and/or a recognition sequence for a molecule.

The invention relates to the use of the composition according to the present invention for producing transgenic cells, animals or plants.

Cleavage site and cleaving agent

25 The cleavage site according to the present invention is preferably not found in the genome of the host cell. If this cleavage site exists in the cell genome, the cell genome presents no more than 5, preferably 2, sites. If some other cleavage sites than the flanking sites exist in the vector, they are preferably located outside of the region comprising the polynucleotide to be linearized or excised.

30 The cleavage site according to the invention is a sequence or an element which is specific of the cleaving agent. The cleaving agent can be for example a ribozyme, an endonuclease. Preferably, the cleaving agent is an endonuclease. The cleavage site according to the invention could be a modified nucleotide leading under the appropriate conditions to DNA cleavage. Such modified nucleotide could be for example an abasic

35 nucleotide (Lhomme et al, 1999, *Biopolymer*, **52**, 65-83; the disclosure of which is

incorporated herein by reference). The presence of an abasic nucleoide leads to a cleavage, for example, by the action of AP-endonuclease such as exonuclease III.

In a particular embodiment of the present invention, two or more different endonucleases can be used in the present method. The polynucleotide to be integrated is

5 flanked for example by two different endonuclease sites and the method comprises a step of introducing each endonucleases or one or several vector(s) encoding the used endonucleases.

Preferably, the sites flanking the polynucleotide sequence to be integrated do not generate cohesive ends after the cleavage. Therefore, if the cleavage of the sites could

10 lead to the generation of cohesive ends, the sites have preferably an inverse orientation so as to avoid the multimerization through the cohesive ends.

The endonuclease need to be chosen so as to generate in the genome of the host cell a very low number of cleavages that can be easily repaired by the cell without any damage for this cell. Indeed, the cell without any damage could tolerate a very low

15 number of cleavages in the chromosome, generally less than 5 double-stranded breaks, more preferably less than 2 double-stranded breaks. Preferably, the endonuclease is chosen so as not to generate any double-stranded break in the host cell chromosome.

Commonly used four and six base cutting restriction enzymes are not convenient for the present invention since they would usually lead to cleavage of chromosomal DNA

20 and to cells death due to the existence of many restriction sites within the cellular DNA.

Preferably, the endonuclease sites flanking the polynucleotide sequence to be linearized or excised according to the invention do not naturally occur in the host cell. Preferably, if the vector comprising the polynucleotide sequence to be linearized or excised is a plasmid, the flanking endonuclease sites do not occur in the bacteria used for

25 the plasmid production. Preferably, the flanking endonuclease sites correspond to endonucleases having a recognition site of at least 10, 12, 15, 18, 20, 22 or 25 nucleotides. In a particular embodiment, an endonuclease used in the present invention can tolerate less than 10 % of change in its recognition site, preferably 5 %, more preferably 2 %.

30 A homing endonuclease or meganuclease which recognizes a large DNA sequence is an example of an endonuclease which can be used in the present invention (Dalgaard et al, 1997, *Nucleic Acids Resarch*, **25**, 4626-463 ; Chevalier et Stoddard, 2001, *Nucleic Acids Resarch*, **29**, 3757-3774). An example of a meganuclease enzyme is I-Sce I which recognizes a 18 bp site that does not appear to be represented in murine or human

35 DNA. I-Scel generates a 4 bp staggered cut with 3'OH overhangs. For more information on I-Scel meganuclease, see US 6,238,924, the teaching of which is incorporated herein

by reference. In a preferred embodiment, the method according to the present invention uses I-Sce I endonuclease and the correponding recognition and cleavage site.

Meganucleases constitute a family of very rare-cutting enzymes. See Figure 2 for a list of meganucleases. Homing endonucleases encoded by introns ORF, independent

5   genes or intervening sequences (inteins) are defined now as "meganucleases". They have recognition sequences that span 12–40 bp of DNA, whereas "classical" restriction enzymes recognise much shorter stretches of DNA, in the 3–8 bp range (up to 12 bp for rare-cutter). Meganucleases are rather well characterised structurally and mechanistically. They fall into 4 separated families on the basis of pretty well conserved amino acids

10   motifs.

1 – The dodecapeptide family (dodecamer, DOD, D1-D2, LAGLI-DADG, P1-P2)

This is the largest family of proteins (more than 150 sequences) clustered by their most general conserved sequence motif: one (5 sequences) or two copies (vast majority) of a twelve-residue sequence: the di-dodecapeptide. Meganucleases with one

15   dodecapetide (D) are around 20 kDa in molecular mass and act as homodimer. Those with two copies (DD) range from 25 kDa (230 AA) to 50 kDa (HO, 545 AA) with 70 to 150 residues between each motif and act as monomer.

2 - GIG family

The jointly motif is short but pretty well conserved: KSGIY (10/11 AA) YIGS. For

20   these meganucleases (28 sequences) the cleavage site is different from the recognition sequence.

3 - HC family

Sequences in this group are rich in Histidines and Cysteines residues and the conserved sequence is approximately: "SHLC - G - G - H - C". The most well

25   characterised enzyme: I-*Ppo* I

4 - HNH family and no motif

The last group of sequences, assembled because they don't have any of previous motifs are not structurally well characterised. The consensus sequence (HH - N - H - H in a window of 35 amino acid residues) is rather complicate. Several properties of these

30   proteins are distinctive from other Meganucleases because they leave a 5' extension of 2 bp after a double-stranded break or have a long size of staggered cut of at least 10 nucleotides. (Most of the Meganucleases cleaves the two strands of a double-stranded DNA and leave a 4 base pair 3' protruding end)

The different classes are summarised:

35   CLASSE I which are chraracterized by : One Dodecapeptide motif (D1, LAGLI, P1) or two Dodecapeptide motifs (D1 - D2, LAGLI - GDAG, P1 - P2) ; and, Cleavage inside

the recognition site, leaving 4 nt staggered cut with 3'OH overhangs. There are 8 sub-families including:

One Dodecapeptide motif: Some examples are I-Ceu I, I-Cre I

Two Dodecapeptide motif: Some examples are I-Chu I, I-Csm I, I-Dmo I, I-Pan I, I-

5   Sce I, I-Sce II, I-Sce III, I-Sce IV, F-Sce I, F-Sce II, PI-Aae I, PI-Ape I, PI-Ceu I, PI-Cir I, PI-Ctr I, PI-Dra I, PI-Mav I, PI-Mfl I, PI-Mgo I, PI-Mja I, PI-Mka I, PI-Mle I, PI-Mtu I, PI-MtuH I, PI-Pab III, PI-Pfu I, PI-Pho I, PI-Pko I, PI-Psp I, PI-Rma I, PI-Sce I, PI-Ssp I, PI-Tfu I, PI-Tfu II, PI-Tli I, PI-Tli II, PI-Tsp I, PI-Tsp II, PI-Bsp I, PI-Mch I, PI-Mfa I, PI-Mga I, PI-Mga II, PI-Min I, PI-Mma I, PI-Msh I, PI-Msm II, PI-Mth I, PI-Tag I, PI-Thy II

10   CLASSE II which is characterized by : At least one specific motif $(GIY-N_{10/11}-YIG)$; and, Cleavage outside the recognition site, leaving 2 nt staggered cut with 3'OH overhangs. Some examples are I-Ncr I, I-Ncr II, I-Pan II, I-Tev I

CLASSE III which are chraracterized by : His-Cys box (SHLC-G-H-C) or no define motif; and, either Cleavage inside the recognition site, leaving 4 nt staggered cut with

15 3'OH overhangs (One example is I-Ppo I), or Cleavage of long size of staggered cut of at least 10 nucleotides (Some examples are I-Dir I, I-Hmu I, I-Hmu II)

CLASSE IV which is characterized by : Half of the specific motif $(GIY-N_{10/11}-YIG)$; and Cleavage outside the recognition site, leaving 2 nt staggered cut with 3'OH overhangs. One example is I-Tev II

20   CLASSE V which is characterized by : HNH motif (HH-G-N-CH-H); and Cleavage inside the recognition site, leaving 2 nt staggered cut, with 5'OH overhangs. One example is I-Tev III.

The meganucleases could also be encoded by "Free" Genes. Up to this date, 5 characterised genes (F-Sce I, F-Sce II (HO), F-Suv I, F-Tev I, F-Tev II) are known in

25 Yeast and bacteriophages.

The meganucleases could be Inteins. Up to this date, 120 protein sequences are characterized (35 with experimental demonstration, 75 theoritical) in 46 different species and strains (7 eukaryotes, 23 bacteria,16 archaea). There are more than 200 potential sequences.

30   The endonuclease in the process of the invention can be chosen in the group including in not restrictif title: I-Ceu I, I-Cre I, I-Chu I, I-Csm I, I-Dmo I, I-Pan I, I-Sce I, I-Sce II, I-Sce III, I-Sce IV, F-Sce I, F-Sce II, PI-Aae I, PI-Ape I, PI-Ceu I, PI-Cir I, PI-Ctr I, PI-Dra I, PI-Mav I, PI-Mfl I, PI-Mgo I, PI-Mja I, PI-Mka I, PI-Mle I, PI-Mtu I, PI-MtuH I, PI-Pab III, PI-Pfu I, PI-Pho I, PI-Pko I, PI-Psp I, PI-Rma I, PI-Sce I, PI-Ssp I, PI-Tfu I, PI-Tfu

35 II, PI-Tli I, PI-Tli II, PI-Tsp I, PI-Tsp II, PI-Bsp I, PI-Mch I, PI-Mfa I, PI-Mga I, PI-Mga II, PI-Min I, PI-Mma I, PI-Msh I, PI-Msm II, PI-Mth I, PI-Tag I, PI-Thy II, I-Ncr I, I-Ncr II, I-Pan II, I-Tev I, I-Ppo I, I-Dir I, I-Hmu I, I-Hmu II, I-Tev II, I-Tev III, F-Sce I, F-Sce II (HO), F-Suv I,

F-Tev I, and F-Tev II. Preferably, said endonuclease is chosen from the group consisting of I-Ceu I, I-Cre I, I-Sce I. More preferably, said endonuclease is I-Sce I.

Some endonucleases such as I-Sce I endonuclease could stay non-covalently bound to one half cleavage site after the cleavage (the big site for I-Sce I). This property

5   can be used to protect the free ends of the excised polynucleotide against the exonuclease degradation. Indeed, the cleavage sites are oriented so as the half site having the binding capacity is placed at the extremities of the polynucleotide. Moreover, some restriction endonucleases such as I-Sce I endonuclease further present a NLS-like sequence (Nuclear Localization Sequence) which allows the nucleus targeting. This

10   additional property can be used to facilitate the nuclear targeting of the linearized or excised polynucleotide if the linearization or excision occurs in cytoplasm.

Therefore, in one embodiment of the present invention, the flanking sites are oriented so as the half sites having an endonuclease binding capacity are placed towards the polynucleotide to be linearized or excised. In this embodiment, the free ends of the

15   excised polynucleotide are protected against the exonuclease degradation and/or the transport of the excised polynucleotide to the nucleus is facilitated.

Otherwise, the presence of an endonuclease to the half cleaved site of the excised or linearized polynucleotide masks the free ends and makes these ends less available for recombine with the host cell genome.

20   Therefore, in an alternative embodiment of the present invention, the flanking sites are oriented so as the half sites having an endonuclease binding capacity are placed towards the vector.

Synthetic endonucleases are also considered in the present invention. These synthetic endonucleases comprise a DNA recognition domain and a DNA cleavage

25   domain.

The DNA recognition domain can be derived from naturally occurring proteins presenting a DNA recognition domain such as the recognition domain of the Type IIS restriction endonuclease (for example amino acids 1-382 of FokI, US 5,356,802, the disclosure of which being incorporated herein by reference). Suitable recognition domains

30   include, but are not limited to, the recognition domains of zinc finger motifs; homeo domain motifs; POU domains ; other DNA binding protein domains of respressor such as lambda repressor, lac repressor, cro, ga14; DNA binding protein domains of oncogenes such as myc, jun; and other naturally occurring sequence-specific DNA binding proteins that recognize more than 6 base pairs. The DNA recognition domain could comprise the

35   following motifs : helix-turn-helix, zinc finger, steroid receptor, helix-loop-helix, or other helical motif like leucine zipper. The DNA recognition domain is preferably a combination of existing DNA recognition domains if their recognition site is less than 10 nucleotides

(WO 96/20951, the disclosure of which being incorporated herein by reference), for example a combination of at least three zinc fingers (US 6,013,453 ; US 6,007,988, the disclosures of which being incorporated herein by reference). The existing DNA recognition domain can be modified or natural. However, such DNA recognition domain

5 could also be synthetic. The DNA recognition domain could also be a natural or modified polynucleotide.

The DNA cleavage domain can be derived from proteins containing a DNA cleavage domain such as the cleavage domain of the Type II restriction endonuclease (for example, amino acids 383-578 of FokI, US 5,356,802, the disclosure of which being

10 enclosed herein by reference).

Vector comprising a nucleic acid encoding the cleaving agent, preferably the endonuclease

The action of the cleaving agent on the host cell can be obtained either by

15 administering to the host cell of the cleaving agent, preferably the endonuclease, or by the expression of the cleaving agent, preferably the endonuclease, in the host cell. In this last case, the process of the invention can be realized by transforming the host cell with a nucleic acid encoding the cleaving agent, preferably the endonuclease, under the control of regulation sequences, notably a promoter adapted to the host cell. An other alternative

20 for the expression of the cleaving agent is the introduction of a mRNA encoding the cleaving agent into the host cell. In the present invention, the vector comprising a nucleic acid encoding the cleaving agent also designates a mRNA encoding the cleaving agent.

The invention also contemplates the use of host cell from a transgenic animal or plant expressing the cleaving agent. In this case, the introduction of the cleaving agent or

25 a nucleic acid encoding the cleaving agent is no more necessary. The nucleic acid encoding the cleaving agent is preferably operably linked ta a promoter specific of the germ line such as VASA, promoters of the protein synthesis in ovocyte, strong promoter with expression as soon as stage II.

The vector comprising a nucleic acid encoding the cleaving agent, preferably an

30 endonuclease, contains all or part of the coding sequence for the endonuclease operably linked to one or more expression control sequences whereby the coding sequence is under the control of transcriptional signals to permit production or synthesis of the endonuclease. Therefore, said nucleic acid encoding said endonuclease is comprised in an expression cassette. More particularly, the vector comprises a replication origin, a

35 promoter operatively linked to said encoding nucleic acid, a ribosome-binding site, an RNA-splicing site (when genomic DNA is used), a polyadenylation site and a transcription

termination site. Selection of the promoter will depend upon the desired route for expressing the endonuclease.

Indeed, for an eukaryotic host, the promoter can be a strong promoter such as the metallothionein promoter, an SV40 promoter, a retroviral promoter, the cytomegalovirus

5 (CMV) promoter, an ubiquitus promoter such as villin or actin promoter, a constitutive or inducible promoter, a tissue-specific promoter or tumor-specific promoters such as the α-fetoprotein promoter, the amylase promoter (especially, the murine amylase promoter), the cathepsin E promoter, the M1 muscarinic receptor promoter, or the γ-glutamyl transferase promoter. Optionally, the vector could further contain an enhancer and

10 insulator (Kaffer et al., *Genes Dev.* 2000, **14**, 1908-19 ; EP 859,059 ; WO96/04390, the disclosures of which are incorporated herein by reference). When the host cell is a fertilized egg or a blastula's cell, the cleaving agent, preferably the endonuclease, needs to be expressed at an early stage.

The elements permitting the production or synthesis of the endonuclease can be

15 native, derived from native elements or manufactured *de novo*. The elements can then be isolated and fused together by methods known in the art such as using compatible cloning and restriction sites. One advantage of the method according to the present invention is that only a transient expression of the appropriate endonuclease is necessary to excise the linear polynucleotide to be integrated in the host cell genome.

20   If the host cell is prokaryotic, the vectors of the present invention will thus contain at least one promoter capable of being recognized by a prokaryotic RNA polymerase, and of thus permitting the transcription of a polynucleotide that is operably linked to that promoter. The vector may have multiple prokaryotic promoters if desired. The specific prokaryotic promoter(s) employed will depend upon the prokaryotic cell that is to be the

25 host of the vector. Examples of suitable promoters include constitutive or inducible prokaryotic promoters, such as the λpL or λpR promoters, the T6 polymerase promoter, the T7 polymerase promoter, or other well-characterized promoters (e.g., lac, recA, gal, trp, ara, hut, etc.). Most preferably, the promoter used for expression in prokaryotic cells will be inducible.

30   The preferred eukarotic expression vectors contain both prokaryotic sequences, to facilitate the propagation of the vector in bacteria, and one or more eukaryotic transcription units that are expressed in eukaryotic cells. The various methods employed in the preparation of plasmids and transformation of host organisms are well known in the art. For other suitable expression systems for both prokaryotic and eukaryotic cells, as

35 well as general recombinant procedures, see Molecular Cloning: A Laboratory Manual, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press: 1989) Chapters 16 and 17; the teaching of which is incorporated herein by reference.

Polynucleotide to be linearized or excised into the host cell

The polynucleotide to be linearized or excised with the process of the invention can be any natural or synthetic polynucleotide. Said polynucleotide is preferably exogenous to the host cell. More preferably, said polynucleotide is non-similar to the host cell chromosome sequence. It can comprise a gene sequence or an intergene sequence. It can comprise coding or not coding sequence. Said coding sequence encode any desired product, including peptides, proteins, and RNA. The polynucleotide can encode a reporter protein. It can comprise an encoding sequence for a polypeptide or an antisense, or a regulatory sequence such as a promoter, an enhancer, a silencer, etc... It can comprise a recognition sequence for a molecule such as a hormone, a transcriptional factor, an endonuclease, a polynucleotide, etc...

The polynucleotide to be integrated can comprise a reporter gene such as β-galactosidase, luciferase, alkaline phosphatase, green fluorescent protein, tyrosinase, DsRed proteins. By reporter gene is intended any nucleic acid encoding a product easily assayed. Preferably, the reporter gene is under the control of a strong promoter or a tissue-specific promoter.

More particularly, the polynucleotide to be integrated with the process of the invention can be chosen in the group including in not restrictif title:

Gene encoding proteins: Secreted proteins such as erythropoietin, albumin, growth hormone, α1-antitripsin; Surface proteins; Otherwise, proteins such as hemoglobin (α, β, γ, ε), collagens (type I, II, III, IV, α1 to 5), chemokine receptors, interferons (α, β, γ), caspase, p53, etc...

Sequences coding for Transfert RNA: tRNA for unusual codons (mitochondrial or synthetic tRNA),

Sequences coding for Vaccining gene α, a super antigen, an adjuvant (C3d);

Sequences coding for a peptide such as a reduced CMH peptide;

Sequences coding for a HLA chain;

Sequences coding for cytochrome p450 combination;

Sequences coding for a metabolic pathway (lysine or phenylalanine).

Chromatin regulatory sequences such as gene of chromosome X inactivation (gene *xist*), gene with maternal effect (*ddk* gene), chromatin opening sequence (HNF4), rich CpG sequence with low or high number of SP1 sites (CCCGCC/G or C/GGCGGG), MAR or SAR sequence, eukoryotic, bacterial or viral chromosomic replication origin;

Transcriptional regulatory sequences such as sequences controlling the chromosome opening for transcription (Locus Control Region LCR, Dominant Control Region DCR), eukaryotic constitutive or tissue specific promoters, inducible promoters

(metalloprotein, batcerial operators, T7 promoter, tetracyclin inducible promoter (tet-ON and tet-OFF), enhancers, silencers, RNA polymerase (T7 or viral polymerase), Internal Ribosomal Entry Sites IRES followed by a gene of interest ;

5   Sequences interesting for the genome engineering such as sites of meganucleases (ex : I-*Sce* I, I-*Tev* III, F-*Sce* I, F-*Sce* II, I-*Ceu* I, I-*Dmo* I, I-*Chu* I, PI-*Sce* I, PI-*Psp*I or sites for synthetic meganucleases) ; repeated sequences (ALU, SINES, LINES) ; micro- and minisatellites ; RAG, *loxP*, FRT or β-resolvase sites ; inversed repeat sequences of transposons; transposons ; provirus ; LTR of transposons and virus ;

    Antisense sequences and ribozymes : antisense of mRNA, more particularly of
10 p53, Rb, p16, p21 mRNAs, mRNAs of proteins involved in the genome engineering such as meganucleases, RAG, transcriptases, viral mRNAs.

    In the case of a gene, the polynucleotide preferably includes all the elements necessary for the expression of the gene. Therefore, the encoding polynucleotode is operably linked to one or more expression control sequences whereby the coding
15 sequence is under the control of transcriptional signals to permit production or synthesis of the encoded product. Said a encoding polynucleotide is comprised in an expression cassette. More particularly, the vector comprises a replication origin, a promoter operatively linked to said encoding nucleic acid, a ribosome-binding site, an RNA-splicing site (when genomic DNA is used), a polyadenylation site and a transcription termination
20 site. Selection of the promoter will depend upon the desired route for expressing the encoded product. Indeed, the promoter can be a strong promoter such as the metallothionein promoter, an SV40 promoter, a retroviral promoter, the cytomegalovirus (CMV) promoter, an ubiquitus promoter such as villin or actin promoter, a constitutive or inducible promoter, a tissue-specific promoter (see WO 98/56902, Table 1; the disclosure
25 of which is incorporated herein by reference) or tumor-specific promoters such as the α-fetoprotein promoter, the amylase promoter (especially, the murine amylase promoter), the cathepsin E promoter, the M1 muscarinic receptor promoter, or the γ-glutamyl transferase promoter. Optionally, the vector could further contain an enhancer and insulator (Kaffer et al., *Genes Dev.* 2000, **14**, 1908-19 ; EP 859,059 ; WO96/04390, the
30 disclosures of which are incorporated herein by reference).

    The elements permitting the production or synthesis of the transgene can be native, derived from native elements or manufactured *de novo*. The elements can then be isolated and fused together by methods known in the art such as using compatible cloning and restriction sites.

35

Vector comprising a polynucleotide to be linearized or excised

The preferred eukarotic expression vectors contain both prokaryotic sequences, to facilitate the propagation of the vector in bacteria, and one or more eukaryotic transcription units that are expressed in eukaryotic cells. The various methods employed in the preparation of plasmids and transformation of host organisms are well known in the art. For other suitable expression systems for both prokaryotic and eukaryotic cells, as well as general recombinant procedures, see Molecular Cloning: A Laboratory Manual, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press: 1989) Chapters 16 and 17; the teaching of which is incorporated herei by reference.

In a preferred embodiment of the invention, the vector comprising the polynucleotide sequence to be linearized or excised also comprises the sequence encoding the cleaving agent, preferably the endonuclease, corresponding to the flanking sites. In this quite preferred embodiment, the gene coding for the cleaving agent, preferably the endonuclease, contains its own sequences of regulation and is not thus placed in the same open reading frame than the polynucleotide to be linearized or excised.

The vector comprising the polynucleotide to be linearized or excised in the host cell genome flanking by the cleaving site can be manufactering according to methods generally known in the art. For example, this vector can be manufactering by chemical synthesis, recombinant DNA/RNA technology or a combination of both (Sambrook et al., Eds., Molecular Cloning: A Laboratory Manual, 2nd edition, Cold Spring Harbor University Press, New York (1989); and Ausubel et al., Eds., Current Protocols In Molecular Biology, John Wiley & Sons, New York (1997); the disclosure of which are incorporated herein by reference)

The vector comprising the polynucleotide to be linearized or excised according to the present invention does not present any free 5' or 3' ends. The vector is advantageously a circular DNA molecule such as a plasmid, but can also be a linear molecule having its ends unaccessible to exonuclease. In a first example, the ends adopt a secondary structure, such as a hairpin, which protects the ends against exonucleolyse. In a second example, the ends could be chemically modified in order to block the exonuclease action. Such modifications can be, for example, the use of phosphothioate links which are resistant to exonuclease (Putney 1981; Olsen and Ecktein 1990). In a third example, the ends could be linked to sterically blocking molecules which impede the access for the exonuclease. Such molecules can be for example a protein or a polyamide. This protein could have some nucleus affinity and could trigger the vector to the nucleus. Preferably, the blocking substituant has a molecular weight of no ore than 1 kD. A variety

34

of non-toxic substituents such as biotin, cholesterol or other steroids, or a non-intercalating cationic fluorescent dye can be used.

The Figure 3 discloses some examples of linear vectors according to the present invention. In figure 3A, a linear vector can be obtained from a plasmid comprising the
5   polynucleotide to be linearized or excised flanked by two endonuclease sites such as I-Tev III sites, themselves flanked by other endonuclease such as F-Tev I sites. *In vitro*, the F-Tev I endonuclease cleaves the F-Tev I sites and this cleavage produces protruded ends able to adopt a hairpin structure. Optionally, a ligase can be used in order to covalently close the loop. The host cell can be transfected with this linear construct with
10 the hairpins at its ends, the polynucleotide to be linearized or excised being flanked by two endonuclease sites such as I-tev III sites.

Similarly, in Figure 3B, a linear vector with a hairpin at each ends can be made as following. *In vitro*, the polynucleotide to be linearized or excised is prepared with blunt ends or cohesive ends. This polynucleotide is mixed with oligonucleotides forming a
15 hairpin, such hairpin presenting a cleavage site, preferably an endonuclease site, in its tail. The end of the hairpin tail can be blunt or cohesive. A ligation is done in order to covalently link the hairpin and the polynucleotide to be linearized or excised. (Perrin et al, *EMBO J* 1993 , **12,** 2939-2947; the teaching of which is incorporated herein by reference)

20   Vectors comprising the polynucleotide to be linearized or excised flanked by the cleavage site and/or, the nucleic acid encoding the cleaving agent, preferably the endonuclease, can be introduced into a cell by a variety of methods (e.g., transformation, transfection, direct uptake, projectile bombardment, using liposomes). Example of suitable methods of transfecting or transforming cells include calcium phosphate precipitation,
25 electroporation, microinjection, infection, lipofection and direct uptake. Such methods are described in more detail, for example, in Sambrook et al., Molecular Cloning : A Laboratory Manual, Second Edition, Cold Spring Harbor University Press, New York (1989); and Ausubel, et al., Current Protocols in Molecular Biology, John Wiley & Sons, New York (1998), the teaching of which is incorporated herein by reference. Other
30 suitable methods are also described in the art. In one embodiment of the present invention, the vectors are associated to the substance susceptible to allow or to facilitate the transformation of the host cell. Vectors comprising the polynucleotide to be linearized or excised flanked by the cleavage site and/or, the nucleic acid encoding the cleaving agent, preferably the endonuclease, can also be introduced into a host cell by targeting
35 the vector to the cell membrane phospholipids. For example, targeting vector to a VSV-G protein, a viral protein with affinity for all cell membrane phospholipids. Such a construct can be produced using methods well known to those practiced in the art.

Cleaving agent, preferably endonuclease can be introduced into a cell according to methods generally known in the art which are appropriate for the particular endonuclease and cell type. For example, an endonuclease can be introduced into a cell by direct uptake, microinjection, calcium phosphate precipitation, electroporation, infection, and

5 lipofection. Such methods are described in more detail, for example, in Sambrook et al., Molecular Cloning : A Laboratory Manual, Second Edition, Cold Spring Harbor University Press, New York (1989); and Ausubel, et al., Current Protocols in Molecular Biology, John Wiley & Sons, New York (1998), the teaching of which is incorporated herein by reference. Other suitable methods are also described in the art. The endonuclease can be

10 coupled to a facilitator of protein entry to facilitate introduction of the enzyme into a cell. Examples of facilitators of protein entry include tat, HSV VP22 and anthrax toxin. Coupling of a protein to a facilitator of protein entry can be accomplished using methods well known to those practiced in the art. Protein delivery strategies (e. g., HSV VP22, anthrax toxin) can be evaluated in accordance with the methods of the invention described herein.

15 Once in the cell, the vectors comprising the polynucleotide to be linearized or excised flanked by the cleavage site and, if necessary, the nucleic acid encoding the cleaving agent, preferably the endonuclease, or the cleaving agent itself, preferably the endonuclease, are imported or translocated by the cell from the cytoplasm to the nucleus.

20 ### Host cells and multicellular organisms

As used herein, a cell refers to a prokaryotic cell such as a bacterial cell or a eukaryotic cell, such as an animal, plant or yeast cell. More preferably, the cell is an eukaryotic cell. The cell can be a stem cell (preferably an embryonic stem cell), a somatic cell, a gamete, a blastomer or an egg (preferably a fertilized egg). The host cell can stem

25 from fish, bird, non-human mammals, insect, amphibian, reptile, preferably from medaka, zebrafish, mice, chicken, xenopus, sheep, cow, rabbit, more preferably from fish, chicken and mice. The host cell can have all stage of differentiation, from totipotent to differenciated cells. Examples of mammalian cells include human (such as HeLa cells), bovine, ovine, porcine, murine (such as embryonic stem cells), rabbit and monkey (such

30 as COS1 cells) cells. The cell may be an embryonic cell, bone marrow stem cell or other progenitor cell. Where the cell is a somatic cell, the cell can be, for example, an epithelial cell, fibroblast, smooth muscle cell, blood cell (including a hematopoietic cell, red blood cell, T-cell, B-cell, etc...), tumor cell, cardiac muscle cell, macrophage, dendritic cell, neuronal cell (e.g., a glial cell or astrocyte, or pathogen-infected cell 'e.g., those infected

35 by bacteria, viruses, virusoids, parasites, or prions). Indeed, the method for integrating a polynucleotide in the host cell genome according to the present invention is well adapted for stable transfection.

The process of the invention is thus quite particularly useful for the transgenesis for the gene therapy, production of recombinant protein in animals or plants, production of transgenic animals as models. As a consequence, the invention has for object a method of transgenesis of a pluricellular organism consisting in administering to said organism a

5   composition comprising one or several vector comprising the polynucleotide to be excised or linearized and, optionally the cleaving agent, as defined previously, possibly associated to a substance susceptible to allow or to facilitate the transformation of one or several cellular types of said organism by said vector, then to generate in said cells the linear molecule ready for be integrated by subjecting said cells to the action of the cleaving.

10 The process of the invention is realized *in ovo* or *in vivo* on cells, such as cells in culture, or *in situ* directly on a multicellular tissue, organ or organism.

The process of transgenesis according to the invention is quite particularly intended for animals chosen in the group including in but not restrictive to: cattle (cows), goats, rabbit, rodents, marmots, monkeys, insects (spider, butterflies, fly), fishes, calmar,

15 amphoxius, xenopus, birds, chickens, ascidies and ovine races (sheeps).

Transgenic fish are produced by introducing a vector according to the present invention into cells of fish, gametes such as spermatozoid or embryonic cell, preferably embryonic cells, and more preferably in a single cell embryo. Where the vector is introduced into embryonic cells, the transgenic fish is obtained by allowing the embryonic

20 cell or cells to develop into a fish. Introduction of the vectors into embryonic fish cells, and subsequent development of the fish, are simplified by the fact that embryos develop outside of the parent fish. Alternatively; when the vector is introduced into spermatozoid, the fecondation of the oocytes is done with the transfected spermatozoid and the embryo is allowing to develop into a fish. Preferably, the spermatozoid is transfected with the

25 vector comprising the transgene and with the cleavaing agent.

Transgenic fish is selected in the group consisting of salmon, trout, tuna, halibut, catfish, zebrafish, medaka, carp, sticklebass, astyanax, tilapia, goldfish, sea bass, sticklebass, astyanax, sturgeon and loach. The most prefered are medaka and zebrafish. However, the method according to the present invention could also be applied for

30 producing aquatic transgenic species such as xenopus, shrimp, sea urchin.

The host cell in the process of the invention can be chosen in the group including in but not restrictive to:

Microorganism :

For industrial production purpose:

35 K. lactis, P. pastoris, S. cerevisiæ, S. pombe, A. immersus, P. limus, E. coli

2) Bacterial or eukarotic vaccines:

Listeria monocytognese, bacille Calmette Guerin

S. cerevisiæ, S. pombe, C. albicans, Plasmodium falciparum, Amoebas

Plants :

1) Biomolecule production

Soya, colza, wheat, corn, rice

2) Transgenic plants

Tomatoes, strawberries, apples, citrus fruits, tobacco

3) Food medecine, and vaccine plants

Spinach, cereal

4) Production of biomaterials

Rubber, paper, wood

Animals :

1) Biomolecules and vaccines production

Cows, goats, sheeps, rabbit, rodents, marmots, monkeys, insects

2) Transgenic animals

Fishes, calmar, amphoxius, xenopus, birds, chickens, cattle, ovine races

3) Biomaterials production

Silkworms, spider, butterflies, fly, sheep, bovine, ovine race

The present invention also relates to the resulting cells and their uses, for example for production of proteins or other genes, biomolecules, biomaterials, transgenic plants, vaccines, transgenic animals or for treatment or prophylaxis of a condition or disorder in an individual (e.g. a human or other mammal vertebrate) arising as a result of a genetic defect. For example, cells can be produced (e.g. ex vivo) by the methods described herein and then introduced into an individual using known methods. Alternatively, cells can be modified in the individual (without being removed from the individual).

Therefore, the invention further relates to a method of production of proteins, biomolecules, biomaterials, transgenic plants, vaccines, transgenic animals or to method of treatment or prophylaxis of a genetic disease in an individual in need thereof, wherein such method comprises the process of *in vivo* linearisation or excision according to the present invention.

The present invention concerns a method for producing a non-human transgenic animal, wherein embryonic stem cells are tranfected by the method according to the present invention and optionally screened for random integration event, the cells are injected into embryos at a stage at which they are capable of integrating the transfected cells, for example at the blastocyst stage, the embryos are then reimplanted in a surrogate mother, and the chimeric individuals obtained at the end of gestation, and in which colonization by embryonic stem cells of the germ line is observed are mated to obtain transgenic animals.

38

Furthermore, the invention also concerns a method for producing a non-human transgenic animal, wherein fertilized eggs are tranfected by the method according to the present invention. If necessary, the eggs are reimplanted in a surrogate mother, and the transgenic individuals obtained at the end of gestation. Optionally, the eggs can be screen
5 for randon integration event before the reimplantation in a surrogate mother. Otherwise, the eggs are incubated in condition allowing the growth of the embryo and the generation of the transgenic animal.

For example, fish embryos or embryonic cells can generally be obtained by collecting eggs immediately after they are laid. Depending on the type of fish, it is
10 generally preferred that the eggs be fertilized prior to or at the time of collection. This is preferably accomplished by placing a male ans female fish together in a tank that allows egg collection under conditions that stimulate mating. After collecting eggs, it is preferred that the chorion is removed from the embryo before the embryo exposition for introducing the vectors. This can be done manually or, preferably, by using a protease such as
15 pronase. A fertilized egg prior to the first cell division is considered as one cell embryo. Therefore, the fertilized egg is considered as an embryonic cell. Preferably, a host cell according to the present invention is an embryonic cell of fish, preferably a fertilzed egg or a blastomere from an embryo at the blastula stage. Most preferably, the host cell is a fertilized egg.

20 Vectors comprising the polynucleotide to be linearized or excised according to the present invention can be introduced into embryonic fish cells using any suitable technique (e.g., transformation, transfection, direct uptake, projectile bombardment, using liposomes). Many techniques for such introduction of exogenous genetic material have been demonstrated in fish and other animals. These include microinjection (Culp, 1991,
25 supra), electroporation (Inoue et al., 1990, Cell. Differ. Develop. 29, 123-128; Müller et al. 1993, FEBS Lett. 324, 27-32 ; Murakami et al. J. Biotechnol. 34, 35-42 ; Müller et al. 1992, Mol. Mar. Biol. Biotechnol. 1, 276-281 ; Symonds et al. 1994, Aquaculture 119, 313-327), particle gun bombardment (Zelennin et al., 1991 FEBS Lett. 287, 118-120), and the use of liposomes (Szelei et al., 1994, Transgenic Res. 3, 116-119) (the teaching of these
30 documents is incorporated herein by reference). Preferably, the vectors are introduced by microinjection.

Cleaving agent, preferably endonuclease can be introduced into an embryonic fish cell according to methods generally known in the art which are appropriate for the particular endonuclease and cell type. Preferably, the cleaving agent is co-injected with
35 the vector comprising the transgene.

After introduction of the vector(s) according to the present invention, the embryo is allowed to develop into fish. This generally involves no more than incubating the embryos

under the same conditions used for incubating egg. However, the embryonic cells can be also be briefly incubated in an isotonic buffer. If appropriate, expression of the transgene can be observed during the embryo development.

Fish or fish embryo harboring the transgene can be identified by any suitable

5   means. For example, the fish genome can be probed for the presence of the transgene by Southern or Nothern blotting. The presence of the transgene can also be detected using PCR or other sequence-specific nucleic acid amplification techniques. In order to identify the expression product of the transgene, the presence of this product can be assayed.

The transgenic fish can provide a fish with the desired characteristic such that the

10 fish or a descendent of the fish has that desired characteristic. Examples of a desired characteristic include enhanced and/or novel nutritional value, disease resistance, growth enhancement (faster growth, increased body size or increased litter size) or production of a desired protein. By desired protein is meant a protein that bestows a desired trait on the fish in which it is produced or a protein which when isolated from the fish is desirable for

15 uses outside of the fish. The desired protein may be produced in a specific tissue, a subset of tissues or in wide range of tissues. Examples of desired proteins include proteins which correct an abnormal condition in the fish or therapeutic proteins for the fish or some other animal.

Transgenic fish can be used as study model. The pattern of expression of a gene,

20 more particularly a fish gene can be study with such transgenic fish by measuring or identifying expression of the transgene in different tissue (tissue-specific expression), at different times during development (developmentally regulated expression or developmental stage-specific expression), in different cell lineages (cell lineage-specific expression). Transgenic fishes are also useful for identifying compounds or genes that

25 affect the expression of a gene which has been introduced by transgenesis, preferably a fish gene. One way to analyze the expression in the different fish tissue is to dissect the fish and assay in the separate tissue sample. RNA can be detected using any of numerous nucleic acid detection techniques. The use of reporter proteins is preferred since these proteins are easily detected. A preferred way of assaying the expression

30 pattern of the transgene during the development is to use an expression product that can be detected in living embryos or animals such as GFP or DsRed.

In one application of the present invention, introduction of the transgene can create an insertional mutation in the fish. Such mutated genes are easely cloned because the inserted transgene serves as a tag for cloning. Genes affecting any process of interest

35 that can be detected can be identified by such insertional mutagenesis. Moreover, in another application, the transgene is a reporter gene used in a gene trap construct. By gene trap is meant a reporter gene that can only be expressed after the DNA integrates

40

into an active gene in the host cell, in this case a fish cell. Gene trap vectors are particularly useful for identifying insertions into active genes. (Kitajima et al., *Biochem Cell Biol* 1998 **76**:1029-37 ; Zambrowicz et al, *Int J Dev Biol* 1998 **42** 1025-36 ; Cecconi et al, *FEBS Lett.* 2000 **480** 63-71 ; the teaching of these documents are incorporated herein by
5   reference)

The process of the invention is thus quite particularly useful for the production of recombinant protein fish, production of transgenic fishes as models. As a consequence, the invention has for object a method of fish transgenesis consisting in administering to said embryonic fish cell a composition comprising one or several vector comprising the
10 polynucleotide to be excised and, optionally the cleaving agent, as defined previously, possibly associated to a substance susceptible to allow or to facilitate the transformation of one or several cellular types of said organism by said vector, then to generate in said cells the linear molecule ready for be integrated by subjecting said cells to the action of the cleaving.

15 The present invention also relates to the resulting cells and their uses, for example for production of transgenic fish as model or food, or for producing proteins or other genes, biomolecules, biomaterials, and vaccines.

Therefore, the invention further relates to a method of production of proteins, biomolecules, biomaterials, transgenic plants, vaccines, transgenic fish, wherein such
20 method comprises the process of *in vivo* or *in ovo* linearisation or excision according to the present invention.

Furthermore, the invention also concerns a method for producing a transgenic fish, wherein fertilized eggs are tranfected by the method according to the present invention and the eggs are incubating in conditions allowing its development in transgenic fish.
25 Optionally, the eggs can be screen for randon integration event before their incubation.

The informations detailed for the fish are applicable to the non mammalian species such as xenopus, shrimp, sea urchin, acsidies, birds, chicken, amphoxius.

The invention also contemplates a method of transgenesis in which a complete genetic material (the DNA contained in a nucleus) is transferred together with the vector
30 comprising the polynucleotide to be integrated in the host genome and the cleaving agent in an unfertilized egg whose own nucleus has been removed.

The invention still has thus for object a pharmaceutical composition comprising as active agent a vector comprising the polynucleotide to be linearized or excised according
35 to the present invention, possibly associated in the pharmaceutical composition to a substance susceptible to allow or to facilitate the transformation of one or several cellular types by said vector. One embodiment of the pharmaceutical composition according to the

invention consists in associating said composition in a transgenesis kit with a composition containing the cleaving agent, preferably a restriction endonuclease.

The invention has finally for object the use of a vector comprising the polynucleotide to be linearized or excised according to the present invention, for the
5   preparation of a pharmaceutical composition intended to transform one or several cellular types of a subject in which said composition is administrated, said cells being exposed after transformation to the action of the cleaving agent, preferably a restriction endonuclease. As indicated previously said vector is possibly associated in the pharmaceutical composition according to the invention to a substance susceptible to allow
10 or to facilitate the transformation of one or several cellular types by said molecule.

The invention concerns a method of treating or prophylaxis of a genetic disease in an individual in need thereof by random integration of a polynucleotide into the individual genome according to the present invention. A genetic disease is often due to a low expression of a gene or by the expression of a mutated gene which is not functional. For
15 example, by the integration of a correct gene with the appropriate elements for its expression, the gene can be expressed and the genetic defect can be compensated. In case of a genetic disease due to an overexpression of a gene, the integration of a polynucleotide allowing the expression of an antisense, for example, could impede the expression of this gene and compensate the genetic defect due to the overexpression.
20 The method according to the present invention can be applied to any one of contemplated gene therapy.

Other advantages and characteristics of the invention will appear in the following examples

25   **EXAMPLES**

EXAMPLE 1 : Method of in vivo linearization in human cells

(see Figures 4 and 5)

Construction of pVirtkU$_3$R□geo (see Figure 4)

pU$_3$R plasmid was obtained by the insertion in pBS KS+ plasmid (Stratagene, #
30 X52328) of the 990 bp fragment from the segment 7840-8330 of the murine leukemia Moloney virus (MoMULV) comprising the U$_3$ sequence from the 3' Invert Repeat sequence following by the R sequence. The U$_3$R fragment comprised the enhancer and promoter sequences of MoMULV and the polyadenylation signal AATAAA contained in the R sequence. The pVirU$_3$R plasmid was obtained in two steps. Firstly, the sequence 5'-
35 ATTACCCTGTTATCCCTAATGAAAGA-3' comprising the recognition and cleavage site of the I-Sce I endonuclease followed by the Invert Repeat of MoMULV virus was inserted

upstream of the U$_3$R sequence of the pU$_3$R plasmid. This I-Sce I endonuclease site is called in this example " 5' I-Sce I endonuclease site". Secondly, the sequence 5'-TAGGGATAACAGGGTAATTTACTTTCA-3' comprising the Invert repeat of MoMULV virus followed by the recognition and cleavage site of the I-Sce I endonuclease was

5   inserted downstream of the U$_3$R sequence of the intermediate plasmid. This I-Sce I endonuclease site is called in this example " 3' I-Sce I endonuclease site". The orientation of the I-Sce I endonuclease sites was done so as to keep the small site (TAGGGAT or ATCCCAT) towards the U$_3$R sequence. Therefore, the site orientation was opposite in order to avoid the generation of cohesive ends. This orientation is called "the small site

10 towards the insert center". pVirU$_3$Rβgeo plasmid was obtained by the insertion in the plasmid pVirU$_3$R of a 5,3 kb *Xho I* restriction fragment comprising the fusion gene *β–geo* from pRSVβgeo plasmid (Friedrich and Soriano, 1991, *Genes Dev*, 5, 1513-1523; the disclosure of which is incorporated herein by reference) (*lacZ* gene fusion, encoding the bacterial β-galactosidase gene fused in phase at its 3' end with the *neo* resistance gene

15 encoding the neomycine phosphotransferase) following by the polyadenylation site of SV40 virus at the *Xho I* site located between the R sequence and the Invert Repeat of "3' I-Sce I endonuclease site". In the pVirU$_3$Rβgeo plasmid, the βgeo expression was controled by U$_3$ promoter of MoMULV. The SV40 polyadenylation site is directed towards the "3' I-Sce I endonuclease site". The pVirtkU$_3$Rβgeo plasmid was obtained by the

20 insertion in pVirU$_3$Rβgeo upstream of the "5' I-Sce I endonuclease site" of 2.5 kb *BamHI – NsiI* restriction fragment of pHSVTK plasmid (Mansour et al, 1988, *Nature*, 336, 348-352; the disclosure of which is incorporated herein by reference) comprising a gene encoding the thymidine kinase (tk) of Herpes simplex virus HHV1 under the control of its own promoter and without its polyadenylation site. The expression of the *tk* gene uses the

25 polyadenylation site of the MoMULV virus R sequence. See figure 4

Use of pVirtkU$_3$Rβgeo system for the validation of the in vivo linearization method

pVirtkU$_3$Rβgeo was cotransfected in the target organism with the I-Sce I endonuclease expression vector pCMV-I Sce I + or pCMV-I Sce I - (Choulika et al, 1995

30 *Mol Cel Biol*, 15, 1968-1973; the disclosure of which is incorporated herein by reference). pCMVI-SceI+ is a vector in which the ORF encoding the endonuclease has a good orientation relative to the CMV promoter for the expression of I-Sce I endonuclease. pCMVI-SceI- refers to a plasmid construct in which the ORF encoding the endonuclease has a reverse orientation relative the CMV promoter and, therefore, does not lead to the

35 endonuclease expression.

If pVirtkU$_3$Rβgeo plasmid was cleaved *in vivo* by I Sce I endonuclease, the cleavage occurred upstream and dowstream of the Invert Repeat sites, namely at the

recognition and cleavage sites of the I-Sce I endonuclease. If pVirtkU$_3$Rβgeo plasmid was cleaved *in vivo* by I-Sce I, the MoMULV promoter contained by U$_3$ activated the βgeo gene transcription. If pVirtkU$_3$Rβgeo plasmid was not cleaved, the transcription of the cistron expressing *tk* gene ended by the polyadenylation in the R sequence located

5   downstream of U3 and prevented the βgeo expression. The integration selection of the *in vivo* linearized fragment was done either in two steps or only one step. In two steps selection; firstly, a selection with 1 μM of ganciclovir in order to eliminate clones presenting an integration of the whole plasmid ; and secondly, a selection with 50 μg/ml of G418 in order to select the clones expressing βgeo . In one step selection, a selection

10 with 50 μg/ml of G418 in order to select the clones expressing βgeo. See figure 5.

Construction of pU$_3$Rβgeo control plasmid

pU$_3$Rβgeo plasmid was obtained by the insertion in pU$_3$R of a 5,3 kb *Xho I* restriction fragment comprising the fusion gene *β–geo* from pRSVβgeo plasmid (Friedrich

15 and Soriano, *supra*) following by the polyadenylation site of SV40 virus at the *Xho I* site located downstream the R sequence. In the pU$_3$Rβgeo plasmid, the βgeo expression was controled by U$_3$ promoter of MoMULV. The SV40 polyadenylation site is directed towards the 3' I-Sce I endonuclease site. See figure 4.

Preparation of the linear fragment U$_3$Rβgeo as linear control

20   The linear U$_3$Rβgeo fragment was isolated after digestion of 10 μg of ptkU$_3$Rβgeo plasmid by *Xba I* and *Not I* by electrophoresis on 0.8 % agarose gel with TAE 1X. A 9.7 kb fragment was isolated and purified on glass beads with Gene-Clean II's protocol. The purified U$_3$Rβgeo fragment was resuspended in H$_2$O pH 7 for transfection. See figure 4.

25 ### Transfection protocol of human 293T cell

20 h before transfection, 293T cells were seeded at a density of 5x10^4 cells per 35-mm dish on DMEM (Dulbecco Modified eagle Medium) medium with 10 % inactivated fœtal veal serum (SVFi). Cells are incubated in an incubator of 1 atm, 5 % CO$_2$, 100 % humidity, 37°C. The transfection was done by phosphate calcium precipitation with a

30 mixture of 125 mM CaCl$_2$, NaCl 8.18 ‰ (W/V), HEPES 5.94 ‰, Na$_2$HPO$_4$ 0,2 ‰ pH 7.12 at 25°C in presence of 2 μg DNA. Precipitates were maintained 16 h on cells. Then, the culture medium with the precipitate was replaced by fresh DMEM 10 % SVFi medium.

48 h after transfection, cells were fixed in a paraformaldehyde 4 % solution during 5 min at 25 °C. Cells were washed twice in PBS x1 (Phosphate buffer saline). Then cells

35 were placed in X Gal coloration medium during 24 h at 37°C for hystochemical staining by β-galactosidase expression of βgeo gene. Otherwise, cells were placed in selection

medium DMEM 10 % SVFi with 60 µg/ml of G418. The selection medium was changed by new one every 48 h. The selection was maintained during 15 days until the emergence of isolated cellular clones. These clones were isolated on plates of 24 wells of 1 cm diameter, and were amplified in order to be freezed and analyzed.

5   ***Results of transfection of 293T cells by pVirtkU₃Rβgeo***

2 µg of DNA mixture (pVirtkU$_3$Rβgeo + pCMV-I Sce I) were transfected at a density 5×10^4 cells per 35-mm dish with the following ratios and in presence of the following transfection controls which are disclosed in Table 1.

ADN / Transfection Number	pCMVI-SceI+	pCMVI-SceI-	pVirtkU₃Rβgeo	pU₃Rβgeo	U₃Rβgeo
# 1.0	0,5µg	-	1,5µg	-	-
# 1.1	-	0,5µg	1,5µg	-	-
# 2.0	0,5µg	-	-	1,5µg	-
# 2.1	-	0,5µg	-	1,5µg	-
# 3.0	-	-	-	-	1,5µg
# 3.1	0,5µg	-	-	-	1,5µg
# 3.2	-	0,5µg	-	-	1,5µg

10   Table 1 : 35mm-dish were transfected by DNA concentration and in mixture according to the table. Each assay was done four times, two times for histochemical staining and two times for clonal selection. Each # number corresponds to a transfection mixture, columns which indicates the plasmid constructs used in the mixture.

15   48 h after transfection, cells were fixed and β-galactosidase expression was revelated by histochemical staining. Clones showing β-galactosidase expression (β-gal clones) were counted (see Table 2). Furthermore, cells were selectionned on a culture medium comprising 60 µg/ml of G418. Cell culture medium were changed by fresh medium each 48 h of incubation. After 15 days of post-transfection, clones resistant to
20 G418 were isolated and amplified (see Table 2).

Transfection number #	1.0	1.1	2.0	2.1	3.0	3.1	3.2
X-gal							
Cells + / dish	38454	10742	11023	10908	7439	8243	8899
Clones + / dish	10751	2684	2536	2923	2613	2522	2866
G418							
Clones / dish	74	6	3	9	2	2	3

Table 2 : Results of 293T transfection by the mixtures according to Table 1.

Results show a significant increase of the construct integration for Transfection
25 number 1.0. This transfection number corresponds to the transfection of a plasmid

45

comprising the polynucleotide to be integrated, namely U₃R-βgeo-polyA$_{SV40}$, flanked by two I-Sce I endonuclease sites with a plasmid comprising an expression cassette for I-Sce I endonuclease. The integration of the plasmid with *in vivo* linearization was about 4 to 10 times more effective than the linear fragment U₃Rβgeo.

5

Genomic DNA of the isolated clones was extracted and analyzed by Southern hybridization (digestion of genomic DNA with *EcoR* V) with probes comprising a restriction fragment *EcoRI* of 3 kb containing *lacZ* gene from plasmid pGemnlslacZ (pGem vector from Promega in which has been inserted the lacZ gene and a nuclear localization

10 sequence). See figure 6. Southern analysis showed that the integration is unique and not in tandem. Indeed, a tandem integration would show a 2,6 kb band which was never seen. Sometimes, like in line 12, several integration events occurred. Therefore, the *in vivo* linearization method makes the integration unique.

Some experiments were done with a cotransfection with a plasmid expressing an

15 integrase (+IN). The integrase used in this experiment is the active fragment of the MoMULV virus. Indeed, the nucleic acid encoding the active fragment has been cloned in an expression vector. The integration showed an increase of about 6 times compared with the *in vivo* linearization method without any integrase. Therefore, optionally the method for integrating a polynucleotide into the host cell genome according to the invention also

20 comprises the introduction into host cell of integrase itself or a nucleic acid encoding integrase.

The integration site of the integrated constructs in 24 different isolated clones was cloned by reverse PCR and analyzed by sequencing. The junctions between the integrated constructs and the genomic DNA were caracterized. These junctions showed

25 that the integrity of the polynucleotide to be integrated is conserved and that the insertion is unique.

EXAMPLE 2 : Method of in vivo linearization in COS cells
Construction of ppSV40NeoIRESegfp plasmids

30 The ppSV40NeoIRESegfp intermediate plasmid was obtained in two steps. First, the cassette containing the Simian virus 40 promoter (pSV40) and the *neo* resistance gene encoding the neomycin phopshotransferase was excised from the pcDNA3.1+ plasmid (Clonetech, Palo Alto, CA, USA) and cloned in the pIRES2-EGFP plasmid (Clonetech, Palo Alto, CA, USA) at the SmaI site, resulting in the insertion of the pSV40-

35 Neo cassette upstream of the IRES bicistronic element and EGFP gene. Second, the PyTknlslacz cassette from the pPytknlslacz plasmid (Henry *et al., C. R. Acad. Sci. III*,

322(12): 1061-1070 (1999)) was excised and replaced by the entire pSV40-Neo-IRES-EGFP (ppSNIE) cassette and was inserted before the SV40 polyA site.

From this initial construct, a plasmid was derived by inserting the sequence corresponding to the I-Scel cleavage site both upstream of the SV40 promoter (5') and
5   downstream of the SV40 polyA site (3'). See Figure 7A.

Transfection protocol of COS cells

A day before transfection, 5×10^4 COS cells were seeded in 6-well plate in DMEM (Dulbecco's Modified Eagle Media) media supplemented with 10% Fetal Calf Serum, 2
10   mM L-Glutamine and incubated at 37°C in a 7% CO_2 atmosphere. The transfection was done by the Effectene method (Qiagen, Chatsworth, CA, USA). The transfection protocol has been set up according to the manufacturer recommendations. Briefly, 1 µg of plasmid DNA was diluted in 100 µl DNA-condensation buffer in the presence of Enhancer reagent (ratio DNA/Enhancer is 1/8). After a 5 min. incubation period, 10 µl of Effectene is added
15   to the DNA (ratio DNA/Effectene is 1/10). The mix is vortexed and incubated for 10 min. at room temperature. 600µl of complete media is then added to the DNA/Effectene complexe, mixed and sprinkled over cells. Cells are washed the next day and incubated for a total period of 48h at 37°C, 7% CO_2.

A FACS ™ analysis is performed on an aliquot of the transfected cells using a
20   FACscan flow cytometer and CellQuest software (Becton-Dickinson, Franklin Lakes, NJ, USA). The remaining cells are plated on a 10 cm dish and selection media containing G418 (Life Technologies,) at the concentration of 400 µg/ml is added. Alternatively, cells were resuspended at the density of 2000 cells/ml and 100µl of suspension were seeded in 96-well plates. After 3 weeks, resistant clones were isolated and amplified in 24-well
25   plates for further analysis.

Genomic DNA extraction, quantitation and Southern blot hybridization.

Genomic DNA was extracted from cultured cells using the High Pure DNA Prep Kit (Roche, Mannheim, Germany) and was quantified using the Picogreen™ dsDNA
30   quantitation Kit (Molecular Probes, Eugene, ON, USA) according to manufacturers' protocols.

1µg of genomic DNA from independent clones was digested with HindIII restriction enzyme (NEB, Beverly, MA, USA). After DNA electrophoresis and blotting on Hybond-N+ membranes (Amersham, Pisscataway, NJ, USA), blots were probed with the NEO probe
35   using a non-radioactive DIG based nucleic acid detection protocol (Roche, Mannheim, Germany).

47

Results of transfection of COS cells by ppSNIE plasmids

ppSNIE plasmid was co-transfected with the I-SceI endonuclease expression vector pCMV-I-SceI+ (Choulika et al, 1995 Mol Cel Biol, 15, 1968-1973). As negative controls, ppSNIE plasmid was cotransfected with pCMV-I-SceI- in which the ORF

5 encoding the endonuclease has a reverse orientation and, therefore, does not allow the I-SceI expression. 1 µg of DNA mixture (0.5 µg of ppSNIE plasmid + 0.5 µg pCMV-I-SceI-) or (0.5 µg of ppSNIE plasmid + 0.5 µg pCMV-I-SceI+) were used to transfect 5×10^4 cells in 6-well plates. 48h post-transfection, cells were harvested and analyzed by flow cytometry. Reproducibly, transfection rate of COS cells was between 35 and 45% of

10 GFP+ cells. Thus, co-transfection of ppSNIE plasmid with I-SceI expression (pCMV-I-SceI+) or control (pCMV-I-SceI-) vectors did not affect the transient expression of EGFP reporter.

Remaining cells were plated on 10cm dishes and selecting media was added. Three weeks later, resistant clones were isolated and amplified. Genomic DNA from

15 independent clones was extracted and analyzed by Southern hybridization. 0.5 µg of genomic DNAs were digested with HindIII restriction enzyme, separated on agarose gel and transferred on nylon membrane. Blot was then probed with the Neo probe.

The results are shown in Figure 7B. As expected, all but clones #3 and #26 showed a 1.5 kb band corresponding to the HindIII fragment within the ppSNIE cassette.

20 The intensity of the 1.5 kb band is highly variable in clones from I-SceI- co-transfection with the presence of huge concatemers (see clones #5, 6, 8 and in a lesser extent clones # 11, 15, 16). In contrast, bands in clones from I-SceI+ co-transfection are less intense suggesting the presence of a few copy number.

25

EXAMPLE 3 : In ovo linearization by the I-SceI meganuclease boosts transgene integration in fish eggs.

Circular plasmidic DNA, bearing a muscle specific promoter followed by a GFP

30 reporter gene and flanked by two I-Sce I recognition sites, was injected in one-cell stage fish embryos. When this construct was co-injected with the I-Sce I meganuclease in a magnesium free buffer (so that the enzyme remains inactive extracellularly), the expected GFP muscular expression was spectacularly improved (80% of the embryos exhibited strong fluorescence in the trunk musculature, as compared to 20% when the injections

35 were performed without the meganuclease). Even more striking was the dramatic increase in germline transmission. Whereas in classical egg injection experiments, it does not, in most cases, exceed a few percents, due to late chimeric integrations, the frequency

of germline transmission in fish co-injected with the meganuclease was boosted to 50%, suggesting that a single insertion occurred at one-cell stage of the founder fish. Further Southern analysis confirmed that single low copy integrations occurred in most cases. To assay whether early integration events may be due to *in vivo* cleavage of DNA, and not to

5   its targeting to the nucleus by the meganuclease, control constructs bearing deleted or mutated recognition sites (bound, but not cleaved, by the I-*Sce* I meganuclease) were injected with the meganuclease. We propose that meganuclease-induced *in ovo* linearisation, by limiting degradation of preintegrative DNA free ends, is a simple and efficient process to improve transgenesis by egg injection.

10   This example shows that meganuclease-mediated transgenesis is indeed a very simple technique that is spectacularly more efficient than the other methods currently reported in fish.

Materials and methods

Plasmid constructs

15   The paact-GFPM2 was constructed with two I-SceI recognition sequences in a 7,8kb plasmid bearing an EGFP reporter gene driven by an α-actin muscle specific promoter (paact-GFP, gift of Dr. Higashijima, [Higashijima, 1997 #698]). Briefly, two sub-cloning steps were used to insert « megalinkers » at the Eco RI and KpnI sites, located on both sides of the α-actin/GFP/polyA cassette in the Bluescript polylinker. Megalinkers

20 were generated by annealing complementary oligonucleotides containing the I-SceI recognition site (CCGCTAGGGATAACAGGGTAATATA) flanked by free ends compatible with either of the EcoRI or KpnI digest products. After ligation of the megalinker with the linearized plasmid, DNA was digested again by the enzyme used to linearize the plasmid and transformed by heat pulse into XL1 Blue ultra-competent E. coli (Stratagene). Clones

25 were sequenced to determine the orientation of the non-palindromic I-SceI site.

Several other constructs were obtained by inserting different linkers at the KpnI site: paact-GFPM with only one I-SceI recognition site, paact-GFPDM with a shortened recognition site (GGGTAATATA), and paact-GFPMM containing a mutated (TAGGGtTAACAGGGTAAT) version of the I-SceI site. These later two sites bind the

30 meganuclease but are not cleaved efficiently [Colleaux, 1986; Colleaux, 1988].

Microinjection of plasmid DNA with meganuclease

Medaka embryos and adults of an Orange-Red strain (kindly provided by A. Shima, Tokyo University, Japan and Y. Ishikawa, Chiba, Japan) were used in all experiments. Fish were raised in 20 litre tanks at 26°C. Adults were placed under a

35 reproduction regime (14h light/10h dark). Eggs were collected immediately after spawning (at the onset of light), cleaned and placed in Yamamoto's embryo rearing medium [Yamamoto, 1975]. For injection, one-cell stage embryos with a recently formed blastodisc

49

were selected [Iwamatsu; 1994], and transferred at 4°C to stop development. Approximately ten embryos at a time were positioned, cytoplasmic discs to the top, in agarose slots made with a plastic mold as described in [Westerfield; 1993]. In all experiments, an Olympus SZX12 stereomicroscope equipped with a micromanipulator

5 MM3 (Fine Science Tools, Germany) and a pressure injector (FemtoJet, Eppendorf, Germany) were used. Borosillicate glass capillaries (1mm outer diameter, GC100T, Clark Electromedical Instruments, UK) were pulled using a vertical puller (PC-10, Narashige, Japan). Capillaries were backfilled with the injection solution (DNA: 10 µg/ml; meganuclease buffer (Roche Diagnostic, Germany): 0,5x; meganuclease I-*Sce*I: 1unit/µl;

10 0,1% phenol red) using sterile microloaders (Femtotips, Eppendorf, Germany). Immediately before injection, the micropipettes tips were broken to about one micrometer in diameter with fine forceps. DNA was injected through the chorion by inserting the pipette directly into the thin blastodisc. One pressure pulse resulted in the injection of a droplet (estimated volume 300pl) visualised by the phenol red. Embryos were then

15 removed from agarose and placed in Petri dishes at 26°C. GFP expression was first screened at three days after injection. Embryos were then raised to sexual maturity and pair-wise crosses were done to identify fish that had transmitted the transgene to their progeny.

Southern blot analysis

20   For southern blotting, genomic DNA from F1 fish was extracted using proteinase K and phenol [Sambrook and Russel; 2001]. To get enough DNA, whole one month old fish were used. To avoid DNAses, muscles from older fish were dissected out for DNA extraction. DNA was digested to completion with BamHI, EcoRI, Xbal, XhoI. DNA standards were prepared by digesting the pαact-GFP plasmid by the corresponding

25 enzymes. Using an estimated DNA content of 10^9 bp per haploid genome, 50pg of digested pαact-GFP plasmid were loaded. Samples (10□g per lane) were electrophoresed in 0,9% agarose in 1xTBE and blotted onto a nylon Zetaprobe membrane (Biorad) by upward capillary transfer, as recommended by the Zetaprobe manufacturer. Filters were hybridized in a roller bottle overnight at 65°C with random-primed

30 radiolabeled probes (XhoI/EcoRV fragment) corresponding to ãactin/GFP/polyA .

Epifluorescence microscopy

       Embryos were observed and scored using a MZFLIII Leica dissecting microscope using a 370 to 420nm excitation filter and a 455nm LP emission filter. Photographs were acquired using a Nikon DXM1200 digital camera. For photography, embryos were

35 dechorionated with hatching enzyme [Yasumasu; 1994], following the procedure described in [Wakamatsu; 1993] and placed in 80% glycerol in PBS or in Vectashield mounting medium.

Results

Transient expression of GFP in muscles of injected embryos is improved when the meganuclease is co-injected with the construct

Circular plasmid DNA, bearing the pαact-GFPM2 actin promoter followed by a
5   GFP reporter gene and flanked by two I-SceI recognition sites (pαact-GFPM2) was first tested for expression by injecting it in one-cell stage embryos (stage 2a, [Iwamatsu; 1994]). The DNA concentration (13ng/µl) was chosen as the highest one which did not lead to a significant mortality of the injected eggs.

GFP reporter gene expression in the embryos was examined at a number of
10 distinct developmental stages using fluorescence binocular microscopy. Onset of GFP expression was first observed in a few embryos after two days (stage x).

Generally, embryos were scored after three days of development when the muscular α-actin GFP expression was well underway. The embryos were grouped according to the intensity of fluorescence, in order to quantitatively estimate the level and
15 distribution of transgene expression in each experiment.

About 50 percent of the embryos showed no muscular fluorescence and were classified as negative (N) (Fig. 8B). In the other embryos, the number of GFP positive cells ranged from a few cells (classified as weak, W), to almost an almost ubiquitous muscle cells labelling (strong, S). When the expression was detected in a large domain of
20 the tail, but nowhere else, the expression was qualified as moderate (M). Expression in individual muscle cells was always strong enough to be easily detectable, and no ectopic expression was observed.

When pαact-GFPM2 was linearized in vitro with I-SceI, purified from an agarose gel and injected in the same conditions as reported above, results were similar to those
25 obtained with circular pαact-GFPM2 (Fig. 8B).

When pαact-GFPM2 was co-injected with the commercially available I-SceI meganuclease in a magnesium free buffer (so that the enzyme remains inactive extracellularly), the GFP muscular expression was spectacularly improved. About 80% of the embryos exhibited a moderate or strong expression in the trunk musculature, as
30 compared to 20% when the injection was performed without the enzyme.

Thus, transitory transgene expression in F0 fish was readily and efficiently improved by the meganuclease protocol, by a mechanism that presumably involves an in vivo linearisation step of the injected circular plasmid.

Improvement in F0 transgene expression is linked to in ovo linearisation

35   I-SceI meganuclease was known to have some NLS activity. To assess whether early integration events were indeed due to an in vivo cleavage of DNA, and not to its targeting to the nucleus by the meganuclease, control constructs bearing deleted (pαact-

GFPDM) or mutated (pαact-GFPMM) recognition sites (bound, but not cleaved by the meganuclease) were co-injected with the meganuclease.

Following injection, embryos were grouped according to the criteria described above (Fig. 8B). The distribution of the transient GFP expression in F0 embryos co-
5 injected with pαact-GFPDM or pαact-GFPMM and the meganuclease was similar to that obtained in experiments involving the injection of the pαact-GFPM2 construct with no enzyme. Therefore, the meganuclease *in ovo* linearization is the key process that improves transient expression in the injected one cell-stage fish embryos.

Generation of germline-transmitting fish using meganucleases

10   GFP muscular expression persisted in some adults resulting from the injection experiments. An important point was therefore to investigate if and how the transgene was transmitted to the next generation, and, in particular, whether there was any correlation between levels of expression in adult muscles and the frequency of germline transmission.

15   In a first screen on 30 injected adults, those exhibiting no GFP expression in muscles all turned out to be negative for germline transgene transmission. Thus, adult GFP-negative fish were *a priori* considered as negative, and subsequently discarded from the F0 mating procedures. The observation of GFP expression in adult F0 thus greatly simplified the screen for transmitting fish.

20   To investigate the effects of meganuclease on the frequency of appearance of germline transmitting founders, injected fish were raised to sexual maturity and mated to wild-type partners. The level of muscular fluorescence in their three-day-old progeny was then scored as described above (results summarised in Figure 9A).

As expected, founder fish resulting from the injection without the meganuclease
25 had mosaic germlines, and, in most lines, rates of F1 inheritance (as assayed by GFP expression) was dramatically low (transmission to only a few percents of the siblings). In a few lines (with high copy number insertions; see below) the transgene was transmitted at a moderate rate (30 to 50 percents).

In sharp contrast, the frequency of germline transmission for founders co-injected
30 with the circular plasmid and the meganuclease was boosted to 50% in most lines (Fig. 2A), corresponding to a 5- to 15-fold increase in the mean frequency of positive embryos in the progeny.

Further Southern analysis confirmed that single integrations at a low copy number occurred in most cases (Fig. 9B).

35   DNA was first digested with EcoRI or XhoI, enzymes that does not cut or cuts once, respectively, into the plasmid construct (not shown). All the animals analyzed contained a single fragment of large size which hybridizes with the insert probe,

52

suggestive of a single integration event. We cannot exclude, however, the existence of several such events, because BamHI or XbaI digestions yielded multiple junction fragments. With BamHI digestion, as represented in Figure 9B, two (1 and 2kb) fragments hybridising with a probe corresponding to the insert were observed in lane 1 to 4, in which

5   DNAs from fish injected with the meganuclease were loaded. These fragments corresponded respectively to the GFP/pA and to the downstream region of the actin promoter (3'αp). The presence of a 4,8kb band, diagnostic of the upstream region of the promoter linked with the plasmid sequence (5'αp+pBluSK) suggested that some transgenes were integrated as direct tandem repeats including the plasmid. This type of

10 integration is expected to occur if only one of the two I-SceI recognition site is cut and the insert readily integrated (with no further cut).

Discussion

This example reports on a meganuclease-mediated technology which leads to spectacular improvements in fish transgenesis at several levels.

15   First, the transitory expression (in F0) of an injected DNA construct is improved, due to a large increase in the number of expressing cells. Thus, mosaic expression of the reporter gene in injected embryos is greatly diminished, partly overcoming one of the main pitfalls of transgenesis in fish. This result opens the way, for example, to easier and more reliable analysis of promoter constructs in fish species, a task that often involves their

20 preliminary transient introduction by microinjection.

Second, meganuclease-mediated technology significantly increases the rates of germline transmission in fish. Moreover, with the muscle specific promoter used, there was a good correlation between the maintenance of the reporter expression in adult injected fish and the ability to produce transgenic descendants. Therefore, the tools

25 presented in this example greatlty simplify the much time-consuming task of selecting the F0 individuals able to found a transgenic family.

Third and maybe most striking, the rate of transmission in each transmitting family was dramatically boosted in these experiments. In practice, that means that large F1 families can generated from a reasonable number of eggs. Importantly, the transmition

30 rate of 50% observed in most crosses between transmitting F0 and wild type fish strongly suggests that a single insertion event occurred in a non-mosaic way in F0.

The control experiments allow us to propose that *in vivo* linearisation is instrumental in the increase of the integration rates, albeit not necessarily the only factor. For example, a nuclear targeting effect of the I-*SceI* enzyme may be at play in these

35 experiments. We propose that, in a first step, the absence of free DNA ends in the circular plasmid reduces the rate of cytoplasmic degradation of the injected DNA. The late

53

liberation of pre-integrative free ends then leads to the efficiency of integration typical of linear plasmids.

Finally, transgene integration in long concatemers, a feature often encountered in fish, never occurred in these experiments. Instead, we found that the transgene was integrated as short direct tandem repeats, a beneficial effect that is likely to be the reason for the spectacular increase in GFP expression levels.

This kind of insertion was expected because, in the plasmidic contruct, the two ISce-I sites were in the same orientation. Southern blot data also suggest that integration of similar tandemly repeated elements occurred and comprised some plasmid sequences. This indicates that these tandems are composed of several plasmids cut at only one site. Since, in vitro, ISce-I cuts at a similar kinetic at both its sites, one possible explanation for that phenomenon is that I-Scel enzyme amount was limiting with respect to DNA concentration, and that integration occurred concomitantly with the first cut.

To further improve the quality of integration, it can be proposed to change the orientations of I-Scel sites.

Applications of this technique are numerous, first of all in several fish species but also more generally. For example, it paves the route to more sophisticated experiments such as knock-out experiments by DNA injection in early embryos or oocytes.

EXAMPLE 4 : In ovo linearization by the I-Scel meganuclease in mice.

In order to improve transgenesis efficiency, we developped a novel method based on the in ovo linearization of a supercoiled plasmid by meganuclases.

Plasmid constructs have been designed to bear one or two recognition/cleaving sites for meganucleases flanking the desired transgene to be integrated. Meganuclease source was provided either by purified proteins or by *in ovo* production from a meganuclease-expressing vector. Improvement of transgenesis (quantitative and qualitative) was compared with experiments of "classical' transgenesis (i.e.: injection of a prelinearized and purified transgene fragment). Negative control was performed using the same supercoil preparation of transgene DNA, free of any meganuclease source.

In ovo linearization mediated by I-Scel expressing vector

Injections were performed at the one-cell stage in male pronuclei from B6SJL mouse eggs. The I-Scel expression vector was checked in ovo by microinjecting the pI-Scel/EGFP plasmid at the one-cell stage in male pronuclei from B6SJL mouse eggs. EGFP fluorescence was already vizualized at the 2-cell stage in injected embryos on an inverted microscope with EGFP filters. Due to the nature of the construct bearing an IRES in between the two I-Scel and EGFP ORFs, it assured us in about 95% of cases that

when EGFP fluorescence is vizualized, I-SceI is coordinately produced. Megafluo (Figure 10A) and pI-SceI/EGFP (Figure 10B) plasmids were coinjected at five hundred copies/pl. Injected eggs were transferred in B6CBA foster female mouse. Birth was checked, still-born animals were recovered and their DNA extracted in order to be genotyped for a

5   putative integration of the injected transgene. New-born mice were typed at 3 weeks by cutting a small piece of tail. Then, DNA extraction was performed, followed with genotyping by PCR and /or Southern experiments. From this experiment, 5 newborn and 2 still-born were recovered. Following DNA extraction, DNA recovery was checked on agarose gel (Figure 11A). In order to check the quality of the genomic DNA extracted, a

10   PCR amplifying a 494bp sub-fragment of the ß-globin gene was performed (Right panel of Figure 11B). All the DNA samples seemed qualitatively correct as the expected PCR ß-globin product was visualized from all DNA samples tested. In order to test for the integration of the transgene, a PCR amplifying a 484bp of the reporter gene (DSRed1-E5) was performed (Left panel of Figure 11B). Among the 5 three-weeks old mice progeny

15   and the 2 still-born, 2 (number 1 and 25) and 1 (number A) of them were transgenic, respectively. All together, these results showed that 40% of progeny were transgenic when considering alive animals and 43% when considering alltogether alive and still born injected mice.

Usually, when a coinjection of plasmid is performed in mouse one-cell stage eggs,

20   cointegration of both plasmid is detected in injected animals and often at the same location. In order to test for the integration of the I-SceI expressing vector (pI-SceI/EGFP), a PCR amplifying a sub-fragment of this plasmid was performed as well. All the animals recovered (whether the transgenic ones for Megafluo and the non-transgenic) were negative for the presence of the I-SceI expressing vector (Figure 11C).

25   In order to test for transmission of transgene to F1 progeny, transgenic animals 1 and 25 were backcrossed with B6SJL uninjected animals. Among the 13 F1 and the 54 F1 animals derived from female 1 and male 25, respectively, only one F1 animal for each back-cross was transgenic as attested by Southern Blot experiments using a probe hybridizing to the promoter pGas 5. These results demonstrated that a transgenic animals

30   obtained by meganuclease mediated transgenesis were fertile and can transmit the transgene to their progeny. The numbers indicated a transmission rate ranging from 2 to 8% and suggested that founders were mosaic for the integration. This result was not surprising as the production of I-SceI from its corresponding expressing vectors is not suppose to start before the 2-cell-stage in mice eggs when zygotic transcription starts and

35   CMV promoter starts to be active in mouse eggs.

In ovo linearization mediated by I-SceI purified protein

Protocol was equivalent in this experiment except that I-SceI was provided as a purified protein at a 1 :5 DNA/protein ratio. 750 copies of Megafluo supercoiled plasmid (Figure 10A) was coinjected with I-SceI purified protein (18μM). Injected eggs were transferred in B6CBA foster female mouse. Among nine new-born mice, two of them were
5 transgenic as detected by Southern Blot analysis using either a probe corresponding to the Gas5 promoter or the plasmid backbone sequences. This result showed transgenic animals have been produced in this experiment by this method with a yield of 22%.

"Classical transgenesis"

10   The same plasmid construct, Megafluo, was used in a classical protocol for mice transgenesis, i.e., injection of a prelinearized transgene. Two types of experiments were performed, one using NdeI subfragment of 1796bp and the other using I-SceI subfragment of 1824bp, both subfragments bearing the promoter Gas5 and the reporter fluorecent gene. Two series of injection were performed for the I-SceI fragment. The first
15 one corresponding to the injection of 500 copies/pl of the DNA fragment has leaded to the birth of 2 mice who were negative by Red5 PCR. The second one corresponding to the injection of 750 copies/pl of the I-Sce fragment has led to the birth of 8 mice negative for the integration. The injection of 750 copies/pl of the NdeI fragment has led to the birth of 8. Among the 8 newborn, only one was transgenic (number 18) (Figure 12). On this
20 experiment, the number of transgenic was 12.5%. On the three experiments, 1 transgenic out of 18 born animals drop the number to 5.5% of transgenic animals obtained when using the same DNA sequence when compared to numbers obtained from meganuclease mediated transgenesis. Transmission to F1 from the transgenic male 18 was tested by crossing this male with B6SJL females. Among the 25 F1 animals, 1 out of 6 and 3 out of
25 19 obtained were transgenic as detected on a Southern Blot analysis using a promoter probe. These result indicated that transmission using « classical transgenesis » is about 16%.

Control experiments

30   Integration of 750 copies/pl of the supercoiled Megafluo plasmid by itself was also checked. Two independent experiments were performed. 10 and 11 mice are obtained, respectively. Using both PCR and Southern Blot analysis in order to genotype the mice, no transgenic animal was detected.

35   ***In ovo linearization mediated by PI-SceI purified protein***

In order to test the efficiency of meganuclease mediated transgenesis by an other meganuclease, we performed experiments using a purified PI-SceI protein. The

transgenic construct, PIFFLago, is described in Figure 13. The Gas5 promoter leading the expression of the reporter gene LagoZ (a CPG-free LacZ sequence) are flanked by two PI-SceI recognition/cleaving sites on the same orientation separating the backbone sequences of the vector. A single I-SceI site is present in between the promoter and the

5 reporter gene.

Coinjection of 750 copies/pl of PIFFLAgo with purified PI-SceI protein at a target DNA/protein ratio of 1 :25 was performed. Following transfer of the injected-eggs, foster mother was sacrified at day 12 of gestation and injected embryos were genotyped both by PCR (Lago) and Southern Blot analysis with a probe corresponding to the reporter gene

10 sequence. Among the 10 embryos genotyped, 3 of them were transgenic (Figure 14, right panel). A 7 kb fragment was detected in the genomic DNA of those embryos by the Lago probe corresponding to a tandem inverted repeat intagration. This result showed that PI-SceI mediated transgensis efficiency was about 30%.

Using the same transgenic construct, I-SceI mediated transgenesis was tested by

15 coinjecting PIFFLago at 750 copies/pl with a 1 :169 I-SceI DNA target/protein molecule. Following transfer of the injected-eggs, foster mother was sacrified at day 12 of gestation and injected embryos were genotyped both by PCR (Lago) and Southern Blot analysis with a probe corresponding to the reporter gene sequence. Among the 9 embryos obtained, 3 of them were transgenic (Figure 14, left panel; Figure 15). This result shows

20 that an I-SceI mediated linearization of a transgene in this experiment led to one third of transgenic embryos.

Detailed protocol for meganuclease mediated transgenesis in mice

Sample preparation for injections

25 Plasmid DNA (supercoiled) solutions are produced in Qiagen Endo-free columns. After precipitation, plasmids were then resuspended at the desired concentration in Brinster buffer, 10mM TRIS, 0,25 mM EDTA. The I-SceI stock solution contained 150 µg/ml (10 units/µl) in 25mM HEPES, pH 8, 5% glycerol. The table 3 indicates quantities of I-SceI for a series of protein:DNA ratio.

Plasmid solution	I-SceI volume (from stock at 150 µg/ml)	X buffer (to adjust the glycerol conentration to 1%)	Final volume of sample; with 500 plasmid copies per pl	Ratio protein:DNA
10 µl	0.12 µl	9.9 µl	20 µl	1:1
10 µl	0.6 µl	9.4 µl	20 µl	**5:1**
10 µl	1.2 µl	8.8 µl	20 µl	10:1

30 Table 3

Samples for injections should be prepared considering that glycerol was present in the I-SceI stock solution. Depending on the amount of I-SceI solution used, some amount of glycerol-containing solution (X buffer) should be added to have in every case a unique, constant final glycerol concentration. For injections, we used samples at 1% glycerol and

5 a protein:DNA ratio of 5:1.

(500 copies of plasmids is equivalent to $3,01\ 10^{-20}$ mol. I-SceI is 29.3 KDa)

A similar protocol was used for transgenesis mediated by PI-SceI, MM 51kDa. On the basis of the condition of activity of PI-SceI, a stock solution of 1.15 mg/ml of PI-SceI containing 20% glycerol was prepared. Two final DNA/Ratio for injections were used: 50x

10 (307,3μg/ml of PI-SceI in 5,35% glycerol and 7mM HEPES) and 94x (307,3μg/ml of PI-SceI in 6,25% glycerol) They were prepared from dilutions of the PI-SceI stock solution in dilution buffer.

Dilution buffer:

25mM HEPES,5% glycerol,pH8 at 25°c

15

I-SceI-mediated transgenesis

The same supercoiled construction was injected either with an-expressing meganuclease vector or with the purified protein. As well, the pre-linearized and purified in vitro form of the same transgene was used to compare "classical transgenesis" with

20 meganuclease mediated transgenesis.

Supercoil plasmid preparation :

Production of plasmid was performed using QIAgen Endo-free kit (QIAGEN). Recovered DNA was precipitated with NaAcetate, washed with Ethanol 70% and

25 resuspended in Brinster buffer (10mM TRIS-0,25 mM EDTA) at the desired concentration for injection.

Linearized transgene preparation :

1. Megafluo was digested to liberate transgene from plasmid vector sequences.

30 Two different enzymes were used for that purpose : NdeI or I-SceI enzymes leading to 1796 bp and 1824 bp, respectively.

2. Restriction digest products were separated on agarose gel using 0,8% TAE.

3. Gel was placed on transilluminator and the desired band was cut out and then purified with QIAquick gel extraction kit (QIAGEN).

35   4. Following DNA elution in Tris (pH 7.6), DNA was precipitated, washed with Ethanol 70% and resuspended in Brinster buffer Tampon Brinster at a concentration of 500 copies/pl.

58

5. DNA was conserved at -20 degrees Centigrade.

Recovery of one-cell stage mouse eggs

a) Superovulation :

- intra-peritoneal injection of 3-weeks-old B6SJL females with PMS (folligon from Intervet) at 5 UI/mouse, 3 days prior injection

- intra-peritoneal injection of 3-weeks-old B6SJL females with HCG (Chorulon from Intervet) at 5 UI/mouse, one day prior injection and cross them with B6SJL males

- Cross foster B6CBA females with B6CBA vasectomized males.

b) Oocytes recovery:

-Sacrifice B6SJL plugged B6SJL females, recover oviducts and piece them in PBI medium

-Separate oocytes in 100 μl of Hyaluronidase at 0,5 mg/ml by cutting the ampulae with forceps for 2'

- Rince eggs in PBI

-Transfer eggs in Whitten covered with siliconized oil and keep them at 37°C, 5% CO_2.

c) Egg injections :

-fertilized eggs were microçinjected in PBI using an Eppendorf microinjector (Femtojet 5247) with micromanipulators (transferman NK2 5188) on a Nikon inverted microscope (TE2000-U) using contention and injection capillaries (GC 100-10 and GC 100F-10 from Phymep)

d) Injected eggs transfer into foster females:

-Anesthetized B6CBA females by intra-peritoneal injection of Avertin at 150 μg/ml with 400 μl/mouse

-Transfer injected eggs into dissected oviducts

Genomic DNA extraction

1)   Add 750 μL of tail buffer and 40 μL of Proteinase K per 1cm tail

        or 500 μL of tail buffer and 30 μL of Proteinase K per embryo membranes

2)   Incubate at 56 °C overnight

3)   Mix 5' on Eppendorf mixer

4)   Add 250 μL of saturated NaCl (≈ 6M) per tail

        or 170 μL saturated NaCl (≈ 6M) per embryo membrane

5)   Mix 5' on Eppendorf mixer

6)   Centrifuge 10' at 13000 rpm

7)   Collect 750 μL of supernatant in a new eppendorf tube per tail

or 500 µL of supernatant per embryo membrane

8)   Add 500 µL of Isopropanol per tail

or 350 µL of Isopropanol per embryo membrane

9)   Mix 2' on Eppendorf mixer

5   10)   Centrifuge 1' at 13000 rpm

11)   Wash pellet twice with 500 µL of EtOH 70% at RT

Genotyping Transgenic Mice by Southern Blot experiments

a)   Labelling of 1µg of purified plasmidic DNA sequences o/n using the Dig-labelling kit

10   from Roche. Probes were purified on G-50 sephadex. Quantification of probe was checked on a dot blot reaction with the Dig-control from the Roche kit.

b)   From 1 to 7µg of genomic DNA were digested o/n with EcoRI restriction enzyme (NEB)

c)   Digested DNA were loaded on 0,8% agarose TAE gels

15 d)   Gels were denatured for 30' in denaturating solution (NaOH (0,5M) + NaCl (1,5M)) then neutralized in neutralizing solution (Tris (pH7.4, 0,5M) + NaCl (1,5M))

e)   DNA transfer was made by capillarity in 10xSSC o/n on a nylon Hybond-N+ membrane(Amersham)

f)   DNA was fixed on membrane by UV-crosslinking (Stratalinker from Stratagene) and

20   bake for 2h in a 80°c oven

g)   Membranes were prehybridized in hybridization buffer for an hour at 68°c in a rotating oven

h)   Hybridization was performed in hybridization buffer with a pre-denatured Dig-labelled probe (10ng/ml) o/n at 68°c in a rotating oven

25 i)   Two washes 5' at RT with a pre-heated washing solution on a rotating plateform

j)   One wash 5' at RT in Buffer I (1x) / 0,3% tween

k)   Blocking of the membranes in Blocking solution for 30' in sealed plastic bags rotated at 250 rpm

l)   Incubation of the Anti-Digoxigenin-Ap Fab fragments (Roche)at 0,0375 U/ml for 30' in

30   blocking solution

m)   Two washes in Buffer I (1x) / 0,3% tween of 30' each

n)   5' in revelation buffer (Buffer III)

o)   Revelation of membranes with chemioluminescent substrate CDP-Star from Roche

35   Buffer compositions:

Hybridization buffer: 0,5M NaPi, 7% SDS, 1mM EDTA

Washing solution: 40mM NaPi, 1% SDS

60

Buffer I (10x): 1M Maleic acid, 1.5M NaCl, pH 7.5

Blocking solution: Blocking reagent (Roche) at 10% w/v in buffer I

Buffer III (1x): Per liter: 100mM Tris (pH9.5), 100mM NaCl, 50mM MgCl2

5   Genotyping Transgenic Mice by PCR

PCR reactions

100ng of genomic DNA or 1ng of plasmid DNA

1µM of each primer

RED Taq Ready mix (Sigma)

10   DMSO from 5 to 10% depending on the pripmer used (see PCR conditions)

Qsp with autoclaved water to 10µl

Overlay with mineral oil and place into thermal cycler.

PCR conditions :

15   Denaturation 3' at 94°c, 1 cycle

Denaturation 20" at 94°c

Annealing 30" at variable temperature depending on the primers Tm   35

Amplification 20" at 72°c   cycles

Final round of amplification at 72°c for 3'

20   PCR products are analyzed on 2,5% TAE agarose gel

ß-globin PCR

10% DMSO, annealing at 51°c

Oligonucleotides used for the mouse ß-globin gene PCR amplification leading to the amplification of a 494bp fragment

25   **Red** PCR amplification

10% DMSO, annealing at 55°c

Oligonucleotides used for the reporter PCR amplification of megafluo sequence leading to the amplification of a 484bp fragment

I-Scel PCR amplification

30   10% DMSO, annealing at 55°c

Oligonucleotides used for the PCR amplification of a sub-fragment of the pI-SceI/EGFP sequences leading to the amplification of a 400bp fragment

Lago PCR amplification

5% DMSO, annealing at 51°c

35   Oligonucleotides used for the reporter PCR amplification of PIFFLago sequence leading to the amplification of a 370bp fragment

PCR controls :

Control of contamination : identical PCR reaction except without DNA but water only (C) Positive control : B6SJL mouse DNA (wt) to ensure that the DNA sample is "amplifiable with the use of primers for mouse beta globin demonstrating that a single copy gene can be amplified with the genomic DNAs to be genotyped.

Positive/negative supercoiled plasmid DNA were systematically used as controls in all PCR experiments.

The reference in this specification to any prior publication (or information derived from it), or to any matter which is known, is not, and should not be taken as an acknowledgment or admission or any form of suggestion that that prior publication (or information derived from it) or known matter forms part of the common general knowledge in the field of endeavour to which this specification relates.

Throughout this specification and the claims which follow, unless the context requires otherwise, the word "comprise", and variations such as "comprises" and "comprising", will be understood to imply the inclusion of a stated integer or step or group of integers or steps but not the exclusion of any other integer or step or group of integers or steps.

CLAIMS:

1.   A method for randomly integrating a polynucleotide into a host cell genome by preparation into the host cell of said linear polynucleotide having free 5' and 3' ends from a vector, wherein said method comprises:

a)   introducing in said host cell a vector having no free 5' and 3' ends and comprising said polynucleotide, said vector comprising at least one cleavage site which is not found in the host cell genome;

b)   causing cleavage of said site(s) in said host cell, thereby creating or releasing said polynucleotide in a linear form having free 5' and 3' ends from said vector into said host cell; and,

c)   maintaining the host cell under conditions and for a period of time sufficient to cause the random integration of said released polynucleotide into said host cell genome.

2.   Method according to claim 1, wherein said method further comprises, prior to step (b), an additional step of introducing into said host cell a cleaving agent or a vector comprising a nucleic acid encoding said cleaving agent.

3.   Method according to claim 2 wherein said method further comprises, prior to step (b), an additional step of introducing into said host cell a cleaving agent.

4.   Method according to claim 1-3, wherein said cleavage site is found in the host cell genome at less than 2 copies.

5.   Method according to claim 1-4 wherein said polynucleotide is flanked by at least one cleavage site.

6.   Method according to claim 5 wherein said polynucleotide is flanked by at two cleavage sites.

7.   Method according to claim 1-6, wherein said cleavage site is an endonuclease site and said cleaving agent is the corresponding endonuclease.

8. Method according to claim 7, wherein said endonuclease has a recognition site of at least 12 nucleotides.

9. Method according to claim 8, wherein said endonuclease is a meganuclease.

10. Method according to claim 9, wherein said meganuclease is selected from the group consisting of I-Ceu I, I-Cre I, I-Chu I, I-Csm I, I-Dmo I, I-Pan I, I-Sce I, I-Sce II, I-Sce III, I-Sce IV, F-Sce I, F-Sce II, PI-Aae I, PI-Ape I, PI-Ceu I, PI-Cir I, PI-Ctr I, PI-Dra I, PI-Mav I, PI-Mfl I, PI-Mgo I, PI-Mja I, PI-Mka I, PI-Mle I, PI-Mtu I, PI-MtuH I, PI-Pab III, PI-Pfu I, PI-Pho I, PI-Pko I, PI-Psp I, PI-Rma I, PI-Sce I, PI-Ssp I, PI-Tfu I, PI-Tfu II, PI-Tli I, PI-Tli II, PI-Tsp I, PI-Tsp II, PI-Bsp I, PI-Mch I, PI-Mfa I, PI-Mga I, PI-Mga II, PI-Min I, PI-Mma I, PI-Msh I, PI-Msm II, PI-Mth I, PI-Tag I, PI-Thy II, I-Ncr I, I-Ncr II, I-Pan II, I-Tev I, I-Ppo I, I-Dir I, I-Hmu I, I-Hmu II, I-Tev II, I-Tev III, F-Sce I, F-Sce II (HO), F-Suv I, F-Tev I, and F-Tev II.

11. Method according to claim 10, wherein said meganuclease is selected from the group consisting of I-Ceu I, I-Cre I, I-Sce I.

12. Method according to claim 8, wherein said endonuclease is synthetic.

13. Method according to claim 1-12, wherein said polynucleotide is unable to undergo homologous recombination with the host cell genome.

14. Method according to claim 13, wherein the 5' and 3' sequences of said polynucleotide have no homology with a host cell genomic sequence.

15. Method according to claim 1-14, wherein said vector is a plasmid.

16. Method according to claim 1-15, wherein said host cell is selected from the group consisting of a stem cell, a somatic cell, a gamete, a blastomer, and an egg.

17. Method according to claim 16, wherein said host cell is a stem cell.

18. Method according to claim 16, wherein said host cell is a somatic cell.

19. Method according to claim 16, wherein said host cell is an egg.

20. Method according to claim 1-17, wherein said polynucleotide sequence is a sequence encoding a polypeptide or an antisense, a regulatory sequence, or a recognition sequence for a molecule.

21. Method according to claim 17, wherein said host cell stems from fish, bird, non-human mammals, insect, amphibian, or reptile.

22. Method according to claim 21, wherein said host cell stem from medaka, zebrafish, sticklebass, astyanax, mice, chicken, xenopus, sheep, cow, rabbit.

23. A cell resulting from a method according to any one of claims 1-22.

24. A non-human transgenic animal comprising a cell according to claim 23.

25. A transgenic plant comprising a cell resulting from a method according any one of claims 1-20.

26. Use of a cell according to claim 23, of transgenic animal according to claim 24 or of a plant according to claim 25 for production of protein, biomolecule, biomaterial, or vaccine.

27. Use of a cell according to claim 23 for treatment or prophylaxis of a condition or disorder in an individual.

28. A method for producing a non-human transgenic animal, wherein embryonic stem cells are transfected by the method according to claim 1-22, the cells are injected into embryos at a stage at which they are capable of integrating the transfected cells the embryos are then reimplanted in a surrogate mother, and the chimeric individuals

obtained at the end of gestation, and in which colonization by embryonic stem cells of the germ line is observed are mated to obtain transgenic animals.

29.   A method for producing a non-human transgenic animal, wherein a fertilized egg is transfected by the method according to any one of claims 1 to 22, the egg is reimplanted in a surrogate mother, and the transgenic animal is obtained at the end of gestation.

30.   A method for producing a non-human transgenic animal, wherein a fertilized egg is transfected by the method according to any one of claims 1 to 22, and the egg is incubated under conditions appropriate for developing said transgenic animal.

31.   A composition when used in a method according to any one of claims 1 to 22, comprising a vector having no 5' and 3' free ends and comprising a transgene to be integrated in a host cell genome, said vector comprising at least one cleavage site which is not found in said host cell genome; and,

a cleaving agent or a vector comprising a nucleic acid encoding said cleaving agent.

32.   Composition according to claim 31, wherein said polynucleotide is flanked by at least one cleavage site.

33.   Composition according to claim 32, wherein said polynucleotide is flanked by at least two cleavage sites.

34.   Composition according to any one of claims 31 to 33, wherein said cleavage site is an endonuclease site and said cleaving agent is the corresponding endonuclease.

35.   Composition according to claim 34, wherein said endonuclease has a recognition site of at least 12 nucleotides.

36.   Composition according to claim 34 or claim 35, wherein said endonuclease is a meganuclease.

37.   Composition according to claim 36, wherein said meganuclease is selected from the group consisting of I-Ceu I, I-Cre I, I-Chu I, I-Csm I, I-Dmo I, I-Pan I, I-Sce I, I-Sce II, I-Sce III, I-Sce IV, F-Sce I, F-Sce II, PI-Aae I, PI-Ape I, PI-Ceu I, PI-Cir I, PI-Ctr I, PI-Dra I, PI-Mav I, PI-Mfl I, PI-Mgo I, PI-Mja I, PI-Mka I, PI-Mle I, PI-Mtu I, PI-MtuH I, PI-Pab III, PI-Pfu I, PI-Pho I, PI-Pko I, PI-Psp I, PI-Rma I, PI-Sce I, PI-Ssp I, PI-Tfu I, PI-Tfu II, PI-Tli I, PI-Tli II, PI-Tsp I, PI-Tsp II, PI-Bsp I, PI-Mch I, PI-Mfa I, PI-Mga I, PI-Mga II, PI-Min I, PI-Mma I, PI-Msh I, PI-Msm II, PI-Mth I, PI-Tag I, PI-Thy II, I-Ncr I, I-Ncr II, I-Pan II, I-Tev I, I-Ppo I, I-Dir I, I-Hmu I, I-Hmu II, I-Tev II, I-Tev III, F-Sce I, F-Sce II (HO), F-Suv I, F-Tev I, and F-Tev II.

38.   Composition according to claim 37, wherein said meganuclease is selected from the group consisting of I-Ceu I, I-Cre I, I-Sce I.

39.   Composition according to claim 36, wherein said endonuclease is synthetic.

40.   Composition according to any one of claims 31 to 39, wherein said vector is a plasmid.

41.   Composition according to any one of claims 31 to 40, wherein said vector comprising said polynucleotide sequence to be linearized further comprises said nucleic acid encoding said cleaving agent.

42.   Composition according to any one of claims 31 to 41, wherein said polynucleotide is unable to undergo homologous recombination with the host cell genome.

43.   Composition according to claim 42, wherein the 5' and 3' sequences of said polynucleotide have no homology with a host cell genomic sequence.

44.   A method according to any one of claims 1 to 22 substantially as herein described with reference to any one of the examples or figures.
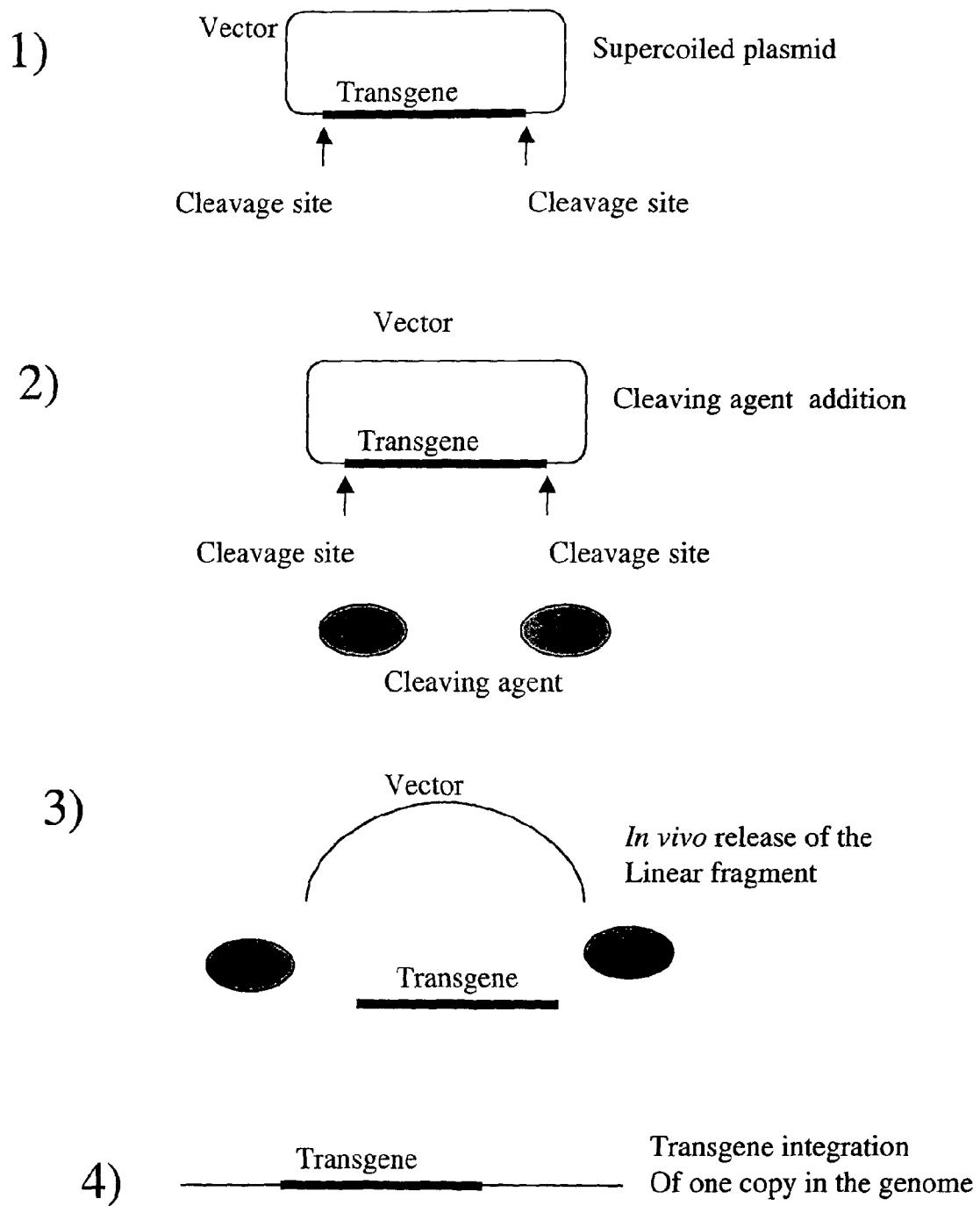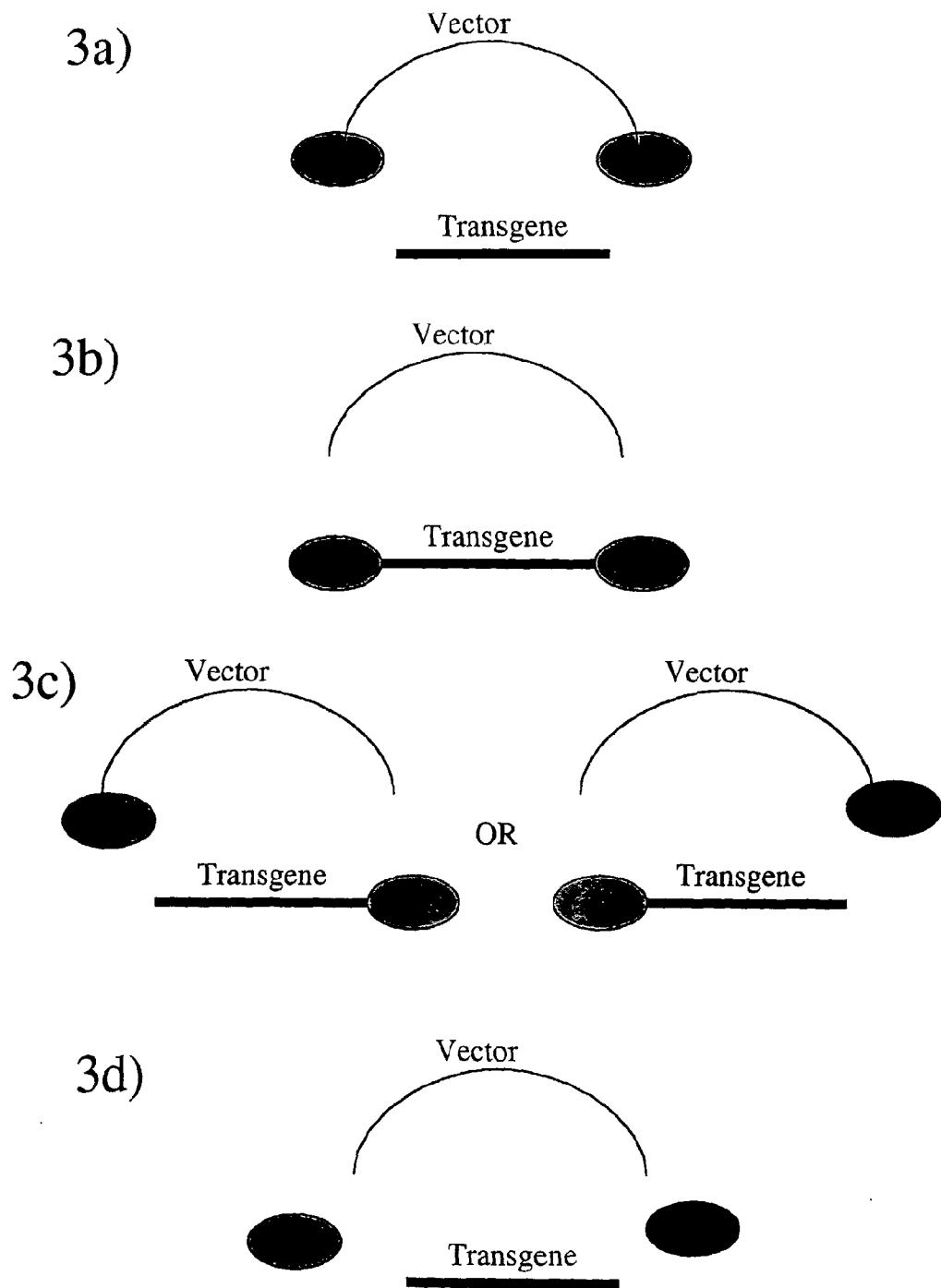
FIGURE 1

FIGURE 1 (suite)

Protein		Organism	species	Size (AA)	Motif(AA)	Accès (h)	Year	Authors	Journal
I-Ama I	Th.	Allomyces	macrogynus	359		U02039	1994	Paquin, B., Laforest, M.-J., Lang, BF.	PNAS 91: 11807-10
I-Ani I	Th.	Aspergillus (Emericella)	nidulans	319			1982	Waring, R.B., Davies, R.W., Scazzocchio, C., Brown, T.A.	PNAS 79: 6332-36
I-Bmo I	Exp	Bacillus	mojavensis	266	GG	AF321518	2001	Edgell, D.R., Shub, D.A.	PNAS 98: 7898-7903
I-Ceu I	Exp	Chlamydomonas	eugametos	218	D	S15138	1991	Gauthier, A., Turmel, M., Lemieux, C.	Curr Genet 19: 43-7
I-Ceu II	Th.	Chlamydomonas	eugametos	283	D		1998	Denovan-Wright, E.M., Nedelcu, A.M., Lee, R.W.	Plant Mol. Biol. 36: 285-95
I-Chu I	Exp	Chlamydomonas	humicola	218	dd	L06107	1993	Coté, V., Mercier, JP., Lemieux, C., Turmel, M.	Gene 129: 69-76
I-Cmoe I	Pot	Chlamydomonas	moewusii		HNH		2000	Drouin, M., Lucas, P., Otis, C., Lemieux, C., Turmel, M.	NAR 28: 4566-72
I-Cpa I	Pot	Chlamydomonas	pallidostigmatica				1995	Turmel, M., Cote, V., Otis, C., Mercier, J.-P., Gray, M.W., Lonergan, K.M., Lemieux, C.	Mol. Biol. Evol. 12: 533-45
I-Cpa II	Pot	Chlamydomonas	pallidostigmatica				1995	Turmel, M., Mercier, J.-P., Cote, V., Otis, C., Lemieux, C.	NAR 23: 2519-25
I-Cre I	Exp	Chlamydomonas	reinhardtii	163	D	X01977	1985	Rochaix, JD., Rahire, M., Michel, F.	NAR 13: 975-84
I-Cre II	Pot	Chlamydomonas	reinhardtii				1999	Holloway, S.P., Deshpande, N.N., Herrin, D.L.	Curr. Genet. 36: 69-78
I-Cre V	Pot	Chlamydomonas	reinhardtii		HNH		1999	Holloway, S.P., Deshpande, N.N., Herrin, D.L.	Curr. Genet. 36: 69-78
I-Csm I	Exp	Chlamydomonas	smithii	237	dd	X55305	1990	Colleaux, L., Michel-Wolwertz, M-R., Matagne, RF., Dujon, B.	Mol Gen Genet 223: 288-296
I-Cvu I	Pot	Chlorella	vulgaris				1998	Watanabe, K.I., Ehara, M., Inagaki, Y., Ohama, T.	Gene 213: 1-7
I-Ddi I	Pot	Dictyostelium	discoideum				1997	Ogawa, S., Naito, K., Angata, K., Morio, T., Urushihara, H., Tanaka, Y.	Gene 191: 115-21
I-Ddi II	Pot	Dictyostelium	discoideum				1997	Yoshino, R., Morio, T., Tanaka, Y.	Unpublished observations
I-Dir I	Exp	Didymium	iridis	244	-	X71792	1993	Johansen, S., Embley, TM., Willassen, N.P.	NAR 21: 4405
I-Dmo I	Exp	Desulfurococcus	mobilis	194	DD	P21505	1985	Kjems, J., Garrett, R.A.	Nature 318: 675-77
I-Hmu I	Exp	Bacteriophage	SPO1	174	-	M37686	1990	Goodrich-Blair, H., Scarlato, V., Gott, JM., Xu, MQ., Shub, DA.	Cell 63: 417-424
I-Hmu II	Exp	Bacteriophage	SP82	187	-	-	1993	Lambowitz, A.M., Belfort, M.	Annu Rev Biochem 62: 587-622
I-Hsp I	Pot	Halobacterium	sp.				1998	Ng, W.V., Ciufo, S.A., Smith, T.M., Bumgarner, R.E., Baskin, D., Faust, J., Hall, B., Loretz, C., Seto, J., Slagel, J., Hood, L., DasSarma, S.	Genome Res. 8: 1131-41
I-Lla I	Pot	Lactococcus	lactis				2001	http://rebase.neb.com/rebase/rebase.homing.html	
I-Naa I	Pot	Naegleria	andersoni				1992	Embley, TM., Dyal, P., Kilvington, S.	NAR 20: 6411
I-Nan I	Pot	Naegleria	andersoni				1994	De Jonckheere, J.F.	J. Eukaryot. Microbiol. 41: 457-63
I-Ncl I	Pot	Naegleria	gruberi				2001	http://rebase.neb.com/rebase/rebase.homing.html	
I-Ncr I	Exp	Neurospora	crassa	304	GG	S06367	1988	Mota, EM., Collins, R.A.	Nature 332: 654-656

FIGURE 2

I-Ncr II	Th.	Neurospora	crassa	317	GG	X06884	1988	Collins et al.	Nucleic Acids Res 16: 1125-34
I-Ngr I	Pot	Naegleria	gruberi				2001	http://rebase.neb.com/rebase/rebase.homing.html	
I-Nit I	Pot	Naegleria	italica				2001	http://rebase.neb.com/rebase/rebase.homing.html	
I-Nja I	Pot	Naegleria	jamiesoni				2001	http://rebase.neb.com/rebase/rebase.homing.html	
I-Nsp I	Pot	Naegleria	sp.				2001	http://rebase.neb.com/rebase/rebase.homing.html	
I-Pan I	Th.	Podospora	anserina	243	dd	X55026	1985	Cummings, DJ., MacNeil, IA., Domenico, J., Matsuura, ET.	J Mol Biol 185: 659-80
I-Pbo I	Pot	Protosiphon	botryoides				1998	Watanabe, K.I., Ehara, M., Inagaki, Y., Ohama, T.	Gene 213: 1-7
I-Pcu I	Pot	Peperomia	cubensis				2001	http://rebase.neb.com/rebase/rebase.homing.html	
I-PcuA I	Pot	Peperomia	cubensis				2001	http://rebase.neb.com/rebase/rebase.homing.html	
I-PcuV I	Pot	Peperomia	cubensis				2001	http://rebase.neb.com/rebase/rebase.homing.html	
I-Pgr I	Pot	Peperomia	griseoargentea				2001	http://rebase.neb.com/rebase/rebase.homing.html	
I-Pob I	Pot	Peperomia	obtusifola				2001	http://rebase.neb.com/rebase/rebase.homing.html	
I-Por I	Pot	Pyrobaculum	organotrophum				2001	http://rebase.neb.com/rebase/rebase.homing.html	
I-Por II	Pot	Pyrobaculum	organotrophum				2001	http://rebase.neb.com/rebase/rebase.homing.html	
I-Ppb I	Pot	Peperomia	polybotrya				2001	http://rebase.neb.com/rebase/rebase.homing.html	
I-Ppo I	Exp	Physarum	polycephalum	185	-	M38131	1990	Muscarella, DE., Ellison, EL., Ruoff, BM., Vogt, VM.	Mol Cell Biol 10: 3386-96
I-Sca I	Th.	Saccharomyces	capensis	250	DD	X95974	2000	Monteilhet, C., Dziadkowiec, D., Szczepanek, T., Lazowska, J.	Nucleic Acids Res. 28: 1245-1251
I-Sce I	Exp	Saccharomyces	cerevisiæ	235	dd	P03882	1985	Jacquier & Dujon	
I-Sce II	Exp	Saccharomyces	cerevisiæ	316	DD	P03878	1980	Bonitz, S. G., Coruzzi, G., Thalenfeld, B. E., Tzagoloff, A., Macino, G.	J Biol Chem 255: 11927-41
I-Sce III	Exp	Saccharomyces	cerevisiæ	335	DD	P03877	1980	Bonitz, S. G., Coruzzi, G., Thalenfeld, B. E., Tzagoloff, A., Macino, G.	J Biol Chem 255: 11927-41
I-Sce IV	Exp	Saccharomyces	cerevisiæ	307	dd	-	1992	Séraphin et al.	Gene 113: 1-8
I-Sex I	Pot	Saccharomyces	exiguus				2001	http://rebase.neb.com/rebase/rebase.homing.html	
I-Sne I	Pot	Simkania	negevensis				2001	http://rebase.neb.com/rebase/rebase.homing.html	

FIGURE 2

I-Spo I	Pot	Schizosaccharomyces	pombe				2001	http://rebase.neb.com/rebase/rebase.homing.html	
I-SPβ I	Pot	Bacteriophage	SPβ				1998	Lazarevic, V., Soldo, B., Dusterhoft, A., Hilbert, H., Mauel, C., Karamata, D.	PNAS 95: 1692-97
I-Squ I	Pot	Scenedesmus	quadricauda				2001	http://rebase.neb.com/rebase/rebase.homing.html	
I-Tde I	Pot	Torulaspora	delbrueckii				2001	http://rebase.neb.com/rebase/rebase.homing.html	
I-Tev I	Exp	Bacteriophage	T4	245	GIG	M12742	1989	West et al.	J Biol Chem 264: 10343-10346
I-Tev II	Exp	Bacteriophage	T4	258	GIG	-	1987	Shub et al.	Cold Spring H. Symp Quant Biol: 193-200
I-Tev III	Exp	Bacteriophage	RB3	269	-	X59078	1991	Eddy, S.R., Gold, L.	Genes Dev 5: 1032-1041
I-Uar A	Pot	archaeon	Uncultured			AB027540	2001	http://rebase.neb.com/rebase/rebase.homing.html	
I-Vin I	Pot	Venturia	inaequalis				2001	http://rebase.neb.com/rebase/rebase.homing.html	
I-Zbi I	Pot	Zygosaccharomyces	bisporus				2001	http://rebase.neb.com/rebase/rebase.homing.html	
F-Sce I	Exp	Saccharomyces	cerevisiæ	476	DD	M63839	1991	Nakagawa et al.	J. Biol. Chem. 266: 1977-1984
F-Sce II	Exp	Saccharomyces	cerevisiæ	586	DD	M14678	1983	Kostriken, R., JN. Strathern, AJS. Klar, JB. Hicks, F. Heffron	Cell 35: 167-174
F-Tev I	Exp	Bacteriophage	T4	221			1994	Sharma, M., Hinton, D.M.	J. Bacteriol. 176: 6439-6448
F-Tev II	Th.	Bacteriophage	T4	205			1997	Kadyrov, F.A., Shlyapnikov, M.G., Kryukov, V.M.	FEBS Lett. 415: 75-80
PI-Aae I	Th.	Aquifex	aeolicus	347	DD		1998	Deckert,G. et al.	Nature 392 (6674), 353-358
PI-Ape I	Th.	Aeropyrum	pernix	468	DD	B72665	1999	Kawarabayasi,Y. et al	DNA Res. 6 (2), 83-101
PI-APSE I	Th.	Bacteriophage	Acyrthosiphon pisum	306	-		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Bsp I	Th.	Bacteriophage	prophage A	385	DD		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Ceu I	Exp	Chlamydomonas	eugametos	457	DD		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Cir I	Th.	Chilo	iridescent	339	DD		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Cry I	Th.	Filobasidiella (Cryptococcus)	neoformans	172	-		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Ctr I	Exp	Candida	tropicalis	471	DD	M64984	1993	Gu, HH., Xu, J., Gallagher, M., Dean, GE.	J Biol Chem 268: 7372-81
PI-Dra I	Th.	Deinococcus	radiodurans	367	DD		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Gth I	Th.	Guillardia	theta	160	-		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	

FIGURE 2

PI-Hsp I	Th.	Halobacterium	sp.	183	-		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Mav I	Exp	Mycobacterium	avium	337	DD		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Mch I	Th.	Mycobacterium	chitae	365	DD		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Meth I	Exp	Methanobacterium	thermoautotrophi cum	134	-		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Mfa I	Th.	Mycobacterium	fallax	364	DD		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Mfl I	Exp	Mycobacterium	flavescens	421	DD		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Mfl II	Exp	Mycobacterium	flavescens	364	DD		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Mfl II 144	Th.	Mycobacterium	flavescens	365	DD		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Mga I	Th.	Mycobacterium	gastri	420	DD		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Mga III	Th.	Mycobacterium	gastri	369	DD		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Mgo I	Th.	Mycobacterium	gordonae	420	DD		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Min I	Th.	Mycobacterium	intracellulare	335	DD		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Mja I	Th.	Methanococcus	jannaschii	500	DD		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Mja V	Exp	Methanococcus	jannaschii	169	-		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Mka I	Th.	Mycobacterium	kansasii	420	DD		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Mle I	Exp	Mycobacterium	leprae	365 (366)	Dd	X73822	1994	Davis EO, Thangaraj HS, Brooks PC, Colston MJ.	EMBO J. Feb 1;13(3):699-703.
PI-Mma I	Th.	Mycobacterium	malmoense	420	DD		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Msh I	Th.	Mycobacterium	shimoidei	365	DD		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Msm I	Th.	Mycobacterium	smegmatis	140	-		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Msm II	Th.	Mycobacterium	smegmatis	426	DD		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Mth I	Th.	Mycobacterium	thermoresistibile	366	DD		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	

FIGURE 2

PI-Mtu	Exp	Mycobacterium	tuberculosis	441	DD		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Mtu I	Exp	Mycobacterium	tuberculosis	439 (441)	DD	X58485	1992	Davis EO, Jenner PJ, Brooks PC, Colston MJ, Sedgwick SG.	Cell Oct 16;71(2):201-10
PI-MtuH I	Exp	Mycobacterium	tuberculosis	415	DD		2001	http://rebase.neb.com/rebase/rebase.homing.html	
PI-Mxe I	Exp	Mycobacterium	xenopi	198	-		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Pab I	Th.	Pyrococcus	abyssi	164	-		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Pab III	Th.	Pyrococcus	abyssi	394	DD		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Pfu I	Exp	Pyrococcus	furiosus	454			1999	Komori, K., Fujita, N., Ichiyanagi, K., Shinagawa, H., Morikawa, K., Ishino, Y.	NAR 27: 4167-74
PI-Pfu II	Exp	Pyrococcus	furiosus	525	DD		1999	Komori, K., Fujita, N., Ichiyanagi, K., Shinagawa, H., Morikawa, K., Ishino, Y.	NAR 27: 4167-74
PI-Pho I	Th.	Pyrococcus	horikoshii	167	-		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Pho II	Th.	Pyrococcus	horikoshii	377	DD		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Pko	Exp	Pyrococcus	kodakaraensis	360	DD		1997	Takagi, M., Nishioka, M., Kakihara, H., Kitabayashi, M., Inoue, H., Kawakami, B., Oka, M., Imanaka, T.	Appl. Environ. Microbiol. 63: 4504-10
PI-Pko II	Th.	Pyrococcus	kodakaraensis	537	DD		1997	Takagi, M., Nishioka, M., Kakihara, H., Kitabayashi, M., Inoue, H., Kawakami, B., Oka, M., Imanaka, T.	Appl. Environ. Microbiol. 63: 4504-10
PI-Ppu I	Th.	Porphyra	purpurea	150	-		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Psp I	Exp	Pyrococcus	sp.	537	DD		1993	Xu, M.-Q., Southworth, MW., Mersha, FB., Hornstra, LJ., Perler, FB.	Cell 75: 1371-77
PI-Rma I	Exp	Rhodothermus	marinus	428	DD		1997	Liu, X.-Q., Hu, Z.	PNAS 94: 7851-56
PI-Sce I	Exp	Saccharomyces	cerevisiæ	454	DD	M21609	1990	Hirata, R. et al.	J Biol Chem 265: 6726-6733
PI-SPβ I	Th.	Bacteriophage	SPβ	385	DD		1998	Lazarevic, V., Soldo, B., Dusterhoft, A., Hilbert, H., Mauel, C., Karamata, D.	PNAS 95: 1692-97
PI-Ssp I	Exp	Synechocystis	sp.	429	DD		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Ssp III	Th.	Synechocystis	sp.	123 +36	-		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Ssp IV	Th.	Synechocystis	sp.	436	HNH		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Tac I	Th.	Thermoplasma	acidophilum	173	-		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	

FIGURE 2

PI-Tag I	Exp	Thermococcus	aggregans	360	DD		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Tag II	Exp	Thermococcus	aggregans	538	DD		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Tfu I	Exp	Thermococcus	fumicolans	360	DD		2000	Saves, I., Ozanne, V., Dietrich, J., Masson, J.M.	JBC 275: 2335-41
PI-Tfu II	Exp	Thermococcus	fumicolans	389	DD		2000	Saves, I., Ozanne, V., Dietrich, J., Masson, J.M.	JBC 275: 2335-41
PI-Thy II	Exp	Thermococcus	hydrothermalis	538	DD		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Thy III	Exp	Thermococcus	hydrothermalis	390	DD		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Tli I	Exp	Thermococcus	litoralis	390	DD	M74198	1992	Perler et al.	PNAS 89: 5577-5581
PI-Tli II	Exp	Thermococcus	litoralis	541	DD	M74198	1993	Lambowitz, A.M., Belfort, M.	Annu Rev Biochem 62: 587-622
PI-Tsp II	Exp	Thermococcus	sp.	536	DD		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Tsp III	Exp	Thermococcus	sp.	390	DD		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	
PI-Tvo I	Th.	Thermoplasma	volcanium	186	-		2001	http://bioinformatics.weizmann.ac.il/~pietro/inteins	

FIGURE 2

FIGURE 3A
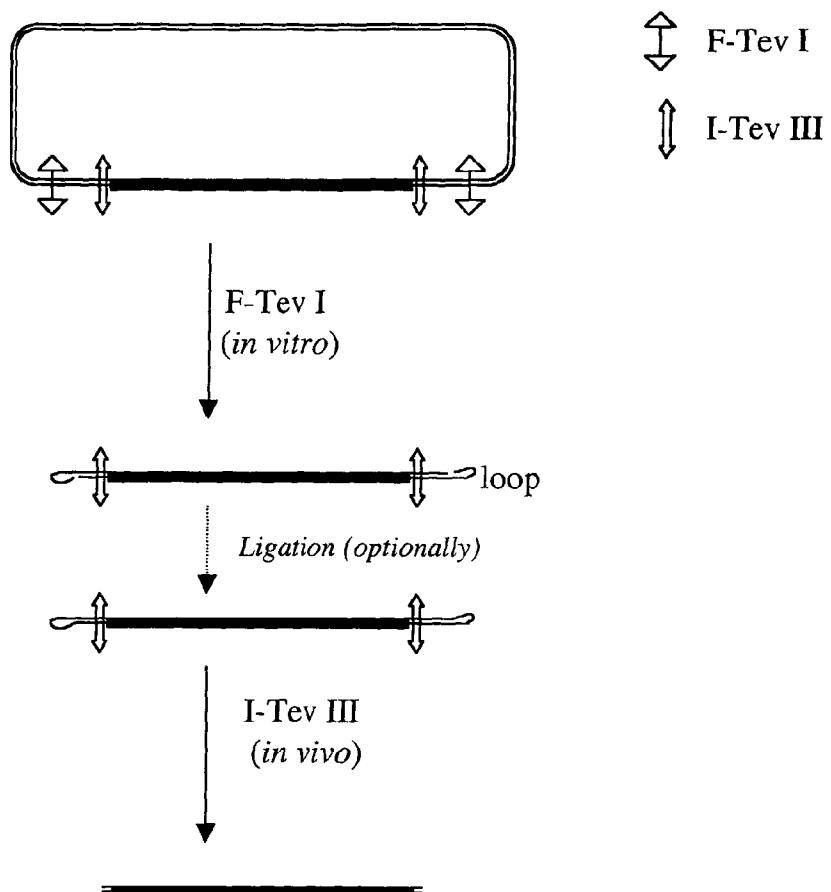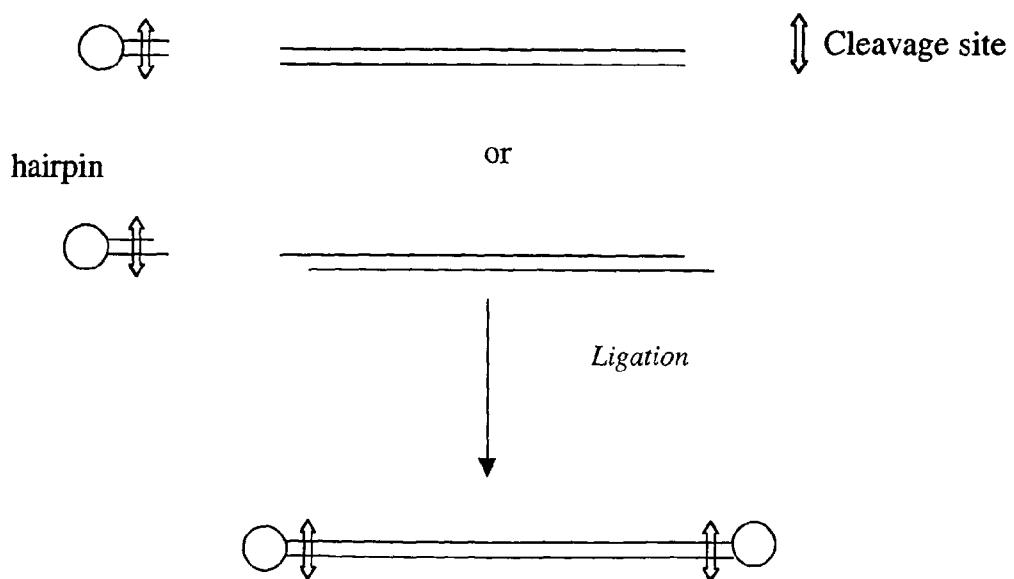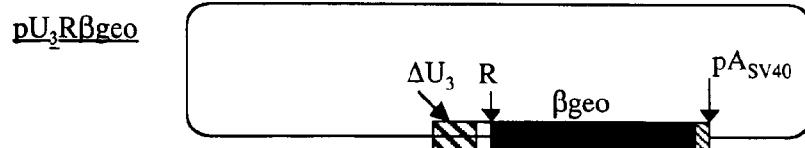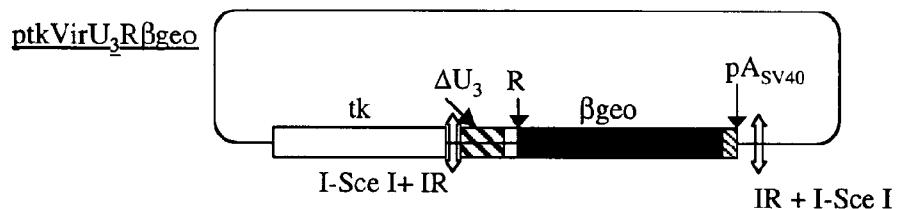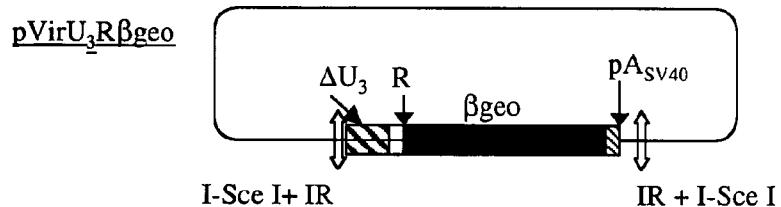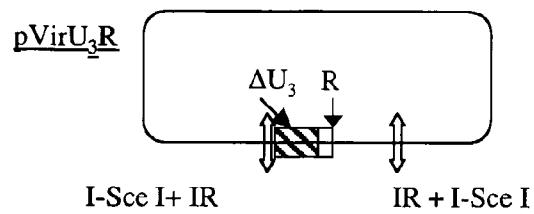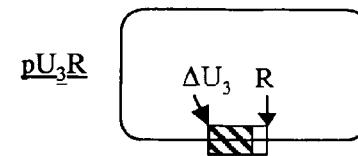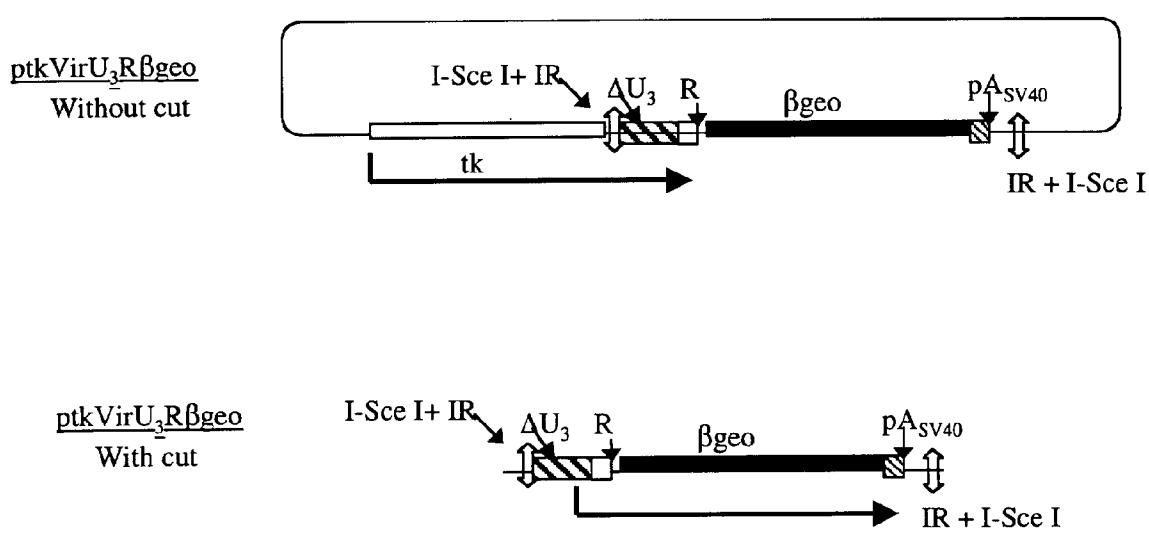
FIGURE 3B

FIGURE 4

FIGURE 5

EcoR V digestion
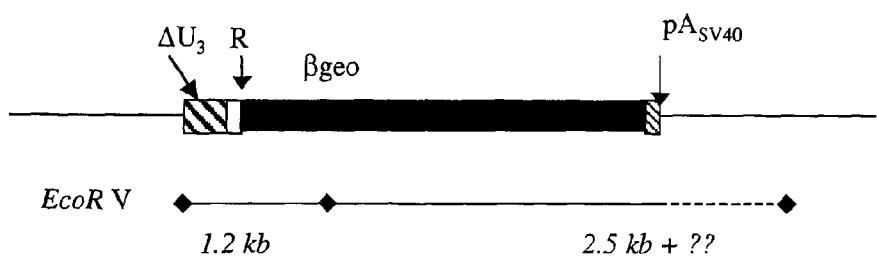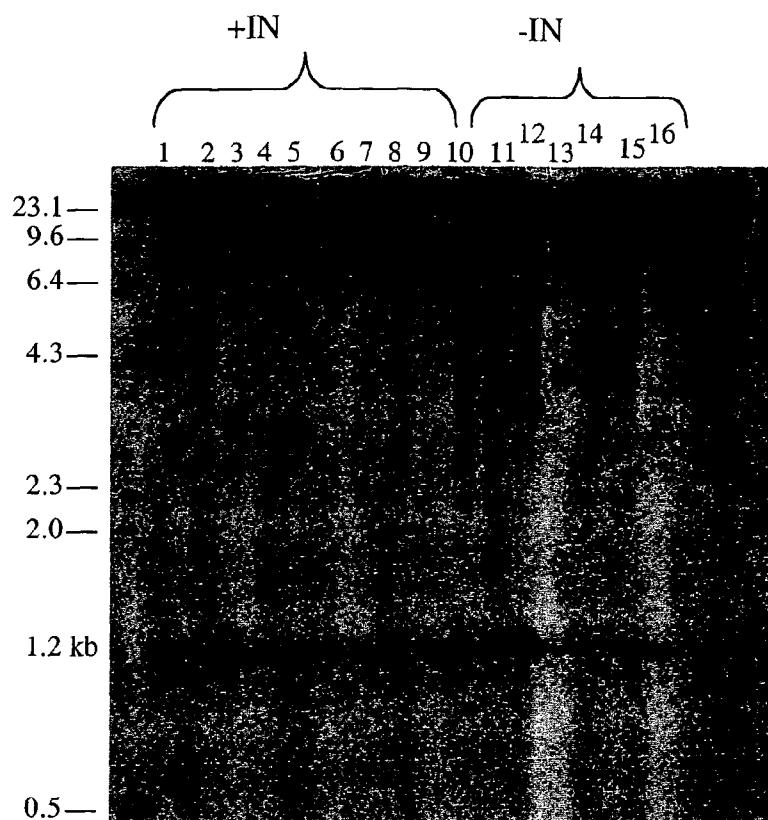
FIGURE 6

Figure 7A

Figure 7B

ppSNIE + pCMV-I-*Sce*I- co-transfection   ppSNIE + pCMV-I-*Sce*I+ co-transfection

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16   17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32   -   +
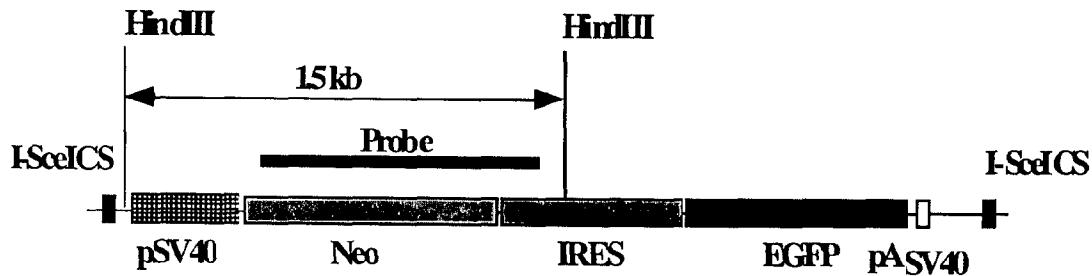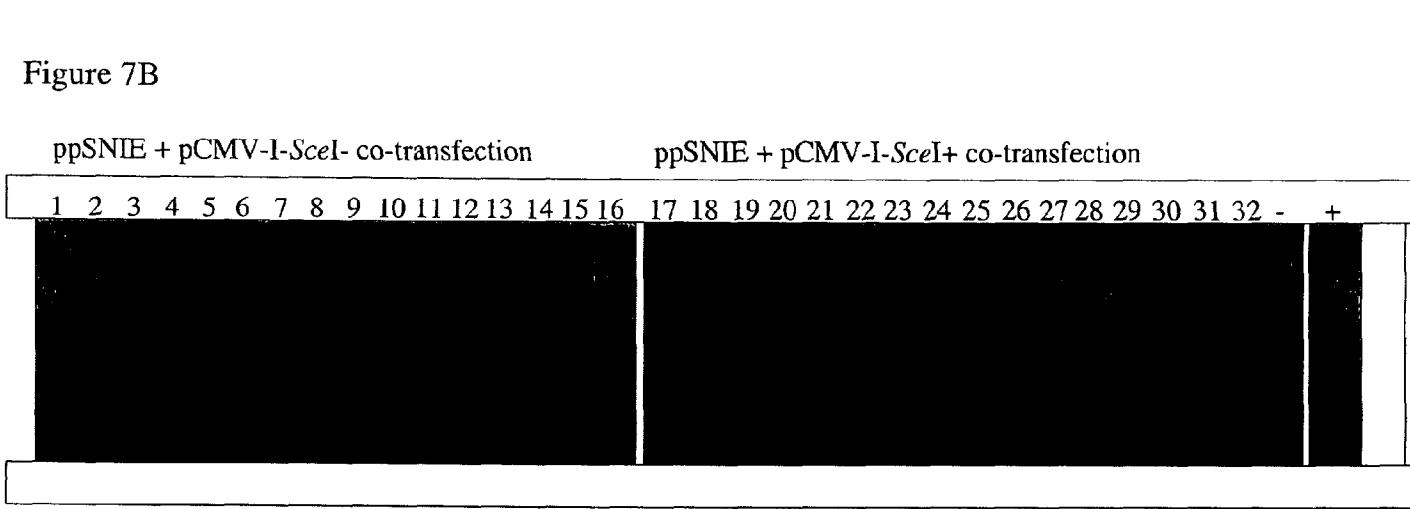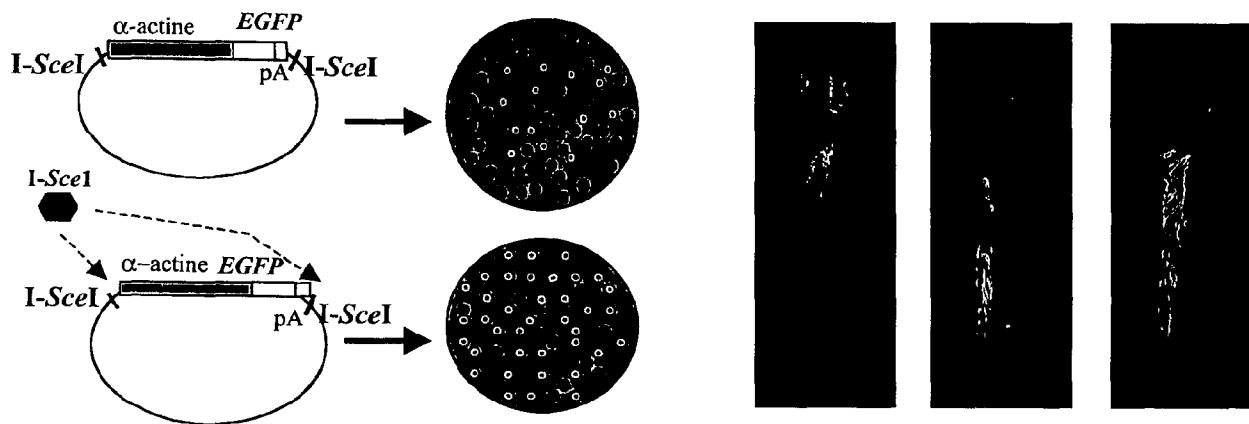
Figure 8A

Figure 8B

	N ●	W ●	M ◉	S ◉
Plasmid without enzyme	56%	16%	19%	9%
Plasmid with enzyme	13%	9%	39%	38%
Plasmid (mutated site) with enzyme	72%	14%	8%	6%
Plasmid (deleted site) with enzyme	69%	14%	12%	5%

Figure 9A

|	Rate of Fo transgenics (%)	Average transmission rate	Transmission rate : σ
Linear plasmid without enzyme	2/35 (5.9)	15.1%	17.7%
Circular plasmid without enzyme	7/32 (21.8)	17.6%	22.2%
Circular plasmid with enzyme	11/36 (30.5)	48.4%	9.1%

Figure 9B

FIGURE 10A.

FIGURE 10B.

FIGURE 11A.

1 25 2 3 4 A B

FIGURE 11B.

ß-Globin   Red5

1 25 2 3 4 A B - *wt* C 1 25 2 3 4 A B + *wt* C

FIGURE 11C.

1 25 2 3 4 *wt* + C

FIGURE 12

FIGURE 13

Figure 14

Figure 15

Gas5 promoter
Probe

Lago
Probe