論文 / 著書情報
Article / Book Information

題目(和文)	枯草菌ゲノムベクターシステムを用いたクラスⅠ嗅覚受容体遺伝子の発現制御領域の解析
Title(English)	Studies on the cis-element for mouse class I odorant receptor genes using the Bacillus subtilis genome vector system
著者(和文)	岩田哲郎
Author(English)	Tetsuo Iwata
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第9686号, 授与年月日:2014年12月31日, 学位の種別:課程博士, 審査員:廣田 順二,中村 聡,和地 正明,福居 俊昭,蒲池 利章
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第9686号, Conferred date:2014/12/31, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

Studies on the *cis*-element for mouse class I odorant receptor genes using the *Bacillus subtilis* genome vector system

A DOCTORAL THESIS

Submitted to the
Department of Bioengineering
Tokyo Institute of Technology
2014

Tetsuo IWATA

Department of Bioengineering
Tokyo Institute of Technology
Nagatsuta-cho 4259, Midori-ku
Yokohama 226-8501, Japan

SUMMARY

Technological developments in chromosome engineering are essential for the manipulation and functional analysis of genomic DNA fragments. Artificial chromosomes, such as bacterial and yeast artificial chromosomes (BACs and YACs), have been used for these purposes in combination with transgenesis. However, there exist several technological limitations in cloning size, genetic modification and insert stability. Through this study, I demonstrated the utility of the *Bacillus subtilis* genome (BGM) vector for engineering large DNA fragments to modify and reconstruct genomic structure and to generate transgenic mice.

The BGM vector is a novel cloning system for large DNA fragments, in which the entire 4.2 Mb genome of *B. subtilis* functions as a vector. The BGM vector system has several attractive properties, such as a large cloning capacity of over 3 Mb, stable propagation of cloned DNA and various modification strategies using RecA-mediated homologous recombination. However, genetic modifications using the BGM vector system have not been fully established, and this system has not been applied to transgenesis. Through this study, I provided a complete genetic modification method of cloned DNA fragments, including insertion, deletion, inversion and fusion to elongate a DNA fragment. I further demonstrated that the modified, enlarged genomic DNA fragments could be used to generate transgenic mice.

To explore the potential of the BGM vector, I focused on the mouse class I odorant receptor (class I OR / fish-like OR) family, which consists of 158 genes and forms a single gene cluster in the genome. Although a *cis*-acting locus control region is expected to activate singular OR gene expression, such element has not been

experimentally identified yet. Using the BGM vector system, I modified and reconstructed genomic structure of a part of class I OR gene cluster for transgenic reporter assay. Transgenic mice carrying the enlarged transgene recapitulated the expression and axonal projection patterns of the target class I OR gene in the main olfactory system. The deletion analysis showed that the approximately 13 kb region upstream the class I OR gene contains a *cis*-acting element necessary for the transgene expression. Further functional analyses revealed that the 3.8 kb region in the 13 kb upstream sequence was capable of recapitulating the expression pattern of class I OR gene in the main olfactory epithelium. In addition, bioinformatics analyses showed that the 3.8 kb region contained a highly homologous sequence between human, and which was conserved in at least 13 mammalian species. Thus, using the BGM vector system, I provided the first experimental evidence of a *cis*-acting element for class I OR gene expression. Since this system can be applied to any BAC and other library resources of any species, the BGM vector system provides a novel tool for recapitulation of genomic structure and functional studies of genomic DNA.

ACKNOWLEDGEMENT

I would like to take this opportunity to thank everyone who has helped me over the last six years. First of all I would like to thank my supervisor, Dr. Junji Hirota, for his continuous support, discussion and encouragement since I began my research works at the Tokyo Institute of Technology in 2009. Under his supervision, I thoroughly enjoyed my works and felt my research at his laboratory meaningful. I express my gratitude to Dr. Shinya Kaneko (Tokyo Institute of Technology) for valuable advice to BGM vector experiments and for instruction of handling the *B. subtilis* step by step. I would like to thank Drs. Yuh Shiwa (Tokyo University of Agriculture (current affiliation: Iwate Medical University)) and Hirofumi Yoshikawa (Tokyo University of Agriculture) for great support in the DNA sequencing analysis. I would like to thank Drs. Yoshihito Niimura and Kazushige Touhara (The University of Tokyo) for great support in the large-scale bioinformatics analysis. I would like to thank Dr. Takayuki Enomoto (Tokyo Institute of Technology) for great support in immunohistochemistry and for useful suggestions. I would like to thank Drs. Naoki Kobayashi (Tokyo Institute of Technology) and Kenta Sumiyama (National Institute of Genetics (current affiliation: RIKEN)) for instruction of the pronuclear-microinjection. I thank Dr. Yoshihiro Yoshihara (RIKEN) for the gift of plasmids and for internship for three month. I also thank Dr. Mitsuhiro Itaya (Keio University) for the gift of *B. subtilis* strains and plasmids and for fruitful discussion. Finally, I engrave the kindness of the members in the Hirota laboratory on my mind.

September 2014

Tetsuo IWATA

CONTENTS

ABBREVIATIONS

BAC	bacterial artificial chromosome
BCIP	5-bromo-4-chloro-3-indolyl phosphate
BGM	*Bacillus subtilis* genome
BReT	Bacillus Recombinational Transfer
Bs	blasticidin S (resistance gene: *bsr*)
CHEF	contour-clamped homogeneous electric fields (gel electrophoresis)
ChIP	chromatin immunoprecipitation
CI	lambda (λ) phage CI repressor
Cm	chloramphenicol (resistance gene: *cat*)
Cm	chloramphenicol resistance gene for *E. coli*
CsCl	cesium chloride
CTCF	CCCTC-binding factor
DAPI	4',6-diamidino-2-phenylindole
DNA	deoxyribonucleic acid
EDTA	ethylenediamine -N,N,N',N'-tetraacetic acid, disodium salt
EGFP	enhanced green fluorescent protein
Em	erythromycin
Emx2	empty spiracles homeobox 2
ES cell	embryonic stem cell
EtBr	ethidium bromide
GpBR	genomic pBR322 sequence
HAC	human artificial chromosome
IHC	immunohistochemistry
IRES	internal ribosome entry site
kp	kilo base pair
Lhx2	LIM-homeobox protein 2
Mb	mega base pair
MOE	main olfactory epithelium
MOR	mouse odorant receptor
NBT	nitro-blue tetrazolium chloride
Nm	neomycin (resistance gene: *neo*)
OB	olfactory bulb

OGG	orthologous gene group
OMP	olfactory marker protein
OR	odorant [olfactory] receptor
OSN	olfactory sensory neuron
PAC	P1 derived artificial chromosome
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PFA	paraformaldehyde
Phl	phleomycin (resistance gene: *phl*)
Pr	lambda (λ) phage Pr promoter
RNA	ribonucleic acid
SDS	sodium dodecyl sulfate
Spc	spectinomycin (resistance gene: *spc*)
Tet	tetracycline (resistance gene: *tet*)
Tg	transgene / transgenic mouse
TSS	transcription start site
YAC	yeast artificial chromosome

CHAPTER I

GENERAL INTRODUCTION

I - 1 General overview of manipulation of genomic DNA

DNA cloning is the first step for investigation of genes and their functions. Traditional genetic engineering based on restriction enzymes and DNA ligase to cut and fuse DNA fragments using multi-copy plasmid vectors, has made a great success using *Escherichia coli* (*E. coli*). This technology has been confined to clone and manipulate small DNA fragments, because naked large DNA fragments rarely have unique restriction enzyme sites and are exposed to physical shearing in liquid. Thus, alternative technologies for manipulation and engineering of large genomic DNA molecules have been required for the physical mapping, manipulation and functional analysis of genomic DNA fragments. To manipulate large genomic DNA fragments, several artificial chromosomes have been developed such as yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), P1 derived artificial chromosomes (PACs) (Sternberg, 1990) and human artificial chromosomes (HACs) (Harrington *et al*., 1997). Of these, the YACs and BACs have been often used to generate genomic libraries and to manipulate genomic DNA fragments. These two systems have complementary advantages over each other in terms of cloning capacity and insert stability.

The YAC system is a high-capacity cloning system based on *Saccharomyces cerevisiae* (Burke *et al*., 1987). The YAC is composed of some functional units of yeast chromosomes: a centromere, an autonomously replicating sequence and two sequences that seed telomere formation (Figure 1-1A). Genomic DNA inserts of up to 2 Mb can be used with the YAC system. In yeast, the DNA double-stranded-break and -repair reaction are very efficient, allowing easy modifications of the inserts by homologous recombination. However, the YAC system has some problems in terms of stability of the cloned inserts.

YACs are often chimeric, i.e., single YAC clone often contains genomic fragments derived from multiple loci because of co-cloning events (Green *et al*., 1991). YACs often suffer from unwanted rearrangements due to a potent and constitutive recombinase activity (Kouprina *et al*., 1994). In addition, the isolation of intact YACs is difficult because of their linear structure and contamination with endogenous yeast chromosomes. Thus, the YAC system has an advantage in cloning capacity, and has disadvantages in insert stability and purification (Table 1-1).

The BAC system is based on *E. coli* and its single-copy plasmid F factor (Shizuya *et al*., 1992). Replication of the F plasmid in *E. coli* is strictly controlled by some regulatory genes including *oriS*, *repE*, *parA* and *parB* (Kline, 1985). Because the BAC vector contains these genes, BAC clones are maintained in low copy number (one or two copy per cell) in a *recA* deficient stain, reducing the potential for unwanted recombination events (Shizuya *et al*., 1992) (Figure 1-1B). BAC clones are easy to manipulate and retrieve due to their plasmid form. The BAC system can generally accommodate up to 300 kb genomic inserts. Modifications of BAC inserts require additional recombination components, e.g., RecA of *E. coli* (Yang *et al*., 1997; Gong *et al*., 2002) or phage-derived recombination proteins, RecE and RecT carried by the cryptic *Rac* prophage (called ET cloning) (Zhang *et al*., 1998) and Redα and Redβ carried by the bacteriophage-λ (RedET or defective λ prophage system) (Muyrers *et al*., 1999; Copeland *et al*., 2001; Lee *et al*., 2001). These phage-encoded recombination factors are capable of modifying BAC inserts using PCR-generated linear double strand DNA targeting cassettes (Zhang *et al*., 1998; Copeland *et al*., 2001). The recombination based genetic manipulations of BACs are very attractive techniques. However, introduction of such additional components complicates the

modification procedures. Thus, the BAC system has advantages in handling and insert stability, and has disadvantages in cloning size and insert modification. (Table 1-1)

I - 2 The *Bacillus subtilis* genome vector system

The *Bacillus subtilis* genome (BGM) vector system has been developed as a novel cloning system using a unique concept in which the entire 4.2 Mb genome of *B. subtilis* functions as a vector (Itaya, 1993; Kaneko *et al*., 2005; Itaya *et al*., 2008; Itaya and Tsuge, 2011). The cloning strategy for the BGM vector system is based on natural competence of *B. subtilis* (Dubnau, 1999; Chen *et al*., 2005). *B. subtilis* expresses competence-related genes at the late stage of cell growth, and their products, transformation machinery molecules (Com proteins), are assembled in the cell membrane. The transformation machinery non-specifically binds and imports extracellular DNA fragments into the cytoplasm in single-stranded form. The recombinogenic DNA is incorporated into the *B. subtilis* genome by homologous recombination (Figure 1-2). In the BGM vector, the *B. subtilis* genome is engineered to have homologous cloning site sequences so that a target DNA fragment flanked by homologous sequences can be cloned into the BGM (Itaya and Tsuge, 2011). *B. subtilis* can be cultivated under the same conditions as *E. coli* in LB broth at 37°C. Competent *B. subtilis* cells can be prepared by merely cultivating the cells in a special medium for several hours, and these cells are transformed by simply mixing in DNA fragments without additional heat-shock or electroporation (Itaya and Tsuge, 2011). Thus, the competency of *B. subtilis*/BGM vector enables easy transformation procedures and efficient recombination reactions.

Cloning principles in the BGM vector

To utilize the *B. subtilis* genome as cloning vehicles, it should be engineered to have the cloning site for exogenous DNA, because such exogenous DNA generally would have no homologous sequences with the host *B. subtilis* genome. The first BGM vector series were constructed by inserting an *E. coli* plasmid vector pBR322, which cannot replicate in *B. subtilis* as plasmid, to mediate homologous recombination reaction (Itaya, 1993; Itaya, 1995). The classical BGM vector series can clone pBR322 inserts composed of exogenous DNA fragment and antibiotic resistance genes via homologous recombination at pBR322 sequences (genomic pBR: GpBR). For general use, a counter-selection system was developed to clone DNA segments without antibiotic resistance markers for *B. subtilis* (Itaya, 1999; Itaya *et al*., 2000). This system uses two cassettes. One is a bacteriophage lambda Pr promoter directed neomycin resistance gene, Pr-*neo* cassette. The other is a *c*I-*spc* cassette composed of bacteriophage lambda CI repressor gene and spectinomycin resistance gene. The original BGM strains carrying the *c*I-*spc* cassette are resistant to spectinomycin (Spc) and sensitive to neomycin (Nm), because the constitutively expressed CI repressor protein specifically binds to Pr promoter and repress the promoter activity. Once the *c*I-*spc* cassette is replaced with the incorporated inserts, the recombinant BGM clones become resistant to Nm and sensitive to Spc due to loss of the repressor. In combination with the internal selection system, the specific BGM vector for cloning the BAC insert was established by inserting a BAC vector pBAC108L (Kaneko *et al*., 2005; Kaneko *et al*., 2009) (Figure 1-3). These BGM vectors can utilize BAC library clones in a one-step transfer. In addition, two I-PpoI recognition sequences, ATGACTCTCTTAA /GGTAGCCAAA, are introduced at the both side of the cloning site, allowing simple

evaluation of the inserts.

Genome DNA cloning and gene manipulation

The BGM vector system is notably valuable for cloning giant DNAs (Table 1-1). Several genome DNA have been cloned into the BGM vectors, including 3.5 Mb *Synechocystis* whole genome (Itaya *et al*., 2005), mouse genomic DNAs (Itaya *et al*., 2000; Kaneko *et al*., 2003), 150 kb rice chloroplast DNA (Itaya *et al*., 2008), 16 kb mouse mitochondrial DNA (Itaya *et al*., 2008), lambda phage DNA (Itaya, 1995), mouse BAC clones (Kaneko *et al*., 2009) and *Arabidopsis* BAC clones (Kaneko *et al*., 2005). The cloned inserts in the BGM vector show high structural and genetic stability, because the DNA fragments are directly inserted into the single circular host genome (Kaneko *et al*., 2005). To achieve genome DNA cloning and extension, there are some large DNA reconstruction/bottom-up approaches named as the 'inchworm method' (Itaya *et al*., 2005) and 'domino method' (Itaya *et al*., 2008). In the 'inchworm method', natural guest DNA directly fills a gap between the two separate fragments to elongate DNA (Itaya *et al*., 2005). As the gap ranges about 40-50 kb, the elongation rate is faster. But this method requires long continuous DNA (>100 kb) as template. In contrast, 'the domino method' connects a number of DNA pieces (5-10 kb, called domino clones) one-by-one by homologous recombination between overlapping sequences of pBR322 sequence and guest DNA (Itaya *et al*., 2008). Therefore, the elongation rate in the domino method is slower than that in the inchworm method. However, this method can utilize DNA fragments prepared by PCR or by chemical synthesis, thus allowing application in all organisms and even in generation of totally artificial genomes.

The inserts in the BGM vector can be engineered through various modification strategies based on RecA-mediated homologous recombination. Using the BGM vector system, genetic modifications of insertion of *B. subtilis* marker gene, deleting and fusing cloned inserts have been achieved (Kaneko *et al*., 2003; Kaneko *et al*., 2009). Assembled and modified DNA molecules in the BGM vector are to be isolated for general applications such as biomaterial production and further genetic experiments. Some retrieval methods for BGM inserts were proposed including direct isolation in test tube by I-PpoI digestion (Kaneko *et al*., 2009), direct isolation *in vivo* by genome segregation (Itaya and Tanaka, 1997) and copying the insert by Bacillus Recombinational Transfer (BReT) (Tsuge and Itaya, 2001) (Figure 1-4). Overall, the BGM system is a novel and an alternative platform for chromosome engineering (Table 1-1).

I - 3 Mouse transgenesis

The generation of transgenic animals is a routine method in many laboratories. Transgenesis is often used to study gene functions *in vivo* and to provide animal models of genetic diseases (Giraldo and Montoliu, 2001). The size of transgenes is generally ~10 kb because of the capacity of multi-copy plasmid vector. However, mammalian functional gene units composed of exons, introns and regulatory elements often reach as long as several hundred kb. To harbor these genomic regions, artificial chromosome transgenesis, including BAC and YAC transgenesis, have contributed greatly to genome researches, because transgenesis with large DNA molecules allows transgene expression to recapitulate the endogenous gene expression patterns (Schedl *et al*., 1993; Mendez *et al*., 1997; Yang *et al*., 1997; Giraldo and Montoliu, 2001; Kuroiwa *et al*., 2002).

Methods for generating transgenic mouse

The most conventional method of introduction of transgenes is the pronuclear microinjection of purified DNA (Schedl *et al*., 1992; Peterson *et al*., 2007) (Figure 1-5A). In this method, linearized DNA molecules are directly introduced into fertilized mouse eggs with a fine glass capillary, and then are randomly integrated into mouse chromosomes. There is the upper limit in size of DNA, because long DNA molecules are easily damaged by physical shearing during purification and injection. For introduction of larger DNA transgenes such as YACs, other methods were reported. One is lipofection of purified DNA into embryonic stem (ES) cells (Strauss *et al*., 1993; Lee and Jaenisch, 1996). In this method, purified YACs are mixed with cationic lipids, and the YAC-lipid complex is introduced into ES cells by endocytosis. The transfectant ES clones are used to generate chimeric mice. By crossing the chimeric mice with wild type, transgenic lines are established. Another is the yeast spheroplast fusion with ES cells (Jakobovits *et al*., 1993; Li and Blankenstein, 2013). In this strategy, intact Mb-sized YACs can be introduced into the germline without purification of YAC DNA. Because ES cells retain their pluripotency after yeast spheroplast fusion, the fused ES cells can be used to produce chimeric mice. Although spheroplast fusion method reliably yield large numbers of transfectants with intact YAC transgenes, the incorporation of endogenous yeast genomic DNA potentially complicates interpretation of results. In contrast, pronuclear microinjection and lipofection use purified DNA but find high frequency of transgene deletion because of physical shearing.

Tol2-transposon mediated transgenesis

Generally, the overall transgenic efficiency is low, only 2-5% of pronuclear- injected fertilized eggs. To improve transgenesis in mice, new methods have been developed including virus-based vector and transposon systems. Recently, it was reported that the *Tol2* transposon system could be applied to conventional and BAC transgenesis in mice (Suster *et al*., 2009; Sumiyama *et al*., 2010). The *Tol2*-mediated cytoplasmic injection method uses a special plasmid vector including minimal *Tol2* transposable elements (Urasaki *et al*., 2006). A foreign DNA cloned in the *Tol2* vector is injected into the cytoplasm of fertilized eggs with the transposase mRNA, reducing physical damages to eggs (Figure 1-5B). The foreign DNA is efficiently transposed from plasmid to the genome by synthesized *Tol2* transposase. The efficient DNA integration and small physical damage to injected egg greatly improve the overall transgenic efficiency: more than 20% (Sumiyama *et al*., 2010). The *Tol2* transposon system could offer following possible advantages; a transgene is integrated as a single copy but not as a concatemer; transposition does not cause gross rearrangements around integration sites; the integration sites are identifiable. Note that transgenic mice generated by the *Tol2* method generally have several independent integration sites because of the potent transposition activity, whereas those generated by the pronuclear injection method generally are single integration (Sumiyama *et al*., 2010).

I - 4 Mouse main olfactory system and odorant receptor genes

The mammalian main olfactory system can detect millions of different types of chemicals in the environment. Animals perceive these odorous molecules to find food

sources and offspring and to avoid predators and rotten foods. Olfactory sensory neurons (OSNs), the primary sensory neurons in the main olfactory epithelium (MOE) can detect odorants to initiate the sense of smell (Figure 1-6A). The detection is mediated by odorant receptors (ORs), which are G-protein coupled receptors with a putative seven-transmembrane domain and are located on the cilia of OSNs. To distinguish a vast number of odorants, the vertebrate genome possesses numerous OR genes, which form the largest multi-gene family in vertebrates.

There are two principles underling the functional organization of the olfactory system. One is the one-neuron–one-receptor rule; each OSN expresses only one functional odorant receptor (OR) gene from a huge repertoire of OR genes in a monoallelic manner (Chess *et al*., 1994; Malnic *et al*., 1999). This singular OR gene expression is considered to be established by a negative feedback regulation, in which functional expression of one OR gene inhibits additional OR genes expression (Serizawa *et al*., 2003; Lewcock and Reed, 2004). The other is OR-instructed axonal projection. Targeted mutagenesis demonstrated OSNs expressing the same OR were shown to converge their axons to a specific set of glomeruli in the olfactory bulb (OB) (Vassar *et al*. 1994; Ressler *et al*. 1994; Mombaerts *et al*. 1996) (Figure 1-6B). In addition, the dorsal-ventral arrangement of glomeruli in the OB is correlated with the locations of OSNs in the MOE (Miyamichi *et al*., 2005). Therefore, whereas OSNs are scattered in the MOE, olfactory signals from OSNs are reorganized and projected as a topographical map in the OB.

Two classes of OR genes

In mouse, ~1400 OR genes have been identified (Zhang *et al*., 2007). Phylogenetic analysis classified OR family into two broad but distinct groups, class I and class II (Zhang and Firestein, 2002). The class I (also called fish-like) OR family is a phylogenetically ancient mammalian OR family (Glusman *et al*., 2001) and includes 158 genes, forming an approximately 3 Mb gene cluster on chromosome 7 (Zhang *et al*., 2007) (Figure 1-6C, 1-7). Class I OR genes are expressed almost exclusively by OSNs in the dorsal MOE, and these OSNs project their axons to a specific subset of glomeruli in the dorsal domain of the OB (Tsuboi *et al*., 2006; Bozza *et al*., 2009). There are at least two exceptions called as non-classical class I OR, which are expressed in a more ventral region (Tsuboi *et al*., 2006). In contrast, class II OR genes are terrestrial specific and are distributed throughout the mouse genome (Niimura and Nei, 2005; Zhang *et al*., 2007) (Figure 1-6C). Class II OR genes are expressed throughout the dorsal and ventral MOE, and the expression areas are specific to each class II OR genes (Miyamichi *et al*., 2005).

***Cis*- and *trans*-regulatory elements for OR genes**

The molecular mechanisms underlying OR genes expression remain unclear. To uncover this issue, several *cis*-regulatory elements have been identified for class II OR genes expression. Transgenic analyses using minimal OR gene regions showed that a few hundred bp of upstream regions of the transcription start site (TSS) are sufficient for some class II OR transgene expressions including *M71*, *MOR23* and *MOR262-12* (Vassalli *et al*., 2002; Rothman *et al*., 2005; Zhang *et al*., 2007) (Figure 1-6D). Indeed, the ~300 bp promoter segments of some class II OR genes are capable of driving expression of a

reporter gene (Vassalli *et al*., 2011). In contrast, some class II OR genes need more distal regulatory regions to activate OR genes; H region at ~60 kb upstream of the *MOR28* cluster on chromosome 14 (Serizawa *et al*., 2003) and P element at position between *P3* and *P4* OR genes on chromosome 7 (Bozza *et al*., 2009) (Figure1-6D). Mutagenesis of the regulator regions revealed that *cis*-element in proximal position activates only one proximal OR gene (Rothman *et al*., 2005), whereas the distal *cis*-elements, such as H region and P element can activate one of several class II OR genes in a cluster (Serizawa *et al*., 2003; Fuss *et al*., 2007; Nishizumi *et al*., 2007; Khan *et al*., 2011). Thus, there are two type regulatory regions in proximal and distal positions depending on OR genes.

The class II OR regulatory regions were shown to contain conserved motifs of homeodomain and O/E binding sites (Vassalli *et al*., 2002; Rothman *et al*., 2005; Nishizumi *et al*., 2007; Vassalli *et al*., 2011). The comprehensive promoter analyses suggested involvements of the two motifs in a large proportion of OR gene promoters (Clowney *et al*., 2011; Plessy *et al*., 2012). Mutagenesis of theses motifs in *M71* promoter (proximal) and H region (distal) demonstrated that the homeodomain and O/E-like motifs played important roles in the class II OR gene expression (Rothman *et al*., 2005; Nishizumi *et al*., 2007).

Regarding to the *trans*-elements for OR gene expression, LIM-homeobox protein 2 (Lhx2) and empty spiracles homeobox 2 (Emx2) were identified as binding proteins in homeodomain site of *M71* promoter region (Hirota and Mombaerts, 2004). Analysis of *Lhx2*-deficient mutant mice demonstrated that Lhx2 was required for all tested class II OR gene expression and OSN development (Hirota and Mombaerts, 2004). Further analysis showed that class II OR and non-classical class I OR expression were abolished in the

mutant mice, but class I OR expression was largely spared, suggesting mechanistic differences in class I versus class II OR gene choice (Hirota *et al*., 2007). Loss of Emx2 decreased many OR genes expression irrespective of their classes and increased some class II OR genes expression (McIntyre *et al*., 2008). Studies on OR gene expression have been so far conducted on mainly class II OR genes, and there are several findings of regulatory elements for class II OR genes. By contrast, neither *cis*- nor *trans*-elements for class I OR genes have been identified yet.

I - 5 Purpose of this study

The BGM vector system has several attractive properties including large cloning capacity, insert stability and amenability of various modification strategies. However, this system is a developing technology, i.e., genetic modifications have not been fully established and this system has not been applied to transgenesis. Therefore, I explored the potential utility of the BGM system in chromosome engineering and mouse transgenesis. For this purpose, I focused on the class I OR gene family (Figure 1-7A, B). Although a *cis*-acting locus control region is expected to regulate transcription, such a region has not been found (Hirota *et al*., 2007; Rodriguez, 2007) (Figure 1-7C).

In this study, I performed two new genetic modifications in the BGM system to complete the BGM-based gene manipulation methods including insertion, deletion, inversion and fusion (Chapter II). Using these techniques, I constructed some transgenes including a 252 kb reconstructed transgene, which was generated by fusing two BAC clones whose inserts were initially oriented in opposite direction. Mutations during DNA manipulation was examined by next-generation DNA sequencing.

To evaluate the potential utility of the BGM system in transgenesis, I applied this system to generate transgenic mice using two transgenes constructed in Chapter II (Chapter III). I established a protocol to retrieve the BGM insert and to generate transgenic mice by microinjection into fertilized mouse eggs. By analyzing the transgenic mice, I showed a first experimental evidence of a *cis*-acting element of a class I OR gene.

Finally, by combining the functional analyses of transgenes and the bioinformatics analyses, I identified the *cis*-element for the class I OR gene, *MOR42-3* (Chapter IV). Comparative analyses revealed that the *cis*-element consists of some functional elements to control the class I OR gene expression.

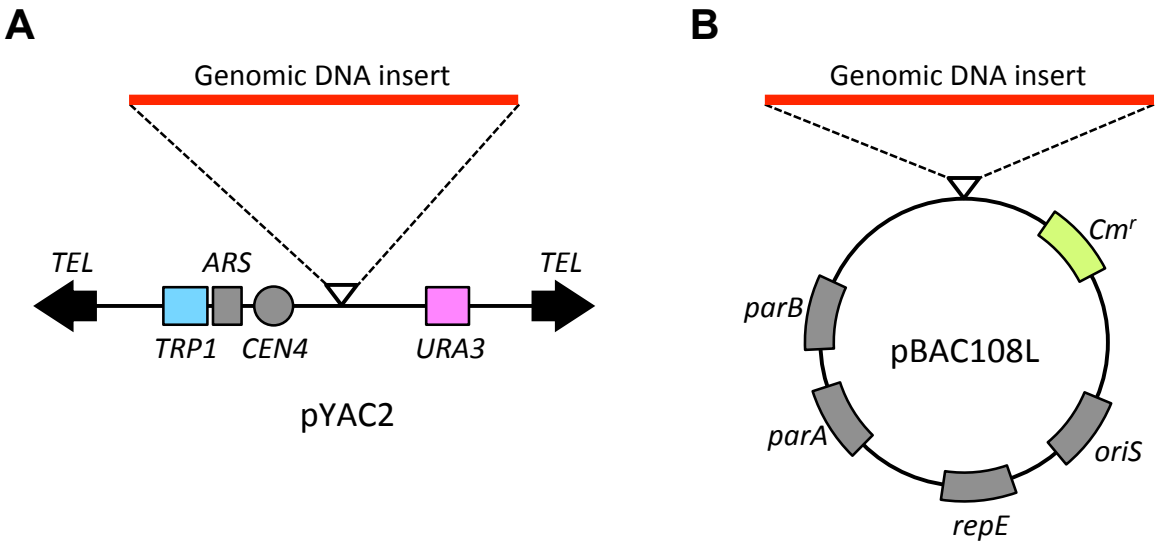

Figure 1-1. Structures of the YAC and BAC vectors.

(A) Schematic structure of a YAC clone (pYAC2, Burke *et al.*, 1987). The YAC vector contains an autonomously replicating sequence (*ARS*), a centromere (*CEN4*) and two telomere seed sequences (*TEL*) for replication and maintenance of YAC clone; and two selectable marker genes (*TRP1* and *URA3*). **(B)** Schematic structure of a BAC clone (pBAC108L, Shizuya *et al.*, 1992). The BAC vector contains several replication-related genes including *oriS*, *repE*, *parA* and *parB*, and a chloramphenicol resistance gene (*Cm*). The triangles indicate cloning sites.

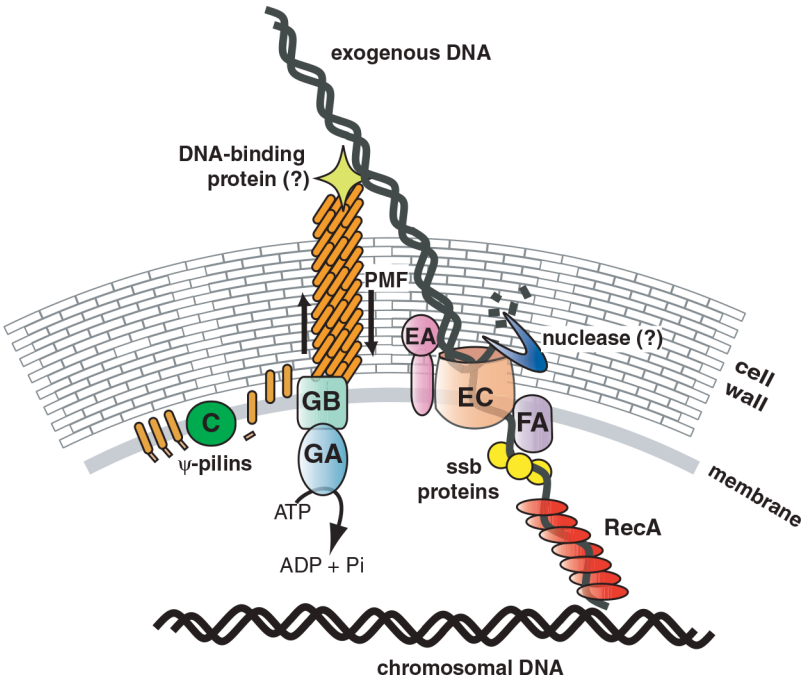

Figure 1-2. A model for the molecular mechanism of transformation in *B. subtilis*. (from Chen *et al.*, 2005)

The transformation machinery molecules are preferentially located at the cell poles. The pseudoprepilins (ψ-pilins) are processed by the peptidase (ComC) and translocate to the outer face of the membrane. With the aid of the other proteins (ComA and ComB), the pseudopilins assemble into the psudopilus, which attaches exogenous DNA via a hypothetical DNA binding protein. Incorporation of the DNA requires retraction of the psudopilus, driven by the proton motive force (PMF), and DNA binding to the receptor (ComEA). The one strand of DNA is transported through the membrane channel (ComEC) while the other is degraded by a nuclease existing the cell. The helicase/DNA translocase (ComFA) assists the process, along with single-strand DNA binding (ssb) proteins interact with the incoming DNA. RecA forms a filament around the single-strand DNA, and mediates a search for homology with chromosomal DNA.

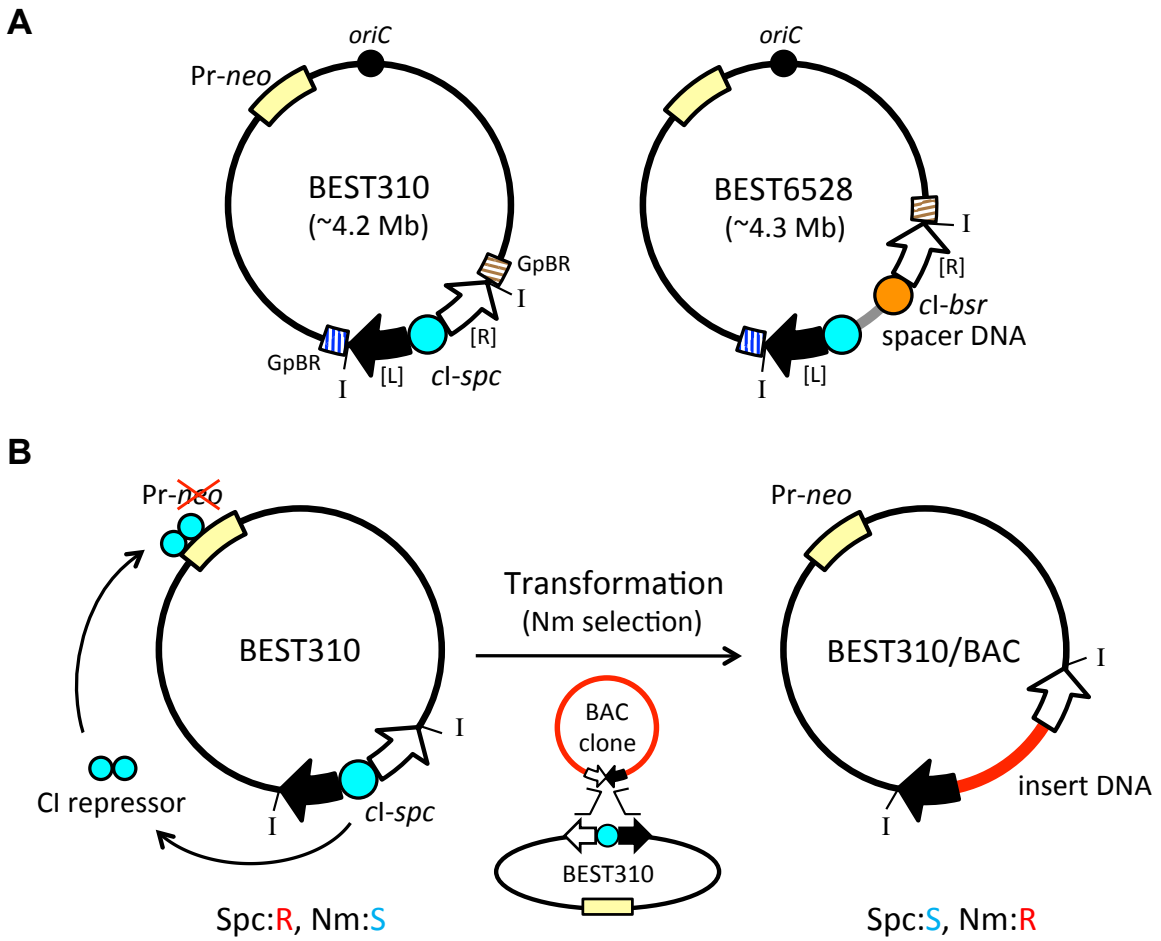

Figure 1-3. Strategies of BAC cloning into the BGM vectors.

(A) Structures of two BGM vectors for BAC cloning. BEST310 has two extra DNA segments (Kaneko *et al.*, 2005). One is a 4.3 kb pBR322 sequence, the genomic pBR (GpBR) sequence, in the NotI site of the *proB* locus at the position of 1,378 kb (*amp*-half, blue stripe; *tet*-half, brawn stripe) (Itaya, 1993; Kunst *et al.*, 1997). A modified pBAC108 plasmid was also inserted in this position with *c*I-*spc* cassette, which is composed of λ-phage *c*I repressor gene and spectinomycin resistance gene (cyan circles). The other segment is a Pr-*neo* cassette, which is composed of λ-phage Pr promoter and neomycin resistance gene, in the NotI site of *yvfC-yveP* locus at the position of 3,516 kb (Itaya, 1999). BEST6528 is another BGM vector for large BAC handling, contains a 100 kb spacer DNA (mtDNA of *Arabidopsis thaliana*) and two *c*I gene cassettes (*spc* and blasticidin S resistance gene, *bsr*) (Kaneko *et al.*, 2009). The BAC vector sequences are indicated by closed and open arrows (3.5 kb each), and the right half contains chloramphenicol resistance gene for *E.coli* (open arrow). The I-PpoI sites are introduced to flank the BAC vector sequences (designated as "I"). **(B)** Schematic illustration of BAC cloning. Before transformation, the BGM vector is resistant to Spc and sensitive to Nm (Spc:R, Nm:S) due to repression of the Pr-*neo* cassette by the CI repressor. The BAC inserts are integrated into the BGM vector via double crossings-over recombination with the BAC vector sequences. The recombinants become resistant to Nm and sensitive to Spc (Spc:S. Nm:R), because of replacement of the *c*I-*spc* cassette with the BAC insert.

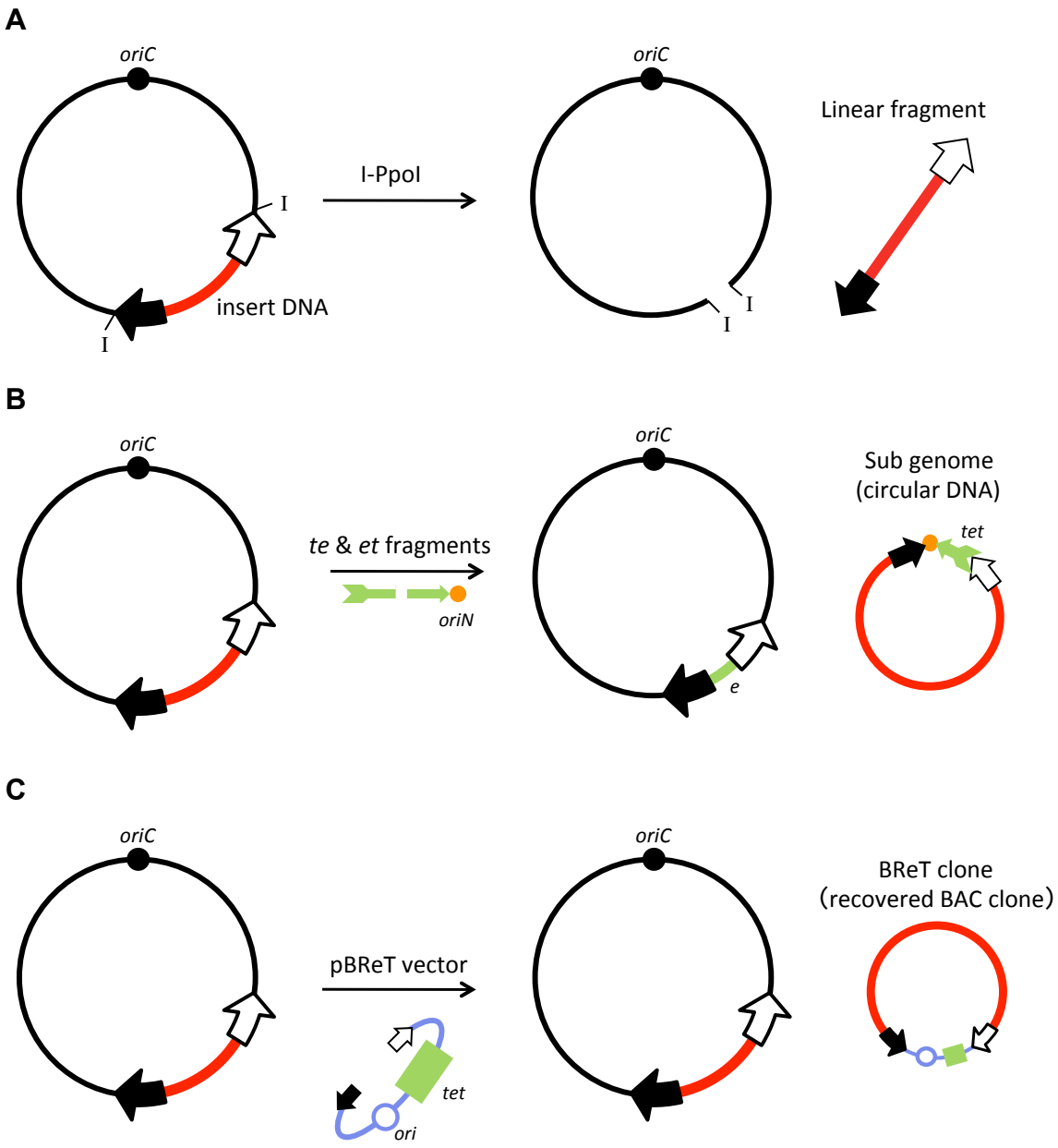

Figure 1-4. Methods for the insert retrieval from the BGM vector.

The BGM vector system has several retrieve methods. **(A)** A liner insert DNA can be retrieved by I-PpoI digestion (Kaneko *et al.*, 2009, Chapter III). The I-PpoI sites are represented by "I". **(B)** An insert DNA can be separated from the BGM vector in a circular form by the targeted insertion of the *te* and *et* cassettes into the ends of BAC vector sequences (Itaya and Tanaka, 1997). **(C)** BGM inserts can be reformed into a plasmid by transformation with a linearized pBReT vector (Tsuge and Itaya, 2001).

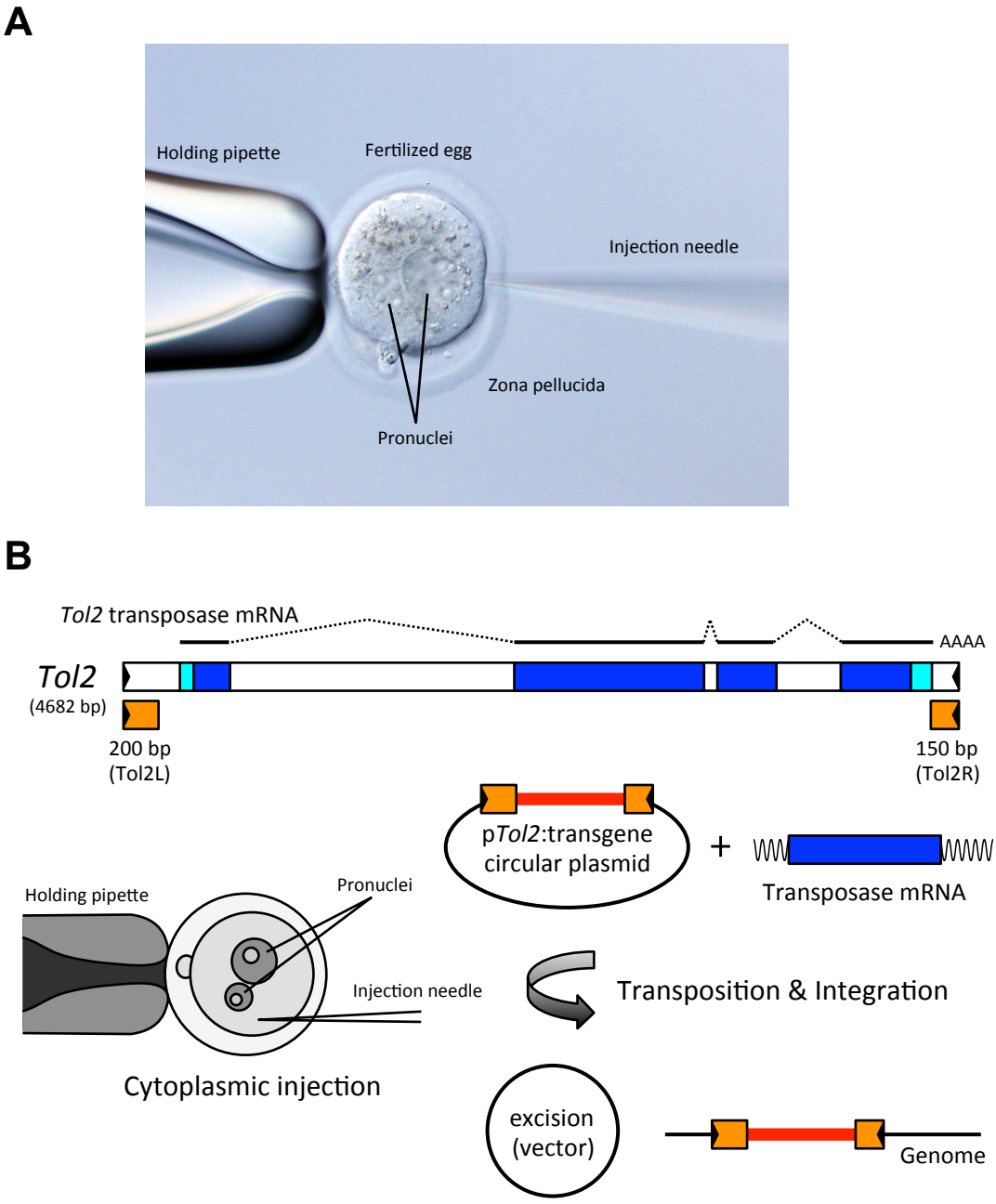

Figure 1-5. Methods for microinjections in the mouse transgenesis.

(A) A view of pronuclear microinjection. A linearized transgene is introduced into a pronucleus of a fertilized mouse egg using a fine glass needle. **(B)** *Tol2*-transgenesis system. The top illustration shows the structure of the full-length *Tol2* transposable element and the minimal *Tol2* sequences (orange, L and R) for *Tol2* vector. RNA transcribed from *Tol2* that encodes a transposase protein is depicted by lines (exons) and dotted lines (introns); coding regions, blue boxes; untranslated regions, cyan boxes; 12 bp terminal inverted repeats (TIRs), black triangles. The lower portion shows the *Tol2*-mediated cytoplasmic injection method. A transgene cloned in a *Tol2* transposon vector together with the transposase mRNA is injected into cytoplasm of fertilized eggs. The transgene is transposed from the plasmid to the host genome.

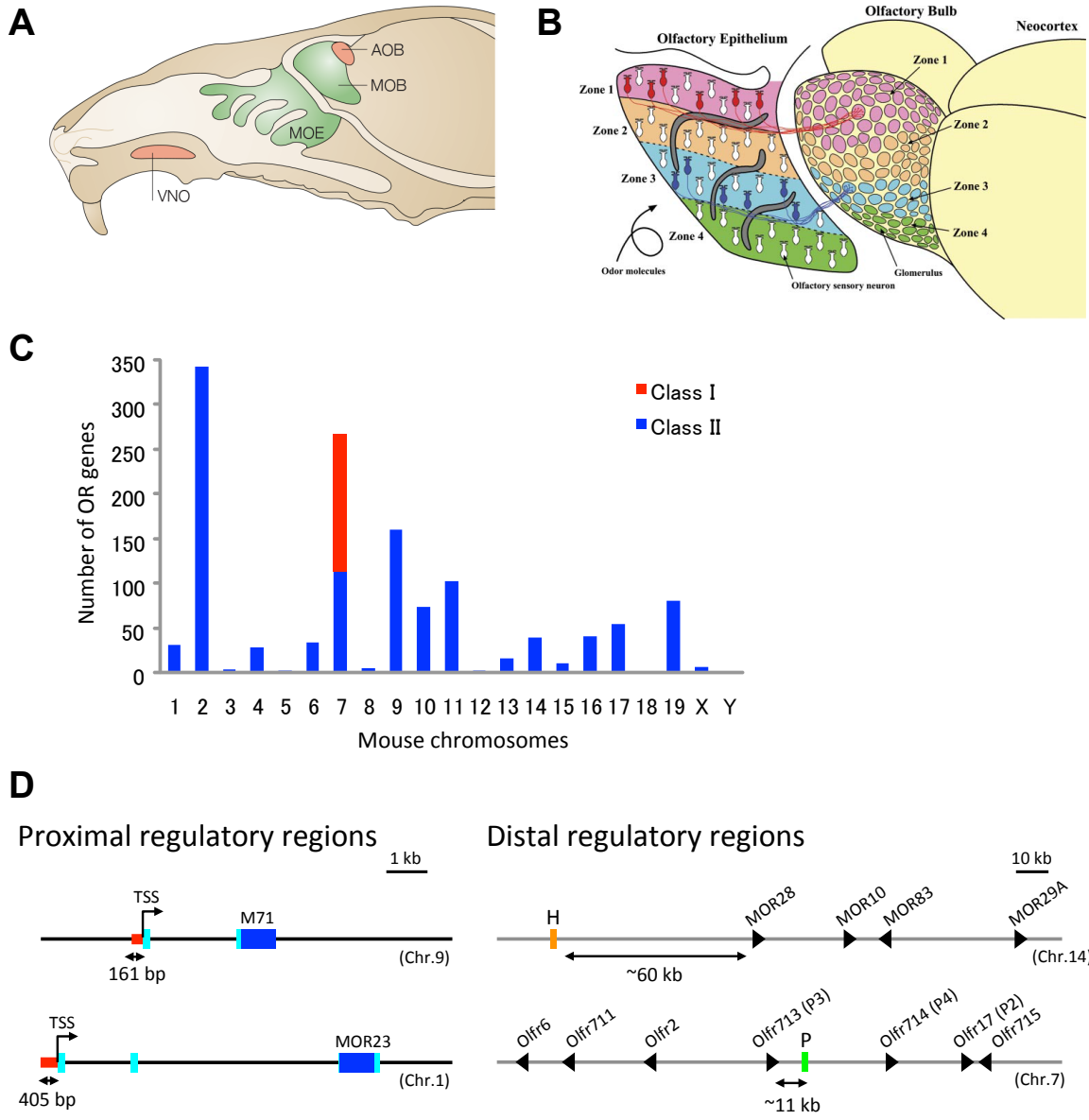

Figure 1-6. Mouse olfactory system and OR genes.

(A) Two major olfactory systems in mouse (from Mombaerts, 2004). Green color indicates the main olfactory system for detection of odorants; main olfactory epithelium (MOE) and main olfactory bulb (MOB). Red color indicates the vomeronasal system for detection of pheromones; vomeronasal organ (VNO) and accessory olfactory bulb (AOB). **(B)** Schematic illustration of the axonal connectivity pattern between MOE and OB (from Mori *et al.*, 2006). OSNs expressing a given OR are distributed widely in the MOE and converge their axons onto a few topographically fixed glomeruli that are located within a corresponding region of the OB. "Zone1-4" are traditional names of compartments of the MOE. **(C)** Distribution of OR genes in mouse genome (Zhang *et al.*, 2007). **(D)** Identified regulatory regions for class II OR genes. In proximal types (left), two mini-OR transgene regions of *M71* and *MOR23* are depicted (Vassalli *et al.*, 2002; Rothman *et al.*, 2005). Red bars, regulatory regions; Cyan, non-coding regions; Blue, coding regions. In distal types (right), H region in *MOR28* cluster and P element in P centromeric cluster are depicted (Serizawa *et al.*, 2003; Bozza *et al.*, 2009).

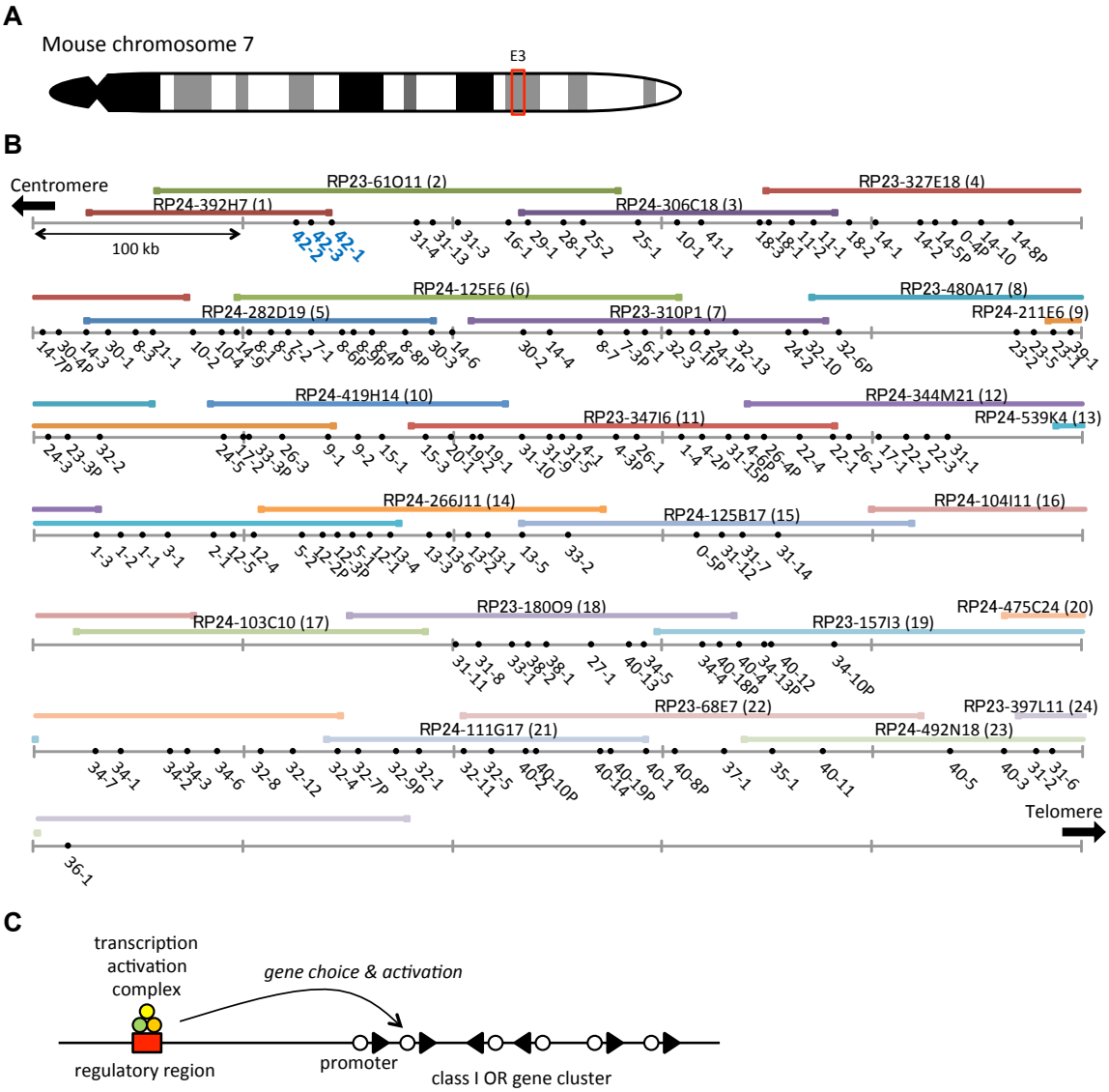

Figure 1-7. Mouse class I OR gene cluster and BAC clones.

(A) The ideogram of mouse chromosome 7. The class I OR gene cluster is located in the E3 region. **(B)** Structure of ~3 Mb mouse class I OR gene cluster. BAC clones, whose full-length insert sequences are available, are represented. In this study, the MOR42 family at most centromere position (blue) was chosen as a target for the transgenic analysis. Numbers in parentheses are BAC numbers in this study (ex. first BAC clone RP24-392H7 means 'BAC1'). **(C)** A hypothetical model for the mechanism of class I OR gene regulation. One regulatory region controls expression of multi-class I OR genes such as the locus control region.

Table 1-1. Properties of cloning platforms for the large genomic DNA.

System	Vector (host cell)	Cloning capacity	Advantages	Disadvantages
B. subtilis genome vector (BGM)	genome (*B. subtilis*) foreign DNA fragment  *B. subtilis* genome	over 3 Mb	- targeted cloning - easy manipulation - intrinsic homologous recombination-based modification - stability of insert - preservation	- developing tool
Bacterial artificial chromosome (BAC)	plasmid (*E.coli*)  BAC vector (including *ori*)	~ 300 kb (generally ~200 kb)	- genome DNA library - easy manipulation - stability of insert - isolation	- requirement of extra factors for modification - potentiality of loss - relatively small cloning size
Yeast artificial chromosome (YAC)	partial genome (*S. cerevisiae*) telomere telomere  ARS, centromere	~ 2 Mb	- genome DNA library - intrinsic homologous recombination-based modification	- instability of inset - isolation

Properties, advantages and disadvantages in the BGM, BAC and YAC systems are summarized. The BAC and YAC systems have complementary advantages over each other in term of cloning capacity and insert stability. The BGM system is an attractive alternative of the manipulation of large DNA, but this system is a developing technology.

CHAPTER II

DEVELOPMENT OF A COMPLETE GENETIC MANIPULATION METHOD FOR THE BGM VECTOR SYSTEM

II - 1 Introduction

The *Bacillus subtilis* genome (BGM) vector system has been developed as a novel cloning system using a unique concept in which the entire genome of *B. subtilis* functions as a vector. The specific features of the BGM vector system include a cloning capacity over 3 Mb, the stable propagation of cloned DNA fragments in a single copy per cell and the amenability of various modification strategies based on RecA-mediated homologous recombination. These advantages make it an attractive alternative for the manipulation of large DNA. It has been used to clone genomic DNA from several species, including cyanobacteria (Itaya *et al*. 2005), *Arabidopsis* (Kaneko *et al*. 2005) and mouse (Itaya *et al*. 2000; Kaneko *et al*. 2003). Modifications consisting of deleting and fusing cloned inserts have also been achieved (Kaneko *et al*. 2003; Kaneko *et al*. 2009). However, genetic modification methods have not been fully established in the BGM vector system; targeted insertion and inversion modifications remain to be demonstrated, and the fusion of two contiguous DNA fragments is limited to clones that are orientated in the same direction (Kaneko *et al*. 2009). Thus, the BGM vector system is a developing technology with attractive potential, which includes its megabase-cloning capacity and homologous recombination-based genetic modification.

In this study, I have added two new genetic modifications to the BGM system tool kit to establish BGM-based gene manipulation methods (Figure 2-1). One is targeted insertion of a reporter gene without introducing an unwanted selectable marker at the target site when constructing the transgene. Another is inversion of the cloned insert to align its orientation so that two contiguous BAC clones in opposite directions can be fused to reconstruct the genomic structure. In fact, inversion of a cloned insert has not been

reported in any other system. Through all experiments in Chapter II, I finally constructed transgenes for analyzing *cis*-elements for a class I OR gene.

II - 2 Experimental procedures

II-2-1) Strains and preparation of competent *B. subtilis cells*

The *B. subtilis* strains and the BAC-specific BGM vectors BEST310 and BEST6528 (Kaneko *et al*., 2005; Kaneko *et al*., 2009) were kindly provided by Dr. Itaya. These strains are derived from the restriction modification-deficient strain RM125 to possess a *c*I repressor gene cassette flanked by pBAC108-based BAC vector sequences (Shizuya *et al*., 1992; Kaneko *et al*., 2005) and are identical with the exception that BEST6528 contains a 100 kb spacer sequence and an additional *c*I repressor gene cassette (Kaneko *et al*., 2009). For defining insert orientation, the half of the BAC vector containing the chloramphenicol resistance gene for *E. coli* is defined as right side (open arrows in Figures). The preparation of competent cells and transformation of *B. subtilis* were performed as described elsewhere (Itaya and Tsuge, 2011). Strains containing multiple antibiotic-resistance genes were tested using the replica plating method. *B. subtilis* was routinely grown in 1-5 ml Luria-Bertani (LB) broth at 37°C by rotating (>50 rpm) or shaking (200 rpm). Antibiotic Medium 3 (Difco) was used for the selection of the BGM transformants with the following antibiotics, as appropriate: neomycin (Nm, 5 μg/ml, Sigma), spectinomycin (Spc, 50 μg/ml, Sigma), erythromycin (Em, 5 μg/ml, Sigma), phleomycin (Phl, 0.5 μg/ml, Sigma), tetracycline (Tet, 10 μg/ml, Nacalai) and blasticidin S (Bs, 250 μg/ml, Funakoshi).

II-2-2) One-step transfer of BAC inserts to the BGM vector

Two contiguous BAC clones, RP24-392H7 and RP23-61O11 (designated BAC1 and BAC2, respectively), were purchased from the Children's Hospital Oakland Research Institute. The BAC DNA was prepared by the alkaline lysis method and subsequent

equilibrium centrifugation in a CsCl-ethidium bromide gradient (Kaneko *et al*., 2003). BGM strains were transformed with the purified BAC DNAs (Kaneko *et al*., 2005). BAC1 and BAC2 were cloned into BEST310 and BEST6528, respectively.

II-2-3) I-PpoI/CHEF analysis

The cloned inserts in the BGM vectors were analyzed by I-PpoI digestion followed by CHEF electrophoresis. Agarose plugs containing BGM clones were prepared as described (Itaya and Tsuge, 2011). A sliced block of the agarose plug was soaked in 200 µl of I-PpoI digestion buffer for 15 min to replace the TE buffer in the agarose plug. After discarding the soaking buffer, the block was immersed in 100 µl of digestion buffer containing 20 U of I-PpoI (Promega) and incubated for 1-1.5 hours at 37°C. The plug containing the digested DNA was embedded in a well of a 1% (w/v) agarose gel and separated by CHEF (Bio Craft) in 0.5× TBE buffer (50 mM Tris–borate (pH 8.0), 1.0 mM EDTA). Electrophoresis was performed at 4 V/cm at 14°C under the following conditions: 12 sec for 21 h (BAC1, Tg-110), 30 sec for 22 h (BAC2, Tg-220) and 25 sec for 22 h (Tg-250SB).

II-2-4) Southern blot analysis of the BGM clones

Genomic DNA from the BGM clones was prepared using the liquid isolation method (Itaya and Tsuge, 2011). The genomic DNA was digested with EcoRI, HindIII or BamHI (TaKaRa). The digested DNA was separated in pulse-field gels at 3 V/cm, 18 sec for 14 h at 14°C, and the DNA was transferred onto a Hybond-N (GE Healthcare) membrane filter. The preparation of digoxigenin (DIG)-labeled DNA probes, Southern hybridization and

detection with NBT/BCIP were performed using a DNA labeling and detection kit (Roche). A BAC probe was prepared by the random prime labeling method using the target BAC clone as a template. A full-length R arm probe and *MOR42-3* coding probe (190-1091 from NM_020289) were prepared by PCR labeling method.

II-2-5) Construction of plasmids

To construct the reporter cassette insertion plasmid, a polylinker containing AscI, SpeI, BglII, NdeI and SphI sites was inserted between the XbaI and SacI sites of pBluescript II SK(+) (Stratagene) to construct the pT1 vector. An EcoRI-XbaI fragment of *IRES-tauEGFP* from the *iTGFP-ACNF* plasmid (Bozza *et al*., 2002) was cloned into pT1 to generate pT1-iTGFP. The 1.0 kb left (L) and right 1.0 kb (R) arms for the targeted insertion of *IRES-tauEGFP* into the 3 bp downstream of the *MOR42-3* stop codon were prepared by PCR and contained sequences that were homologous to the upstream and downstream *MOR42-3* insertion sites, respectively. The L arm was first cloned into the SalI-EcoRI site of pT1-iTGFP, and then the R arm was cloned into the SpeI-BglII site to generate the plasmid piTG423. The piTG423-CISP plasmid was constructed by inserting a selection marker-containing the *c*I-*spc* cassette into the AscI-SpeI site of the piTG423 plasmid.

To construct the inversion plasmid, the new BAC plasmid p108IPpoI-HPNSB was generated by inserting a linker containing PmlI, NotI and SphI sites between the HindIII and BamHI sites of the BAC plasmid p108NHBN-MIM (Kaneko *et al*., 2005). The *te-erm* cassette, which was obtained from a NotI-digested fragment from pBEAZ191 (Kuroki *et al*., 2008), was inserted into the PmlI site of this plasmid after blunt-end treatment to

generate the plasmid p108Term. Similarly, the *phl-et* cassette, a NotI fragment from pBEAZ195 (Kuroki *et al*., 2008), was inserted to generate the plasmid p108Phlet. A 0.9 kb fragment homologous to the left end of Tg-110 was cloned into the NotI site of p108Term. A 0.8 kb fragment homologous to the right side of Tg-110 (1.3 kb from the right end) was then cloned into the HindIII site of p108Phlet.

Fusion plasmids were also constructed based on p108IPpoI-HPNSB. The p108CISP plasmid was constructed by inserting the *c*I-*spc* cassette, a HindIII-BamHI fragment excised from pCISP310B (Kaneko *et al*., 2005), into the PmlI site after blunt-ending. Similarly, p108CIBS was constructed by cloning the *c*I-*bsr* cassette, a PstI fragment derived from pBEST10007. A 1.3 kb PCR fragment homologous to the left end of Tg-220 was cloned into the NotI site of p108CISP, and a 1.3 kb fragment homologous to the right end of the inverted Tg-110 was cloned into the MluI site of p108CIBS.

All of the plasmids were linearized with the appropriate restriction enzymes (TaKaRa or Toyobo) and used for transformation. The all homologous arms (approximately 1 kb) and *c*I-*spc* cassette were amplified by PCR (PrimeSTAR HS DNA polymerase, TaKaRa, or KOD plus DNA polymerase, Toyobo) using the BAC clones and pCISP310B as templates, respectively. Primer sequences and PCR conditions are summarized in Table 2-1. The orientations of the inserts were determined by restriction enzyme digestion or PCR. The accuracy of the sequences generated by PCR was confirmed by DNA sequencing.

II-2-6) DNA sequencing

Sequences of the original BACs and modified BGM clones (Tg-110CIBS,

Tg-220CISP and Tg-250SB) were verified using next-generation sequencing. Briefly, genomic DNA (3 μg) from the above clones was fragmented to an average length of 300 bp using the Covaris S2 system (Covaris). After purification, end-repairing, A-tailing, paired-end adapter ligation and 12-cycle PCR were performed using the NEBNext DNA Library Prep Master Mix Set and NEBNext Multiplex Oligos for Illumina (New England Biolabs). All libraries were quantified using an Agilent Bioanalyzer 2100 (Agilent Technologies) and pooled to provide equal genome coverage from each library. Pooled libraries were sequenced in a single lane of the Illumina Genome Analyzer IIx (Illumina), which produced 102 paired-end reads, in accordance with the manufacturer's instructions.

Reads obtained from the Illumina Genome Analyzer IIx were analyzed using CLC Genomics Workbench 5.5 (CLC Bio). Reads were trimmed and mapped to each reference sequence with default parameters. The reference sequences used in this study included the following: BAC1, original BAC clone RP24-392H7 [GenBank: AC132096]; BAC2, original BAC clone RP23-61O11 [GenBank: AC102535]; B1TgCIBS and B2TgCISP, predicted modified mouse genomic Tg-110CIBS and Tg-220CISP insert sequences; and Tg250SB, a predicted fused mouse genomic Tg-250SB insert sequence. The mapping results are detailed in Table 2-2. After mapping the reads, variant calling was performed using the probabilistic variant detection function with default parameters, and variants with frequencies of at least 90% were considered. All read data have been deposited in the DDBJ Sequence Read Archive (DRA) under accession number [DRA000859].

Table 2-1. Primer sequences and PCR conditions used in Chapter II.

Primer name	Strand	Sequence	Annealing temp Extension time	Purpose	PCR condition
***MOR42-3* targeting**					
J530_FF	Forward	TTTGTCGACATGTCAGGGTGGAGCAATG	59.2 °C	homology arm (L-arm)	PrimeSTAR HS
J531_FR	Reverse	TTGAATTCAATCATCCAGGAAGCTCTC	1 min	to 5' side of the insertion site	
J532_RF	Forward	TACTAGTCCAGAATTTGTGGGTCTCA	55.5 °C	homology arm (R-arm)	PrimeSTAR HS
J533_RR	Reverse	TTTAGATCTAGAGTTGAGTGCAAGCTG	1 min	to 3' side of the insertion site	
J554_CISPF	Forward	GGCGCGCCAGCCGTCATCCTCGGCAC	55 °C	subcloning of *c*I-*spc* cassette	PrimeSTAR HS
J555_CISPR	Reverse	TACTAGTAACCAATTCCGACAGGCTTTG	2 min 30 sec	check (*c*I-*spc* cassette: F3/R3)	Genotyping
J476_MOR42-3	Forward	GCAATGGCACCTACAATGAGT	60 °C	check (*MOR42-3* coding: F1/R1)	Genotyping
J477_MOR42-3	Reverse	CTGCCTGATTTCCTTGGTTC	1 min	*MOR42-3* coding probe	
GFP-F	Forward	GGCATCAAGGTGAACTTCAAGATCC	62 °C	check (GFP: F2/R2)	Genotyping
GFP-R3	Reverse	CTTTACTTGTACAGCTCGTCCATGC	30 sec		
J554_R seq.	Forward	ACATGGTCCTGCTGGAGTTC	55 °C	check (GFP_42-3 R-arm: F4/R4)	Genotyping
J533_RR	Reverse	TTTAGATCTAGAGTTGAGTGCAAGCTG	3 min 30 sec		
Inversion of Tg-110					
J591_BAC1_LF	Forward	AATGCGGCCGCGATCTCAGCTCCTCAAACTC	54.5 °C	homology arm (L-arm)	KOD plus
J592_BAC1_LR	Reverse	AATGCGGCCGCCTGGATACATTCCTCTTCC	1 min	to left-end of Tg-110	
J593_BAC1_RF	Forward	CCCAAGCTTGGATTTCCAGCTTGCTTCC	59.2 °C	homology arm (R-arm)	KOD plus
J594_BAC1_RR	Reverse	CCCAAGCTTAGGCAGTGATGACCTTTGC	1 min	to right-end of Tg-110	
191F-teA2	Forward	GATAGCCCATTATATTCATTGGTCC	61 °C	check of *tet* gene	Genotyping
195F-etA	Reverse	CAAGTATCCCACCAATGTAGCCG	1 min 40 sec	*tet* probe	
191R-ermB	Forward	GTAAAATGGCCTTTTCCTGAGCCG	61 °C	check of *erm-e-phl*	Genotyping
195R-PhlB	Reverse	GGCAATGCCGGGATAGACTG	1 min 40 sec		
Fusion of Tg-220 and Inverted Tg-110					
J629_BAC2_LNotF	Forward	ATGCGGCCGCGAATTCCTGATCTTTCCAAGAC	58.4 °C	homology arm (L-arm)	KOD plus
J630_BAC2_LNotR	Reverse	ATGCGGCCGCAGACCACCATGGCCTAATG	1 min 10 sec	to left-end of Tg-220	
J631_BAC1Inv_RMluF	Forward	CCGACGCGTAGCACACCTGAAAGCTCTAG	55.5 °C	homology arm (R-arm)	KOD plus
J632_BAC1Inv_RMluR	Reverse	CCGACGCGTGATCTCAGCTCCTCAAACTC	1 min 10 sec	to right-end of Tg-220	
Poly linkers					
Linker 1	Sense	CTAGAGGCGCGCCACTAGTAGATCTCATATGGCATGCGAGCT		linker for pT1 vector	
Linker 2	Antisense	TCCGCGCGGTGATCATCTAGAGTATACCGTACGC		containing AscI-SpeI-BglII-NdeI-SphI	
HPNSB-1	Sense	AGCTTCACGTGCGGCCGCATGCG		linker for p108IPpoI-HPNSB pasmid	
HPNSB-2	Antisense	GATCCGCATGCGGCCGCACGTGA		containing PmlI-NotI-SphI	

PrimeSTAR HS (TaKaRa)	**KOD plus ver.2 (Toyobo)**	**Genotyping, Check of BGM clones**
94 °C 5 min	94 °C 2 min	94 °C 5 min
[30 cycles]	[30 cycles]	[30 cycles]
98 °C 10 sec	98 °C 10 sec	94 °C 30 sec
(Annealing temp.)* 5 sec	(Annealing temp.)* 30 sec	(Annealing temp.)* 30 sec
72 °C (Extension time)*	68 °C (Extension time)*	72 °C (Extension time)*
72 °C 5 min	68 °C 5 min	72 °C 5 min

Primers and oligonucleotides used in Chapter II for constructing homology arms, cassettes and new BAC plasmids, screening of BGM clones are summarized. PCRs with PrimeSTAR and KOD plus kits were basically performed according to the manufacturer's recommendations; all PCR conditions are summarized in bottom of Table. Genotyping PCR used a recombinant Taq DNA polymerase. The annealing temperature and extension time were described in Table.

II - 3 Results

II-3-1) Cloning of genomic DNA fragments into the BGM vector

Mouse genomic BAC libraries have been constructed that cover nearly the entire mouse genome, and each clone contains a genomic DNA fragment with an average size of approximately 150 kb (Osoegawa *et al*., 2000). The BGM vector system can utilize these valuable DNA resources in a one-step transfer (Kaneko *et al*., 2005). I prepared two BAC clones, designated BAC1 and BAC2, which overlapped each other via 82 kb region and carried 114 kb and 220 kb mouse genomic DNA fragments containing two and ten class I OR genes, respectively (Figure 2-2). I transferred these BAC inserts into BGM vectors, which harbor BAC vector sequences, to clone the BAC inserts by homologous recombination (Figure 2-1) (Kaneko *et al*., 2005; Kaneko *et al*., 2009). Briefly, the original BGM vector is resistant to spectinomycin (Spc) and sensitive to neomycin (Nm) due to repression of the Pr-*neo* cassette by the CI repressor. The BAC inserts taken up by the *B. subtilis* are integrated directly into the genome (BGM vector) via double crossings-over recombination with the BAC vector sequences of the BGM vector. Once a BAC insert is cloned into the BGM vector, the recombinants become resistant to Nm and sensitive to Spc, because the BAC insert replaces the *c*I-*spc* cassette. This selection mechanism can be used in *c*I cassette-mediating modifications, such as the targeted insertion technique. Because two 23-bp sequences "ATGACTCTCTTAA/GGTAGCCAAA" recognized by the rare-cutting endonuclease, I-PpoI are introduced at the both sides of the BAC cloning site, I-PpoI digestion enables to excise out the cloned insert from the BGM vector. The inserts were evaluated by digesting the genomic DNA of the recombinant vectors with I-PpoI followed by contour-clamped homologous electric field (CHEF) gel electrophoresis

(Figure 2-3). Three of seven candidate clones (resistant to Nm and sensitive to Spc) in BAC1 and one of 29 candidate clones in BAC2 contained BAC inserts. The efficiency of cloning the BAC insert into the BGM vector differs between BAC clones used (Table 2-3). I cloned 10 different BAC clones with 3 to 100% efficiency (average 40%). The cloned inserts were further confirmed by Southern blot analyses using the original BAC clones as probes; the results indicated that the digest patterns of the BGM recombinants were identical to those of the original BAC clones, except for the fragments derived from the ends of the insert (Figure 2-3). Thus, the BGM vector system imported the BAC insert by simple transformation, which enabled the use of already established valuable BAC resources.

II-3-2) Targeted insertion of a reporter gene

To construct transgenes for the analysis of *cis*-acting elements of class I OR genes, I inserted an *IRES-tauEGFP* reporter gene cassette (Bozza *et al*., 2002), which consisted of an internal ribosome entry site and the tauEGFP fusion protein coding sequence, into the class I OR gene *MOR42-3* locus by two-step recombination using the *c*I-*spc* cassette (Figure 2-4A). In the first step, the BGM clones containing the BAC inserts were transformed with a reporter cassette with the selection marker *c*I-*spc* flanked by 1.0 kb homology regions to the target OR locus (L and R arms) to generate Tg-110CISP and Tg-220CISP. Four of 15 Spc-resistant Tg-110CISP clones and six of 15 Spc-resistant Tg-220CISP clones were sensitive to Nm. In the second step, the counter-selection was performed to remove selection marker using a reporter gene cassette without a *c*I-*spc* cassette to generate two transgenes, Tg-110 and Tg-220. Nm-resistant and Spc-sensitive

clones were screened by PCR to amplify the region between *EGFP* and R arm (Figure 2-5). The *c*I-*spc* cassettes were correctly removed in four of 15 tested clones for Tg-110 and two of 48 clones for Tg-220. Negative clones were no PCR product, suggesting that insert DNA sequences had been also deleted along with the *c*I-*spc* cassettes. Relatively low success of the recombination events in the counter-selection was previously reported in the BGM system (Itaya *et al*., 2000) as well as the BAC system (Zhang *et al*., 1998). Accordingly, two transgenes, Tg-110 and Tg-220, which were derived from BAC1 and BAC2, respectively, were constructed by inserting *IRES-tauEGFP* 3 bp downstream of the *MOR42-3* stop codon. Southern blot analyses of representative clones with the *MOR42-3* coding or R arm probes indicated the correct insertion of the reporter cassette (Figure 2-4 B, C). Thus, I established a targeted insertion method using the BGM vector system with homologous recombination using several antibiotic resistance genes and the *c*I gene without leaving any trace of selection markers or site-specific recombination sites, such as loxP and FRT, in the target sequence (Figure 2-1B) (Zhang *et al*., 1998; Lee *et al*., 2001).

II-3-3) Inversion of the insert in the BGM vector

The Tg-110 and Tg-220 shared approximately 82 kb of overlapping sequence (Figure 2-2), which would enable the fusion of these two fragments to create a longer transgenic construct in the BGM vector system if the two clones were cloned in the same direction. This technique extends one insert by homologous recombination between the overlapping region of another insert and the common vector region (Kaneko *et al*., 2009). However, the Tg-110 and Tg-220 inserts were cloned in opposite directions in the BGM vector. In this case, the two clones could not be fused to extend the insert by the double

crossing-over recombination. Insert inversion techniques have not been reported in the BGM or other conventional systems. To fuse these two fragments, I first inverted the Tg-110 insert to orient it in the same direction as Tg-220 using two incomplete fragments of the tetracycline resistance gene (*tet*), *te* (5´ side) and *et* (3´ side), which shared an approximately 1.1 kb overlap region designated *e* (Kuroki *et al*., 2008). The erythromycin resistance (*erm*) and phleomycin resistance (*phl*) genes were added as selection markers to the *te* and *et* fragments, respectively. The two inversion cassettes, *te-erm* and *phl-et*, were sequentially inserted at the ends of the Tg-110 insert in the reverse direction (Figure 2-6A). Briefly, *phl-et* cassette was inserted to the right end of the Tg-110 insert by transforming Tg-110 with the linearized BAC plasmid containing the *phl-et* cassette and a 0.8 kb homology region. The *phl-et* cassette was inserted in all five Phl-resistant clones assessed by PCR. Subsequently, I inserted the *te-erm* to the left end of the insert using the linearized BAC plasmid containing the *te-erm* cassette with a 0.9 kb homology region, and obtained Tg-110 T/P (1 clone / 1 tested). Intrachromosomal homologous recombination between *te* and *et* via the overlapping region *e* (Tg-110 T/P) resulted in an inversion of the insert and the formation of a complete *tet* gene, enabling the selection of the inverted clone (Tg-110-Inverted) using Tet (Figure 2-6A). I obtained over 100 Tet-resistant colonies from 20 μl of Tg-110 T/P overnight culture. The formation of *tet* and *erm-e-phl* cassettes was confirmed by PCR in all four randomly picked clones from these colonies. As expected, Southern blot analysis using a *tet* probe revealed changes in size of the *tet* cassette fragments, indicating the successful inversion of the Tg-110 insert (Figure 2-6B). Because the direction of the inserts is not coordinated in the BAC library, the inversion technique for insert fragments is essential to fuse and reconstruct genomic structures using BAC

clones.

II-3-4) Fusion of two inserts to reconstruct genomic structures

I then fused Tg-220 and the inverted Tg-110 to enlarge the mouse genomic DNA fragment (Figure 2-7A). The *c*I-*spc* and *c*I-*bsr* marker gene cassettes were inserted into the left end of the Tg-220 insert and the right end of the inverted Tg-110 insert, respectively. I transformed Tg-220 with the linearized BAC plasmid containing the *c*I-*spc* and a 1.3 kb homology region, and obtained Tg-220CISP clones (9 clones from 10 Spc-resistant clones). Tg-110-Inverted was similarly transformed with the *c*I-*bsr* cassette fragment containing a 1.3 kb homology region, and obtained Tg-110CIBS clones (2 clones from 2 Bs-resistant clones). The two transgenes were fused by genetic transformation of Tg-220CISP with the purified genomic DNA of Tg-110CIBS. Homologous recombination of the 82 kb overlapping region and the common sequence of the BGM vectors resulted in the extension of the insert to generate Tg-250SB. Twenty-six of 29 Spc- and Bs-resistant colony recombinants were sensitive to Nm. I selected two representative clones for further examinations. The two Tg-250SB clones turned to be sensitive to Tet, whose resistance gene exists in the left end of the Tg-110-Inverted insert and was removed by the fusion, suggesting correct recombination occurs to extend the insert. I-PpoI/CHEF analyses of the Tg-250SB genomic DNA demonstrated a larger band corresponding to the 252 kb fused insert fragment (Figure 2-7B). Southern blot analysis showed that Tg-250SB was composed of two transgenes, Tg-220 and Tg-110 (Figure 2-7C), indicating that the two inserts were successfully fused to extend the insert size.

II-3-5) DNA sequencing of modified and reconstructed BGM clones

To verify the reliability of the BGM vector system in manipulating genomic DNA fragments and its accuracy in genetic engineering, I performed DNA sequencing of the original BAC clones (BAC1 and BAC2), the modified BGM clones (Tg-110CIBS and Tg-220CISP) and the fused BGM clone (Tg-250SB). The read mapping results are summarized in Table 2-2. No mutation was observed in the cloned and modified mouse genomic inserts during manipulations of the BGM vector, and all targeted modifications were accurately introduced. A few nucleotides differed from the reference sequence data in both BAC1 and BAC2. Two single-nucleotide differences were found between BAC1 and the reference sequence [GenBank: AC132096.4]: the replacement of G at positions 10349 and 113530 by A and T, respectively. Four differences were found between BAC2 and the reference data [GenBank: AC102535.17]: G at 210780, G at 210788, A at 210815 and G at 210840 were replaced by A, A, G and A, respectively. Therefore, starting from targeted insertion, all possible genetic modifications, i.e., insertion, deletion, inversion and fusion to the targeted site, could now be introduced using the BGM system without leaving unnecessary sequences in the insert DNA.

II - 4 Discussion

In this chapter, I developed a complete genetic manipulation method for the BGM vector system by demonstration of a target insertion and an inversion modification. The DNA sequencing analysis of the constructed transgenes revealed all targeted modification were precisely conducted. Because the reliable inversion modification was not demonstrated in any other system including BAC or YAC, the BGM vector system has many advantages in not only the property of a vector system but also the availability of modification approaches.

A concern is the efficiencies of each modification. Overall, the efficiency of the insertion of a fragment with a selection maker was high (~100%), whereas that of the removal of a selection marker was low in BAC cloning or targeted insertion modification. The major cause of this phenomenon is considered to be mutations to *c*I gene, which increase the number of Nm-resistant false-positive colonies. These clones are resistant to both Nm and Spc. I also observed another false-positive colonies, which clones are resistant to both Nm and Spc. This is likely to be caused by an accidentally deletion of genomic region including a selection marker cassette *c*I-*spc*. This unwanted recombination may result from an unexpected recombination between neighbor regions of the selection marker in *B. subtilis* genome or mouse genomic sequences (Bird *et al*., 2011). Because the current BGM vector uses the *proB* locus for cloning, the change of the cloning site to alternative locus, such as *leuB* may increase cloning efficiencies of counter-selection steps. But it is note that the efficiency of counter-selection using the BGM vector in this study is comparable in the BAC system (Zhang *et al*., 1998; Bird *et al*., 2011).

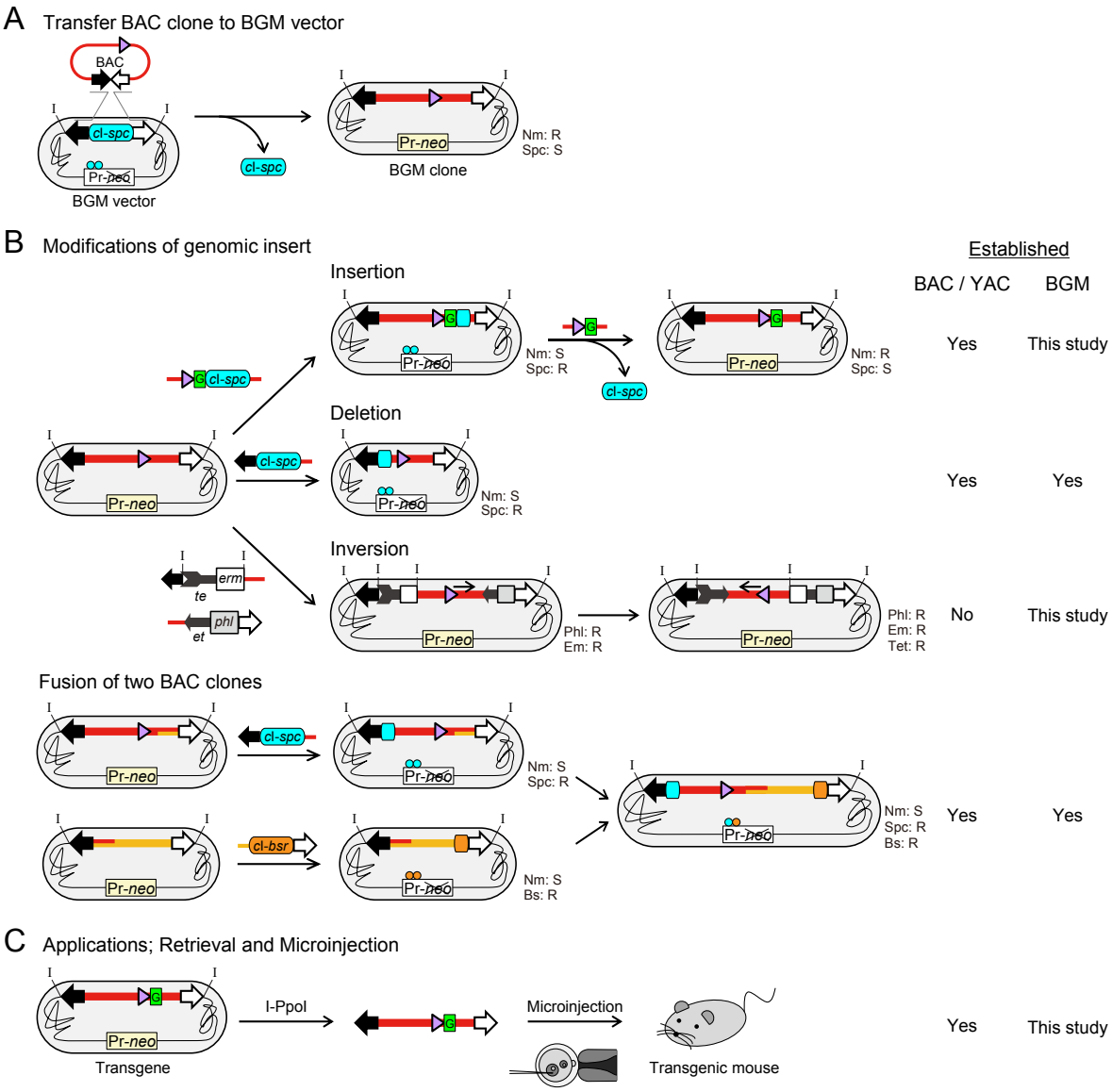

Figure 2-1. Cloning and various modification methods of the BGM vector system.

(A) The cloning strategy used to introduce BAC inserts into the BGM vector. BGM vectors possess two antibiotic gene cassettes: Pr-*neo*, a lambda Pr promoter fused to the neomycin resistance gene (*neo*), and c*I*-*spc*, which contains *c*I encoding the CI repressor protein, which binds to the Pr promoter, fused to the spectinomycin resistance gene (*spc*). The c*I*-*spc* is flanked by two BAC vector sequences that function as the cloning site for BAC inserts. The original BGM strains are resistant to Spc and sensitive to Nm. Once a BAC insert is cloned into the BGM vector, the recombinant becomes resistant to Nm and sensitive to Spc because the c*I*-*spc* cassette is replaced by the BAC insert. **(B)** Modifications of large DNA fragments. The BGM vector system allows various manipulations, including insertion (e.g., EGFP is designated as "G" in the box), deletion and the inversion and fusion of two overlapping clones. **(C)** The BGM inserts can be retrieved by several methods, and the recombinant DNA fragments can be used for the generation of transgenic animals by microinjection (described in Chapter III). The right column represents the summary of achievements for each modification using the BAC, YAC and BGM vector systems. The closed triangles represent a target gene and its direction. The BAC vector sequences are indicated by the closed and open arrows. The half part of BAC vector containing chloramphenicol resistance gene for *E. coli* is depicted by open arrow (defined as right side). The I-PpoI sites, designated as "I" are introduced to flank the BAC vector sequences.

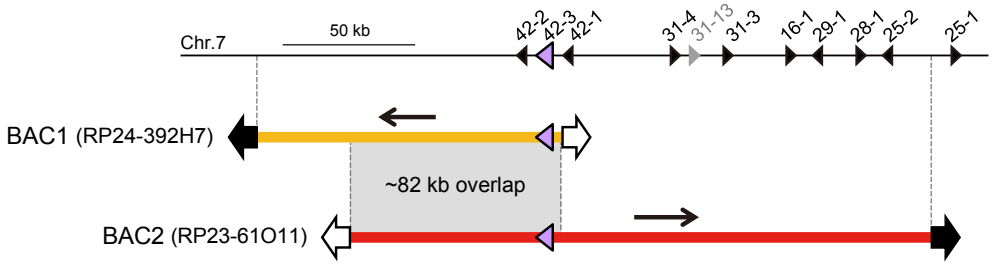

Figure 2-2. Genomic structures of mouse chromosome 7 and the BAC clones used in this study.

The BAC clones cover a portion of the class I OR locus on chromosome 7, share 82 kb overlapping region and are oriented in opposite directions in the BAC vector. A large triangle depicts *MOR42-3*, black triangles indicate other OR genes, and the gray triangle represents the pseudo-OR gene (Zhang *et al.*, 2007).

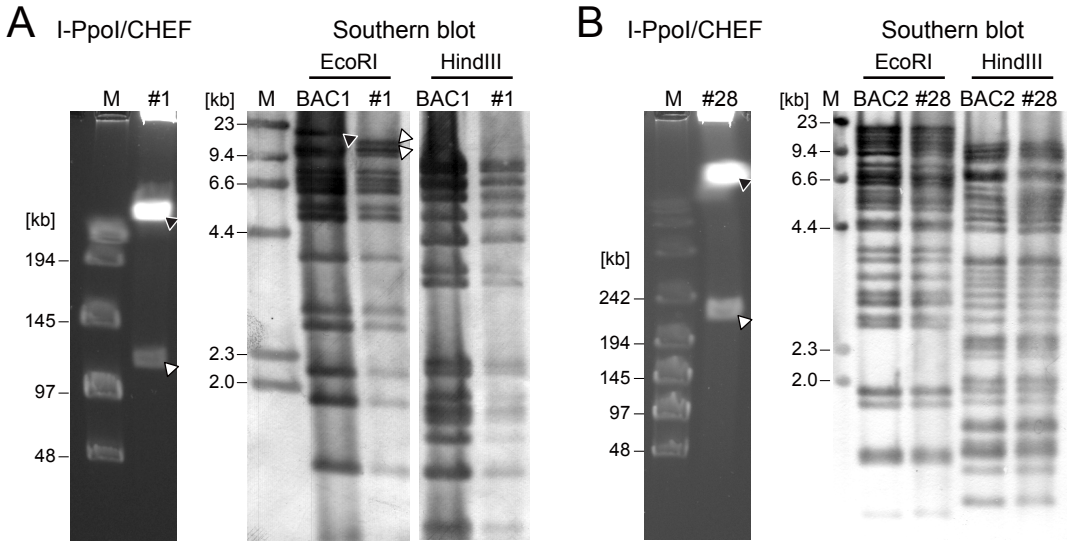

Figure 2-3. Cloning of BAC inserts into the BGM vector.

The cloning of BAC1 and BAC2 into the BGM vector was confirmed by I-PpoI/CHEF (left panels in **A** and **B**) and Southern blotting (right panels in **A** and **B**). The representative BGM vectors with transferred BAC1 and BAC2 inserts were digested with I-PpoI and resolved by CHEF. The open arrowheads indicate BAC insert bands, and the closed arrowheads indicate the BGM vector. Purified original BAC clones and genomic DNA from representative BGM clones were digested with EcoRI or HindIII and hybridized with the original BAC clones as probes. The BGM recombinants showed a band pattern identical to the original BAC clones (lane: BAC), except for the bands corresponding to the BAC end sequences from EcoRI-digested BAC1; the closed arrowhead indicates a BAC-specific signal, and the open arrowheads indicate BGM-specific signals from EcoRI-digested BAC1. In other digestion patterns, the bands of the BAC end sequences were overlapping and indistinguishable from insert signals predicted from restriction maps of BAC clones and the BGM vector. Numbers above lanes are the BGM clone numbers. In lane M, lambda/HindIII fragments or a lambda DNA concatemer was used as a size marker.

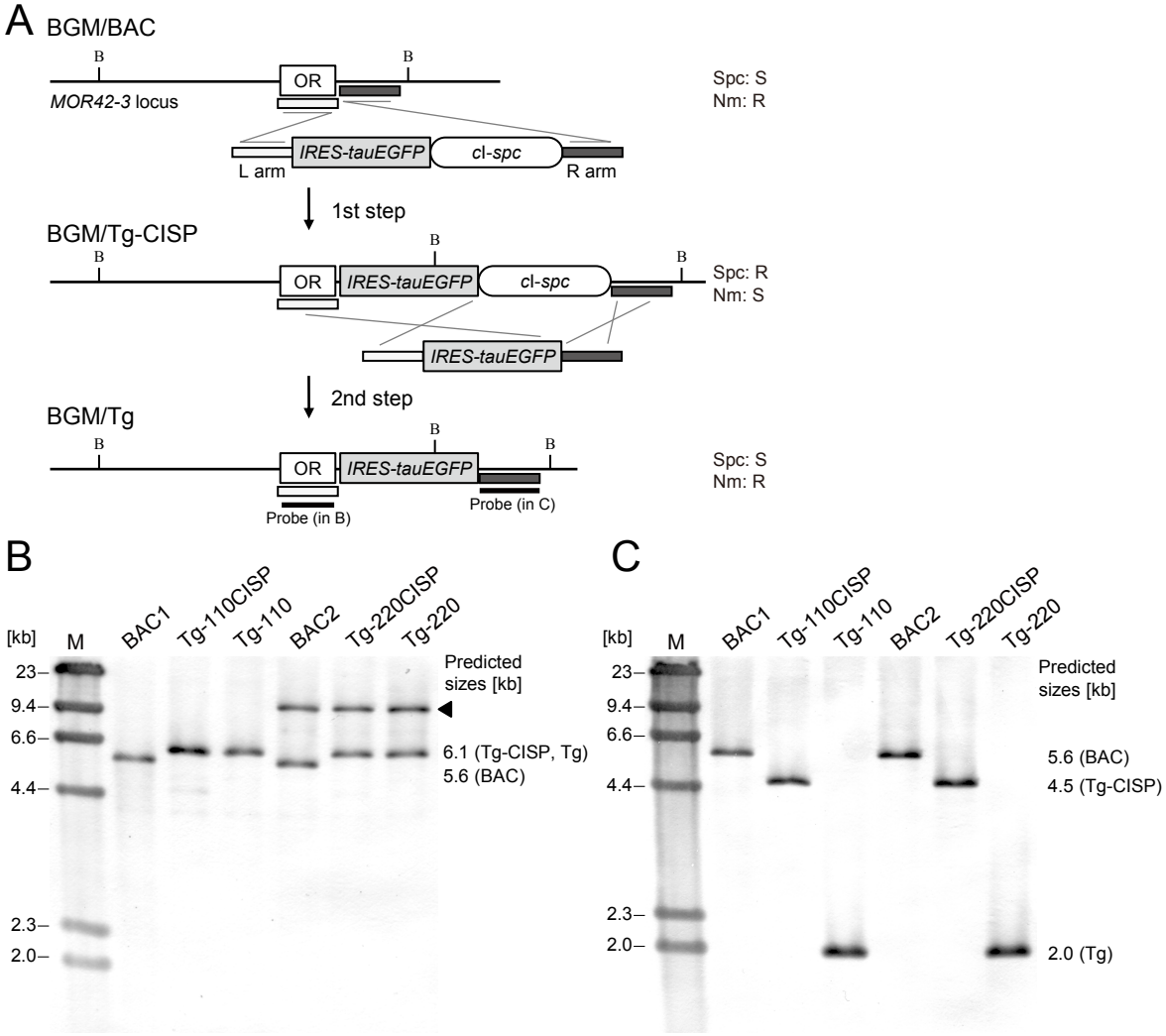

Figure 2-4. Targeted insertion of a reporter gene.

(A) In the first step, *IRES-tauEGFP* was inserted into the targeted site 3 bp downstream of the *MOR42-3* stop codon by transformation with a purified DNA fragment consisting of L and R arms (1.0 kb each) homologous to *MOR42-3* and the *c*I-*spc* selection cassette. In the second step, the selection cassette was removed by transformation with a fragment lacking the selection marker. **(B, C)** The targeted insertion of the reporter gene was confirmed by Southern blot analysis using a *MOR42-3* coding region or R homology region as a probe. The genomic DNA of each BGM clone was digested with BamHI and resolved by CHEF, and the changes in the sizes of the BamHI fragments were as expected, indicating the successful targeted insertion of *IRES-tauEGFP*. The sensitivity to antibiotics is shown. A bold arrowhead in panel B represents bands corresponding to the cross-hybridized *MOR42-1* coding region, which sequence is contained in BAC2 but not in BAC1. R, resistant; S, sensitive. The BamHI sites are represented by "B". Lane M, lambda/HindIII fragments were used as a size marker.

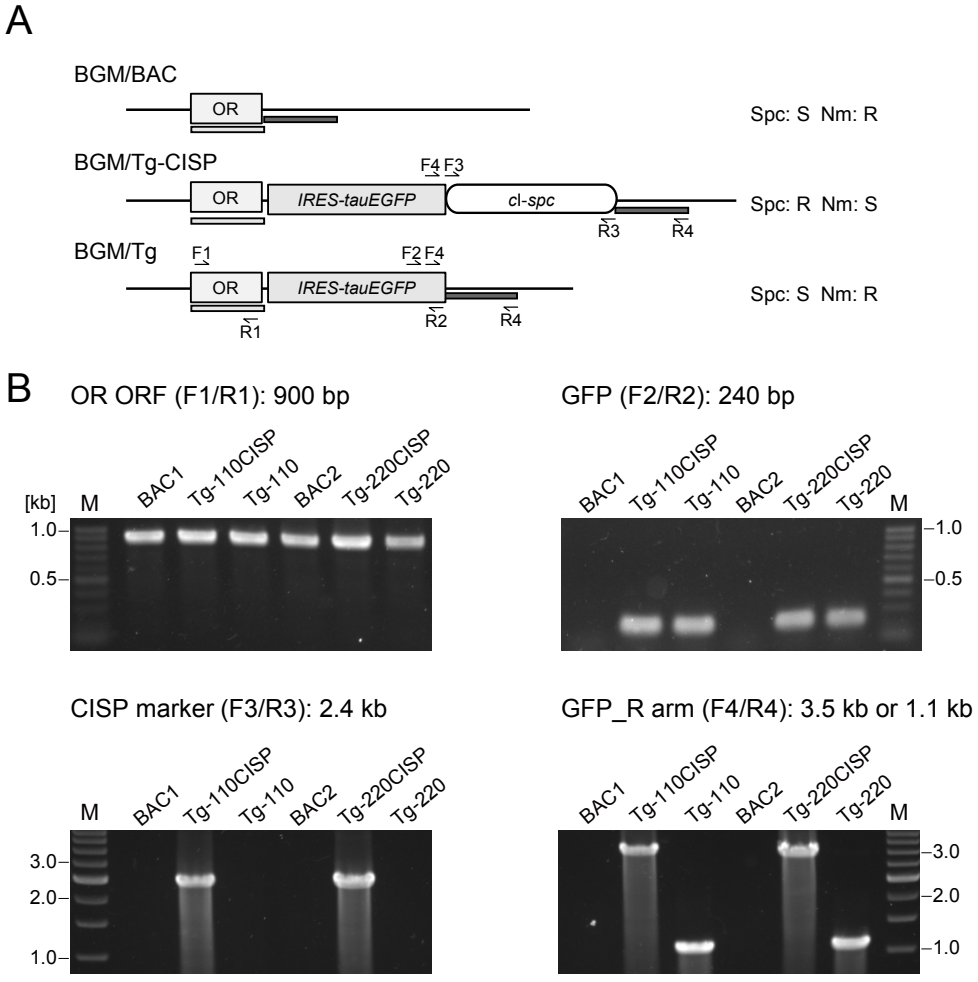

Figure 2-5. Verification of each targeted insertion step by PCR.

(A) Primer binding positions to verify recombinant BGM clones are as indicated. Primer sets were used as follows; F1/R1 for *MOR42-3* coding sequence, F2/R2 for the 3´ side of *EGFP*, F3/R3 for the *c*I-*spc* cassette and F4/R4 for the region between inside *EGFP* to the 3´ end of the R. **(B)** PCR analyses of each insertion step confirm the insertion and removal of the reporter and marker genes in the Tg-110 and Tg-220 constructs. Representative results of BGM/BAC clones (BAC1 and BAC2), BGM/TgCISP clones (Tg-110CISP and Tg-220CISP) and BGM/Tg clones (Tg-110 and Tg-220) are shown, and verifying the correct targeted insertion of the reporter cassette. A F4/R4 primer set was used to screen of Nm-resistant and Spc-sensitive clones in removal step of a *c*I-*spc* cassette.

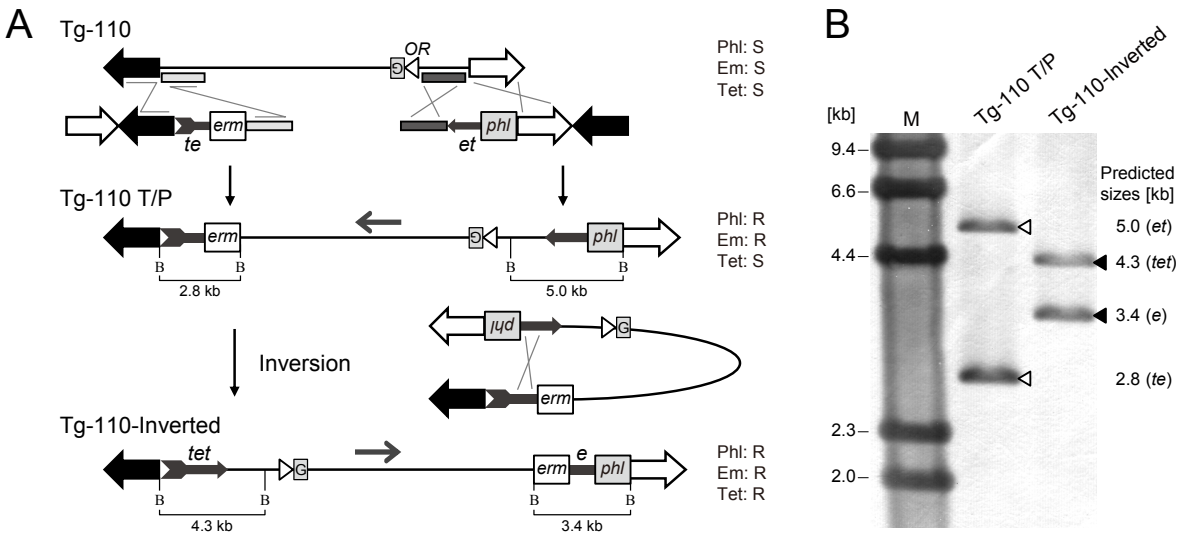

Figure 2-6. Inversion of the Tg-110 insert.

(A) Schematic diagram of inversion strategy. Two incomplete fragments of the tetracycline resistance gene (*tet*), *te* (5' side) and *et* (3' side), which contained an overlapping region *e*, were used for inversion. The inversion cassettes, *phl-et*, consisting of *et* and the phleomycin resistance gene (*phl*), and *te-erm*, consisting of *te* and the erythromycin resistance gene (*erm*), were sequentially inserted into the ends of the Tg-110 insert. A 0.9 kb fragment in the left end and a 0.8 kb fragment in the right end were used as homology arms. Homologous recombination between *te* and *et* in Tg-110 T/P results in the inversion of the insert to form an intact *tet* gene. The about positions of BamHI sites are represented by "B". **(B)** Inversion was confirmed by Southern blotting using the *tet* gene as a probe. The genomic DNA of the represented clones was digested with BamHI. Changes in the size of *tet* cassette fragments represented inversion by intrachromosomal homologous recombination. The open (2.8 kb *te-erm* cassette and 2.3 kb *phl-et* plus 2.7 kb Tg-110 fragment) and closed (3.4 kb *erm-e-phl* cassette and 1.6 kb *tet* plus 2.7 kb Tg-110 fragment) arrowheads indicate signals before and after inversion, respectively.

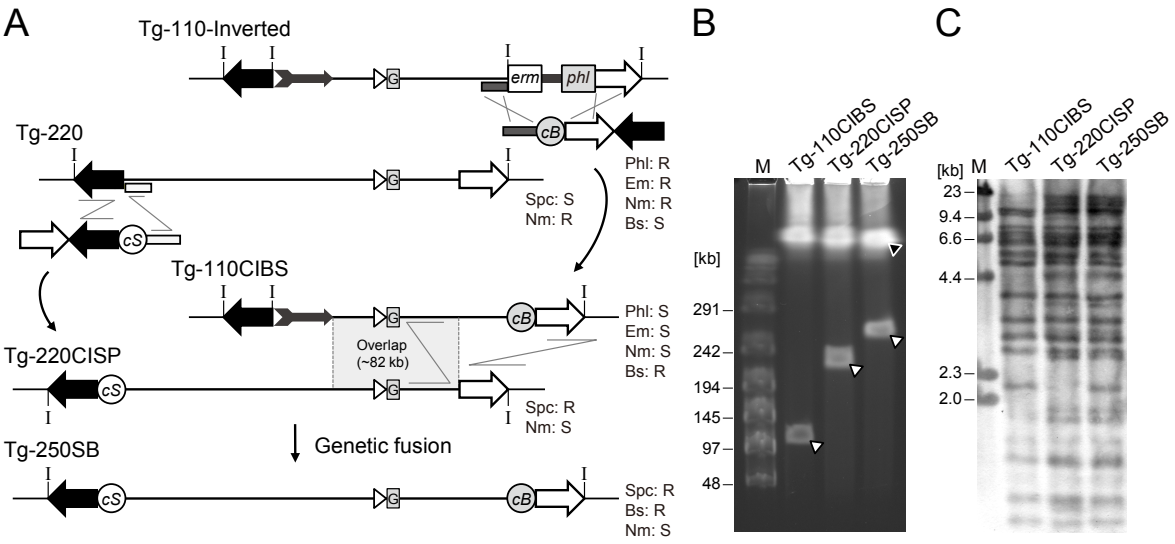

Figure 2-7. The genetic fusion of the inverted Tg-110 and Tg-220 fragments to reconstruct the genomic structure of the *MOR42-3* locus.

(A) Schematic diagram of fusion of Tg-110-Inverted and Tg-220 fragments. The selection marker cassettes *c*I-*bsr*, consisting of the *c*I and blasticidin S resistance genes (depicted as *c*B), and *c*I-*spc* (depicted as *c*S) were inserted to the right end of the inverted Tg-110 and to the left end of Tg-220, respectively. 1.3 kb fragments in the left end of Tg-220 and in the right end of Tg-110-Inverted were used as homology arms. Genomic DNA was isolated from the *bsr*-labeled Tg-110 clone (Tg-110CIBS) and added to the *spc*-labeled Tg-220 competent cells (Tg-220CISP). The Tg-220 insert was extended to 252 kb by homologous recombination at the overlap region. **(B)** The fused Tg-250SB insert was evaluated by I-PpoI digestion followed by CHEF. Tg-250SB shows a larger band that corresponds to the fused insert fragment. The closed and open arrowheads represent bands from the 4.2 Mb BGM vector and inserts, respectively. **(C)** The structure of the Tg-250SB insert was confirmed by Southern blot analysis. Genomic DNA was digested with EcoRI and hybridized with Tg-110 and Tg-220 probes. The genetic fusion of the two inserts is shown as the sum of the respective Tg-250SB bands. The I-PpoI sites are represented by "I". Lane M, lambda/HindIII fragments or a concatemer of lambda DNA was used as a size marker.

Table 2-2. Summary of the read mapping in DNA sequencing.

	BAC1		BAC2		Tg110CIBS		Tg220CISP		Tg250SB	
	# reads	%	# reads	%	# reads	%	# reads	%	# reads	%
Total reference length	114,183		220,003		115,372		222,532		254,658	
Mapped reads	8,341,236	91.1	8,604,424	95.9	1,028,479	2.7	1,737,127	5.2	504,510	5.8
Unmapped reads	814,824	8.9	370,790	4.1	37,166,207	97.3	31,564,383	94.8	8,227,736	94.2
Reads in pairs	8,327,110	91.0	8,594,722	95.8	960,578	2.5	1,714,062	5.2	498,152	5.7
Broken paired reads	14,126	0.2	9,702	0.1	67,901	0.2	23,065	0.1	6,358	0.1
Total number of reads	9,156,060		8,975,214		38,194,686		33,301,510		8,732,246	
Total read length (Mb)	915.6		897.5		3,819.5		3,330.2		873.2	
Fraction of reference covered	1.00		1.00		1.00		1.00		1.00	
Minimum coverage	241		125		28		22		6	
Maximum coverage	15147		5631		2211		1514		574	
Average coverage	7100.7		3812.3		851.5		752.2		191.0	

Reads obtained from the Illumina Genome Analyzer IIx were analyzed and mapped to each reference sequence using CLC Genomics Workbench 5.5 (CLC Bio). The number of mapped reads in each sample was enough to evaluate mutations. Analyzed DNA samples are follows: original BAC clones of BAC1 and BAC2 and BGM clones of Tg110CIBS, Tg220CISP and Tg250SB.

Table 2-3. Cloning efficiency of BAC inserts into the BGM vector.

BAC #	clone ID	insert (kb)	BGM vector	total tested clones (Nm-resistant, Spc-sensitive)	insert positive	cloning efficiency (%)
BAC1	RP24-392H7	114	BEST310	7	3	43%
BAC2	RP23-61O11	220	BEST6528	29	1	3%
BAC3	RP24-306C18	149	BEST6528	3	3	100%
BAC5	RP24-282D19	165	BEST6528	15	6	40%
BAC7	RP23-310P1	165	BEST6528	32	3	9%
BAC8	RP23-480A17	177	BEST6528	10	3	30%
BAC9	RP24-211E6	159	BEST6528	4	4	100%
BAC10	RP24-419H14	141	BEST310	10	2	20%
BAC11	RP23-347I6	202	BEST6528	25	7	28%
BAC14	RP24-266J11	163	BEST6528	24	6	25%
					Average	40%

Efficiency of cloning of BAC clones into the BGM vector are summarized; clone names and ID, length of BAC inserts, BGM vector used for cloning and the numbers of tested and insert-positive clones.

CHAPTER III

THE BGM VECTOR-BASED MOUSE TRANSGENESIS

III - 1 Introduction

The BAC and YAC systems have been recognized as powerful tools for functional analysis of genomic DNA fragments in combination with transgenesis, and they have contributed greatly to life science researches, including functional analyses of genomic DNA, production of recombinant proteins and studies on genetic diseases (Schedl *et al.,* 1993; Mendez *et al.,* 1997; Yang *et al.,* 1997; Giraldo and Montoliu, 2001).

The BGM vector is now available for cloning and manipulation of various DNA fragments including cyanobacteria genome, mouse mitochondrial genome and rice chloroplast genome DNA (Itaya *et al*., 2005; Itaya *et al*., 2008). Although some retrieve methods for BGM inserts were previously proposed (Itaya *et al*., 2008; Kaneko *et al*., 2009), the BGM system has not been applied to transgenesis in any species. Because of many advantages such as the Mb cloning capacity and amenability of modifications, the establishment of the BGM vector-based transgenesis (BGM transgenesis) must provide a new approach in various research fields. In this study, I established the method to generate transgenic mice using the BGM vector system.

III - 2 Experimental procedures

III-2-1) Preparation of the transgenes for pronuclear injection

Genomic DNA carrying the transgene in an agarose plug was prepared, digested with I-PpoI and resolved by CHEF in a 1% (w/v) agarose gel using sterile 0.5× TBE, as performed for the I-PpoI analysis. To concentrate the transgene fragment, the band was excised from the agarose gel, turned vertically and embedded in 4% (w/v) agarose; electrophoresis was then performed at 3.3 V/cm for more than 10 hours in 0.5× TBE. The excised concentrated transgene band was placed in a prepared dialysis tube hydrated with 0.5× TBE, and the transgene was electroeluted from the gel under the same conditions. The eluate was dialyzed with a high salt injection buffer (10 mM Tris-HCl (pH 8.0), 0.1 mM EDTA, 100 mM NaCl) at 4-6°C overnight. The concentration and integrity were estimated based on the control band intensity in the CHEF analysis. The purified transgenes were used immediately. Alternatively, 75 μM spermidine and 30 μM spermine were added for long-term storage. The stock transgenes were diluted to 0.3-1.5 ng/μl with injection buffer before use.

III-2-2) Production and analyses of transgenic mice

The purified transgenes were microinjected into the pronucleus of B6C3F1 (C3H/HeSlc male × C57BL/6NCrSlc female) mouse zygotes in a standard protocol (Sumiyama *et al*., 2010). Olympus IX-71 microscope equipped with the micromanipulator transgenic system (Narishige) and Eppendorf FemtoJet system were used for microinjection. Injected eggs were transferred to the oviducts of pseudopregnant female ICRs. The founders were screened by PCR with the following three primer sets that

specifically amplified the internal and left and right ends of the transgenes: EGFP: GFP-F and GFP-R; the left end of the BAC sequence: SacB-F and SacB-R, or BAC108L-F and BAC108L-R; and for the right end of the BAC sequence: CmR-F and CmR-R (Table 3-1). All founder mice positive for these three primer sets were selected as candidates containing intact transgenes. The founder mice were crossed with C57BL/6 mice. The transgenic mice were dissected and fixed in 4% paraformaldehyde (PFA)/PBS for 10-15 min on ice. Line #8 of two EGFP-expressing Tg-250SB transgenic lines was used for the expression analyses in Figure 3-5.

III-2-3) Southern blot analysis of transgenic mice

10 μg of genomic DNA extracted from a tail was digested with EcoRI (TaKaRa) and separated on a 1% agarose gel. The DNA was transferred onto a membrane filter, and hybridized with a DIG-labeled *MOR42-3* coding probe (nucleotides 190 – 1091 from NM_020289). The signals were detected by chemiluminescence (CSPD, Roche) using a CCD camera (ChemiDoc™ XRS, Bio-Rad). The copy numbers were estimated by comparing the intensities of transgenic and endogenous signals.

III-2-4) Immunohistochemistry (IHC)

Mice were fixed by perfusion with 4% PFA/PBS and infiltration in the same fixative solution at 4°C for 30 min. After cryoprotection with 15% and 30% sucrose/PBS, tissues were embedded in Frozen Section Compound (Surgipath FSC22, Leica microsystems). Serial cryosections (20 μm) were collected on MAS-coated microscope glass slides (Matsunami) and dried for 1 h at room temperature. IHC was performed by a standard

protocol (Enomoto *et al*., 2011). After post-fixation, permeabilization and antigen-retrieval pretreatment, sections were blocked with 10% normal horse serum and incubated with primary antibody at 4°C overnight. The following primary antibodies and dilutions were used: rat anti-GFP (1:2000, catalog #04404-84, Nacalai) and goat anti-OMP (1:5000, catalog #544-10001, Wako). After incubation of the primary antibodies, sections were washed in PBS containing 0.01% Tween 20 and stained by the following Alexa Fluor-conjugated secondary antibodies (1:500, Invitrogen): Alexa Fluor 488-conjugated donkey anti-rat IgG and Alexa Fluor 546-conjugated donkey anti-goat IgG. Nuclear staining was performed with DAPI (1:1000, Vector Laboratories), and CC/Mount (Diagnostic BioSystems) was used for mounting.

Fluorescent images of endogenous GFP and IHC signals were taken with Olympus SZX10 fluorescent stereomicroscope with a DP71 digital CCD camera and Leica SPE confocal microscope. Confocal images were collected as *z*-stacks and projected into a single image for display. Images were analyzed and adjusted using Photoshop CS4 (Adobe). All of the mouse studies were approved by the Institutional Animal Experiment Committee of the Tokyo Institute of Technology and performed in accordance with institutional and governmental guidelines.

Table 3-1. Primer sequences and PCR conditions used in Chapter III

Primer name	Strand	Sequence	Annealing temp Extension time	Purpose	PCR condition
GFP-F	Forward	GGCATCAAGGTGAACTTCAAGATCC	62 °C	genotype	Genotyping
GFP-R3	Reverse	CTTTACTTGTACAGCTCGTCCATGC	30 sec	EGFP (intermediate of transgenes)	
J619_SacB_F	Forward	GCTGAATACAACGGCTATCACG	62 °C	genotype	Genotyping
J620_SacB_R	Reverse	TCTCTCAGCGTATGGTTGTCG	30 sec	Left end of Tg-110	
J621_CmR_F	Forward	GAGGCATTTCAGTCAGTTGCTC	62 °C	genotype	Genotyping
J622_CmR_R	Reverse	CGGCATGATGAACCTGAATCG	30 sec	Right end of Tg-250SB and Tg-110	
J709_BAC108L_F	Forward	CGTATTCAGTGTCGCTGATTTG	60 °C	genotype	Genotyping
J710_BAC108L_R	Reverse	TTAGCGATGAGCTCGGACTTC	30 sec	Left end of Tg-250SB	

Genotyping
94 °C 5 min

[30 cycles]
94 °C 30 sec
(Annealing temp.)* 30 sec
72 °C (Extension time)*

72 °C 5 min

Primers and PCR conditions used for genotyping Tg-110 and Tg-250SB transgenic mice are summarized. PCR with a recombinant Taq DNA polymerase was performed according to the condition described under the Table.

III - 3 Results

III-3-1) Generation of the BGM transgenic mice

In contrast to BAC and YAC systems, the BGM vector system utilizes the entire endogenous genome as a vector. Therefore, the transgene can be obtained by digesting the genome with I-PpoI, whose recognition sites are introduced to flank the BAC vector sequence (in Chapter II, Figure 2-1). But its concentration is expected to be low because of the single copy of the transgene per genome/cell. Thus, I prepared linearized transgenes by excision from the BGM vector using I-PpoI digestion and CHEF and concentrated the transgenes by conventional electrophoresis followed by electroelution and dialysis with a high-salt injection buffer (Figure 3-1). Using the method established in this study, linearized microinjection-grade transgenes were readily prepared from the BGM clones. The purified transgenes were microinjected into fertilized mouse eggs to obtain founder transgenic mice. Newborn mice were genotyped by PCR to amplify transgene-specific sequences. I designed three primer sets to amplify BAC vector sequences (left and right end of the transgene) and *EGFP* (middle of the transgene) to assess the integration of the transgene into chromosome (Figure 3-2). Assuming that all three PCR positive mouse would carry an intact transgene, I obtained six Tg-110 founders from 45 newborn mice. Partial integrations of the transgene were also observed in 2 mice, which were PCR positive for the left end of the transgene only. For Tg-250SB, four of 26 pups carried an intact transgene, and one lacked the left end sequence (Table 3-2). These transgenic ratios are comparable to a conventional BAC transgenesis (Yang *et al.,* 1997). Germline transmission of the transgenes was obtained in 5 transgenic founders for Tg-110 (from 6 founders) and two for Tg-250SB (from 4 founders).

III-3-2) Experimental evidence of a *cis*-acting element for a class I OR gene

An individual olfactory sensory neuron (OSN) expresses only one OR gene in a monoallelic and mutually exclusive manner (Chess *et al*., 1994). Class I ORs are expressed almost exclusively by OSNs in the dorsal main olfactory epithelium (MOE), and these OSNs project their axons to a specific subset of glomeruli in the dorsal domain of the olfactory bulb (OB) (Tsuboi *et al*., 2006; Bozza *et al*., 2009). The transgenes of the *MOR42-3* locus are designed to activate the bicistronic expression of *MOR42-3* and *tauEGFP*, a fusion of the microtubule-associated protein tau with EGFP, to visualize the *MOR42-3* transgene expression and OSN axonal projections. Thus, if the transgenes contain a *MOR42-3 cis*-acting element, *MOR42-3* expression can be monitored by EGFP fluorescence. For Tg-110, I analyzed 5 transgenic lines and 1 founder. None of the Tg-110 transgenic mice displayed labeled OSNs. For Tg-250SB, two transgenic lines established from two founders were EGFP positive (Figure 3-3), indicating expression of *MOR42-3* transgene. Remaining two founders were EGFP negative. Whole-mount images of the MOE and the OB of Tg-250SB showed a punctate EGFP expression pattern within the dorsal region of the MOE and axonal projections into the dorsomedial and dorsolateral glomerulus of each OB (Figure 3-3B and 3-5A). To assess the copy number and integrity of the transgene, I analyzed Tg-250SB by Southern blot using *MOR42-3* coding region as a probe (Figure 3-4), which cross-hybridizes to highly homologous gene *MOR42-1* (97% identical). Southern blot analysis showed that the specific band corresponding to the transgene was observed in expected size and the bands corresponding to endogenous *MOR42-3* were the same between control and Tg-250SB, indicating that the intact transgenes were integrated into chromosome, and this is not due to a gene targeting event.

Ratio of signal intensities (transgene / endogenous gene) were 1.05 for line #5 (n=1) and 0.46 ± 0.02 for line #8 (n = 4), indicating the copy number of transgene of two and one for line#5 and line#8, respectively. Because the Tg-250SB transgene carries *MOR42-1*, intensities of the band corresponding to *MOR42-1* increased 1.61 ± 0.18 fold in Tg-250SB line #8 with reference to the control, confirming one copy of the transgene was integrated.

OSNs expressing highly homologous ORs tend to project their axons to near but distinct subsets of glomeruli in the OB (Feinstein and Mombaerts, 2004). Because MOR42-3 and MOR42-1, members of the MOR42 subfamily, share 97% amino acid similarity in their coding sequences, the axonal projection site of *MOR42-3*-expressing OSNs was expected to be close to that of *MOR42-1*. Using axonal projection of OSNs expressing endogenous *MOR42-1* as control, we examined the MOR42-1-iTG gene-targeted mouse (Bozza *et al*., 2009) in which an *IRES-tauEGFP* reporter was inserted downstream of *MOR42-1*. The glomerular positions of transgenic-MOR42-3 (Tg-MOR42-3) in Tg-250SB transgenic mice were similar to those of endogenous MOR42-1 (Figure 3-3B, 3-5A). I further analyzed the detailed structure of the Tg-MOR42-3 glomeruli. Glomeruli are innervated exclusively by axons from OSNs expressing the same OR (Treloar *et al*., 2002). I performed double-label immunohistochemistry on coronal cryosections with antibodies against olfactory marker protein (OMP) to stain all mature OSN and antibodies against EGFP to label *MOR42-3* transgene- or *MOR42-1* endogenous gene-expressing OSNs. Tg-MOR42-3 glomeruli contained a mixture of EGFP-negative and EGFP-positive axons, whereas MOR42-1 glomeruli of homozygous MOR42-1-iTG mice were completely EGFP positive (Figure 3-5B), suggesting that axons of OSNs expressing Tg-*MOR42-3* formed glomeruli together

with those of OSNs expressing endogenous *MOR42-3* (Vassalli *et al*., 2002). The projection site and converged glomerular structure of OSNs expressing the EGFP reporter suggest proper expression of the *MOR42-3* transgene. Taken together, these results indicate that a *cis*-acting element for *MOR42-3* is present in the extended region from Tg-110 and provide the first experimental evidence of a *cis*-acting element for a class I OR gene.

III - 4 Discussion

In this chapter, I established the BGM transgenesis using the transgenes, Tg-110 and Tg-250SB constructed in Chapter II. This is the first time that modified BGM inserts were transferred into a mammalian genome. The purified DNA from BGM vector can be introduced into not only mouse but also other species. Using the BGM transgenic mice, I provided the first evidence of a *cis*-element for a class I OR gene.

The transgenic efficiency (the ratio of the number of transgene-positive pups in total number of pups) was comparable previous BAC transgenesis methods (Yang *et al*., 1997). The BGM vectors used in this study are derived from the restriction modification-deficient strain RM125 (Uozumi *et al*., 1977; Kaneko *et al*., 2005). Thus, it is conceivable that the difference of host cells between *B. subtilis* and *E. coli* does not affect the transgenic efficiency.

In the YAC system, a similar pronuclear microinjection method succeeded in a generation of ~460 kb YAC transgenic mouse (Serizawa *et al*., 2000). Thus, it is possible to speculate that my protocols can be applied to several hundreds kb transgene. However, as a long DNA molecule is easy to shear, the upper limit of size for microinjection is expected to be approximately 500-800 kb (Peterson, 2007; Li and Blankenstein, 2013). In addition, the purification of an intact long DNA fragment is extremely difficult. Thus generation of transgenic animals carrying ~1 Mb transgene must require different approaches for DNA introduction such as lipofection of purified transgene into embryonic stem cells following chimera mouse production (Lamb *et al*., 1999) or sperm-mediated transfer methods (Garrels *et al*., 2012).

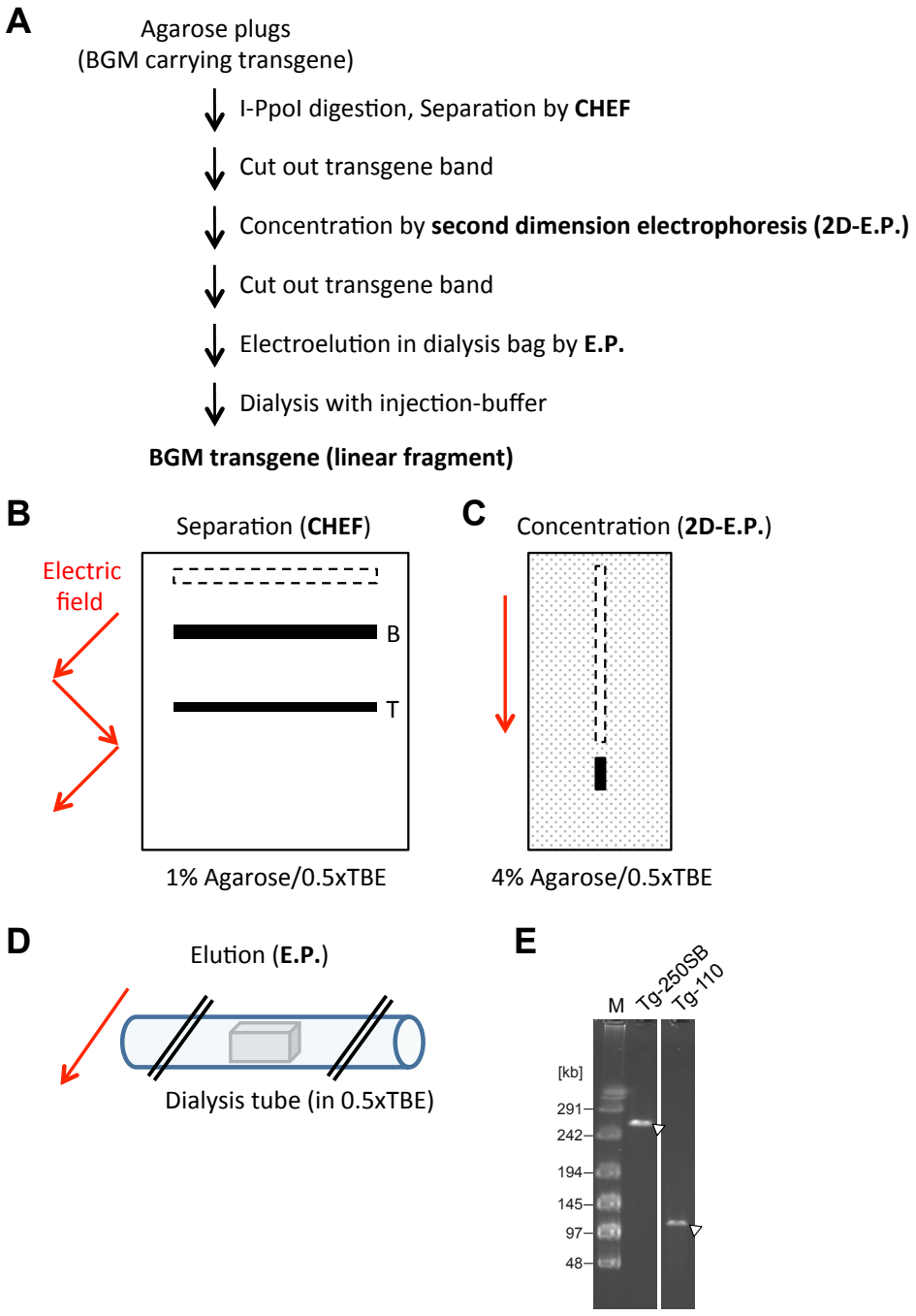

Figure 3-1. Preparation of transgenes from BGM clones.

(A) Procedures of the preparation method of transgene for pronuclear-microinjection for the BGM transgenesis. Linearized transgene is isolated from agarose plugs in 4-steps. First, the transgene is separated from the BGM vector (band B) by CHEF electrophoresis of I-PpoI digested agarose plugs (illustrated in **B**). Second, the transgene band (band T) is excised, and the agarose strip is placed vertically and embedded in a high-percentage (4%) agarose gel. The transgene is concentrated by a second dimension (2D) electrophoresis (illustrated in **C**). Third, the concentrated transgene band is excised and placed into a dialysis bag with electrophoresis buffer. By electrophoresis, the transgene is eluted from a 4% agarose plug (illustrated in **D**). Dotted or solid bands represent positions of DNA bands before or after electrophoresis. Finally, the transgene solution in the dialysis bag is dialyzed with a high-salt injection buffer. **(E)** Quality and yield of the purified transgene for microinjection are confirmed by CHEF. The concentration of retrieved transgenes ranged from 0.3 to 3 ng/µl.

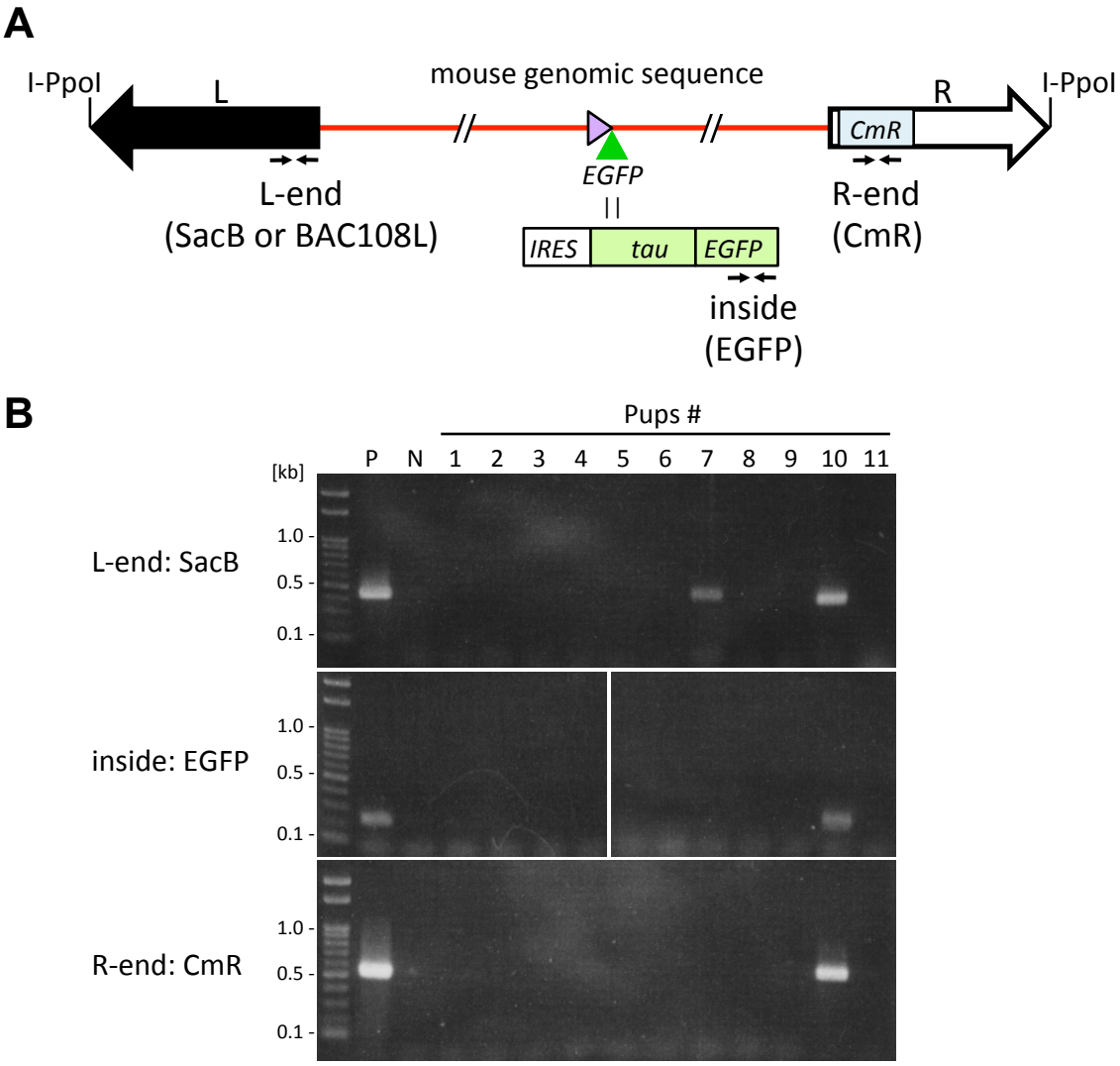

Figure 3-2. Screening of founder mice.

(A) Schematic representation of transgenic construct and targets for genotyping PCR. To assess the integration of the transgene, three primer sets are used; SacB or BAC108L vector sequence for left end, EGFP for middle, CmR sequence for right end of the transgene. **(B)** A representative result of genotyping of newborn mice for Tg-110 founder. Mouse#10 was all three PCR positive, would carry an intact transgene. In contrast, Mouse#7 was PCR positive for the left end only, indicating a partial integration of the transgene.

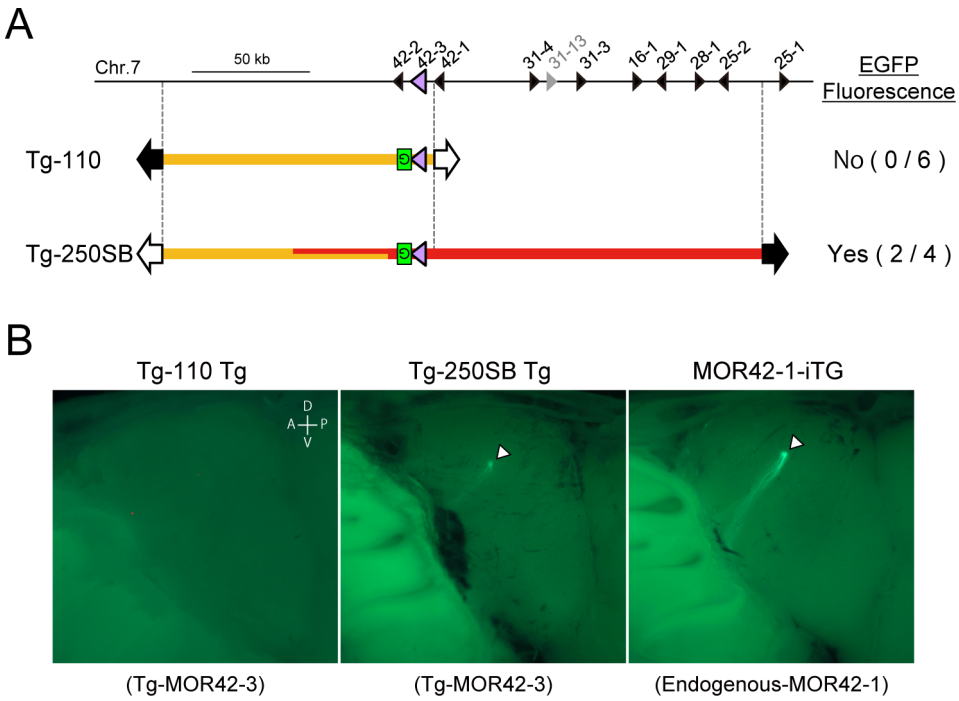

Figure 3-3. Expression of the *EGFP* reporter gene in transgenic mice.

(A) Schematic diagram of Tg-110 and Tg-250SB. Tg-250SB transgenic mice express the *EGFP* reporter gene, whereas Tg-110 transgenic mice do not. The number of EGFP-positive independent founders and lines among the total analyzed is shown in parentheses. **(B)** Whole-mount images of the medial aspect of the OB in transgenic Tg-110 (line #10, left panel) and Tg-250SB mice (line #8, medial panel) and the MOR42-1-iTG (Bozza *et al.* 2009) gene-targeted mouse (right panel). Tg-110 transgenic mice do not show EGFP fluorescence, whereas Tg-250SB transgenic mice do, allowing the visualization of axonal projections of transgene-expressing OSNs to the OB. EGFP-labeled axons in Tg-250SB converge on the dorsal side of the OB and form a glomerulus. Arrowheads indicate glomeruli. D, dorsal; V, ventral; A, anterior; P, posterior.

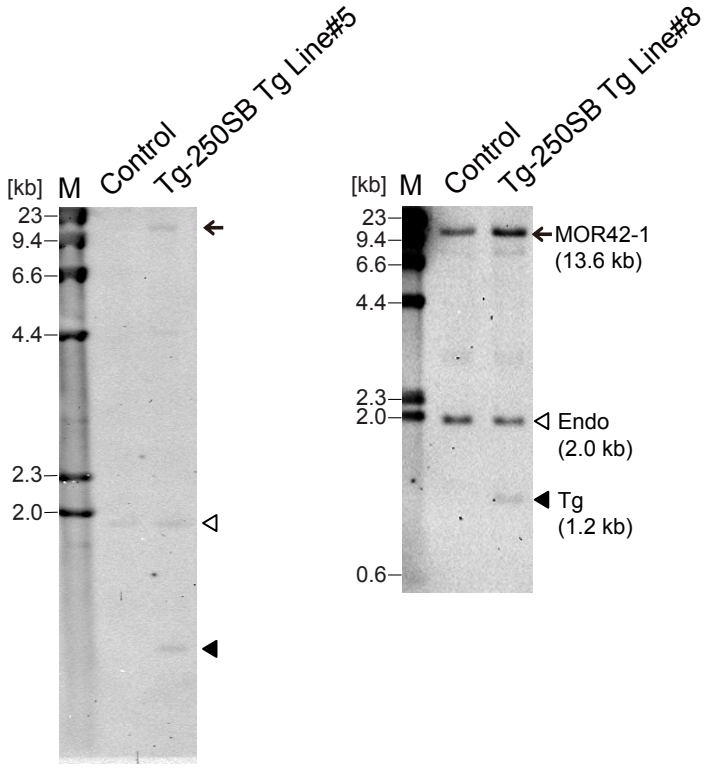

Figure 3-4. Southern blot analysis of two Tg-250SB lines.

Southern blot analysis of transgenic mouse of Tg-250SB line #8 and littermate control mouse. The *MOR42-3* probe detected the endogenous (2.0 kb, open arrowhead) and the transgenic (1.2 kb, closed arrowhead) *MOR42-3* genes, and also detect both endogenous and transgenic *MOR42-1* gene (13.6 kb, arrow) due to cross-hybridization. By comparing the intensities of the hybridized signals, the copy numbers of the transgene in Tg-250SB line #5 and #8 was estimated to be two and one, respectively.

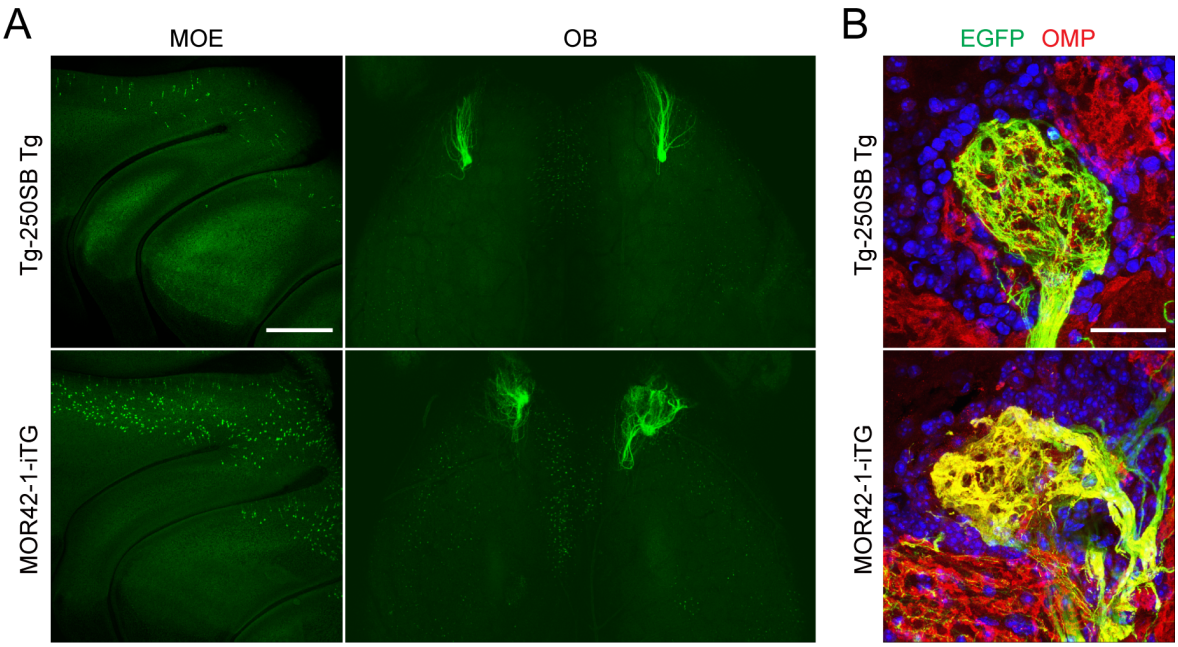

Figure 3-5. Analysis of the *EGFP*-expression pattern in Tg-250SB.
(A) Confocal images of Tg-250SB line #8 transgenic and MOR42-1-iTG gene-targeted mice (Bozza *et al.* 2009). Mosaic patterns of *EGFP*-expressing OSNs in the dorsal MOE (medial view, left panel) and glomeruli formed by axons in the OB (dorsal view, right panel) were observed, indicating the functional expression of transgenic *MOR42-3*. **(B)** Tg-MOR42-3 (top panel) or endogenous-MOR42-1 glomeruli (bottom panel) were detected by immunostaining coronal cryosections. Red, OMP immunoreactivity, which designates olfactory sensory axons. Green, EGFP immunoreactivity, which designates Tg-MOR42-3 or endogenous-MOR42-1 axons. Blue, DAPI nuclear staining, which marks the glomeruli structure. The glomeruli in the Tg-250SB transgenic mouse showed an intermingling of red (EGFP-negative) and yellow (EGFP-positive) axons, whereas the glomeruli in the homozygous MOR42-1-iTG mouse were fully doubly labeled. Scale bars: A, 500 μm; B, 30 μm.

Table 3-2. Summary of screening of Tg-110 and Tg-250SB founder mice.

Tg-110 Founder	Sex	Marker genes *EGFP*	*SacB*	*CmR*	Germline transmission (Tg+ / pups)
# 7	♂	-	+	-	N.D.
# 10	♀	+	+	+	2 / 9
# 34	♀	+	+	+	6 / 8
# 35	♀	-	+	-	N.D.
# 37	♀	+	+	+	0 /16
# 41	♂	+	+	+	1 / 3
# 44	♀	+	+	+	4 / 9
# 45	♂	+	+	+	5 / 17

Tg-250SB Founder	Sex	Marker genes *EGFP*	*BAC108L*	*CmR*	Germline transmission (Tg+ / pups)
#1	♂	+	-	+	N.D.
# 3	♀	+	+	+	0 / 31
# 5	♀	+	+	+	3 / 20
# 8	♂	+	+	+	9 / 18
# 11	♂	+	+	+	0 / 26

Results of the genotyping PCR and germline transmission of founder mice are summarized. Totally, eight Tg-110 and five Tg-250SB transgenic founders were obtained from 45 and 26 newborn mice, respectively; the efficiencies of founder in newborn mice were 18% for Tg-110 and 19% for Tg-250SB. +, PCR-positive; -, PCR-negative; N.D., not determined.

CHAPTER IV

IDENTIFICATION AND FUNCTIONAL ANALYSIS OF *CIS*-ELEMENT FOR A CLASS I ODORANT RECEPTOR GENE

IV - 1 Introduction

C*is*-elements are regulatory sequences to control spatial and temporal gene expression, and also provide important clues for exploration of the *trans*-elements. Several regulatory regions including *cis*-elements have been identified for class II OR genes expression (Vassalli *et al*., 2002; Serizawa *et al*., 2003; Rothman *et al*., 2005; Bozza *et al*., 2009). Some of these have been shown to activate single class II OR gene in proximal position of each *cis*-element, others can select and activate one of class II OR genes among the gene cluster in distal position (Vassalli *et al*., 2002; Serizawa *et al*., 2003; Rothman *et al*., 2005; Fuss *et al*., 2007; Nishizumi *et al*., 2007; Khan *et al*., 2011). Regarding to the *trans*-elements for class II OR genes, some transcription factors have been identified, including Lhx2 (Hirota and Mombaerts, 2004) and Emx2 (McIntyre *et al*., 2008). The loss of function mutation in Lhx2 precludes expression of all tested class II OR genes. By contrast most of class I OR genes are still expressed in Lhx2-deficient mutants, indicating a different regulatory mechanism between two classes of OR genes (Hirota *et al*., 2007). Loss of Emx2 decreases many class I and class II OR genes expression and increases some class II OR genes expression, suggesting class-independent regulatory function of Emx2. Whereas, the *cis*-elements and critical *trans*-elements for class I OR genes have not been identified.

In the previous chapter, I demonstrated that the *cis*-element for the class I OR gene, *MOR42-3* exists in the Tg-220 mouse genomic region. Identification of *cis*-element for class I OR genes will allow us to understand the regulatory mechanism of this gene family. In this chapter, I identified the *cis*-element for *MOR42-3* by a deletion mutagenesis of Tg-220 in combination with bioinformatics analyses, and investigated its function.

IV - 2 Experimental procedures

IV-2-1) Nucleotide identity plots

The nucleotide sequence of an approximately 220 kb mouse genomic region (BAC2: RP23-61O11, Accession #: AC102535), in which the *cis*-element for the *MOR42-3* gene exists, was compared with those of the class I OR cluster region in other species by using the web-based VISTA program (http:// genome.lbl.gov/vista/mvista/submit.shtml). I used the Shuffle-LAGAN for the alignment program and the following settings of VISTA parameters; conservation identity: 70%; calculation window and minimal conserved width: 100 bp for mammals or 30 bp for other vertebrates. The 220 kb mouse genomic region contains a part of class I OR gene cluster, and non-OR genes *Rrm1* and *Stim1*. Because class I OR gene cluster and non-OR genes compose a syntenic region among mammals, I used approximately 2 Mb genomic sequences from *Stim1* locus to the class I OR gene cluster to analyze homology between mammalian species. For zebrafish, western clawed frog and fugu, I used genomic regions or contigs containing orthologous genes of mouse MOR42 subfamily (β group) up to 2 Mbp. I compared the following sequences; mouse (*Mus musculus*, mm9) chromosome 7: 109,558,891 - 109,778,893; human (*Homo sapiens*, GRCh37) chromosome 11: 3,873,375 - 5,867,373; rhesus macaque (*Macaca mulatta*, rheMac2) chromosome 14: 67,553,852 - 69,553,842; horse (*Equus caballus*, equCab2) chromosome 7: 71,917,076 - 73,917,066; dog (*Canis familiaris*, canFam3) chromosome 21: 26,492,740 - 28,491,838; cow (*Bos taurus*, bosTau7) chromosome 15: 48,532,134 - 50,532,124; opossum (*Monodelphis domestica*, monDom5) chromosome 4: 348,999,211 - 350,999,201; zebrafish (*Danio rerio*, danRer7) chromosome 15: 4,465,100 - 6,464,551; western clawed frog (*Xenopus tropicalis*, xenTro3) contig GL173612: 1 - 241,636; fugu

(*Takifugu rubripes*, fr3) contig HE593413: 1 - 9,721. Repetitive elements of the mouse base sequence were masked with RepeatMasker in VISTA server.

IV-2-2) Construction of a deletion series of Tg-220

To construct upstream deletion plasmids, I used a BAC plasmid p108BGMC (provided from Dr. S. Kaneko), which contains a *cat* gene for *B. subtilis*. 1.2 kb homology arms homologous to -79, -29 or -16 kb position of Tg-220 were cloned into p108BGMC as HindIII-BamHI fragments. To construct downstream deletion plasmid, the new BAC plasmid p108Tet was generated by inserting a *tet* gene of a PstI-EcoRI fragment derived from pBEST307 (Kuroki *et al*., 2008) into the PmlI site of p108IPpoI-HPNSB after blunt-ending treatment. A 1.2 kb homology arm to +19 kb position of Tg-220 was cloned into p108Tet as a HindIII fragment. All homology arms were amplified by PCR (KOD plus DNA polymerase) using the BAC2 DNA as a template. Primer sequence and PCR conditions are summarized in Table 4-1. All deletion plasmids were linearized with BamHI or MluI, and used for transformation.

Transformation and the I-PpoI/CHEF analysis of the BGM clones were performed as described in Chapter II. The transformants were screened with chloramphenicol (Cm, 5 μg/ml, Wako) or 10 μg/ml tetracycline. Electrophoresis was performed at 4 V/cm at 14°C under the following conditions: 20 sec for 24 h (single deletion clones) and 12 sec for 21 h (double deletion clones).

IV-2-3) Characterization of the identified 13 kb region

To characterize the 13 kb region containing the *cis*-element for the *MOR42-3* class I

OR gene, I used the UCSC Genome Browser (UCSC web page, http://genome.ucsc.edu/cgi-bin/hgGateway). I analyzed the 13 kb region locating the positions of 109,653,017 - 109,666,061 on chromosome 7 (NCBI/mm9) with the deposited "Expression and Regulation" datasets including chromatin immunoprecipitation sequencing (ChIP-seq) results of CTCF (ENCODE mouse project, Mouse ENCODE Consortium *et al*., 2012), and compared the information from mouse genome data with my nucleotide identity plot of human.

IV-2-4) Construction of a series of 13 kb reporter transgenes

To construct a vector plasmid, a linker containing EcoRI, PacI and AscI sites was inserted in the SacI site of the PAKI3 plasmid, in which the KpnI site of the pBluescript II SK(+) (Stratagene) was replaced with AscI and PmeI sites. Then, another linker containing XbaI, SacI and SacII sites was inserted between the HindIII and BamHI sites. The *gapVenus-pA* fragment (provided from Dr. Y. Yoshihara) was inserted into the BamHI-PacI sites of the vector plasmid, and then the PCR-amplified 423p promoter fragment was inserted into the XbaI-SacII sites to construct the 423pgVpA plasmid. Following the reconstructed 13 kb region and its deletion fragments were inserted into this plasmid to construct a series of reporter transgenes.

To clone the identified 13 kb region, the 2.4 kb 5´-end and the 0.9 kb 3´-end of the 13 kb region were subcloned into pBluescript II SK vector as SalI-BamHI and BamHI-XbaI fragments, respectively. Both fragments have an endogenous PacI site. The 10 kb PacI fragment from BAC2 clone was inserted into the subcloned plasmid to generate the pBS-13k plasmid, which contains a full-length 13 kb region as a SalI-XbaI fragment.

The 9.1 kb fragment was generated by deleting a SalI-PmlI fragment of the pBS-13k plasmid. Similarly, the 7.9k and 3.9k fragments were generated by deleting SalI-NheI and SalI-NcoI fragments of pBS-13k, respectively. The ΔCore fragment was generated by self-ligating the vector part after NcoI digestion of the pBS-13k plasmid to remove the 3.8 kb NcoI fragment, designated as the Core (Nco) containing the Core region. The 13k, 9.1k, 7.9k, 3.9k and ΔCore fragments were transferred to the 423pgVpA plasmid as XhoI-XbaI fragments to construct transgenes. The Core only transgene, the Core (Nco) was constructed by directly inserting the Nco fragment containing the Core region into the XhoI and XbaI sites after blunt-ending. The 13k-423pgV, 9.1k-423pgV, Core (Nco)-423pgV and ΔCore-423pgV were transferred into a *Tol2* donor plasmid, pTol2-EPA. The vector was constructed by inserting a linker containing EcoRI, AscI and PacI sites between SalI and BglII sites of the pT2 vector (provided from Dr. H. Nishihara). The 423 promoter, the 5´-end and 3´-end fragments of the 13 kb region were amplified by PCR (PrimeSTAR HS) using the BAC2 DNA as a template (Table 4-1).

IV-2-5) Generation and analysis of transgenic mice

Transgenic mice carrying Tg-220 and its deletion mutants were generated as described in Chapter III. The founder mice of Tg-220 were screened by PCR with the same primer sets for Tg-110. The same primer set for Tg-250SB was used of deletion series of Tg-220 (Table 4-1).

For a series of 13 kb reporter transgenes, I took two microinjection methods of the conventional pronuclear microinjection and the *Tol2*-mediated cytoplasmic microinjection method (Sumiyama *et al*., 2010). For 7.9k-423pgV and 3.9k-423pgV, the linear transgenes

were excised from the vector plasmids by AscI digestion, and purified with the QIAquick Gel Extraction Kit (QIAGEN), phenol-chloroform extraction and ethanol precipitation. The purified transgenes were dialyzed with the TG buffer (5 mM Tris-HCl (pH 8.0), 0.1 mM EDTA) at 4-6°C overnight. The transgenes were diluted to 5 - 10 ng/μl with injection buffer before use, and were microinjected into the pronuclei of fertilized eggs (pronuclear microinjection). For high efficient transgenesis, I used the *Tol2* cytoplasmic microinjection method as described elsewhere (Sumiyama *et al*., 2010). I took this method to generate the 13k-423pgV, 9.1k-423pgV, Core (Nco)-423pgV and ΔCore-423pgV transgenic mice. Briefly, the *Tol2* transgene plasmids were purified by phenol-chloroform extraction and ethanol precipitation, and then dissolved a TG buffer prepared with DEPC-treated water. *Tol2* transposase mRNA was synthesized *in vitro* using the NotI-linearized pCS-TP (Kawakami *et al*., 2004) with the mMESSAGE mMACHINE SP6 Kit (Ambion). DNA solution containing 20 ng/μl circular *Tol2*-trangene plasmid and 25 ng/μl transposase mRNA was injected into the cytoplasm of fertilized eggs (cytoplasmic microinjection). In both methods, the newborn mice were screened by PCR with a primer set for the *Venus* gene (Table 4-1).

The founders and lines obtained by crossing with C57BL/6 mice were used for analysis. The procedures of expression analysis of the *Venus* reporter are described in Chapter III.

IV-2-6) Bioinformatics analysis of the Core region in other species

A genome alignment was performed using 59 vertebrate genomes provided from UCSC regarding the mouse genome (*Mus musculus*, mm10, released in December 2011)

as a reference. Two continuous regions in the mouse genome were picked up from alignable regions between group β and α class I OR genes among mouse, human and dog, and defined as core1 and core2. Then, I mapped the two core sequences to 13 placental mammalian genomes. The class I OR gene cluster of cow was divided because of incomplete assembly. Names of orthologous gene groups (OGGs) were used as described in a previous report of the analysis of OR genes among 13 placental mammals (Niimura *et al*., 2014). Briefly, The OGG1-60 and OGG1-105 belong to group β class I OR genes (*MOR42* family in mouse), whereas other OGGs including OGG1-47, OGG1-107 and OGG1-120 belong to group α class I OR genes.

Table 4-1. Primer sequences and PCR conditions used in Chapter IV.

Primer name	Strand	Sequence	Annealing temp Extension time	Purpose	PCR condition
Deletion of Tg-220					
J631_B2Del1_F	Forward	CCCAAGCTTATCACCAGGGACAACTTGG	55.5 °C	homology arm	KOD plus
J632_B2Del1_R	Reverse	CGGGATCCAGAGGTGCAAGAAACTGAC	1 min 20 sec	to -79 kb of *MOR42-3* (ΔTg-U80)	
J633_B2Del2_F	Forward	CTCAAGCTTGTCTGCCCTCTCTGAC	60.6 °C	homology arm	KOD plus
J634_B2Del2_R	Reverse	CGGGATCCTTGACGGTGGCTCTGATTG	1 min 20 sec	to -16 kb of *MOR42-3* (ΔTg-U20)	
J635_ B2Del3_F	Forward	CCCAAGCTTGGCTCGCTGAGACATTAGC	61 °C	homology arm	KOD plus
J636_ B2Del3_R	Reverse	CCCAAGCTTCCATCTTCTGCTCCCTAGAC	1 min 20 sec	to +19 kb of *MOR42-3* (ΔTg-140Δ3')	
J675_B2Del4_F	Forward	GGCCATTCAAGCTTCTACTCTCTAC	58 °C	homology arm	KOD plus
J676_B2Del4_R	Reverse	CGGGATCCACTTCCATGACCTGAGTC	1 min 20 sec	to -29 kb of *MOR42-3* (ΔTg-U30)	
Short transgene for reporter assay					
J734_13k_5F	Forward	ACGCGTCGACTCAGTATCAGGCCATTC	55 °C	5'-region of 13 kb region	PrimeSTAR HS
J735_13k_5R	Reverse	CGGGATCCTAGTTCAGCTGTCGCTTTCC	2 min 30 sec	for reconstruction	
J736_13k_3F	Forward	CGGGATCCATAGATCCGCCTACTTCCAC	57 °C	3'-region of 13 kb region	PrimeSTAR HS
J737_13k_3R	Reverse	GCTCTAGAGAGGGCTCAACATTGCTTTG	1 min	for reconstruction	
J738_423p-F	Forward	GCTCTAGAGCACACCCCTTCAGCTCCTC	62 °C	42-3 promoter region	PrimeSTAR HS
J739_423p-R	Reverse	TCCCCGCGGCAGGACTCACACAGGCAGGG	1 min		
Poly linkers					
J740_Hind-XSS-Bam_F	Sense	AGCTTTCTAGAGAGCTCCCGCGGG		linker for short-Tg vector	
J741_Hind-XSS-Bam_R	Antisense	GATCCCCGCGGGAGCTCTCTAGAA		containing XbaI-SacI-SacII	
J731_Linker 3	Sense	CTAGGAATTCTTAATTAAGGCGCGCCAGCT		linker for short Tg vector	
J732_Linker 4	Antisense	GGCGCGCCTTAATTAAGAATTC		containing EcoRI-PacI-AscI	
J761_dSalEAPdBgl	Sense	TCGAGAATTCGGCGCGCCGGTTAATTAAC		linker for pTol2-EPA	
J762_dBglPAEdSal	Antisense	GATCGTTAATTAACCGGCGCGCCGAATTC		containing EcoRI-AscI-PacI	
Genotyping of Tg mice					
GFP-F	Forward	GGCATCAAGGTGAACTTCAAGATCC	62 °C	genotype	Genotyping
GFP-R3	Reverse	CTTTACTTGTACAGCTCGTCCATGC	30 sec	EGFP (intermediate of transgenes)	
J619_SacB_F	Forward	GCTGAATACAACGGCTATCACG	62 °C	genotype	Genotyping
J620_SacB_R	Reverse	TCTCTCAGCGTATGGTTGTCG	30 sec	Left end of Tg-220	
J709_BAC108L_F	Forward	CGTATTCAGTGTCGCTGATTTG	60 °C	genotype	Genotyping
J710_BAC108L_R	Reverse	TTAGCGATGAGCTCGGACTTC	30 sec	Left end of deletion-Tgs	
J621_CmR_F	Forward	GAGGCATTTCAGTCAGTTGCTC	62 °C	genotype	Genotyping
J622_CmR_R	Reverse	CGGCATGATGAACCTGAATCG	30 sec	Right end of Tg-220, deletion-Tgs	
Venus_F	Forward	GCAAGCTGACCCTGAAGCTG	60 °C	genotype	Genotyping
Venus_R	Reverse	TTGCTCAGGGCGGACTGGTA	30 sec	Short transgene series	

PrimeSTAR HS (TaKaRa)	**KOD plus ver.2 (Toyobo)**	**Genotyping**
94 °C 5 min	94 °C 2 min	94 °C 5 min
[30 cycles]	[30 cycles]	[30 cycles]
98 °C 10 sec	98 °C 10 sec	94 °C 30 sec
(Annealing temp.)* 5 sec	(Annealing temp.)* 30 sec	(Annealing temp.)* 30 sec
72 °C (Extension time)*	68 °C (Extension time)*	72 °C (Extension time)*
72 °C 5 min	68 °C 5 min	72 °C 5 min

Sequences of primers and oligonucleotides used in Chapter IV are summarized. They were used for constructing homology arms, transgenes for reporter assay, and genotyping of transgenic mice. PCRs with PrimeSTAR and KOD plus kits were basically performed according to the manufacturer's recommendations; all PCR conditions are summarized in bottom of Table. Genotyping PCR used a recombinant Taq DNA polymerase. The annealing temperature and extension time were described in Table.

IV - 3 Results

IV-3-1) Comparative analysis of the Tg-220 mouse genomic sequence

Transgenic analysis using the Tg-110 and Tg-250SB demonstrates that the genomic region of 4 to 142 kb upstream the *MOR42-3* transcription start site (TSS) contains a *cis*-acting element necessary for the expression of the transgenic OR gene. Because this genomic region was included in Tg-220, I compared the mouse genomic sequence of Tg-220 with the class I OR cluster regions in other mammals including human, rhesus macaque, horse, dog, cow, and opossum, to estimate the position of the *cis*-element. Functional genomic regions such as coding sequences, exons and intergenic regulatory regions tend to be conserved during evolution. A nucleotide identity plots of mammals revealed several conserved intergenic regions in addition to coding sequences of OR or non-OR genes and exons (Figure 4-1A). Note that many OR genes have a 5´ non-coding exon, which are not annotated in the database. Thus, I focused a 13 kb region of 16 to 29 kb upstream the *MOR42-3* TSS. This region has a cluster of homology peaks and is far from the neighboring class I OR genes, *MOR42-1* and *MOR31-4*. No significant homology peak was found in cow and opossum, probably due to an incomplete assembly of genomic sequences and to a far evolutional distance.

The mouse *MOR42* family is a unique OR gene subfamily that is orthologous to some fish OR genes (Niimura, 2009). Therefore, it is conceivable that the *cis*-element for *MOR42-3* is conserved in fish or amphibian genome, which contains orthologous genes of *MOR42-3*. To examine this possibility, I also generated a homology plot of zebrafish, western clawed frog and fugu (human as control, Figure 4-1B). I found no significant conservation in intergenic region, suggesting that the highly conserved regions in

mammals might be acquired during evolution.

IV-3-2) Construction of a series of deletion mutants of the Tg-220

According to the comparative analysis of the nucleotide identity plots, I designed a series of deletion constructs of Tg-220 at the four positions of 79 kb, 29 kb, 16 kb upstream and 19 kb downstream the *MOR42-3* TSS (Figure 4-1, 4-2). The deletion mutagenesis was performed by replacing the targeted regions with one of antibiotic resistance genes of *tet* or chloramphenicol resistance genes (*cat*) using the BGM vector system (Figure 4-2A). I constructed three mutant transgenes by upstream deletion, ΔTg-U80, ΔTg-U30 and ΔTg-U20 (Figure 4-2B). I obtained 22 colonies for ΔTg-U80, 51 colonies for ΔTg-U30 and 34 colonies for ΔTg-U20 resistant to chloramphenicol (Cm) in single transformation. I further constructed four deletion transgenes by downstream deletion (Figure 4-2C), and obtained 25 colonies for ΔTg-U140Δ3´, 321 colonies for ΔTg-U80Δ3´, 339 colonies for ΔTg-U30Δ3´ and 145 colonies for ΔTg-U20Δ3´resistant to Tet. I-PpoI/CHEF analyses of each three representative clones of these constructs showed bands correspond to the expected deletion transgene inserts (Figure 4-2D). The deletion of BGM insert was previously described using a *c*I-*spc* cassette (Kaneko *et al*., 2003). Several plasmids enable the simultaneous deletion of two regions, for example, upstream and downstream of the targeted region.

IV-3-3) The 13 kb region is necessary for transgenic *MOR42-3* expression

Using the deletion constructs, I generated transgenic mice carrying Tg-220, ΔTg-U80, ΔTg-U30 or ΔTg-U20 (Figure 4-3A). Briefly, transgenes were prepared from

BGM clones with same method in Chapter III and were microinjected into fertilized mouse eggs. Newborn mice were genotyped by PCR using the same primer sets for Tg-110 and Tg-250SB transgenic mice (Table 4-1). I obtained four Tg-220 founders carrying the intact transgene, and one transgenic mouse carrying the partial transgene, which was PCR positive for right end sequence only. For ΔTg-U80, five pups carried the intact transgene, and one lacked the left end sequence. For ΔTg-U30, three pups carried the intact transgene, and two carried only right end sequence. For ΔTg-U20, one pup carried the intact transgene.

I analyzed expression of the *EGFP* reporter gene in transgenic founders or lines. All four Tg-220 transgenic lines showed EGFP fluorescence (Figure 4-3A). In the deletion mutants, three founders of ΔTg-U80 were EGFP positive, but one line was negative. Two founders and one line of ΔTg-U30 whose upstream region was shorten to 16 kb position from *MOR42-3* TSS were also EGFP positive. No apparent difference in expression levels of the *EGFP* reporter gene and axonal projection patterns among these transgenic mice were observed, indicating that a genomic region carried in ΔTg-U30 is sufficient for proper expression of *MOR42-3* transgene (Figure 4-3B). In contrast, a ΔTg-U20 line did not show EGFP fluorescence. In addition, none of the transgenic mice carrying Tg-110, which included approximately 4 kb sequence upstream the *MOR42-3* TSS, displayed labeled OSNs (in Chapter III, Figure 3-3). Because I analyzed only one founder of ΔTg-U20, the result should be confirmed by further experiments of other ΔTg-U20 transgenic founders or lines. However, considering with the bioinformatics analysis, it is most likely that the 13 kb region contains regulatory sequence to activate expression of the transgenic *MOR42-3* gene.

IV-3-4) Functional analysis of the 13 kb region by conventional transgenesis

Before performing the functional analysis of the identified 13 kb region, I investigated the genomic structure of this region using the UCSC Genome Browser (UCSC web page), in terms of both conservation between human and binding factors. By the nucleotide identity plot, I found that a 1.9 kb sequence displayed high homology peaks (named as Core region) (Figure 4-4). Interestingly, the Core region was flanked by two CTCF (CCCTC-binding factor) binding sites in several tissues (depicted as CTCF/CpG and CTCF), which also displayed homology peaks in the nucleotide identity plot of human (Figure 4-4B, C). Note that the 5´ CTCF binding site locates in a CpG island, suggesting that DNA methylation regulates the binding of CTCF to this site (Figure 4-4B). CTCF is a transcription factor of 11-zinc finger protein, and is known as an insulator element (Chung *et al*., 1993; Gaszner and Felsenfeld, 2006). Recent studies have demonstrated that CTCF is also involved in the chromatin conformation and regulates gene expression or other events, such as stochastic expression of clustered protocadherin (*Pcdh*) genes in neurons (Hirayama *et al*., 2012) and V(D)J recombination in immune cells (Chaumeil and Skok, 2012).

To reveal function of each component in the 13 kb region, I conducted a conventional transgenic reporter assay (Figure 4-5). I constructed a reporter transgene, 13k-423pgV, which consists of the 13 kb region, a promoter sequence including *MOR42-3* TSS, a membrane-targeted *Venus* reporter gene and a polyadenylation sequence, and a deletion series of 13k-423pgV; the 9.1k-423pgV contains all there components of the CTCF/CpG, Core and CTCF; the 7.9k-423pgV contains the Core and CTCF; the 3.9k-423pgV lacks the CTCF/CpG and Core; the Core (Nco)-423pgV contains only the

Core; the ΔCore-423pgV lacks the Core region but contains the CTCF/CpG and CTCF (Figure 4-5A). The transgenic mice carrying 7.9k or 3.9k transgene were generated by the conventional pronuclear-microinjection of a linearized transgene, whereas the 13k, 9.1k, Core (Nco) or ΔCore transgenic mice were generated by the *Tol2* transposon mediated cytoplasmic injection method because of high efficient transgenesis (Sumiyama *et al*., 2010) (Figure 4-5A).

I analyzed expression of the *Venus* reporter gene in transgenic founders or lines (Figure 4-6). Note that these transgene constructs have no OR coding sequence. Thus, it was expected that certain populations of OSN were labeled with Venus fluorescence; the signals are broadly observed in multiple glomeruli not in a specific subset in the olfactory bulb (OB). The 13k, 9.1k and Core (Nco) transgenic mice showed similar *Venus* expression patterns; Venus fluorescence was observed in the most dorsal and dorsomedial OB, and the dorsal MOE, suggesting expression of the *Venus* in class I OR-expressing OSNs. Some signals were also observed in the ventral MOE and the ventromedial OB in the 9.1k or Core (Nco) transgenic mice. These labeled OSNs are considered to express some class II OR genes. It is notable that the Core (Nco) fragment can mainly drive expression in the class I OR-expressing OSN population. On the other hand, transgenic mice carrying the 7.9k fragment, which lacks the CTCF/CpG element, showed a different pattern compared to other Venus-positive transgenic constructs; strong signals were observed in the dorsal and ventral MOE and the ventromedial region of OB. Additionally, bright Venus fluorescence was observed in the vomeronasal sensory nerves in two founders and one line. Different expression pattern between 7.9k and other reporter positive transgenes indicates that the CTCF/CpG sequence or the combination of

CTCF/CpG and CTCF sequences defines the region of transgene expression in the MOE. In contrast, all the 3.9k and majority of the ΔCore transgenic mice, both lacked the Core region, did not show Venus fluorescence. One ΔCore transgenic founder showed strong *Venus* expression in the most dorsal and dorsomedial OB, in addition to moderate expression in the dorsolateral OB, suggesting a possible positional effect of transgene-integration locus. The numbers of reporter positive founders or lines are summarized in Figure 4-5A.

IV-3-5) The Core region is conserved in 13 placental mammals

The OR family, one of chemosensory receptor families, is conserved in many species including mammals, birds, amphibians, fishes and insects (Nei *et al*., 2008). In a large-scale phylogenetic analysis, mammalian OR genes were categorized into three groups; class I genes corresponding to group α or β, whereas all class II genes corresponding to group γ (Niimura and Nei, 2005). The group β OR genes including mouse *MOR42-3* are conserved in many mammals, though this group includes only a few class I OR genes (Niimura, 2009). Therefore, it is conceivable that the Core region is conserved in other mammalian genomes. In fact, nucleotide identity plots revealed that the Core region was conserved in human, rhesus macaque, horse and dog genomes (Figure 4-1A).

To analyze the conservation of the Core region in detail, I performed a large-scale bioinformatics approach based on a sequence comparison analysis, and I found that the Core region consisted of two conserved fragments, designated as core1 and core2, which were conserved in 13 placental mammalian species (Figure 4-7). In mouse, the core1 (0.9

kb) and core2 (1.3 kb) were detected within the 2.3 kb genomic region (Figure 4-7 inset), which is wider than the 1.9 kb Core region; 355 bp larger in upstream and 133 bp larger in downstream. Accordingly, the highly conservation of the Core region was confirmed by the alternative bioinformatics approach.

In addition, I found that the group β OR genes (OGG1-60 and OGG1-105) are consistently present at the end of the class I OR gene cluster in all tested 13 mammals. However, the OGG1-105 has been pseudogenized, and the OGG1-60 has been lost in primate genome including human, chimpanzee, orangutan and macaque. Among analyzed 13 mammals, the core1 and core2 located at the intergenic region between group α and β OR genes, with the exception of cow. The cow class I cluster region is divided into two contigs because of possible assemble error, but the JH122518 contig including two core sequences and a group β OR gene is likely to be inserted into between OGG1-105 and OGG1-109 (group α). Therefore, the bioinformatics analysis demonstrated that the genomic organization was also conserved in 13 placental mammals in addition to the presence of the Core region.

IV - 4 Discussion

In this chapter, I performed the deletion mutagenesis and the bioinformatics analyses and to identify the *cis*-element for the class I OR gene, *MOR42-3*. Starting from the 140 kb region containing the *cis*-element for *MOR42-3* existed (Chapter III), I shortened the putative *cis*-acting element into the 13 kb genomic region by deletion mutagenesis.

Further functional analyses of the 13 kb region demonstrated that the Core (Nco) (3.8 kb NcoI fragment) was capable of activating expression the reporter transgene in the class I OR expressing region in the MOE (Figure 4-5, 4-6). This 3.8 kb region locates approximately 20 kb upstream the *MOR42-3* TSS between *MOR42-1* and *MOR31-4* (Figure 4-5A). Because some *cis*-elements for class II OR genes activate the expression of multiple class II OR genes among the gene cluster in a monogenic manner (Serizawa *et al*., 2003; Fuss *et al*., 2007; Nishizumi *et al*., 2007; Khan *et al*., 2011), the identified region may regulate not only *MOR42-3* but also other two *MOR42* subfamily and/or other class I OR genes. Interestingly, loss of the upstream CTCF binding site (CTCF/CpG) affects the expression pattern of the reporter transgene, suggesting the involvement in regulation of the spatial expression pattern. These results suggest that the 3.8 kb region activates the class I OR gene expression, and the CTCF/CpG alone or the combination of CTCF/CpG and CTCF regulates the spatial and temporal expression patterns of class I OR genes. To examine these possibilities, further experiments are required, such as transgenic analysis of other class I OR gene-tagged transgene like Tg-220 and loss-of-function analyses of the non-coding regions of the Core, and the CTCF/CpG and CTCF.

The large-scale comparative analysis demonstrated the Core region was conserved in all tested 13 placental mammals (Figure 4-7), supporting the idea of an important role of

the Core region as a functional non-coding element. The group β class I OR genes, *MOR42* family, were widely conserved in vertebrates from fish to mammals, whereas the group α class I OR genes present only in mammals (Niimura, 2009). Four tested primates including human, chimpanzee, macaque and orangutan have only one group β class I OR genes (OGG1-105). Although all of them are pseudogenized, the Core region is still conserved in these species (Figure 4-7), suggesting the possible role of the Core region to control both groups of class I OR, α and β class I OR genes. In contrast, although non-mammalian vertebrates, zebrafish, western clawed frog and fugu, had intact group β OR genes, these species did not have the Core region (Figure 4-1B). Note that the genome sequencing and assembly are incomplete in those species, however, it is possible to speculate that the Core region has been acquired in mammals during evolution to regulate mammalian class I OR gene expression.

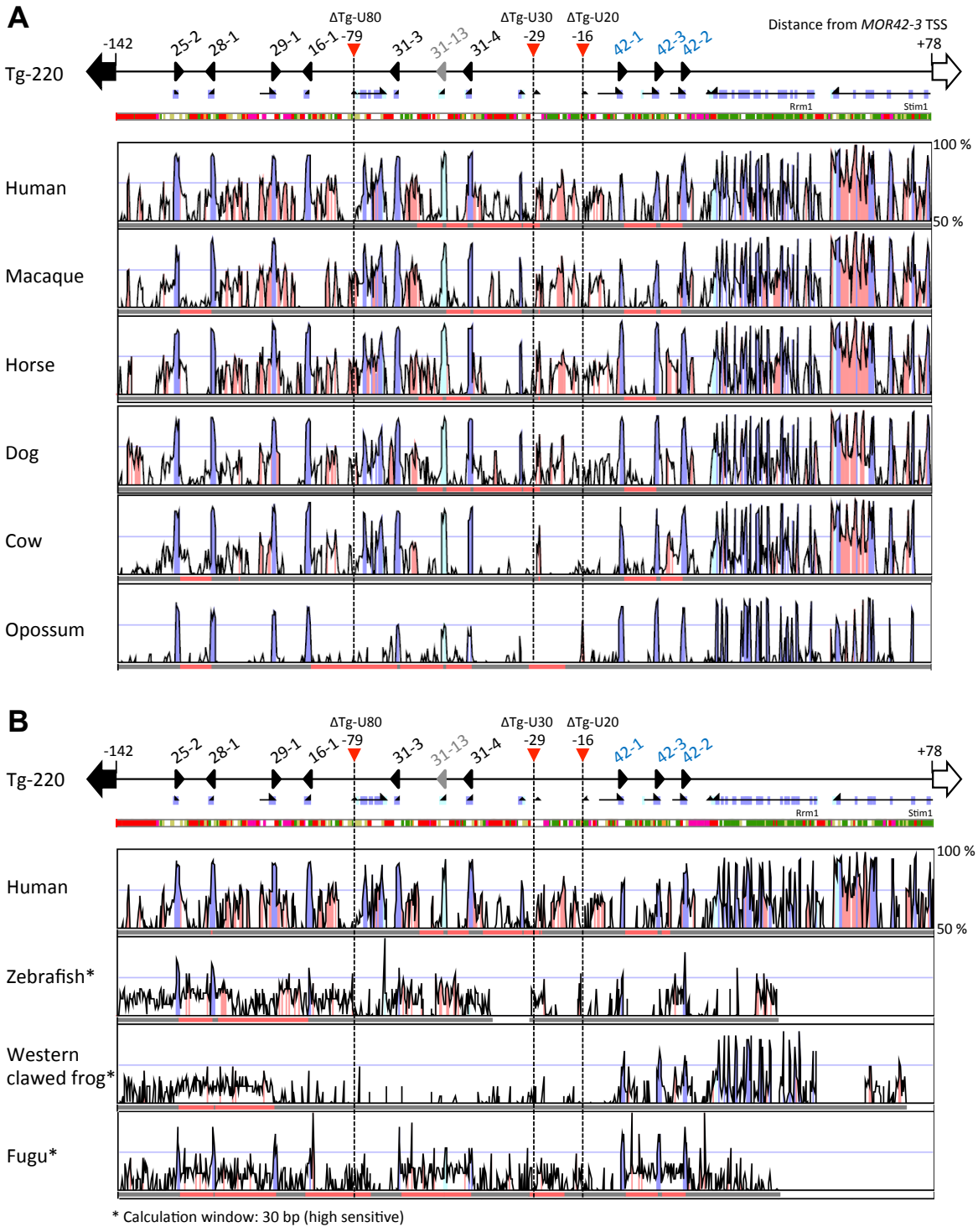

Figure 4-1. Comparative analysis of a Tg-220 mouse genomic sequence with class I OR loci in other species.

(A) Genomic organization of class I OR genes were analyzed by nucleotide identity plots of human, rhesus macaque, horse, dog, cow and opossum with the mouse genomic sequence of Tg-220 using VISTA program. Highly conserved regions (> 70% identity) in intergenic region and coding region are shown in pink and blue, respectively. Triangles represent class I OR genes. An upstream region from -29 to -16 kb positions showed a significant conservation among intergenic regions. **(B)** Nucleotide identity plots of human, zebrafish, western clawed frog and fugu with the mouse genomic sequence. No significant conservations in intergenic region were detected even in the a 30 bp-calculate window for high-sensitive detection.

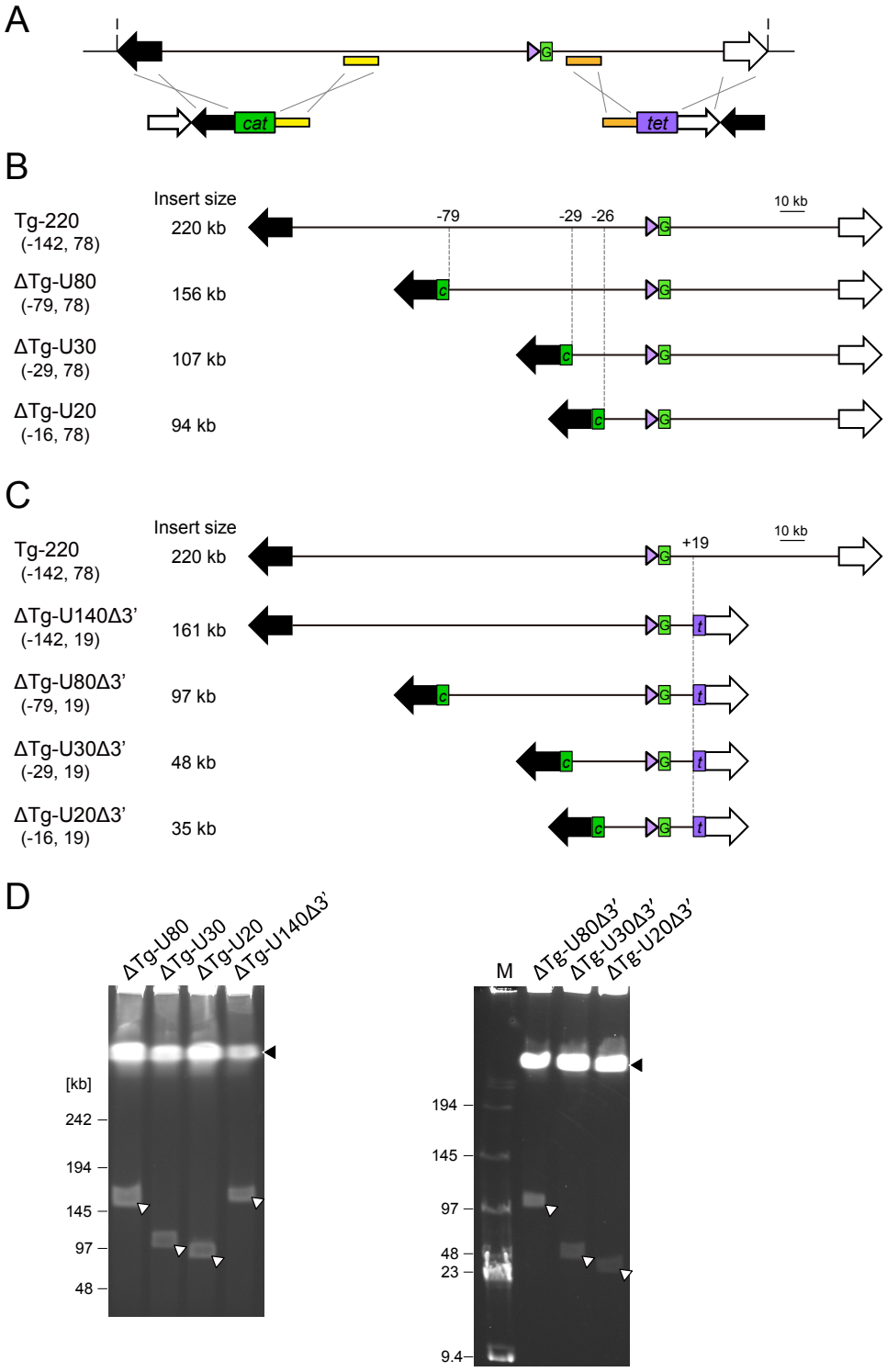

Figure 4-2. Deletion series to identify the *cis*-element for *MOR42-3*

(A) Schematic illustration of a deletion mutagenesis using the BGM vector. A target region is replaced with *tet* and/or *cat* genes by homologous recombination at particular homology regions and BAC vector sequences. The BAC vector sequences (closed and open arrows) were approximately 3.5 kb each. The I-PpoI sites are designated as "I". **(B, C)** Structures of a deletion series of the Tg-220. The *tet* and *cat* genes are designated as *t* and *c* in the boxes, respectively. Distances from the *MOR42-3* TSS to insert ends are shown in parenthesis (5´: -, and 3´:+). **(D)** The insert sizes of the deletion transgenes were confirmed by I-PpoI and CHEF analyses. Representative clones of each deletion constructs showed the expected size of the deletion transgene (open arrowheads). The closed arrowhead indicates the BGM vector.

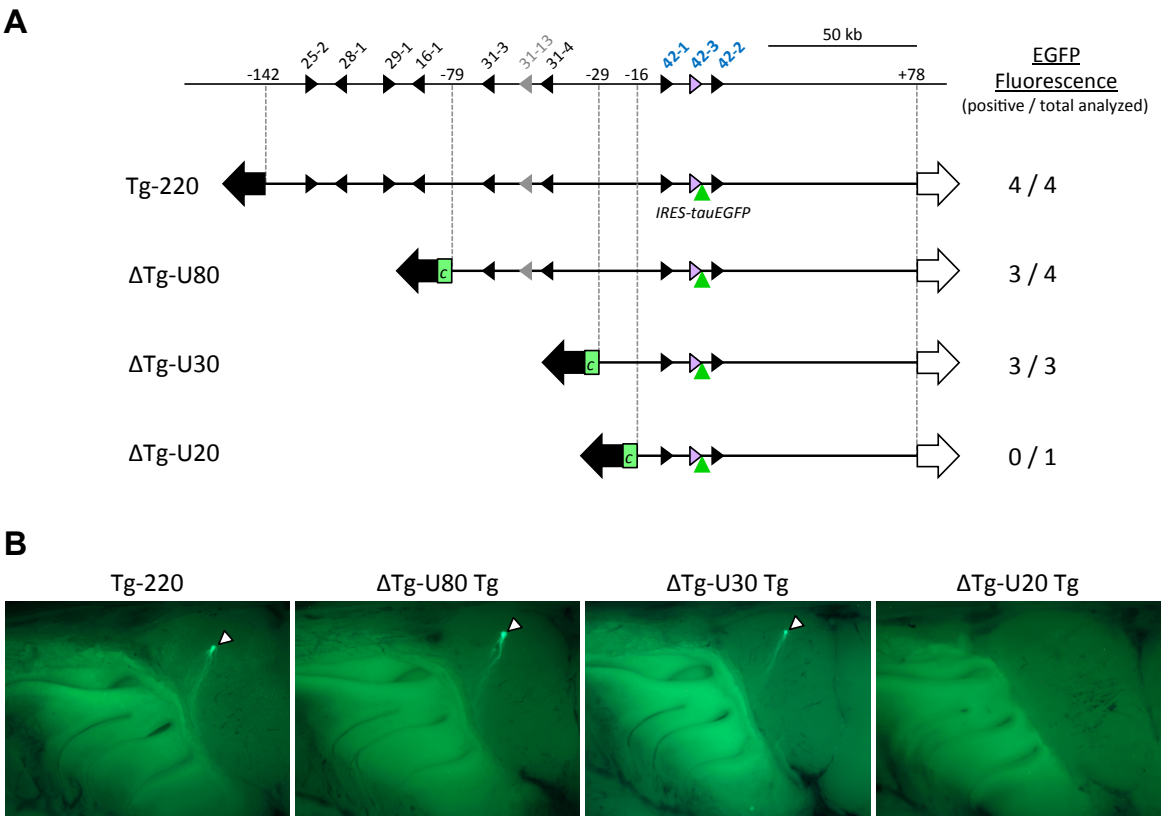

Figure 4-3. Deletion analysis of Tg-220 to identify the *cis*-element for *MOR42-3*.

(A) Schematic illustration of Tg-220 and its deletion series. The number of EGFP-positive founders and lines among the total analyzed is shown in right side. Tg-220, ΔTg-U80 and ΔTg-U30 transgenic mice expressed the *EGFP* reporter gene, where as a ΔTg-U20 transgenic mouse did not. (B) Whole-mount images of the medial aspect of the OB in the transgenic mice. ΔTg-U80 and ΔTg-U30 transgenic mice showed similar axonal projection patterns and formed a glomerulus on the dorsal portion of OB as well as Tg-220. In contrast, ΔTg-U20 whose upstream region was truncated to -16 kb position did not show EGFP fluorescence.

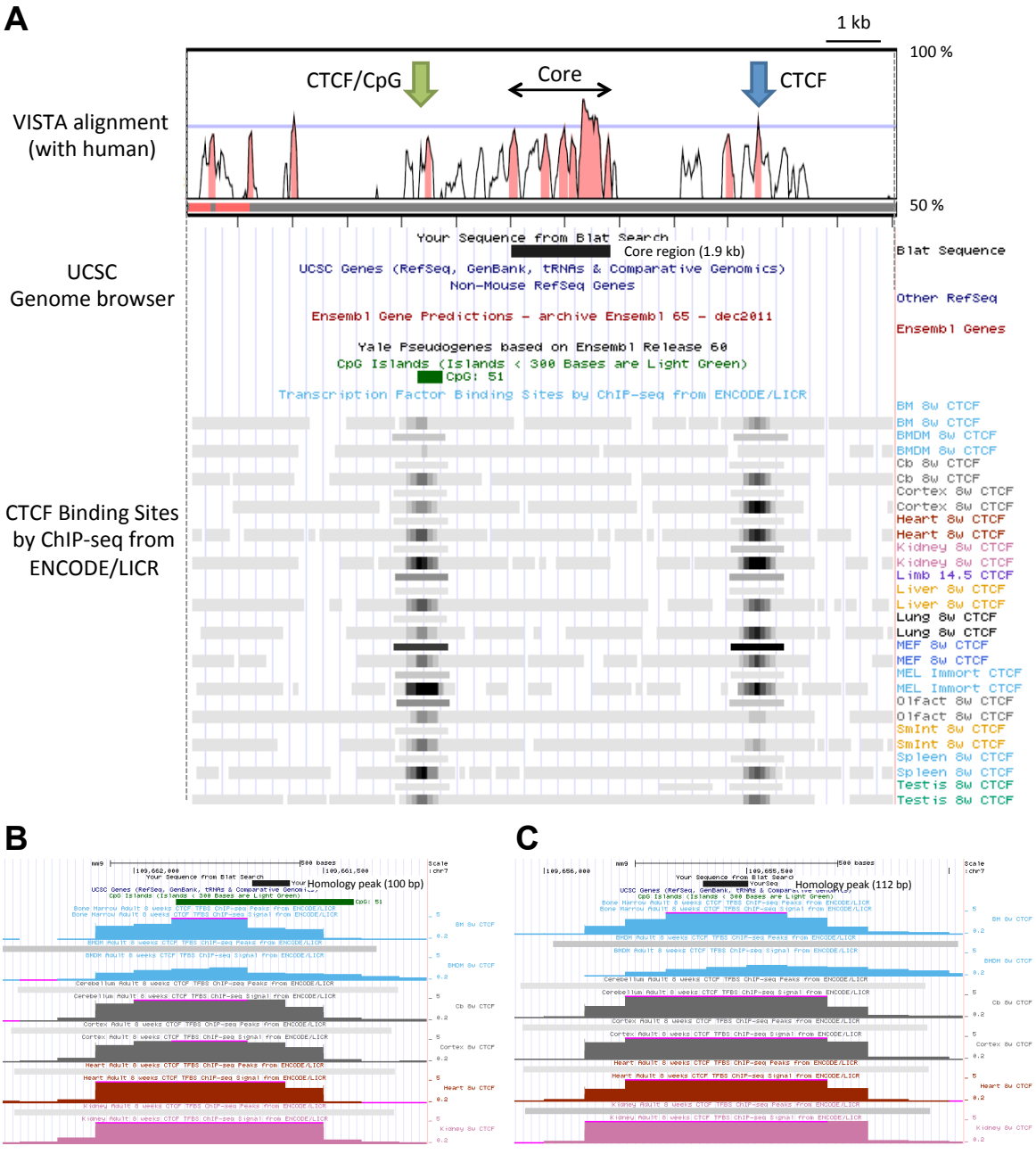

Figure 4-4. The genomic structure of the identified 13 kb region for the *cis*-element for *MOR42-3*.

(A) The nucleotide identity plot of human and UCSC Genome Browser view of mm9 assembly with a CTCF binding profile on several mouse tissues (deposited as ENCODE/LIR) in the 13 kb region. The 13 kb region contains an approximately 1.9 kb conserved region (named as the Core region, black bar) and two CTCF binding sites flanking the Core region. Of these, the 5´-CTCF binding site locates on CpG island (CpG51, green bar). The CTCF binding profile shows intensities of processed ChIP-sequence data of CTCF in grey scale; upper row represents “Peaks” (regions of signal enrichment), lower row represents “Signals” (density graph of signal enrichment). Data from different tissues are depicted in different colors. **(B, C)** Magnified views of CTCF/CpG and CTCF sites. Each conserved region overlaps with CTCF-binding peak.

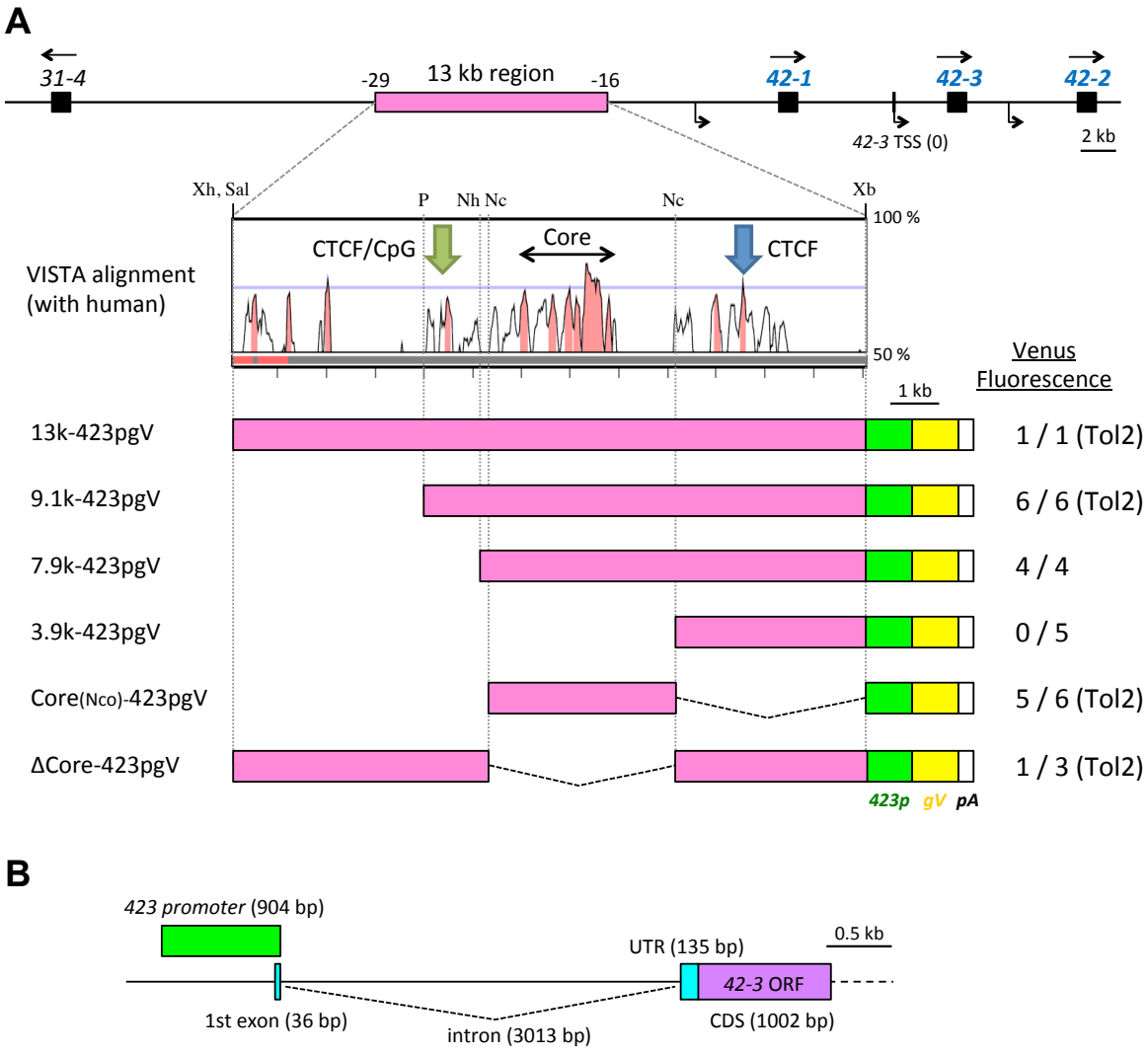

Figure 4-5. Constructs for the functional analysis of the 13 kb region.

(A) Schematic structures of a 13 kb transgene and its deletion series to refine the 13 kb region. Schematic structure of a genomic region around 13 kb region are shown in the top of the panel. Restriction sites used for constructing deletion series of 13 kb region are represented as P, PmlI; Nh, NheI; Nc, NcoI; Xh, XhoI; Sal, SalI; Xb, XbaI. The 9.1k fragment has all three components of CTCF/CpG, Core, and CTCF. The 7.9k fragment lacks the CTCF/CpG. The 3.9k fragment lacks the CTCF/CpG and Core region. The Core (Nco) fragment contains only the Core region. The ΔCore lacks the Core region. The 13k, 9.1k, Core (Nco) and ΔCore transgenic mice were generated by the *Tol2* transposon-mediated cytoplasmic injection method. *423p*, 423 promoter; *gV*, membrane-targeted *Venus* reporter (*gapVenus*); *pA*, poly adenylation sequence. The number of Venus-positive founders and lines among the total analyzed are summarized in parentheses. **(B)** The structure of *MOR42-3* gene (Hoppe *et al.*, 2006). An approximately 0.9 kb fragment (green box) consists of a non-coding exon 1 and a upstream region of TSS used as the 423 promoter.

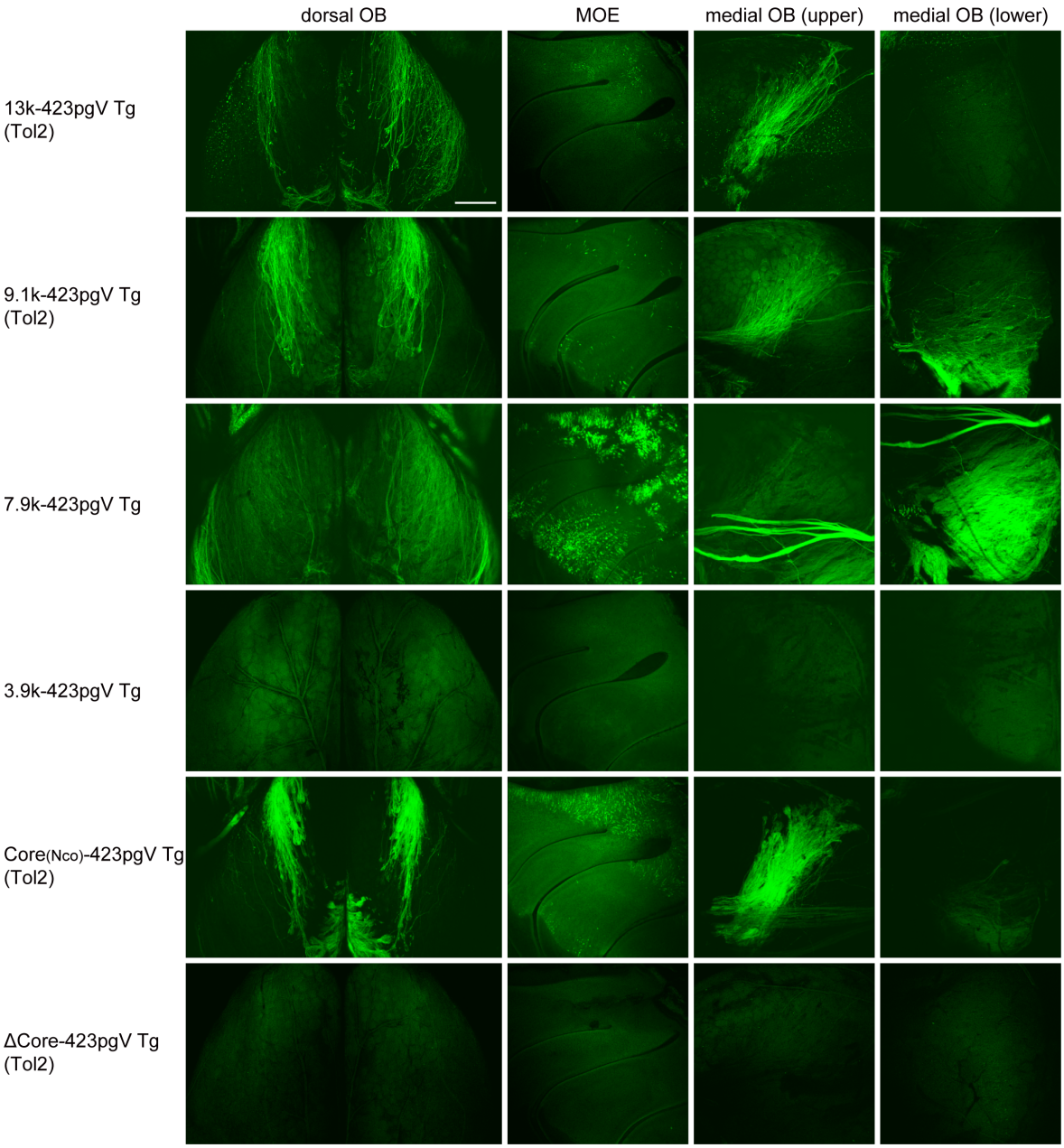

Figure 4-6. *Venus* expression patterns of reporter transgenes in transgenic mice.
Representative confocal images of the dorsal view of the OB, and the medial view of the MOE and the OB (upper or lower parts) of each transgenic mouse. The 13k, 9.1k and Core (Nco) transgenic mice showed similar *Venus* expression patterns; Venus signals were mainly observed in the most dorsal region and the dorsomedial of OB (left and right middle panels), and the dorsal MOE (middle left panel, medial view). Some signals were also observed in the ventral MOE (left middle panel) and the ventromedial region of OB (right panel) in the 9.1k or Core (Nco) transgenic mice. 7.9k transgenic mice showed a different pattern compared to other Venus-positive transgenic constructs; strong signals were observed in the dorsal and ventral MOE and the ventromedial region of OB in addition to the vomeronasal sensory nerves (medial views of the OB). Signals in the most dorsal or dorsomedial region of the OB were also observed, but the intensity is lower than lateral or ventral parts. In contrast, 3.9k and ΔCore transgenic mice did not show Venus fluorescence. Scale bar, 500 μm.

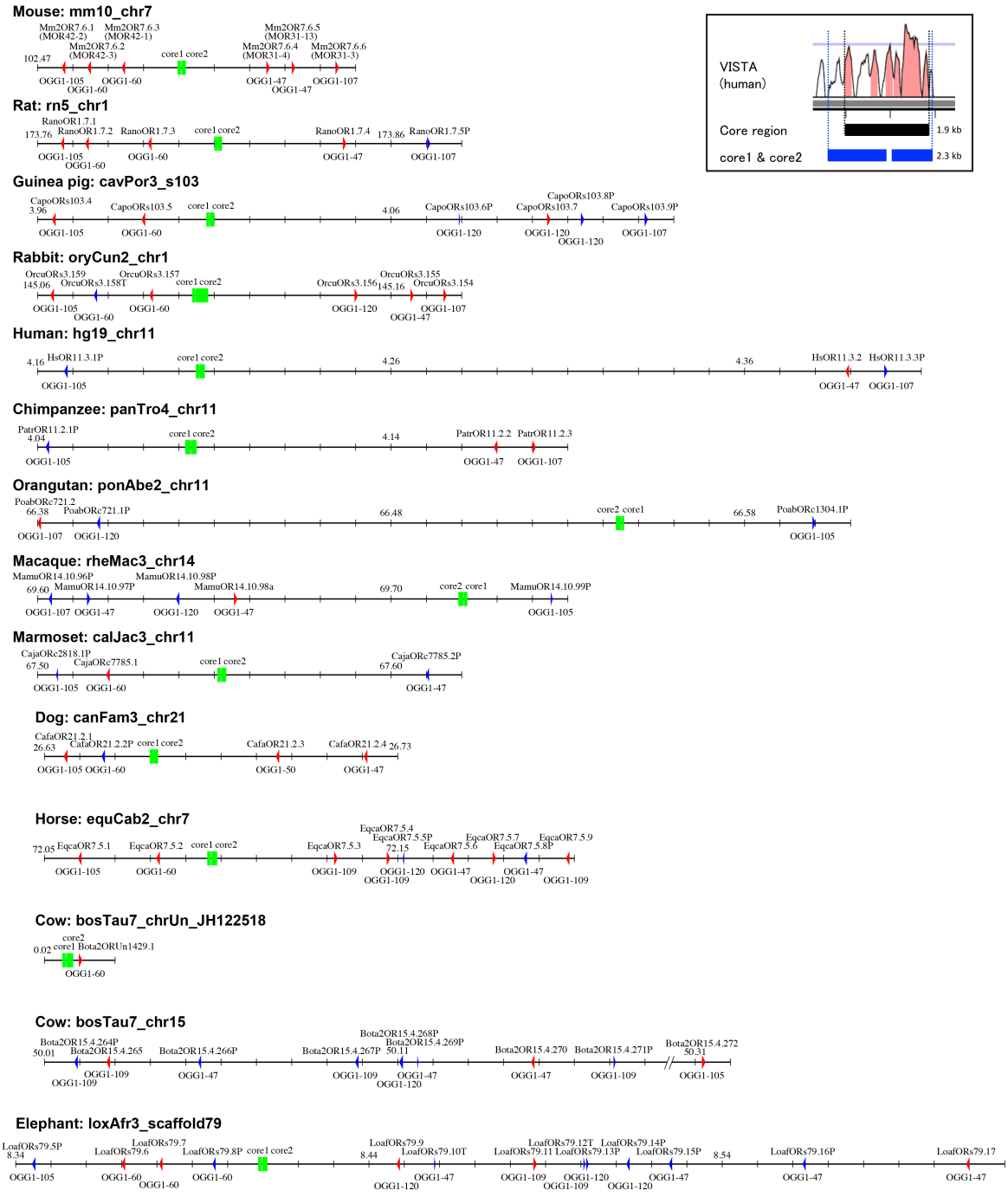

Figure 4-7. Conservation of the Core region in 13 placental mammalian species.

Schematic illustrations of class I OR genes and the Core region in 13 placental mammalian species. Two orthologous gene groups (OGG), OGG1-60 and OGG1-105, belong to group β OR gene family (Niimura, 2009). Mouse *MOR42-1* and *MOR42-3* genes are categorized into OGG1-60, whereas mouse *MOR42-2* gene is categorized into OGG1-105. The Core region (depicted as green boxes, core1 plus core2) is conserved in all tested mammalian genomes, and is localized in intergenic region between group α and β OR genes in the end of the class I OR gene cluster. In cow, class I OR locus was divided into two contigs because of assemble error. Red or blue triangles indicate intact or pseudo OR genes and their direction. The inset represents the relationship between Core region and core1 plus core2 sequence in mouse.

CHAPTER V

CONCLUSION

The BGM vector system is a unique and developing technology for propagation of very large fragments of heterologous DNA. To explore the potential of the BGM vector, I applied this system to identify the *cis*-element for mouse class I OR genes. Through this study, I have demonstrated that the BGM vector system enables the complete genetic modification of large genomic DNA fragments, including targeted insertion, deletion, inversion and fusion. Using these techniques, a 252 kb transgene was reconstructed from two BAC clones whose inserts were initially oriented in opposite direction with reference to the BAC vector sequence (Chapter II). In addition, I demonstrated that this system could be used to generate transgenic mice. Analyzing the transgenic mice, I provided an experimental evidence of a *cis*-element for a class I OR gene (Chapter III). Finally, I performed several transgenic analyses and bioinformatics analyses, and identified the first *cis*-element for class I OR gene (Chapter IV).

The BGM vector system as an alternative tool for chromosome engineering

The BGM vector has several specific features that give advantages over the BAC and YAC systems. First, the megabase-scale cloning capacity of the BGM vector is greater than that of conventional systems. The BGM vector system is capable of cloning the entire 3.5 Mb *Synechocystis* genome (Itaya *et al*., 2005), and the upper limit of the cloning size has not yet been determined. Second, cloned DNA inserts show high structural and genetic stability (Kaneko *et al*., 2005) because of direct insertion into the single circular host genome. In fact, the next-generation DNA sequencing of modified and reconstructed genomic DNA fragments confirmed the structural stability of inserts in the BGM vector even though genome inserts included many repetitive sequences and several similar class I

OR genes. However, it should be noted that recombination is constitutively active in *B. subtilis* like the YAC system; thus, unwanted rearrangements may occur. This issue can be overcome by introducing an inducible RecA system into the BGM vector. Third, various accurate modification approaches are available. The *c*I repressor cassette-mediating modification technique provides desired gene modifications without leaving any traces in the DNA, enabling the repetitive modification of BGM inserts. In addition to insertion and deletion modifications, I demonstrated the elongation of inserts via the fusion of two DNA fragments even though they were initially oriented in opposite directions in the BAC vector, thus providing a method for the construction of giant recombinant DNA fragments. Considering with maximum cloning capacity of the BGM vector of 3 Mb (Itaya *et al*., 2005), the fusion of contiguous genomic fragments can be a powerful technique to reconstruct genomic structures (Kozzamanis *et al*., 2004; Kaneko *et al*., 2009). In addition, my sequencing analyses confirmed that these targeted genetic modifications were accurate and reliable. Fourth, BGM inserts can be retrieved because a single host cell contains a single genome composed of the recombinant insert and the 4.2 Mb BGM vector. As I demonstrated, BAC clones are easily transferred, modified and reconstructed in the BGM vector. It is noteworthy that modified BGM inserts can be restored to a circular BAC form (Kaneko *et al*., 2005) (Chapter I, Figure 1-4), enabling this "shuttle genetic modification" of BAC clones to enhance the utility of BGM vector system (Figure 5-1).

Application of the BGM vector system

I have used mouse genomic DNA to demonstrate the utilities of the BGM system for genetic manipulation and transgenesis. Many BAC libraries have been already established

and/or are under construction for diverse species, including mammals, other vertebrates and plants (Web page: *Other sources of bacterial artificial chromosome libraries*). Because the BGM vector harbors BAC vector sequences for cloning BAC inserts, this system can be applied to other species. Moreover, the BGM vector can be designed to clone other library resources by introducing cloning vector sequences from other systems, e.g., YACs and human artificial chromosomes.

An interesting achievement in bioengineering field is gene assembly involved in biosynthesis of a natural carotenoid, zeaxanthin, for production in *E. coli* (Nishizaki *et al*., 2007). This novel DNA fragment assembly method, named Ordered Gene Assembly in *B. subtilis* (OGAB), produces a DNA in plasmid form via *B. subtilis* transformation (Tsuge *et al*., 2003). This method is applicable to make DNA cassettes with many relevant genes (Nishizaki *et al*., 2007; Itaya and Tsuge, 2011). Gene assembly technologies are useful for efficient productions of natural products in heterologous cells. Here, I demonstrated that the BGM inserts were able to introduce into the mouse eggs to generate transgenic mice. The combination of BGM-based gene manipulation system and mammalian transgenesis, including transgenic animals and transfected cell lines, expands the range of application of the system. Because the BGM vector system can provide large cloning capacity in size and various accurate gene manipulation approaches, the BGM vector system is an attractive cloning tool for the manipulation of large DNA fragments, such as in the functional analysis of genomic DNA, gene expression analysis, material production, recombinant genomes in synthetic biology (Gilbson *et al*., 2008; Itaya *et al*., 2008) and other life science research fields (Figure 5-1).

Function of the identified *cis*-element in class I OR genes expression

I demonstrated the existence of a *cis*-acting element of a mouse class I OR gene in the Tg-220 transgene by the BGM transgenesis. The BGM-based deletion assay and functional analysis suggest that the 3.8 kb region is a *cis*-element activating class I OR gene expression. This is supported by the large-scale comparative analysis; the Core region, which is contained the 3.8 kb region, is conserved in all tested 13 mammalian genomes. In addition, the functional analysis suggested the CTCF/CpG site alone or the combination of CTCF/CpG and CTCF also regulates class I OR gene expression.

Because some identified class II *cis*-elements, H region and P sequence, regulates several distal class II OR genes, the identified 3.8 kb region also regulates not only *MOR42-3* but also other *MOR42* genes and/or other class I OR genes. It is known that *cis*-elements for sonic hedgehog (*Shh*) gene are scattered within several hundred kb of genomic region; the most distal element locates ~840 kb upstream from the *Shh* gene (Sagai *et al*., 2005; Sagai *et al*., 2009). Thus, as previously predicted (Hirota *et al*., 2007; Rodriguez, 2007), the 3.8 kb region may regulate all class I OR genes forming ~3 Mb cluster. However, because the 3.8 kb region locates near the centromeric end of the class I OR gene cluster (Figure 1-7), the distance from the 3.8 kb element to the other end class I OR gene (*MOR36-1*) is approximately 3 Mb long. Thus, I cannot exclude possibility of existence of other *cis*-elements for class I OR genes.

The molecular mechanisms underlying single OR genes expression remain unclear. Recently, the nuclear architecture and chromatin status (euchromatin/heterochromatin) of OR gene loci in OSNs are involved in proper expression of OR genes (Magklara *et al*., 2011; Clowney *et al*., 2012; Lyons *et al*., 2013). The 3.8 kb region is flanked by two CTCF

binding sites. CTCF is known as a regulator of the chromatin structure including DNA loops, which regulates enhancer activity (Phillips and Corces, 2009). Therefore, the 3.8 kb region may be regulated by CTCF binding following a DNA loop formation (Figure 5-2). To examine these possibilities, further experiments are required such as transgenic analysis of other class I OR gene-targeted transgene like Tg-220 and loss-of-function analyses of the non-coding regions of the Core and the CTCF/CpG and CTCF.

Physiological function of class I OR genes

The class I OR subfamily is maintained in vertebrates during evolution, suggesting a physiological importance of the class I OR family. In addition, the group α class I OR genes present only in mammals, whereas the group β class I OR genes are conserved not only in mammals but also amphibians, reptiles and fishes (Niimura 2009; Wang *et al*., 2013). The physiological functions of class I OR family are unclear, although several ligands have been reported (Malnic *et al*., 1999; Abaffy *et al*., 2006; Saito *et al*., 2009). One report showed that a part of the dorsal OB (including class I OR domain) was activated by 2-methylbutyric acid and pentanal, which are a pungent odorants of spoiled foods, and suggested the dorsal OB mediates innate fear responses (Kobayakawa *et al*., 2007). It is conceivable that the class I OR family is involved in some innate behaviors. Because the identified *cis*-element can drive transgene expression in the class I OR-expressing OSNs, this element can be used to label or ablate class I OR-expressing OSNs in future physiological experiments.

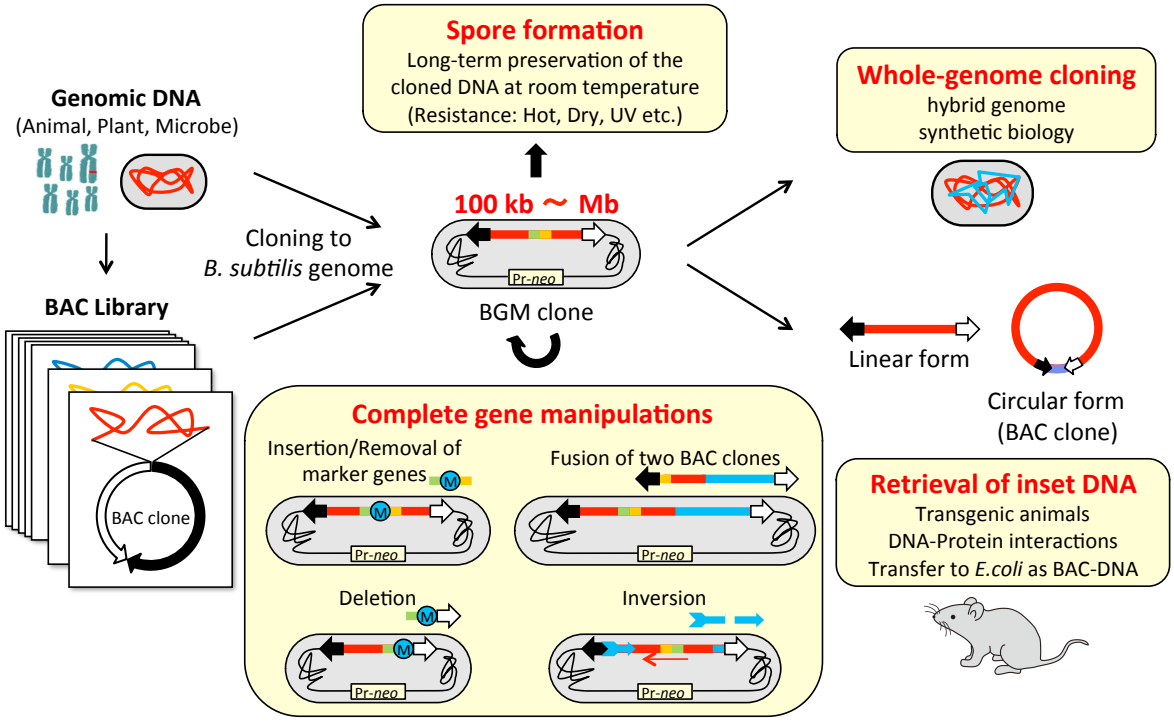

Figure 5-1. Manipulations and applications of the BGM vector system.
The BGM vector can use both natural genomic DNA and BAC library clones. BGM inserts can be modified by several manipulation techniques (Chapter II), and are stably stored in spores (Kaneko *et al.*, 2005). BGM inserts can be retrieved as linear form (Chapter III) or circular form (Kaneko *et al.*, 2005). The BGM vector system and its application to transgenesis offer a new genetic approach for not only systems and synthetic biology but also other life science research fields.

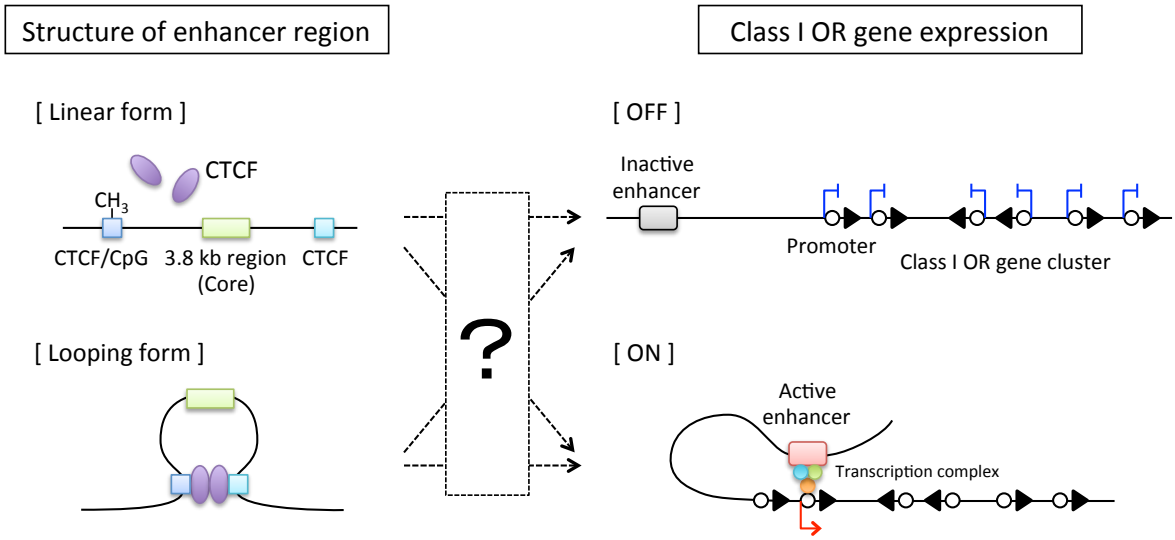

Figure 5-2. A putative regulation mechanism for class I OR gene expression.
Local structure of the identified enhancer region is regulated by CTCF binding status. When CTCF not bind to CTCF sites (CpG island is methylated), the region is linear form. In contrast, binding of CTCF into the sites forms a DNA loop structure. These two enhancer forms define the activity to activate class I OR genes expression. The active enhancer chooses one OR gene from the class I OR gene cluster and activates the transcription.

REFERENCES

Abaffy T, Matsunami H, Luetje CW: Functional analysis of a mammalian odorant receptor subfamily. *J Neurochem* 2006, 97:1506-1518.

Bird AW, Erler A, Fu J, Hériché JK, Maresca M, Zhang Y, Hyman AA, Stewart AF: High-efficiency counterselection recombineering for site-directed mutagenesis in bacterial artificial chromosomes. *Nat Methods* 2011, 9:103-109.

Bozza T, Feinstein P, Zheng C, Mombaerts P: Odorant receptor expression defines functional units in the mouse olfactory system. *J Neurosci* 2002, 22:3033-3043.

Bozza T, Vassalli A, Fuss S, Zhang JJ, Weiland B, Pacifico R, Feinstein P, Mombaerts P: Mapping of class I and class II odorant receptors to glomerular domains by two distinct types of olfactory sensory neurons in the mouse. *Neuron* 2009, 61:220-233.

Burke DT, Carle GF, Olson MV: Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* 1987, 236:806-812.

Chaumeil J, Skok JA: The role of CTCF in regulating V(D)J recombination. *Curr Opin Immunol* 2012, 24:153-159.

Chen I, Christie PJ, Dubnau D: The ins and outs of DNA transfer in bacteria. *Science* 2005, 310:1456-1460.

Chess A, Simon I, Cedar H, Axel R: Allelic inactivation regulates olfactory receptor gene expression. *Cell* 1994, 78:823-834.

Chung JH, Whiteley M, Felsenfeld G: A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila. *Cell* 1993, 74:505-514.

Clowney EJ, Magklara A, Colquitt BM, Pathak N, Lane RP, Lomvardas S: High-throughput mapping of the promoters of the mouse olfactory receptor genes reveals a new type of mammalian promoter and provides insight into olfactory receptor gene regulation. *Genome Res* 2011, 21:1249-1259.

Clowney EJ, LeGros MA, Mosley CP, Clowney FG, Markenskoff-Papadimitriou EC, Myllys M, Barnea G, Larabell CA, Lomvardas S: Nuclear aggregation of olfactory receptor genes governs their monogenic expression. *Cell* 2012, 151:724-737.

Copeland NG, Jenkins NA, Court DL: Recombineering: a powerful new tool for mouse functional genomics. *Nat Rev Genet* 2001, 2:769-779.

Dubnau D: DNA uptake in bacteria. *Annu Rev Microbiol* 1999, 53:217-244.

Enomoto T, Ohmoto M, Iwata T, Uno A, Saitou M, Yamaguchi T, Kominami R, Matsumoto I, Hirota J: *Bcl11b/Ctip2* controls the differentiation of vomeronasal sensory neurons in mice. *J Neurosci* 2011, 31:10159-10173.

Feinstein P, Mombaerts P: A contextual model for axonal sorting into glomeruli in the mouse olfactory system. *Cell* 2004, 117:817-831.

Fuss SH, Omura M, Mombaerts P: Local and *cis* effects of the H element on expression of odorant receptor genes in mouse. *Cell* 2007, 130:373-384.

Garrels W, Ivics Z, Kues WA: Precision genetic engineering in large mammals. *Trends Biotechnol* 2012, 30:386-393.

Gaszner M, Felsenfeld G: Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet* 2006, 7:703-713.

Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA, Baden-Tillson H, Zaveri J, Stockwell TB, Brownley A, Thomas DW, Algire MA *et al.*: Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* 2008, 319:1215-1220.

Giraldo P, Montoliu L: Size matters: use of YACs, BACs and PACs in transgenic animals. *Transgenic Res* 2001, 10:83-103.

Glusman G, Yanai I, Rubin I, Lancet D: The complete human olfactory subgenome. *Genome Res* 2001, 11:685-702.

Gong S, Yang XW, Li C, Heintz N: Highly efficient modification of bacterial artificial chromosomes (BACs) using novel shuttle vectors containing the R6Kgamma origin of replication. *Genome Res*.2002, 12:1992-1998.

Green ED, Riethman HC, Dutchik JE, Olson MV: Detection and characterization of chimeric yeast artificial-chromosome clones. *Genomics* 1991, 11:658-669.

Harrington JJ, Van Bokkelen G, Mays RW, Gustashaw K, Willard HF: Formation of de novo centromeres and construction of first-generation human artificial microchromosomes. *Nat Genet* 1997, 15:345-355.

Hirayama T, Tarusawa E, Yoshimura Y, Galjart N, Yagi T: CTCF is required for neural development and stochastic expression of clustered Pcdh genes in neurons. *Cell Rep* 2012, 2:345-357.

Hirota J, Mombaerts P: The LIM-homeodomain protein Lhx2 is required for complete development of mouse olfactory sensory neurons. *Proc Natl Acad Sci U S A* 2004, 101:8751-8755.

Hirota J, Omura M, Mombaerts P: Differential impact of Lhx2 deficiency on expression of class I and class II odorant receptor genes in mouse. *Mol Cell Neurosci* 2007, 34:679-688.

Itaya M: Integration of repeated sequences (pBR322) in the *Bacillus subtilis* 168 chromosome without affecting the genome structure. *Mol Gen Genet* 1993, 241:287-297.

Itaya M: Toward a bacterial genome technology: integration of the *Escherichia coli* prophage lambda genome into the *Bacillus subtilis* 168 chromosome. *Mol Gen Genet* 1995, 248:9-16.

Itaya M, Tanaka T: Experimental surgery to create subgenomes of *Bacillus subtilis* 168. *Proc Natl Acad Sci U S A* 1997, 94:5378-5382.

Itaya M: Effective cloning of unmarked DNA fragments in the *Bucillus subtilis* 168 genome. *Biosci Biotechnol Biochem* 1999, 63:602-604.

Itaya M, Nagata T, Shiroishi T, Fujita K, Tsuge K: Efficient cloning and engineering of giant DNAs in a novel *Bacillus subtilis* genome vector. *J Biochem* 2000, 128:869-875.

Itaya M, Tsuge K, Koizumi M, Fujita K: Combining two genomes in one cell: stable cloning of the *Synechocystis* PCC6803 genome in the *Bacillus subtilis* 168 genome. *Proc Natl Acad Sci U S A* 2005, 102:15971-15976.

Itaya M, Fujita K, Kuroki A, Tsuge K: Bottom-up genome assembly using the *Bacillus subtilis* genome vector. *Nat Methods* 2008, 5:41-43.

Itaya M, Tsuge K: Construction and manipulation of giant DNA by a genome vector. *Methods Enzymol* 2011, 498:427-447.

Jakobovits A, Moore AL, Green LL, Vergara GJ, Maynard-Currie CE, Austin HA, Klapholz S: Germ-line transmission and expression of a human-derived yeast artificial chromosome. *Nature* 1993, 362:255-258.

Kaneko S, Tsuge K, Takeuchi T, Itaya M: Conversion of sub-megasized DNA to desired structures using a novel *Bacillus subtilis* genome vector. *Nucleic Acids Res* 2003, 31:e112.

Kaneko S, Akioka M, Tsuge K, Itaya M: DNA shuttling between plasmid vectors and a genome vector: systematic conversion and preservation of DNA libraries using the *Bacillus subtilis* genome (BGM) vector. *J Mol Biol* 2005, 349:1036-1044.

Kaneko S, Takeuchi T, Itaya M: Genetic connection of two contiguous bacterial artificial chromosomes using homologous recombination in *Bacillus subtilis* genome vector. *J Biotech* 2009, 139:211-213.

Kawakami K, Takeda H, Kawakami N, Kobayashi M, Matsuda N, Mishina M: A transposon-mediated gene trap approach identifies developmentally regulated genes in zebrafish. *Dev Cell* 2004, 7:133-144.

Khan M, Vaes E, Mombaerts P: Regulation of the probability of mouse odorant receptor gene choice. *Cell* 2011, 147:907-921.

Kline B: A review of mini-F plasmid maintenance. *Plasmid* 1985, 14:1-16.

Kobayakawa K, Kobayakawa R, Matsumoto H, Oka Y, Imai T, Ikawa M, Okabe M, Ikeda T, Itohara S, Kikusui T *et al*.: Innate versus learned odour processing in the mouse olfactory bulb. *Nature* 2007, 450:503-508.

Kotzamanis G, Huxley C: Recombining overlapping BACs into a single larger BAC. *BMC Biotech* 2004, 4:1.

Kouprina N, Eldarov M, Moyzis R, Resnick M, Larionov V: A model system to assess the integrity of mammalian YACs during transformation and propagation in yeast. *Genomics* 1994, 21:7-17.

Kunst F, Ogasawara N, Moszer I, Albertini AM, Alloni G, Azevedo V, Bertero MG, Bessières P, Bolotin A, Borchert S *et al*.: The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* 1997, 390:249-256.

Kuroiwa Y, Kasinathan P, Choi YJ, Naeem R, Tomizuka K, Sullivan EJ, Knott JG, Duteau A, Goldsby RA, Osborne BA *et al*.: Cloned transchromosomic calves producing human immunoglobulin. *Nat Biotechnol* 2002, 20:889-894.

Kuroki A, Toda T, Matsui K, Uotsu-Tomita R, Tomita M, Itaya M: Reshuffling of the *Bacillus subtilis* 168 genome by multifold inversion. *J Biochem* 2008, 143:97-105.

Lamb BT, Bardel KA, Kulnane LS, Anderson JJ, Holtz G, Wagner SL, Sisodia SS, Hoeger EJ: Amyloid production and deposition in mutant amyloid precursor protein and presenilin-1 yeast artificial chromosome transgenic mice. *Nat Neurosci* 1999, 2:695-697.

Lee EC, Yu D, Martinez de Velasco J, Tessarollo L, Swing DA, Court DL, Jenkins NA, Copeland NG: A highly efficient *Escherichia coli*-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics* 2001, 73:56-65.

Lee JT, Jaenisch R: A method for high efficiency YAC lipofection into murine embryonic stem cells. *Nucleic Acids Res* 1996, 24:5054-5055.

Lewcock JW, Reed RR: A feedback mechanism regulates monoallelic odorant receptor expression. *Proc Natl Acad Sci U S A* 2004, 101:1069-1074.

Li L, Blankenstein T: Generation of transgenic mice with megabase-sized human yeast artificial chromosomes by yeast spheroplast-embryonic stem cell fusion. *Nat Protoc* 2013, 8:1567-1582.

Lyons DB, Allen WE, Goh T, Tsai L, Barnea G, Lomvardas S: An epigenetic trap stabilizes singular olfactory receptor expression. *Cell* 2013, 154:325-336.

Magklara A, Yen A, Colquitt BM, Clowney EJ, Allen W, Markenscoff-Papadimitriou E, Evans ZA, Kheradpour P, Mountoufaris G, Carey C *et al*.: An epigenetic signature for monoallelic olfactory receptor expression. *Cell* 2011, 145:555-570.

Malnic B, Hirono J, Sato T, Buck LB: Combinatorial receptor codes for odors. *Cell* 1999, 96:713-723.

McIntyre JC, Bose SC, Stromberg AJ, McClintock TS: Emx2 stimulates odorant receptor gene expression. *Chem Senses* 2008, 33:825-837.

Mendez MJ, Green LL, Corvalan JR, Jia XC, Maynard-Currie CE, Yang XD, Gallo ML, Louie DM, Lee DV, Erickson KL *et al*.: Functional transplant of megabase human immunoglobulin loci recapitulates human antibody response in mice. *Nat Genet* 1997, 15:146-156.

Miyamichi K, Serizawa S, Kimura HM, Sakano H: Continuous and overlapping expression domains of odorant receptor genes in the olfactory epithelium determine the dorsal/ventral positioning of glomeruli in the olfactory bulb. *J Neurosci* 2005, 25:3586-3592.

Mombaerts P, Wang F, Dulac C, Chao SK, Nemes A, Mendelsohn M, Edmondson J, Axel R: Visualizing an olfactory sensory map. *Cell* 1996, 87:675-686.

Mombaerts P: Genes and ligands for odorant, vomeronasal and taste receptors. *Nat Rev Neurosci* 2004, 5:263-278.

Mori K, Takahashi YK, Igarashi KM, Yamaguchi M: Maps of odorant molecular features in the Mammalian olfactory bulb. *Physiol Rev* 2006, 86:409-433.

Muyrers JP, Zhang Y, Testa G, Stewart AF: Rapid modification of bacterial artificial chromosomes by ET-recombination. *Nucleic Acids Res* 1999, 27:1555-1557.

Nei M, Niimura Y, Nozawa M: The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. *Nat Rev Genet* 2008, 9:951-963.

Niimura Y, Nei M: Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. *Proc Natl Acad Sci U S A* 2005, 102:6039-6044.

Niimura Y: On the origin and evolution of vertebrate olfactory receptor genes: comparative genome analysis among 23 chordate species. *Genome Biol Evol* 2009, 1:34-44.

Niimura Y, Matsui A, Touhara K: Extreme expansion of the olfactory receptor gene repertoire in African elephants and evolutionary dynamics of orthologous gene groups in 13 placental mammals. *Genome Res* 2014 24:1485-1496.

Nishizaki T, Tsuge K, Itaya M, Doi N, Yanagawa H: Metabolic engineering of carotenoid biosynthesis in *Escherichia coli* by ordered gene assembly in *Bacillus subtilis*. *Appl Environ Microbiol* 2007, 73:1355-1361.

Nishizumi H, Kumasaka K, Inoue N, Nakashima A, Sakano H: Deletion of the core-*H* region in mice abolishes the expression of three proximal odorant receptor genes in *cis*. *Proc Natl Acad Sci U S A* 2007, 104:20067-20072.

Other Sources of Bacterial Artificial Chromosome (BAC) Libraries [http://www.genome.gov/11008350]

Osoegawa K, Tateno M, Woon PY, Frengen E, Mammoser AG, Catanese JJ, Hayashizaki Y, de Jong PJ: Bacterial artificial chromosome libraries for mouse sequencing and functional analysis. *Genome Res* 2000, 10:116-128.

Peterson KR: Preparation of intact yeast artificial chromosome DNA for transgenesis of mice. *Nat Protoc* 2007, 2:3009-3015.

Phillips JE, Corces VG: CTCF: master weaver of the genome. *Cell* 2009, 137:1194-1211.

Plessy C, Pascarella G, Bertin N, Akalin A, Carrieri C, Vassalli A, Lazarevic D, Severin J, Vlachouli C, Simone R *et al*.: Promoter architecture of mouse olfactory receptor genes. *Genome Res* 2012, 22:486-497.

Ressler KJ, Sullivan SL, Buck LB: Information coding in the olfactory system: evidence for a stereotyped and highly organized epitope map in the olfactory bulb. *Cell* 1994, 79:1245-1255.

Rodriguez I: Odorant and pheromone receptor gene regulation in vertebrates. *Curr Opin Genet Dev* 2007 17:465-470.

Rothman A, Feinstein P, Hirota J, Mombaerts P: The promoter of the mouse odorant receptor gene *M71*. *Mol Cell Neurosci* 2005, 28:535-546.

Saito H, Chi Q, Zhuang H, Matsunami H, Mainland JD: Odor coding by a Mammalian receptor repertoire. *Sci Signal* 2009, 2:ra9.

Sagai T, Hosoya M, Mizushina Y, Tamura M, Shiroishi T: Elimination of a long-range *cis*-regulatory module causes complete loss of limb-specific *Shh* expression and truncation of the mouse limb. *Development* 2005, 132:797-803.

Sagai T, Amano T, Tamura M, Mizushina Y, Sumiyama K, Shiroishi T: A cluster of three long-range enhancers directs regional *Shh* expression in the epithelial linings. *Development* 2009, 136:1665-1674.

Schedl A, Montoliu L, Kelsey G, Schutz G: A yeast artificial chromosome covering the tyrosinase gene confers copy number-dependent expression in transgenic mice. *Nature* 1993, 362:258-261.

Serizawa S, Ishii T, Nakatani H, Tsuboi A, Nagawa F, Asano M, Sudo K, Sakagami J, Sakano H, Ijiri T *et al*.: Mutually exclusive expression of odorant receptor transgenes. *Nat Neurosci* 2000, 3:687-693.

Serizawa S, Miyamichi K, Nakatani H, Suzuki M, Saito M, Yoshihara Y, Sakano H: Negative feedback regulation ensures the one receptor-one olfactory neuron rule in mouse. *Science* 2003, 302:2088-2094.

Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, Simon M: Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc Natl Acad Sci U S A* 1992, 89:8794-8797.

Sumiyama K, Kawakami K, Yagita K: A simple and highly efficient transgenesis method in mice with the *Tol2* transposon system and cytoplasmic microinjection. *Genomics* 2010, 95:306-311.

Sternberg N: Bacteriophage P1 cloning system for the isolation, amplification, and recovery of DNA fragments as large as 100 kilobase pairs. *Proc Natl Acad Sci U S A* 1990, 87:103-107.

Strauss WM, Dausman J, Beard C, Johnson C, Lawrence JB, Jaenisch R: Germ line transmission of a yeast artificial chromosome spanning the murine α_1(I) collagen locus. *Science* 1993, 259:1904-1907.

Suster ML, Sumiyama K, Kawakami K: Transposon-mediated BAC transgenesis in zebrafish and mice. *BMC Genomics* 2009, 10:477.

Treloar HB, Feinstein P, Mombaerts P, Greer CA: Specificity of glomerular targeting by olfactory sensory axons. *J Neurosci* 2002, 22:2469-2477.

Tsuboi A, Miyazaki T, Imai T, Sakano H: Olfactory sensory neurons expressing class I odorant receptors converge their axons on an antero-dorsal domain of the olfactory bulb in the mouse. *Eur J Neurosci* 2006, 23:1436-1444.

Tsuge K, Itaya M: Recombinational transfer of 100-kilobase genomic DNA to plasmid in *Bacillus subtilis* 168. *J Bacteriol* 2001, 183:5453-5458.

Tsuge K, Matsui K, Itaya M: One step assembly of multiple DNA fragments with a designed order and orientation in *Bacillus subtilis* plasmid. *Nucleic Acids Res* 2003, 31:e133.

UCSC Genome Browser: [http://genome.ucsc.edu/cgi-bin/hgGateway]

Uozumi T, Hoshino T, Miwa K, Horinouchi S, Beppu T, and Arima K: Restriction and modification in *Bacillus* species: Genetic transformation of bacteria with DNA from different species, part I. *Mol Gen Genet* 1977, 152:65–69.

Urasaki A, Morvan G, Kawakami K: Functional dissection of the *Tol2* transposable element identified the minimal *cis*-sequence and a highly repetitive sequence in the subterminal region essential for transposition. *Genetics* 2006, 174:639-649.

Vassalli A, Rothman A, Feinstein P, Zapotocky M, Mombaerts P: Minigenes impart odorant receptor-specific axon guidance in the olfactory bulb. *Neuron* 2002, 35:681-696.

Vassalli A, Feinstein P, Mombaerts P: Homeodomain binding motifs modulate the probability of odorant receptor gene choice in transgenic mice. *Mol Cell Neurosci* 2011, 46:381-396.

Vassar R, Chao SK, Sitcheran R, Nuñez JM, Vosshall LB, Axel R: Topographic organization of sensory projections to the olfactory bulb. *Cell* 1994, 79:981-991.

VISTA: [http://genome.lbl.gov/vista/index.shtml]

Wang Z, Pascual-Anaya J, Zadissa A, Li W, Niimura Y, Huang Z, Li C, White S, Xiong Z, Fang D *et al.*: The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan. *Nat Genet* 2013, 45:701-706.

Yang XW, Model P, Heintz N: Homologous recombination based modification in *Escherichia coli* and germline transmission in transgenic mice of a bacterial artificial chromosome. *Nat Biotech* 1997, 15:859-865.

Zhang X, Firestein S: The olfactory receptor gene superfamily of the mouse. *Nat Neurosci* 2002, 5:124-133.

Zhang X, Zhang X, Firestein S: Comparative genomics of odorant and pheromone receptor genes in rodents. *Genomics* 2007, 89:441-450.

Zhang Y, Buchholz F, Muyrers JP, Stewart AF: A new logic for DNA engineering using recombination in *Escherichia coli*. *Nat Genet* 1998, 20:123-128.

Zhang YQ, Breer H, Strotmann J: Promotor elements governing the clustered expression pattern of odorant receptor genes. *Mol Cell Neurosci* 2007, 36:95-107.