Establishment of a practical gene knock-in system and its application in medaka

Yu Murakami

2020

Contents

Chapter1

General introduction

Genetics was born as a novel scientific discipline in the mid-19th century to study the genetic phenomenon transmitting parental traits to their offspring. Genetical research has been continuously conducted until now, and accelerated development of other biological studies such as embryology, physiology and thremmatology. Genetics itself can be progressed owing to development of various analytical methods for understanding the function of individual genes. These methods in genetics are broadly divided into two categories. One is forward genetics that identifies the genetic basis responsible for a phenotype, whereas the other is reverse genetics that analyzes the phonotypic variation caused by the modification of a targeted DNA sequence.

Transgenic (Tg) technology emerged as one of powerful methods for reverse genetics in 1970s. Jaenisch R and Mintz B successfully created the first transgenic mice by Tg technology, showing its usefulness for evaluating the specific gene function in *in vivo* [1]. Moreover, Tg technology enabled to add useful genetic traits to a wide range of organisms including livestocks, crops and seafoods [2,3,4]. In several commercially important fish species, the successful transductions of desirable traits, such as rapid growth and disease resistance, have been reported [5,6,7]. However, Tg technology tended to induce an ectopic expression of transgenes because of random and/or multiple integration of transgenes into genome [8]. There were some problems in strictly regulating expression pattern of transgenes in those days. In 2001, Mario. R Capecchi succeeded in overcoming these shortcomings by establishing the homology-directed

repair (HDR)-mediated gene targeting method allowing to insert a single copy of a transgene into a specific site on genome [9]. However, the gene targeting method was applicable to only a few species such as mice because of the difficulties in establishment of embryonic stem (ES) cells essential for this method. Thus, development of a simple and efficient technology that can precisely integrate foreign genes into any genomic target site in various species was desired.

Over the past few decades, genome editing using targetable nucleases, including systems utilizing clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) nucleases, transcription activator-like effector nucleases (TALENs), and zinc-finger nucleases (ZFNs), have been developed as powerful methods to meet the demands [10]. These nucleases can induce DNA double strand breaks (DSBs) on any specific genomic loci, which are cell-autonomously repaired by either of two DNA repair pathways: non-homologous end-joining (NHEJ) or HDR. NHEJ yields several nucleotide deletions, insertions, or substitutions at the genomic target site resulting in a targeted gene disruption by flame-shift mutations. In contrast, HDR can repair DSBs with DNA templates containing homologous sequences allowing to the precise insertion of the desired DNA sequence (Fig. 1-1). So far, a number of previous studies using the genome editing technologies have reported the successful gene modification mediated by these DNA repair pathways in a wide range of species [11,12,13]. However, the targeted gene integration mediated by HDR is less efficient than the targeted gene disruption mediated by NHEJ due to the lower activity of HDR comparing to NHEJ [14]. Indeed, at the beginning of my Ph.D. study, little was known about an efficient protocol for the HDR-mediated gene integration in teleost fish except zebrafish. The targeted gene integration mediated by HDR is desirable for more precise and complex genome manipulations than

the targeted gene disruption mediated by NHEJ. Thus, it was imperative to establish a novel system for improving the efficiency of targeted gene integration events mediated by HDR, and show its practicality with a view to various fish species.

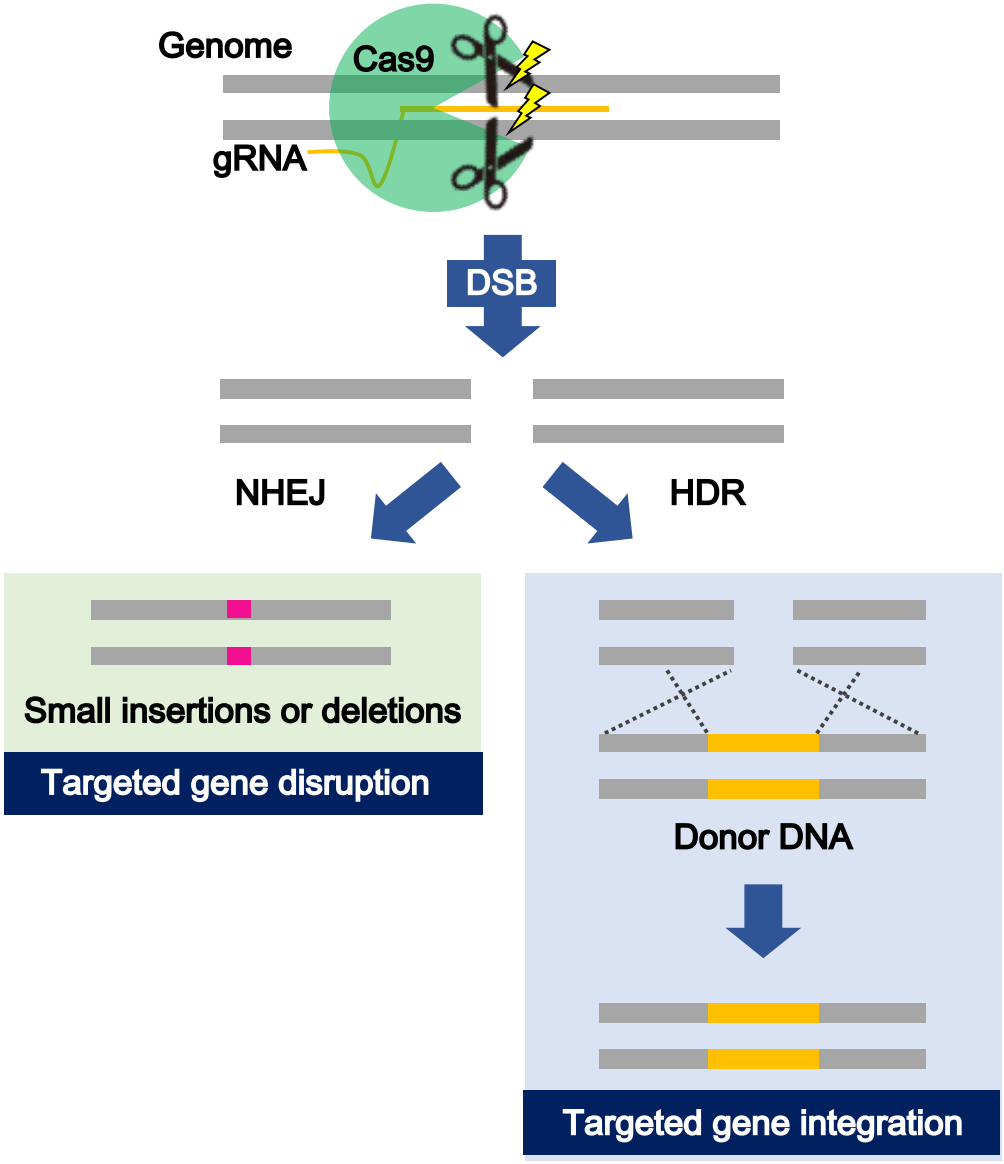


Fig. 1-1. Schematic illustration of CRISPR-Cas9-induced DSB and its sequential DNA repair pathways. NHEJ is an error-prone repair pathway leading to small insertions or deletions (indels) of several nucleotides in the targeted gene, while HDR requires a donor DNA with homologous sequences which is used as a template for high fidelity DSB repair.

Medaka (*Oryzias latipes*) is a small freshwater teleost species, which serves as an excellent vertebrate model for genetics because of their ease of breeding and unique genetic resources, such as spontaneous mutant collections, highly polymorphic inbred strains, and relative species with a number of unique features [15,16]. In medaka, a large number of transgenic strains has been available for molecular genetic analysis [16,17]. However, targeted genome manipulation technology to engineer the endogenous genes, such as genetic tagging by reporter genes and site-specific introduction of single nucleotide polymorphisms, has been hindered because of the difficulty in establishing germline-competent embryonic stem cells.

In chapter2, I aimed to establish an efficient system for HDR-mediated targeted gene integration in medaka by optimizing each component of donor plasmids such as length of homologous sequences or bait sequences. Subsequently, I showed that Cas9 nickase, which is a variant of Cas9, enables to integrate exogenous genes into lethal genes in chapter3. Finally, in chapter4, I successfully established a novel system for transporting foreign proteins into eggs in medaka by using the conventional transgenic technology and the established gene knock-in method in this study.

Chapter1 References

1. Jaenisch, R. & Mintz, B. (1974) Simian virus 40 DNA sequences in DNA of healthy adult mice derived from preimplantation blastocysts injected with viral DNA. Proc Natl Acad Sci U S A. 71: 1250-4.
2. Clark, J. & Whitelaw, B. (2003) A future for transgenic livestock. Nat Rev Genet. 4: 825-33.

3. Raman, R. (2017) The impact of Genetically Modified (GM) crops in modern agriculture: A review. GM Crops Food. 8: 195-208.
4. Chen, T.T., Vrolijk, N.H., Lu, J.K., Lin, C.M., Reimschuessel, R. & Dunham, R.A. (1996) Transgenic fish and its application in basic and applied research. Biotechnol Annu Rev. 2: 205-36.
5. Maclean, N., Penman, D. & Zhu, Z. (1987) Introduction of Novel Genes into Fish. Bio/Technology. 5: 257-261.
6. Du, S.J., Gong, Z., Fletcher, G.L., Shears, M.A., King, M.J., Idler, D.R. & Hew, C.L. (1992) Growth Enhancement in Transgenic Atlantic Salmon by the Use of an "All Fish" Chimeric Growth Hormone Gene Construct. Bio/Technology. 10: 176-181.
7. Chiou, P.P., Chen, M.J., Lin, C.M., Khoo, J., Larson, J., Holt, R., Leong, J.A., Thorgarrd, G. & Chen, T.T. (2014) Production of homozygous transgenic rainbow trout with enhanced disease resistance. Mar Biotechnol (NY), 16: 299-308.
8. Kurauchi, K., Hirata, T. & Kinoshita, M. (2008) Characteristics of ChgH–GFP transgenic medaka lines, an in vivo estrogenic compound detection system. Marine Pollution Bulletin. 57: 441-444.
9. Capecchi, M.R. (2001). Generating mice with targeted mutations. Nat Med, 7: 1086-90.
10. Gaj, T., Gersbach, C.A. & Barbas, C.F. (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol. 31: 397-405.
11. Belhaj, K., Chaparro-Garcia, A., Kamoun, S., Patron, N.J. & Nekrasov, V. (2015) Editing plant genomes with CRISPR/Cas9. Current Opinion in Biotechnology. 32: 76-84.
12. Gratacap, R.L., Wargelius, A., Edvardsen, R.B. & Houston, R.D. (2019) Potential of

Genome Editing to Improve Aquaculture Breeding and Production. Trends in Genetics. 35: 672-684.

13. Van Eenennaam, A.L. (2019) Application of genome editing in farm animals: cattle. Transgenic Res. 28: 93-100.
14. Mao, Z., Bozzella, M., Seluanov, A. & Gorbunova, V. (2008) DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells. Cell Cycle. 7: 2902-6.
15. Takeda H, Shimada A. The art of medaka genetics and genomics: what makes them so unique? (2010) Annu Rev Genet. 44:217–41.
16. Kirchmaier S, Naruse K, Wittbrodt J, Loosli F. (2015) The genomic and genetic toolbox of the teleost medaka (*Oryzias latipes*). Genetics. 199:905–18.
17. Kinoshita M, Murata K, Naruse K, Tanaka M, editors. (2009) Medaka: Biology, Management, and Experimental Protocols. Ames, Iowa, USA: Wiley-Blackwell

Chapter2

An efficient gene knock-in system mediated by homologous recombination

2.1 Introduction

Genome editing using targetable nucleases, including clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) system, transcription activator-like effector nucleases (TALENs), and zinc-finger nucleases (ZFNs), has been established as a powerful method for reverse genetics in a wide range of organisms [1]. These nucleases can induce DNA double-strand-breaks (DSBs) at any genomic target loci, which allows for various types of targeted genome modifications via DNA DSBs repair systems, such as targeted gene disruptions by small insertions and deletions (indels) via non-homologous end-joining (NHEJ) and targeted gene integration by homology-directed repair (HDR) [2]. My research group previously demonstrated targeted gene disruptions mediated by NHEJ, so that efficient methods for targeted mutagenesis using targetable nuclease systems had been established in medaka [3–5]. On the other hand, although targeted gene integration mediated by HDR is desired for more precise and complex genome manipulations, there is only a few reports on the HDR-mediated gene integration in medaka [6]. Therefore, the detailed knowledge is required to establish efficient protocols for the technology.

It has been known that the length of homologous sequences plays an important role in determining which pathways are used for the repair of DNA DSBs [7]. Relatively

long homology arms (0.5–1 kb) can induce HDR that have been commonly used for targeted gene integration by the genome editing technology [2]. On the other hand, short homology arms (2–25 bp) can induce microhomology-mediated end joining (MMEJ), which were recently known as a DSB repair pathway for a high efficient gene knockin method by the genome editing technology [8,9]. Recent studies demonstrated that both pathways can mediate the targeted gene integration in zebrafish embryos [10–12]. However, no studies have directly compared the integration efficiencies mediated by these two pathways in fish embryos, so that the effects of length of homology arms on the integration efficiencies have remained unclear.

Furthermore, previous studies demonstrated that simultaneous cleavage of a circular donor plasmid and the targeted genomic locus by the targetable nucleases can enhance the HDR-mediated gene integration efficiencies in zebrafish and sea urchin embryos [10,13]. For efficient methods of the targeted gene integration by CRISPR/Cas system, guide RNAs (gRNAs) and their targets, called “bait” sequences, were designed for the cleavage of circular donor plasmids. Gbait was a bait sequence designed on the coding sequence of *EGFP* gene and has been used in some zebrafish studies because of its high genome-editing activity and less off-target effects. In addition, PITCh gRNAs were designed for targeted integration by PITCh system with minimized off-target effects in various mammalian genomes [9]. However, there has been no comparative studies on the effect of these sequences on the targeted gene integrate efficiency *in vivo*.

In this study, I aimed to establish an efficient system for targeted gene integration in medaka. Firstly, I examined the effects of length of homology arms on the integration efficiencies at a genomic locus in medaka embryos. Next, I developed novel bait sequences with less off-target effects in fish genomes and validated the effects on

integration efficiencies by comparing with previously reported bait sequences, the Gbait and the PITCh gRNAs. Finally, I demonstrated that gene knockin strains lebeled different color fluorescent protein genes at the targeting locus can be generated by a method established in this study and this labeling system is helpful in maintaining mutant strains without PCR genotyping.

2.2 Materials and Methods

Fish

The cab (closed colony) of medaka was used in this study. The fish were kept under a 14/10-h light/dark cycle at 26 °C. The fish handling and sampling methods were approved by Kyoto University (No.26-71). All efforts were made to minimize suffering.

Design of bait sequences

Candidate bait sequences that are disrupted by the corresponding single guide RNAs (sgRNAs) were designed according to the published data sets of high-throughput screening of the sgRNA activity in mammalian cells [14]. The top 40 sgRNAs with the highest gene disrupting activity (20 from the “non ribo efficient sgRNA” data set and 20 from the “mESC efficient sgRNA” data set) were nominated as candidates. These candidates were screened by following two criteria using an off-line version of Cas-OFFinder [15] with the genome data base of 12 teleost fish species (Table 2-1). First, for each candidate, the total number of potential off-target sequences with ~3 bp mismatches in 18 bp of target sequence and PAM (NGG or NAG) sequence (in total 21 bp) in 12 fish genome date base was calculated. Then, the seven candidates with less number of total

potential off-target sequences were selected. The selected ones and the previously reported bait sequences (Gbait [16] and PITCh-gRNAs [9]) were screened according to the second criterion: genomic sequences with ~2 bp mismatches in the 18-bp target sequence for each sgRNA [5,17]. After the calculation of the total number of off-target in 12 fish genome data base, the eight bait sequences with less the number were used in this study.

Construction of donor plasmids

Backbone fragment 1: Backbone fragments containing a pUC replication origin and an ampicillin resistance gene were amplified from a plasmid pPBIS19-*mgfc*:Tag*BFP*-8x*HSE*:*Cre* [18] by PCR using a primer pair (pUCoriFW-SpeI and pUCoriRV-XhoI) (Table 2-2) and then digested with XhoI and SpeI.

Backbone fragment 2: Backbone fragments containing each bait sequence, a pUC replication origin, and an ampicillin resistance gene were amplified from a plasmid pPBIS19-*mgfc*:Tag*BFP*-8x*HSE*:*Cre* [18] by PCR using a primer pair (SpeI-Bait-FW and XhoI-Bait-RV) with each bait sequence and a restriction site (XhoI or SpeI) (Table 2-2) and then digested with XhoI and SpeI.

GFP cassette: To avoid *EGFP* gene disruption induced by Gbait, *monomeric Azami-Green* (*mAG*) gene was selected as reporter gene expressing green fluorescence protein (GFP) in this study. A GFP reporter cassette, a *mAG* gene with a N-terminal linker and a SV40 polyA signal, was amplified from a plasmid phmAG1-MNLinker (Medical & Biological Laboratories, Aichi, Japan) by PCR using primers mAG-linker-BamHI and polyA-RV-EcoRI (Table 2-2) and then digested with BamHI and EcoRI.

Table 2-1. Reference genome sequences of teleost species used in this study

Common Name	Scientific Name	Assembly	Download source	Reference
Medaka	*Oryzias latipes*	oryLat2 (Oct. 2005)	USCS genome browser (http://hgdownload.soe.ucsc.edu/downloads.html)	(Kasahara et al. 2007)
Zebrafish	*Danio rerio*	danRer10 (Sep. 2014)	USCS genome browser (http://hgdownload.soe.ucsc.edu/downloads.html)	(Howe et al. 2013)
Three-spined stickleback	*Gasterosteus aculeatus*	gasAcu1 (Feb. 2006)	USCS genome browser (http://hgdownload.soe.ucsc.edu/downloads.html)	(Jones et al. 2012)
Nile tilapia	*Oreochromis niloticus*	oreNil2 (Jan. 2011)	USCS genome browser (http://hgdownload.soe.ucsc.edu/downloads.html)	(Brawand et al. 2014)
Tiger puffer	*Takifugu rubripes*	v5 assembly (July 2010)	Fugu Genome Project (http://www.fugu-sg.org/)	(Kai et al. 2011)
Atlantic salman	*Salmo salar*	ICSASG_v2 (June 2015)	NCBI (GCF_000233375.1)	(Lien et al. 2016)
Rainbow trout	*Oncorhynchus mykiss*	(April 2014)	https://www.genoscope.cns.fr/trout/	(Berthelot et al. 2014)
African turquoise killifish	*Nothobranchius furzeri*	NotFur1.1 (April 2016)	http://africanturquoisekillifishbrowser.org/	(Valenzano et al. 2015)
Mexican tetra	*Astyanax mexicanus*	AstMex102 (Apr. 2013)	Ensembl (http://www.ensembl.org/info/data/ftp/index.html)	(McGaugh et al. 2014)
Red sea bream	*Pagrus major*	madai_150714 (July 2015)	-	(Toyoda et al. unpublished)
Pacific bluefin tuna	*Thunnus orientalis*	ver_Ba_1.0 (June 2013)	NCBI (GCA_000418415.1)	(Nakamura et al. 2013)
Tongue sole	*Cynoglossus semilaevis*	Cse_v1.0 (Jan. 2014)	NCBI (GCF_000523025.1)	(Chen et al. 2014)

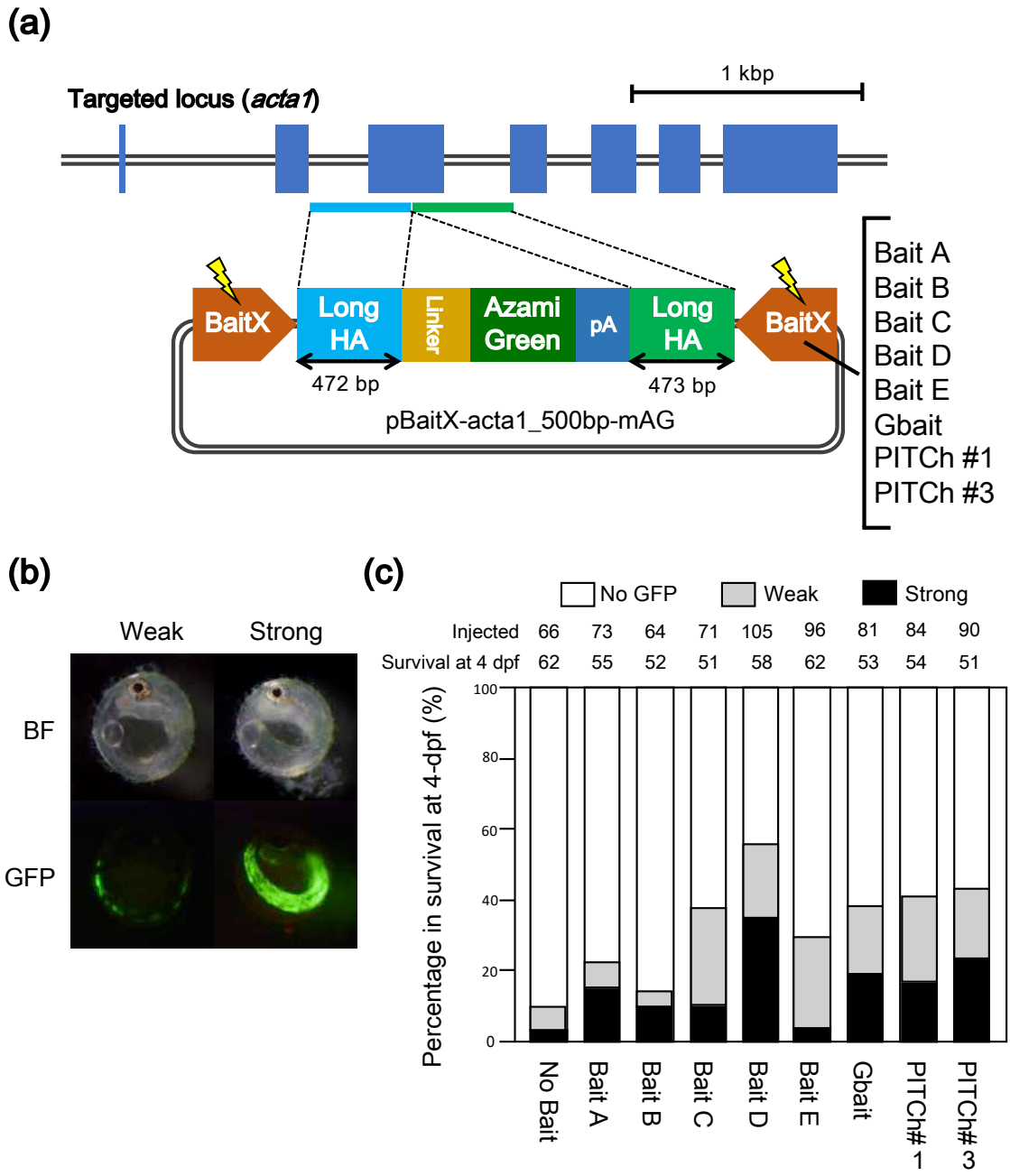


Fig. 2-3. Comparison of the efficiencies of targeted gene integration among bait sequences. (a) Schematic illustration of the genomic targeted locus, acta1, and donor plasmids containing each bait sequence. Long homology arms (Long HAs) are shown in boxes with light blue or green. (b) Representative GFP expression in injected embryos at four days post fertilization (dpf). The injected embryos are categorized by GFP expressing area into the following three groups; "No GFP", "Weak", or "Strong" with no fluorescence, less than 40% fluorescence, or more than 40%

fluorescence of the skeletal muscle, respectively. (c) Evaluation of the efficiencies of targeted gene integration of donor vectors with various bait sequences. The number of injected eggs is shown as "Injected", and the number of survived embryos at 4-days post fertilization (dpf) among the injected eggs is shown as "Survival at 4 dpf".

pBaitX-acta1_500bp-mAG (Fig. 2-3a): To generate donor plasmids with ~500 bp of sequences homologous to *acta1* locus (pBaitX-acta1_500bp-mAG), upstream (472 bp) or downstream (473 bp) genomic region at the target site of sgRNA-acta1 was PCR-amplified using primer pairs with restriction sites, acta1-5FW-XhoI/acta1-5RV-BamHI or acta1-3FW-EcoRI/acta1-3RV-SpeI (Table 2-2) and then digested with XhoI/BamHI or EcoRI/SpeI, respectively. These two fragments were ligated with the "Backbone fragment 2" and the "GFP cassette" simultaneously.

pNoBait-acta1_500bp-mAG (Fig. 2-3a): The "Backbone fragment 1" and the "GFP cassette" are ligated to generate a donor plasmid without bait sequences.

pGbait-acta1_20bp-mAG (Fig. 2-3a): The single-stranded oligo DNAs of acta1-exon3-5S1 and acta1-exon3-5AS1 or acta1-exon3-3S1 and acta1-exon3-3AS1 (Table 2-2) were annealed to generate double-stranded oligo DNAs containing upstream or downstream short homology arms (20 bp). These double-stranded oligo DNAs were ligated with the "GFP cassette" and one of the "Backbone fragment 2" which contains the Gbait to generate the pGbait-acta1-short-mAG donor plasmid.

pGbait-acta1_40bp-mAG (Fig. 2-3a): To generate a donor plasmid containing short homology arms (40 bp), a fragment containing "GFP cassette" and homology arms (40 bp) was amplified from the above plasmid, pGbait-acta1_500bp-mAG, using primer pairs

XhoI-acta1-Fw and acta1-SpeI-Rv. The fragment was digested with XhoI and SpeI, and was ligated with one of the "Backbone fragment 2" which contains the Gbait to generate the pGbait-acta1_40bp-mAG donor plasmid.

pBaitD-gap43_500bp-mAG (Fig. 3a): To construct a donor plasmid for targeting *gap43* locus (pBaitD-gap43_500bp-mAG), an upstream (483 bp) or downstream (496 bp) genomic region at the target site of sgRNA-gap43 was amplified by PCR with the primer pairs, XhoI-GAP43-FW/GAP43-BamHI-RV or EcoRI-GAP43-FW/GAP43-SpeI-RV (Table 2-2) and digested with XhoI/BamHI or EcoRI/SpeI, respectively. These fragments and BamHI/EcoRI-digested GFP cassette were cloned into one of the "Backbone fragment 2" containing BaitD sequence.

pBaitD-gap43_500bp-tdTomato (Fig. 3a): The coding sequence of *tdTomato* gene was amplified from ptdTomato Plasmid (Clontech, California, USA) by PCR using primer pairs BamHI-tdTomato-Fw/tdTomato-XbaI-Rv (Table 2-2). The donor plasmid, pBaitD-gap43_500bp-tdTomato, was generated following the method also used for generating pBaitD-gap43_500bp-mAG using the tdTomato coding sequence instead of the "GFP cassette".

All constructed plasmids were purified by an alkaline lysis method or NucleoSpin Plasmid QuickPure kit (MACHEREY-NAGEL, Düren, Germany).

Preparation of donor plasmids, Cas9 RNA and sgRNAs for microinjection

To eliminate the residual RNase activity of extracted plasmids, the donor plasmids dissolved with 50 μL of 5 mM Tris-HCl buffer (pH 8.5) were incubated with 5 μL of 10 % sodium dodecyl sulfate (SDS) and 2 μL of Proteinase K (20 mg/mL) at 55 °C for 30 min, and then were purified using the NucleoSpin Gel and PCR Clean-up kit

(MACHEREY-NAGEL) with the Buffer NTB supplied by the manufacture.

Cas9 RNA was transcribed from pCS2+hSpCas9 (Addgene Plasmid 51815) [5] using the mMessage mMachine SP6 Kit (Thermo Fisher Scientific, Waltham, MA). Custom-designed sgRNAs for genomic sequence of medaka were designed using a track for the UCSC genome browser of CRISPRscan [19]. Expression plasmids for the custom-designed sgRNAs were constructed by cloning the annealed oligonucleotides into a sgRNA expression plasmid pDR274 (Addgene Plasmid 42250) [20] as described previously [5]. The sgRNAs were transcribed from the DraI-digested template plasmids using the Ampliscribe T7-Flash Transcription Kit (Epicentre, WI). All synthesized RNAs were purified using the RNeasy Plus Mini Kit (Qiagen, Hilden, Germany) to eliminate the template DNA without DNase treatment. Sequences of the genomic target sites and annealed oligonucleotides are listed in Table 2-2.

Microinjection

To evaluate the DSB inducing activity of each sgRNA at the genomic target site, the injection mixture containing 100 ng/μL of Cas9 RNA and 50 ng/μL of each sgRNA were prepared. To evaluate efficiencies of targeted gene integration, the injection mixture containing 2.5 ng/μL of each donor plasmid, 100 ng/μL of Cas9 RNA, and 50 ng/μL of sgRNA corresponding to the donor plasmid were prepared. These injection mixtures were introduced into medaka eggs at the one-cell stage as described previously [21].

Genomic DNA extraction

Embryos were lysed individually at 4 days post fertilization (dpf) in 25 μL of alkaline lysis buffer containing 25 mM NaOH and 0.2 mM EDTA and incubated at 95°C for 15

min after breaking the egg envelope with forceps. Each sample was neutralized with 25 μL of 40 mM Tris-HCl (pH 8.0) and used as a genomic DNA sample.

Heteroduplex mobility assay

Heteroduplex mobility assay (HMA) was performed to investigate the DSB inducing activity of each sgRNA on the genomic target site as described previously [4]. Briefly, a genomic region containing the target sequence of sgRNAs-acta1 or sgRNA-gap43 was amplified by PCR using KOD-FX DNA polymerase (Toyobo, Osaka, Japan) with primers acta1-HMA-FW and acta1-HMA-RV or GAP43-for-HMA-FW and GAP43-for-HMA-RV (Table 2-2), respectively. The resulting amplicons were analyzed using a microchip electrophoresis system (MCE-202 MultiNA; Shimadzu, Kyoto, Japan) with DNA-500 reagent kit.

Microscopic observation

Embryos and larvae injected with the donor plasmids were observed using a fluorescence stereomicroscope MZFLIII (Leica Microsystems, Wetzlar, Germany) with a GFP2 filter set (for GFP) and a DsRed filter set (for RFP). Microscopic images were captured using a digital color-cooled charge-coupled camera and the VB-7010 image control system (Keyence, Osaka, Japan).

Sequence analysis

To evaluate the precise targeted integration, the DNA sequence around the integration site was investigated. The junction regions of the target site on the host genome and introduced gene were amplified by PCR with the following primer pairs, acta1-for-Seq-

Table 2-2. Oligonucleotide sequences used in this study

Name	Sequence (5'-3')	Usage
sgRNA-BaitA-S	TAGGAGTACTTCGGGGCGCAGT	Construction of sgRNA-BaitA
sgRNA-BaitA-AS	AAACACTGCGCCCCGAAGTACT	Construction of sgRNA-BaitA
sgRNA-BaitB-S	TAGGCATGTAGCGCTCGAAGGA	Construction of sgRNA-BaitB
sgRNA-BaitB-AS	AAACTCCTTCGAGCGCTACATG	Construction of sgRNA-BaitB
sgRNA-BaitC-S	TAGGTCAGGCTCGTTCTCGGTA	Construction of sgRNA-BaitC
sgRNA-BaitC-AS	AAACTACCGAGAACGAGCCTGA	Construction of sgRNA-BaitC
sgRNA-BaitD-S	TAGGTCTTCGGCCTAGACTGCG	Construction of sgRNA-BaitD
sgRNA-BaitD-AS	AAACCGCAGTCTAGGCCGAAGA	Construction of sgRNA-BaitD
sgRNA-BaitE-S	TAGGTCAAGAAGGGCTCGACCG	Construction of sgRNA-BaitE
sgRNA-BaitE-AS	AAACCGGTCGAGCCCTTCTTGA	Construction of sgRNA-BaitE
sgRNA-acta1-1S	TAGGAGCTGAGAGTGGCCCCCG	Construction of sgRNA-acta1 No.1
sgRNA-acta1-1AS	AAACCGGGGGCCACTCTCAGCT	Construction of sgRNA-acta1 No.1
sgRNA-acta1-2S	TAGGAGGCCCCCCTGAACCCCA	Construction of sgRNA-acta1 No.2
sgRNA-acta1-2AS	AAACTGGGGTTCAGGGGGGCCT	Construction of sgRNA-acta1 No.2
sgRNA-GAP43-S	TAGGGCGCAGCCACTTCTCCTG	Construction of sgRNA-gap43
sgRNA-GAP43-AS	AAACCAGGAGAAGTGGCTGCGC	Construction of sgRNA-gap43
pUCoriFW-BaitA-SpeI	ACCACTAGTCCGACTGCGCCCCGAAGTACTTCAGCAAAAGGCCAGCAAAAGG	Construction of a vector backbone with BaitA
AmpRV-BaitA-XhoI	TCCCTCGAGCCGACTGCGCCCCGAAGTACTTCCACTACGTGTATCCGCTCAT	Construction of a vector backbone with BaitA
pUCoriFW-BaitB-SpeI	ACCACTAGTCCCTCCTTCGAGCGCTACATGGCAGCAAAAGGCCAGCAAAAGG	Construction of a vector backbone with BaitB
AmpRV-BaitB-XhoI	TCCCTCGAGCCCTCCTTCGAGCGCTACATGGCCACTACGTGTATCCGCTCAT	Construction of a vector backbone with BaitB
pUCoriFW-BaitC-SpeI	ACCACTAGTCCCTACCGAGAACGAGCCTGATTAGCAAAAGGCCAGCAAAAGG	Construction of a vector backbone with BaitC
AmpRV-BaitC-XhoI	TCCCTCGAGCCCTACCGAGAACGAGCCTGATTCACTACGTGTATCCGCTCAT	Construction of a vector backbone with BaitC
pUCoriFW-BaitD-SpeI	ACCACTAGTCCTCGCAGTCTAGGCCGAAGATCAGCAAAAGGCCAGCAAAAGG	Construction of a vector backbone with BaitD
AmpRV-BaitD-XhoI	TCCCTCGAGCCTCGCAGTCTAGGCCGAAGATCCACTACGTGTATCCGCTCAT	Construction of a vector backbone with BaitD
pUCoriFW-BaitE-SpeI	ACCACTAGTCCGCGGTCGAGCCCTTCTTGACGAGCAAAAGGCCAGCAAAAGG	Construction of a vector backbone with BaitE
AmpRV-BaitE-XhoI	TCCCTCGAGCCGCGGTCGAGCCCTTCTTGACGCACTACGTGTATCCGCTCAT	Construction of a vector backbone with BaitE
pUCoriFW-SpeI	ACTGACTAGTAGCAAAAGGCCAGCAAAAGGCC	Construction of a vector backbone without Bait
AmpRV-XhoI	ACTGCTCGAGCACTACGTGTATCCGCTC	Construction of a vector backbone without Bait

Continues.

Table 2-2. Continued

mAG-linker-BamHI	CCCGGATCCAATTCCGCTGACGGCGGC	Cloning of upstream homology arm (long) of acta1
polyA-RV-EcoRI	CTTGAATTCCGCTTACAATTTACGCCT	Cloning of upstream homology arm (long) of acta1
acta1-5FW-XhoI	CGGCTCGAGCGTGTTCCCATCCATTGTC	Cloning of upstream homology arm (long) of acta1
acta1-5RV-BamHI	ATTGGATCCGGGGGCCACTCTCAGCTCAT	Cloning of upstream homology arm (long) of acta1
acta1-3FW-EcoRI	GCGGAATTCCCGAGGAGCACCCAACCC	Cloning of downstream homology arm (long) of acta1
acta1-3RV-SpeI	CGGACTAGTTCACACCATCACCAGCATCC	Cloning of downstream homology arm (long) of acta1
acta1-exon3-5S1	TCGAGCAATGAGCTGAGAGTGGCCCCCG	Cloning of upstream homology arm (short) of acta1
acta1-exon3-5AS1	GATCCGGGGGCCACTCTCAGCTCATTGC	Cloning of upstream homology arm (short) of acta1
acta1-exon3-3S1	AATTCCCGAGGAGCACCCAACCCTGA	Cloning of downstream homology arm (short) of acta1
acta1-exon3-3AS1	CTAGTCAGGGTTGGGTGCTCCTCGGG	Cloning of downstream homology arm (short) of acta1
acta1-HMA-FW	ACTGGGACGACATGGAGAAG	HMA for acta1
acta1-HMA-RV	TGTCTGTTGTCGCTCTCACC	HMA for acta1
XhoI-GAP43-Fw	GTCACTCGAGGAACATCTCAGTCTAAGACTCC	Cloning of upstream homology arm of gap43
GAP43-BamHI-Rv	ACTGGGATCCCGAGAAGTGGCTGCGCCT	Cloning of upstream homology arm of gap43
EcoRI-GAP43-Fw	CATGGAATTCCTGTGGCGCCGGCTCCATCTGCTCCTACAACT	Cloning of downstream homology arm of gap43
GAP43-SpeI-Rv	GATCACTAGTCCTTATAAATGCTTCACTATACACTTACCACATGCACGAC	Cloning of downstream homology arm of gap43
BamHI-tdTomato-Fw	GATCGGATCCATGGTGAGCAAGGGCGAGGA	Cloning of reporter gene for gap43
tdTomato-XbaI-Rv	CGTATCTAGACTACTTGTACAGCTCGTCCATGC	Cloning of reporter gene for gap43
GAP43-for-HMA Fw	CTGCAGCAGACAAACCAGCTAAC	HMA for gap43
GAP43-for-HMA-Rv	CTCTTGGTTGTTGCAGCTGGA	HMA for gap43
GAP43-for-Seq-Fw	GCAATGATCTGTAGTTGTATTCTGGG	Sequencing for gap43
GAP43-for-Seq-Rv	GCACTTAATGTACCTTATAGTTCAGGGC	Sequencing for gap43
GAP43-Rv	GGAGCCGGCGCCACAGGAGA	Investigating for precise integration into gap43
mAG-Rv	ATCCTGGACCTGAACGTGAC	Investigating for precise integration into gap43
tdTomato-Rv	GCCCTTGGTCACCTTCAGCT	Investigating for precise integration into gap43
EF1α-Fw	CAGGACGTCTACAAAATCGG	Confirmation of genomic DNA extraction
EF1α-Rv	AGCTCGTTGAACTTGCAGGC	Confirmation of genomic DNA extraction

Fw and mAG-Rv, mAG-Fw and acta1-for-Seq-Rv, or GAP43-for-Seq-FW and GAP43-for Seq-RV (Table 2-2), and KOD -plus- Neo DNA polymerase (Toyobo). The PCR condition were as follows: one cycle at 94°C for 2 min, followed by 35 cycles of 98°C for 10 sec, 58°C for 30 sec, and 68°C for 1 min. The resulting PCR products were subjected to electrophoresis with a 1% agarose gel. Then, a PCR fragment that was predicted to contain the introduced gene was excised from the gel and purified using NucleoSpin Gel and PCR Clean-up (MACHEREY-NAGEL). The purified fragment was sequenced with the primers acta1-for-Seq-Fw and acta1-for-Seq-Rv (for *acta1*) or GAP43-for-Seq-FW and GAP43-for-Seq-RV (for *gap43*) (Table 2-2).

2.3 Results

Selection of sgRNA targeting to the skeletal muscle-specific actin gene in medaka genome

To evaluate precisely and rapidly the gene knockin efficiency using each donor plasmid, it was needed to select the genomic target locus expressing widely at early-stage embryo and sgRNA possessing the high genome-editing activity at the locus. A skeletal muscle-specific actin gene (*acta1*; Ensembl gene number ENSORLG00000010881) was selected as the target locus because the detection of site-specific integration by observing the green fluorescence in the skeletal muscles is simple [22,23].

At first, to obtain a sgRNA targeting to the *acta1* gene with high DSB inducing activity, two sgRNAs targeting to the third exon of the gene without potential off-target sites in medaka genome were designed (Fig. 2-1). Each sgRNA was injected with a Cas9 RNA into fertilized medaka eggs and the genome-editing activity was evaluated by HMA.

More multiple banding patterns were observed in embryos with sgRNA-acta1 #1 than with sgRNA-acta1 #2, indicating that sgRNA-acta1 #1 has higher DSB inducing activity than sgRNA-acta1 #2 (Fig. 2-1). Thus, the sgRNA-acta1 #1 was used as a sgRNA for targeted genome cleavage in the following experiments.

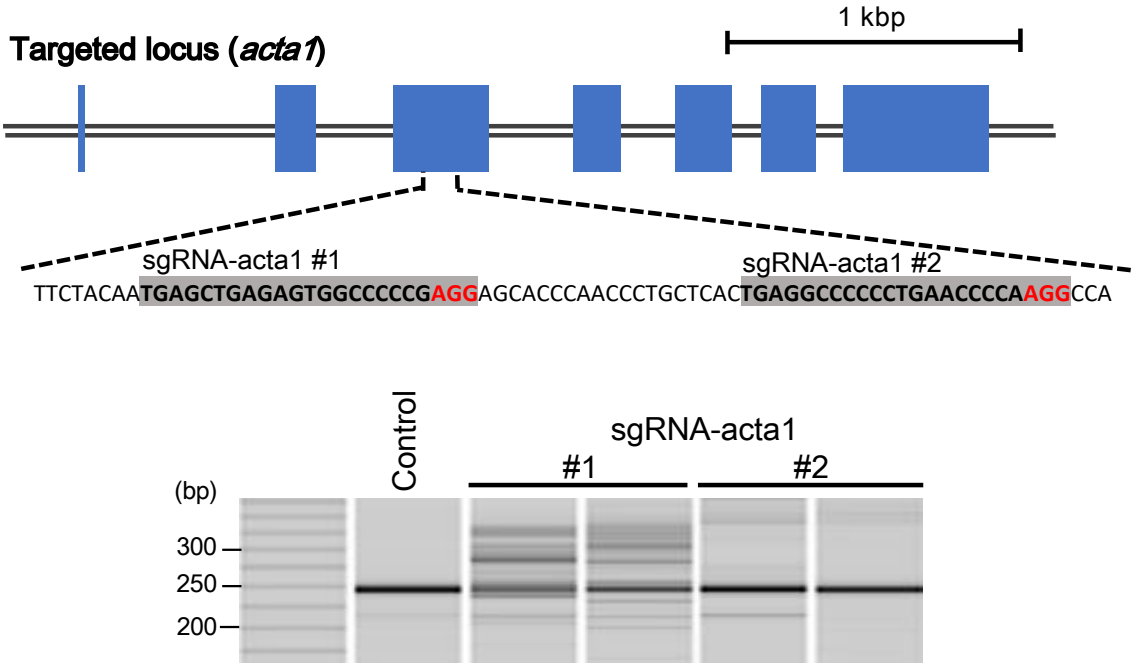


Fig. 2-1. Selection of sgRNA targeting to the skeletal muscle-specific actin gene (acta1) in medaka. (Upper): The design of sgRNAs targeting the acta1 locus. (Lower): An electrophoresis image shows results of the heteroduplex mobility assay (HMA) in embryos injected with 50 ng/μL of a sgRNA (sgRNA-acta1 #1 or #2) and 100 ng/μL of Cas9 mRNA. Control shows a result from an embryo without injection.

Effect of length of homologous sequences and existence of the bait sequence on targeted gene integration events

To assess the effects of lengths of homologous sequences that are placed on both ends of the inserted gene fragment in the donor plasmid, the targeted gene integrating efficiency into the third exon of *acta1* gene was evaluated using three donor plasmids that contain the Gbait sequences [16]: one of the donor plasmids is pGbait-acta1_20bp-mAG, which possesses short homologous sequences (20 bp each) on both ends of the insert gene fragment (mAG-pA). Another is pGbait-acta1_40bp-mAG possessing short homologous

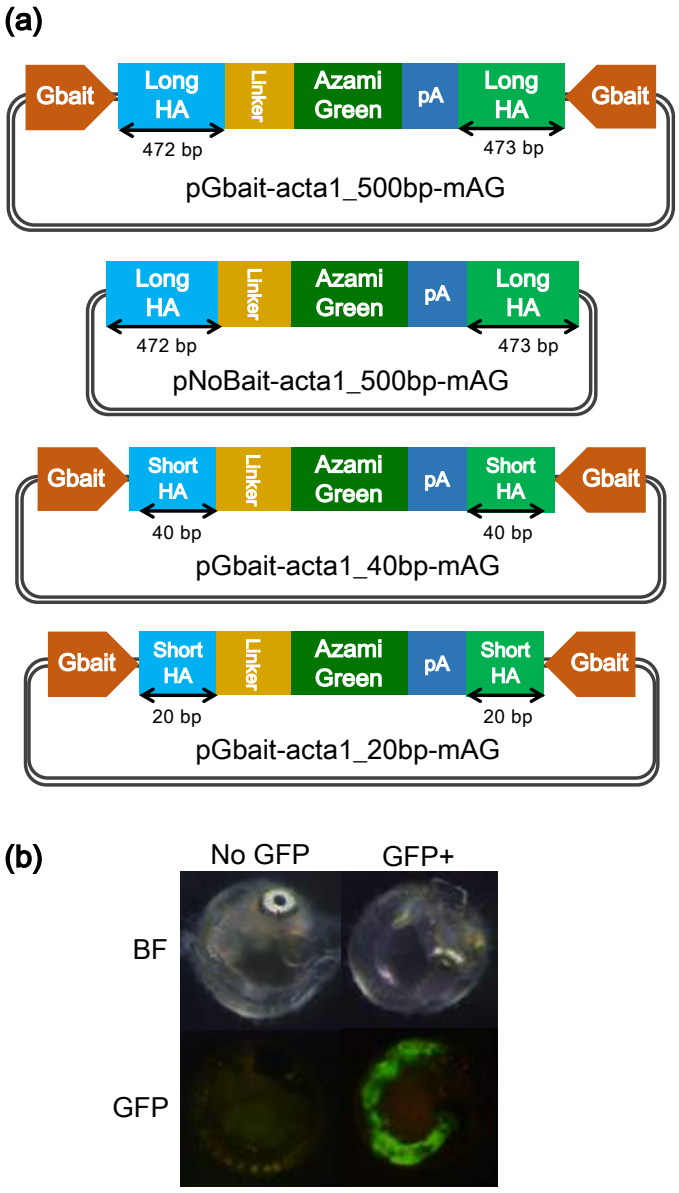


Fig. 2-2. Effects of the length of homologous sequences and presence of bait sequence on the

efficiencies of targeted gene integration. (a) Schematics of each donor plasmid for evaluating the effects of homologous sequences and the bait sequence. The pGbait-acta1_500 bp-mAG plasmid contains long homology arms (Long HAs) and Gbait. The pNoBait-acta1_500 bp-mAG plasmid is the plasmid that removed the Gbait from pGbait-acta1_500 bp-mAG plasmid. The pGbait-acta1_40 bp-mAG plasmid contains 40 bp short homology arms (Short HAs) and Gbait. The pGbait-acta1_20 bp-mAG plasmid contains 20 bp short homology arms (Short HAs) and Gbait. (b) Embryos at 4 days post fertilization (dpf). Embryos injected with each plasmid, Cas9 RNA (100 ng/μl) with or without Gbait-sgRNA were categorized into the following two groups; "No GFP" for embryos without green fluorescence and "GFP+" for embryos with fluorescence in the skeletal muscle, respectively. BF: bright field, GFP: GFP fluorescent image.

sequences (40 bp each) on both ends of the insert gene. The other is pGbait-acta1_500bp-mAG, which possesses longer homologous sequences (472 and 473 bp) on both ends of the insert gene (Fig. 2-2a). When an integration event occurred precisely into the target site, green fluorescence was observed in the skeletal muscle. As shown in Table 2-3 and Fig 2-2b, 17 (24.3%) of 70 injected embryos expressed green fluorescence in the skeletal muscle by injection of the long homologous plasmid. On the other hand, 2 (1.9%) of 110 and 3 (4.7%) of 64 injected embryos expressed green fluorescence in the skeletal muscle with by injection of the short homologous plasmids (20 bp and 40 bp, respectively) (Table 2-3). These indicate that the donor plasmid with longer homologous sequences (ca. 500 bp) can be efficiently integrated at the target site of this locus.

Table 2-3. Comparison of integrate efficiency among each of donor plasmids

homology arms	bait sequences	Survival at 4 dpf	No GFP	GFP+	Integrate efficiency (%)
long	+	70	53	17	24.3
	-	62	56	6	9.7
short	+	50	50	0	0.0

Integrate efficiency (%) = GFP+ / Survival at 4 dpf

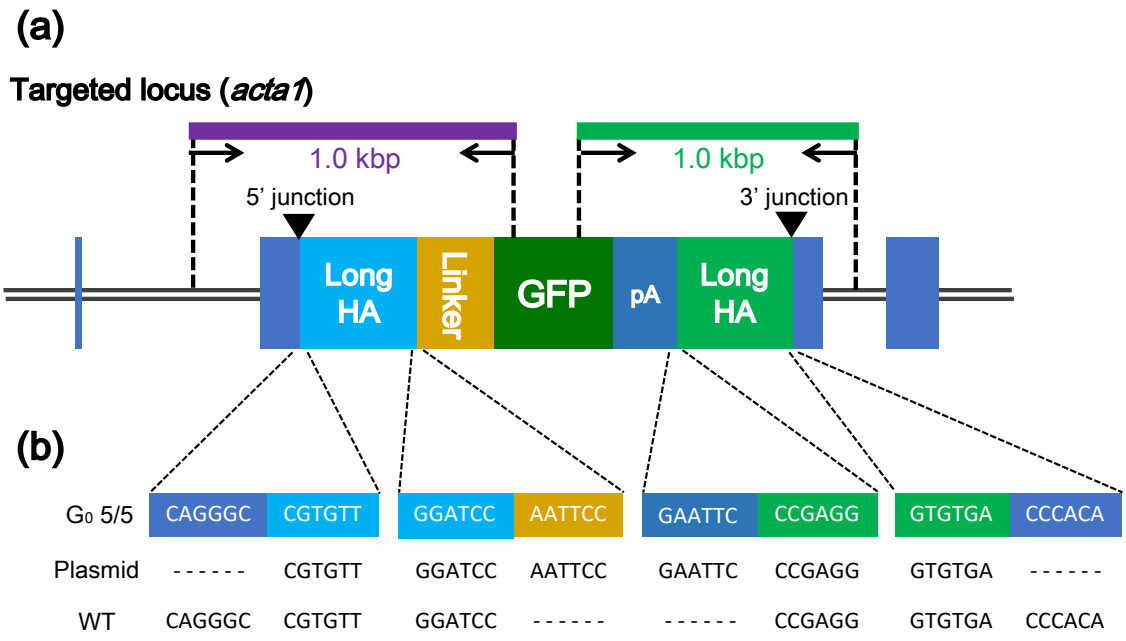


Fig. 2-4. Precise gene integration of a donor vector containing the GFP gene into the acta1 locus. (a) Schematic illustration of the locus. The primer pair (acta1-for-Seq-Fw and mAG-Rv or mAG-Fw and acta1-for-Seq-Rv) was used for investigating the precise integration into the acta1 locus. The amplicons containing 5′ junction or 3′ junction are 1.0 kbp, respectively. (b) The sequence analysis of the PCR amplicons (1.0 kbp) containing 5′ or 3′ junction. (Upper): The sequence observed in the G0 fish with the inserted gene (#1–#5). (Middle): The sequence in the donor plasmid, pBaitD-acta1_500 bp-mAG (Plasmid). (Lower): The sequence in the wild type fish without the insertion (WT).

Previous studies demonstrated that the induction of DSBs on the sgRNA target sequences next to homologous sequences in circular donor plasmids using the CRISPR/Cas9 system can enhance the targeted integration events [10,13]. The targeted gene integrate efficiency was also evaluated using the donor plasmids containing the long homologous sequences for the *acta1* gene with (pGbait-acta1_500bp-mAG) or without (pNoBait-acta1_500bp-mAG) the Gbait sequence with long homologous sequences of *acta1*. Without the bait sequence, the integrate efficiency was reduced to 9.7%, compared to 24.3% with the bait sequence (Table 2-3). This indicates the importance of the bait sequence in efficient targeted gene integration events in medaka (Table 2-3).

To investigate whether the gene knockin events precisely occurred by using the donor plasmid with bait sequences and long homologous sequences (pGbait-acta1_500bp-mAG), genomic DNA was extracted from five G_0 embryos with GFP fluorescence in the skeletal muscle (Fig.2-2b). Each fragment containing upstream or downstream junction between the genomic sequence and the insert gene was PCR-amplified from the genomic DNA (Fig. 2-4a) and was sequenced. In all the five embryos examined, the donor plasmid was precisely integrated on the target site without any indels (Fig. 2-4b), indicating that the gene knockin events were performed via HDR as expected.

Design of bait sequences for targeted gene integration in teleost fish species

To improve the efficiencies of targeted gene integration events in medaka and other teleost species, bait sequences with high DSB inducing activity and without off-target effects in their genome are needed to design. In this study, I nominated 40 sgRNA target sequences with highest DSB inducing activity from the data of high-throughput screening in mammalian cells [14] as candidates for newly designed bait sequences (Table 2-4).

Table 2-4. The numbers of potential offtarget sequences.

#	Sequence (N18-NGG)	*O. latipes*	*D. rerio*	*G. aculeatus*	*O. niloticus*	*T. rubripes*	*S. salar*	*O. mykiss*	*N. fuzeri*	*A. mexicanus*	*P. major*	*T. orientalis*	*C. semilaevis*	Total
1	TACTCCCGCTTACCTCTGCGG	6	0	4	8	1	5	2	6	1	1	1	6	41
2	ACCTCCTTGATCGGAAGGAGG	7	3	2	7	9	12	8	10	5	11	1	3	78
3	AGTACTTCGGGGCGCAGTCGG	0	0	2	1	1	2	3	3	1	5	1	2	21
4	ATATACATGGTTTCTCGTAGG	3	6	3	3	2	16	19	7	6	11	1	2	79
5	GTGACTCCCACATCGTCGTGG	3	2	3	6	3	36	11	4	0	5	3	6	82
6	GTACTTGATGTGCCGGATAGG	0	2	2	1	1	8	15	1	4	2	1	1	38
7	GGTTGATTCTGTCCGCCATGG	8	6	2	7	5	26	16	8	5	9	6	5	103
8	GCCTTTTCTACCTCACGCCGG	5	4	1	1	3	4	7	8	5	5	4	8	55
9	CCAACGACTGGTTACCCCTGG	6	1	5	4	6	11	15	1	4	4	1	2	60
10	CATGTAGCGCTCGAAGGAGGG	2	3	4	2	3	5	6	2	0	2	0	4	33
11	TCAGGCTCGTTCTCGGTAGGG	2	0	2	1	1	9	9	4	0	2	2	0	32
12	CATACCGTCCGGCACTGTGGG	2	2	5	2	4	7	6	4	9	4	3	2	50
13	GTGTACTAACATCAATCCCGG	2	2	0	2	1	1	4	4	1	4	0	0	21
14	CGGATTCGCTCTTCCTGTGGG	3	3	4	5	3	5	13	3	4	2	5	2	52
15	TGGCAATACTGCTCGCTGTGG	8	13	8	8	6	18	17	6	11	6	12	10	123
16	ATTGCTCATAGTGGGCGATGG	1	3	6	6	4	8	10	4	4	7	6	2	61
17	GATTATAAGTATCTGCGCTGG	1	7	6	7	2	5	7	6	6	5	3	4	59
18	CCAACTTTGAGATGCGCTTGG	4	15	0	6	2	13	10	5	3	6	4	1	69
19	AAAATGAGAAACCGACGGCGG	3	4	10	8	9	6	4	7	5	6	2	11	75

Continues.

Table 2-4. Continued

20	GGCTCATGAGATTAGCGAAGG	4	3	5	7	1	8	9	4	2	7	5	5	60
21	TGTTCAGCTCCTCGGCAATGG	10	12	6	9	10	791	578	7	21	8	10	8	1470
22	TGAAGCGGAAGCGCAGTGCGG	10	3	4	6	5	11	6	7	7	5	7	4	75
23	CGACCTCTTGCTGTGCTGCGG	10	333	3	4	10	7	12	8	2	6	7	3	405
24	CATCAAGAGAGTCCAGTTCGG	10	8	2	5	8	17	9	4	10	12	9	3	97
25	CATTGTAGGGCTCGACCACGG	10	9	15	8	2	20	13	20	6	12	5	4	124
26	TCTTGCATGGCCCGAGTGCGG	0	2	2	2	1	1	3	1	0	2	2	0	16
27	AGATGTTGAGAATGACAAGGG	16	14	12	22	11	143	106	28	16	21	17	15	421
28	TCTTCGGCCTAGACTGCGAGG	2	4	1	1	0	2	2	0	0	1	1	3	17
29	CTTCTCCTGGATCAGCGTAGG	8	6	6	6	8	19	20	8	13	17	9	13	133
30	CGGGAAGGGCCTCGGCACTGG	1	5	6	2	3	31	12	4	3	5	3	4	79
31	CTGGGATATTAGCTCCAGCGG	4	26	6	4	6	7	3	1	3	5	4	4	73
32	GCCTCCAACACCTTCTTGAGG	20	6	11	18	12	38	46	14	10	18	15	10	218
33	TCAAGAAGGGCTCGACCGCGG	0	0	1	0	1	4	2	1	2	1	1	0	13
34	ACCGCAAGACCGGCCAGAAGG	2	2	3	2	6	16	20	1	2	3	0	2	59
35	AGAAGCAGCGCGAGCAGATGG	25	23	26	32	31	90	76	25	24	24	24	30	430
36	ACAGTAAGAAATATGGTGTGG	19	70	9	71	5	107	51	25	28	22	24	11	442
37	GTTGGAGAGAAGCCACGAAGG	11	11	15	14	7	25	14	5	8	59	4	5	178
38	TGCAGCGGAGCCCTTAAGCGG	4	31	9	1	5	2	1	5	2	4	2	2	68
39	CTACACCAAGGAGTCCCGCGG	1	3	6	3	3	6	5	1	3	5	1	4	41
40	GGACTCAGGACTCCGAACTGG	2	0	3	4	1	21	22	4	19	3	4	6	89

Numbers of genomic sites with up to 3 bp mismacthes in the total 21 bp sequences are shown.

To screen bait sequences with less off-target effects in wide range of teleost fish species, I investigated the numbers of potential off-target sites of 40 candidates in reference genome sequences of medaka and other 11 teleost species (*Danio rerio*, *Gasterosteus aculeatus*, *Oreochromis niloticus*, *Takifugu rubripes*, *Salmo salar*, *Oncorhynchus mykiss*, *Nothobranchius furzeri*, *Astyanax mexicanus*, *Pagrus major*, *Thunnus orientalis*, and *Cynoglossus semilaevis*) (Table 2-1). The numbers of genomic sequences with up to 3-bp mismatches in a total of 21 bp of the candidates and a NRG PAM were counted (Table 2-4) and then top 7 sequences with the fewest potential off-target numbers (#3, #10, #11, #13, #26, #28, and #33) were selected as candidates. Subsequently, I examined the numbers of mismatch base pairs in all potential off-target sites of these 7 candidates and 4 previously designed bait sequences—Gbait [16] and PITCh gRNAs (the PITCh-gRNA#1-#3) [9] in details. Ten candidates except the PITCh-gRNA#3 had some potential off-target sites with 2-bp mismatches in the 18-bp target sequence but all 11 candidates had no sites with 1 bp or less (Table 2-5). Thereafter, I excluded 3 sequences (#13, #26, the PITCh-gRNA#2) with larger numbers of the genomic site with 2-bp mismatches. Finally, 5 bait candidates (#3, #10, #11, #28, and #33; hereafter designated as BaitA, BaitB, BaitC, BaitD, and BaitE, respectively) and 3 previous reported sequences (Gbait, PITCh-gRNA#1 and PITCh-gRNA#3) were subjected to the following evaluations in medaka embryos.

Table 2-5. Potential off-target sites of 7 cadidates selected in first screening and previously reported bait sequences.

Species	Chromosome	Position	Strand	Sequence	# of mismatches in 18 bp
Candidate #3 (AGTACTTCGGGGCGCAGTCGG)					
O. mykiss	Un	384205847	+	AGTACTTCGGGGCGgAGgTGG	2
A. mexicanus	KB882109.1	2397485	-	AGTACTTCtGGGgGCAGTTGG	2
Candidate #10 (CATGTAGCGCTCGAAGGAGGG)					
O. niloticus	LG14	27746629	+	CATGTAtCGCTCaAAGGATGG	2
S. salar	ssa10	47603974	-	CcTGTAGCtCTCGAAGGATGG	2
	ssa18	29297706	+	CATGTAGCaCTCGAAaGAGGG	2
O. mykiss	chrUn	176512456	+	CcTGTAGCtCTCGAAGGATGG	2
N. furzeri	GapFilledScaffold_2068	114109	-	CATGTtGCtCTCGAAGGAAGG	2
Candidate #11 (TCAGGCTCGTTCTCGGTAGGG)					
S. salar	ssa27	6855250	+	TCAGGCTaGTTCcCGGTACGG	2
	ssa27	6855208	+	TCAGGCTaGTTCcCGGTAAGG	2
	ssa27	6855229	+	TCAGGCTaGTTCcCGGTAAGG	2
O. mykiss	Un	564842094	-	TCAGGtTCaTTCTCGGTATGG	2
Candidate #13 (GTGTACTAACATCAATCCCGG)					
S. salar	ssa27	27256339	-	GTGTAaTAACATCAATaCAGG	2
O. mykiss	Un	7716753	-	GTaTACTAACAcCAATCCAGG	2
	Un	497859850	+	cTGTACTAACAcCAATCCAGG	2
N. furzeri	GapFilledScaffold_9653	7833	-	GTGTACTAcCATCcATCCAGG	2
	GapFilledScaffold_12722	8163	+	GTGTACTAcCATCcATCCAGG	2
	GapFilledScaffold_33097	433	-	GTGTACTAcCATCcATCCAGG	2
P. major	Scaffold0220	98015	-	GTGTgCTcACATCAATCCTGG	2
Candidate #26 (TCTTGCATGGCCCGAGTGCGG)					
D. rerio	19	9422708	-	TagTGCATGGCCCGAGTGAGG	2
G. aculeatus	V	11379995	-	TCTTaCtTGGCCCGAGTGAGG	2
T. rubripes	1	7084094	-	TCTTaCtTGGCCCGAGTGAGG	2
O. mykiss	Un_10	34163178	+	TtTTGCATGGCCCaAGTGTGG	2
N. furzeri	GapFilledScaffold_1998	70883	-	TCTTaCtTGGCCCGAGTGAGG	2
P. major	Scaffold0006	3079823	-	TCTTaCtTGGCCCGAGTGAGG	2

Continues.

Table 2-5. Continued

Candidate #28 (TCTTCGGCCTAGACTGCGAGG)					
D. rerio	11	22151597	-	TCTTCGGgCTgGACTGCGCGG	2
G. aculeatus	XVIII	15001883	-	TCTTCGGCCTcGACgGCGAGG	2
C. semilaevis	6	6259207	+	TCTTaGGCCTAGACTGCtGGG	2
Candidate #33 (TCAAGAAGGGCTCGACCGCGG)					
G. aculeatus	V	1232158	-	TCAAGAtGGGCTCGtCCGCGG	2
T. rubripes	4	11779781	+	TCAAGAAGGGCTCcACaGAGG	2
S. salar	ssa09	116731307	+	TaAAGAAGGGCTCcACCGAGG	2
	ssa20	61101718	+	TaAAGAAGGGCTCcACCGAGG	2
O. mykiss	Un	864968259	+	TaAAGAAGGGCTCcACCGAGG	2
GBait (CGAGGGCGATGCCACCTACGG)					
D.rerio	13	10787556	-	CGAGGGCtATGCCACCTtGGG	2
T. rubripes	17	2778599	-	CGAGGGaGATGCCACCTtTGG	2
C. semilaevis	10	1465660	-	CGAGGGCGggGCCACCTAGGG	2
PITCh-gRNA #1 (GCTTCGATATCGATCGTTTGG)					
O. niloticus	LG5	11429925	+	GCTTCGATATCcATCtTTCGG	2
PITCh-gRNA #2 (CTAATCGAGATCGACTGTTGG)					
G. aculeatus	III	3941302	+	CTAATgGAGATCcACTGTTGG	2
O. niloticus	LG18	3141526	+	CTAATgGAGATCcACTGTTGG	2
O. niloticus	LG7	10529204	-	CTgATgGAGATCGACTGTAGG	2
N. furzeri	GapFilledScaffold_3669	8164	+	CTAATgGAGAgCGACTGTCGG	2
A. mexicanus	KB871663.1	845674	+	CTAATaGAGAgCGACTGTGGG	2
T. orientalis	superscaffoldBa00003326	6700	+	CTAATgGAGATCcACTGTTGG	2
PITCh-gRNA #3 (ATCGTACGCGTACGTGTTTGG)					
No candidate					

Potential off-target sites are defined as genomic sequence harboring up to 2 bp mismacthes in the total 18 bp sequences and a NGG PAM.

Comparison of targeted integration efficiencies among bait sequences in medaka

I compared the targeted gene integration activity of the selected five bait sequences with previously designed bait sequences, Gbait [16] and PITCh gRNAs (the PITCh-gRNA#1 and PITCh-gRNA#3) [9], which were reported to have high activities for targeted gene integration. A mixture of each donor plasmid with a type of bait sequences, a sgRNA corresponding to each bait sequence, Cas9 RNA, and sgRNA-acta1 was injected into the one-cell stage embryos of medaka. Expression of green fluorescence in the skeletal muscle was observed at 4 days post fertilization (dpf). According to the area expressing green fluorescence, the larvae were categorized into the following 3 groups: ‘‘Strong”, green fluorescence was observed in more than 40% of the area of whole embryonic body; “Weak”: less than 40% of the area of whole embryonic body; “No GFP”: no GFP expression in the embryonic body (Fig. 2-3b and 3c).

As shown in Fig. 2-3c, with all donor plasmids, the larvae expressing green fluorescence were observed and all donor plasmids containing bait sequence showed a higher green fluorescence expressing ratio (13.5–55.2%) than those with no bait (without bait sequence) (9.6%). Among the bait sequences, the ratio of larvae with green fluorescence varied and depended on the bait sequences. The highest rate (55.2%) of the larvae with green fluorescence (total larvae with green fluorescence) was observed with BaitD. Additionally, the ratio “Strong” was the highest in BaitD (Fig. 2-3c). These findings indicate that the BaitD system possesses the highest activity to induce the targeted gene integration events among the bait systems tested. Unfortunately, all the fish expressing green fluorescence in the skeletal muscle died before or soon after hatching; however, the cause of death was unclear.

Germline transmission of a DNA fragment integrated at the target locus using the BaitD system

To investigate whether the gene fragments integrated at the target loci using the BaitD system can be transmitted to the progeny, a gene knockin experiment targeting to another gene growth associated protein 43 (*gap43*; Ensembl gene number ENSORLG00000015837) was carried out. The *gap43* transcript is expressed in the central nervous system (CNS) from 4 dpf and contributes to the growth of neuroblasts in medaka [24]. The sgRNA-gap43 without any potential off-target site in the medaka

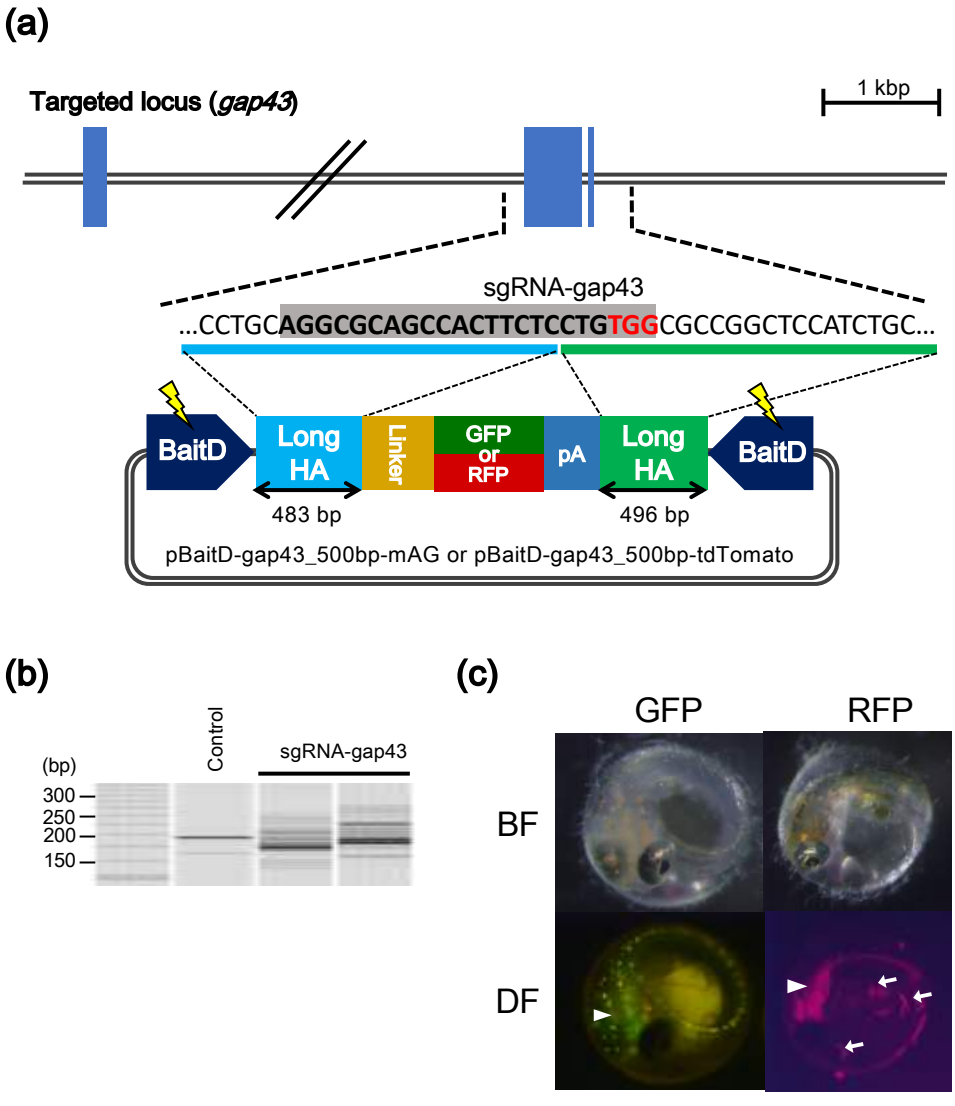


Fig. 2-5. Targeted gene integration of GFP or RFP constructs containing BaitD into the gap43 locus. (a) Design of sgRNA and homology arms to target the second exon of gap43 gene (Ensembl transcript number ENSORLT00000015837). The donor plasmids contain two BaitD sequences, upstream and downstream homology arms, and monomeric Azami-Green (mAG; as GFP) or tandem-dimer-Tomato (tdTomato; as RFP) with a N-terminal linker and a SV40 polyA signal (pA). The upstream or downstream homology arms (483 or 496 bp) for the donor plasmids are shown in boxes with light blue or green color, respectively. (b) The genome-editing activity by the sgRNA targeting to the gap43 locus (sgRNA-gap43), which was investigated using the heteroduplex mobility assay (HMA) of embryos injected with a mixture containing 100 ng/μL of Cas9 mRNA and 50 ng/μL of sgRNA-gap43. Control shows the result from an embryo without injection. (c) GFP and RFP expression in the injected embryos at 4 days post fertilization (dpf). Embryos injected with the donor vectors expressed either GFP or RFP in their central nervous system (CNS). White arrowheads show fluorescence in the CNS, while white arrows show autofluorescence of bacteria on the surface of the egg membrane.

genome was designed on the second exon (Fig. 2-5a). After confirming the sgRNA possessing high DSB inducing activity by HMA (Fig. 2-5b), the mixture of sgRNA-gap43, containing a donor plasmid with two BaitD sequences, a sgRNA for the BaitD sequence, and the Cas9 RNA, was injected into the one-cell stage medaka eggs. I successfully observed that green fluorescence was expressed strongly in the brain of 39 out of 320 injected eggs at 4 dpf (Fig. 2-5c) (Table 2-6). This fluorescent pattern was corresponding to the endogenous expression pattern of *gap43* gene that was reported in a previous study [24]. Twenty-eight of the GFP expressing embryos were raised to adulthood and mated

with wild-type fish. Out of the 28 adult fish, two individuals transmitted the insert sequence to the next generation (34.6% and 26.8% germline transmission rate, respectively) (Table 2-6 and 7). To investigate whether the knockin events were precisely

Table 2-6. Results of microinjection targeting the *gap43* locus with GFP or RFP

	Injected	Survival at 4 dpf	Expressed	Integrate efficiency	Sexually matured	Germ-line transmission rate
GFP	320	158	39	24.7% (39/158)	28	7.1% (2/28)
RFP	103	60	16	26.7% (16/103)	11	9.1% (1/11)

Table 2-7. Germ-line transmition efficiency of F_0 individuals harboring GFP or RFP in the *gap43* locus

	Collected eggs	Expressed embryos	Transmit efficiency (%)
GFP#1	55	19	34.6
GFP#2	56	15	26.8
RFP#1	58	15	25.9

Transmit efficiency (%) = Expressed embryos / Collected eggs

occurred at the target locus, the DNA sequencing around the inserted fragment was performed after the PCR amplification using genomic DNA extracted from five F_1 embryos with GFP fluorescence in the CNS (Fig. 2-6a, b, and c). In all the five examined embryos, the insert fragment was precisely integrated on the target site without any indels (Fig. 2-6b), demonstrating that the BaitD system allows for effective and precise targeted gene integration in the germ cells of medaka.

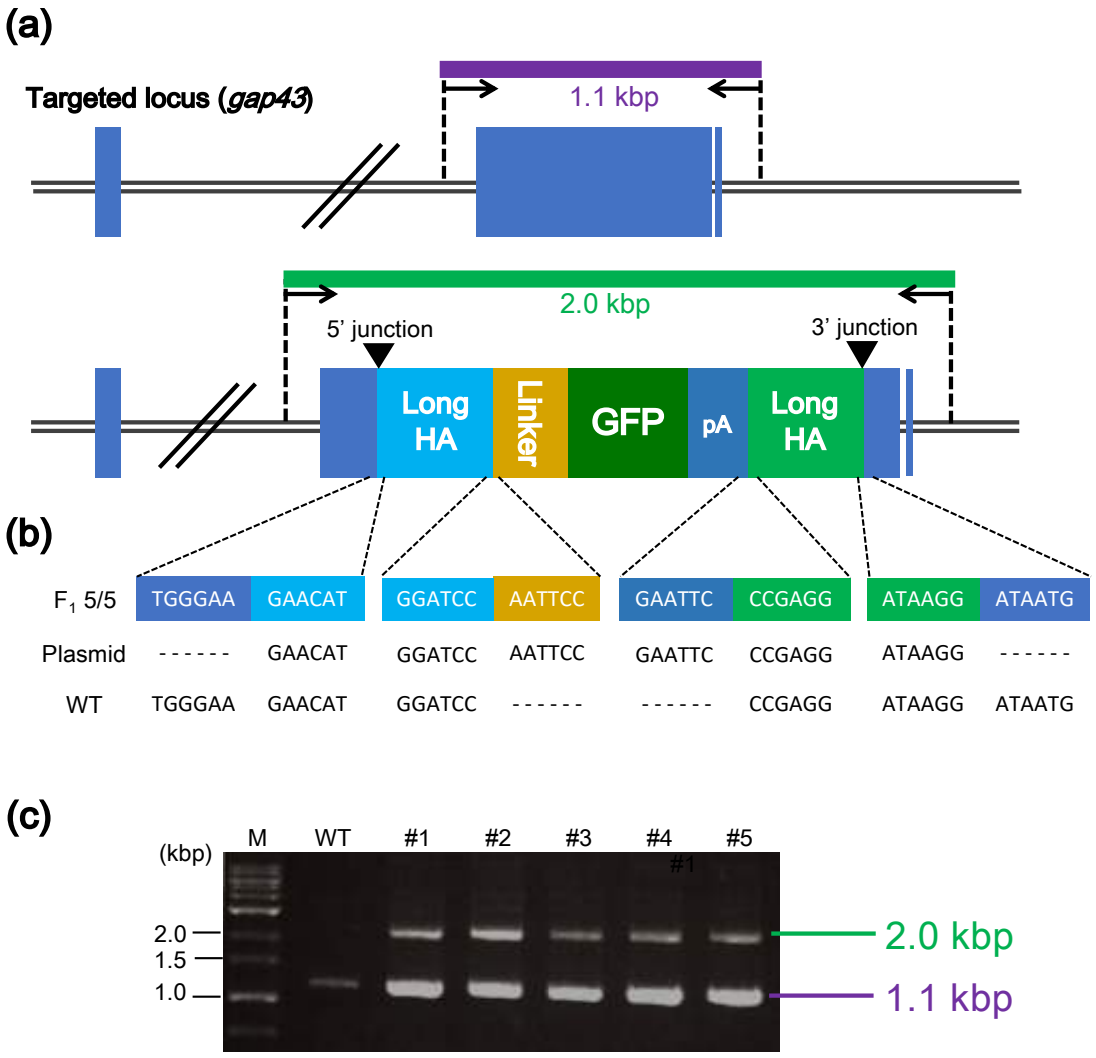


Fig. 2-6. Precise gene integration of a donor vector containing the GFP gene into the gap43 locus. (a) Schematic illustration of genotyping at the locus. The primer pair (GAP43-for-Seq-Fw and GAP43-for-Seq-Rv) was used for investigating the precise integration into the gap43 loci. The amplicons with or without the insert sequence are 2.0 kbp or 1.1 kbp, respectively. (b) Sequence analysis of the PCR amplicons (2.0 kbp) containing GFP gene. (Upper): Sequence observed in the F1 fish with the inserted gene (#1–#5) (Middle): Sequence in the donor plasmid, pBaitD-gap43_500 bp-mAG (Plasmid). (Lower): Sequence in the wild type fish without the insertion (WT). (c) Electrophoresis image of the PCR products. The shorter PCR product (1.1 kbp) amplified from the intact allele was detected in both a wild type (WT) and five F1 fish with the

insertion (#1–#5). The longer PCR product (2.0 kbp) was detected only in the F1 fish with the insertion (#1–#5).

PCR genotyping-free selection of double allelic gene edited fish using two different fluorescent colors

To establish a novel genotyping method using two different fluorescent colors, the donor plasmid (pBaitD-gap43_500bp-tdTomato) containing RFP (*tdTomato* gene) was introduced into fertilized medaka eggs following the same method for the GFP plasmid. RFP expression in the CNS was observed in 16 out of 103 injected embryos and one individual transmitted the insert sequence to the next generation (Table 2-6).

To investigate whether the double allelic gene knockin fish can be selected simply by fluorescence without any PCR genotyping, the F_1 fish harboring the RFP gene in the *gap43* locus was mated with a F_1 individual harboring the GFP gene in the locus. Resulting F_2 individuals were divided into the following four groups by fluorescent color: green fluorescence (G+/R-), red fluorescence (G-/R+), both green and red fluorescence (G+/R+), and no fluorescence (G-/R-) (Fig. 2-7a). Out of 47 F_2 individuals, the number of individuals in each group was 12, 11, 13, and 11, respectively. This indicates that each inserted gene fragment was transmitted to the progeny in a Mendelian manner. With the genomic DNA extracted from the embryo of each group, PCR analysis was carried out to discriminate among the two inserted gene fragments (GFP and RFP) and an intact allele of the *gap43* gene. As shown in Fig. 2-7b, in the group “G+/R-”, a GFP allele and an intact allele were detected, indicating that one of the *gap43* alleles was disrupted by the insertion of the GFP gene. Similarly, one of the *gap43* alleles was disrupted by the insertion of the RFP gene in the group “G-/R+”. In the group “G+/R+”, both the GFP and

RFP genes were detected in the *gap43* locus, indicating that both alleles were disrupted by the inserted genes. These results showed that, by mating two fish harboring different fluorescent color of reporter gene in a targeted locus, the genotypes of their progenies (fish without any mutations, monoalleleic mutants, and biallelic mutants) can be easily determined only by fluorescent observations at the embryonic stage.

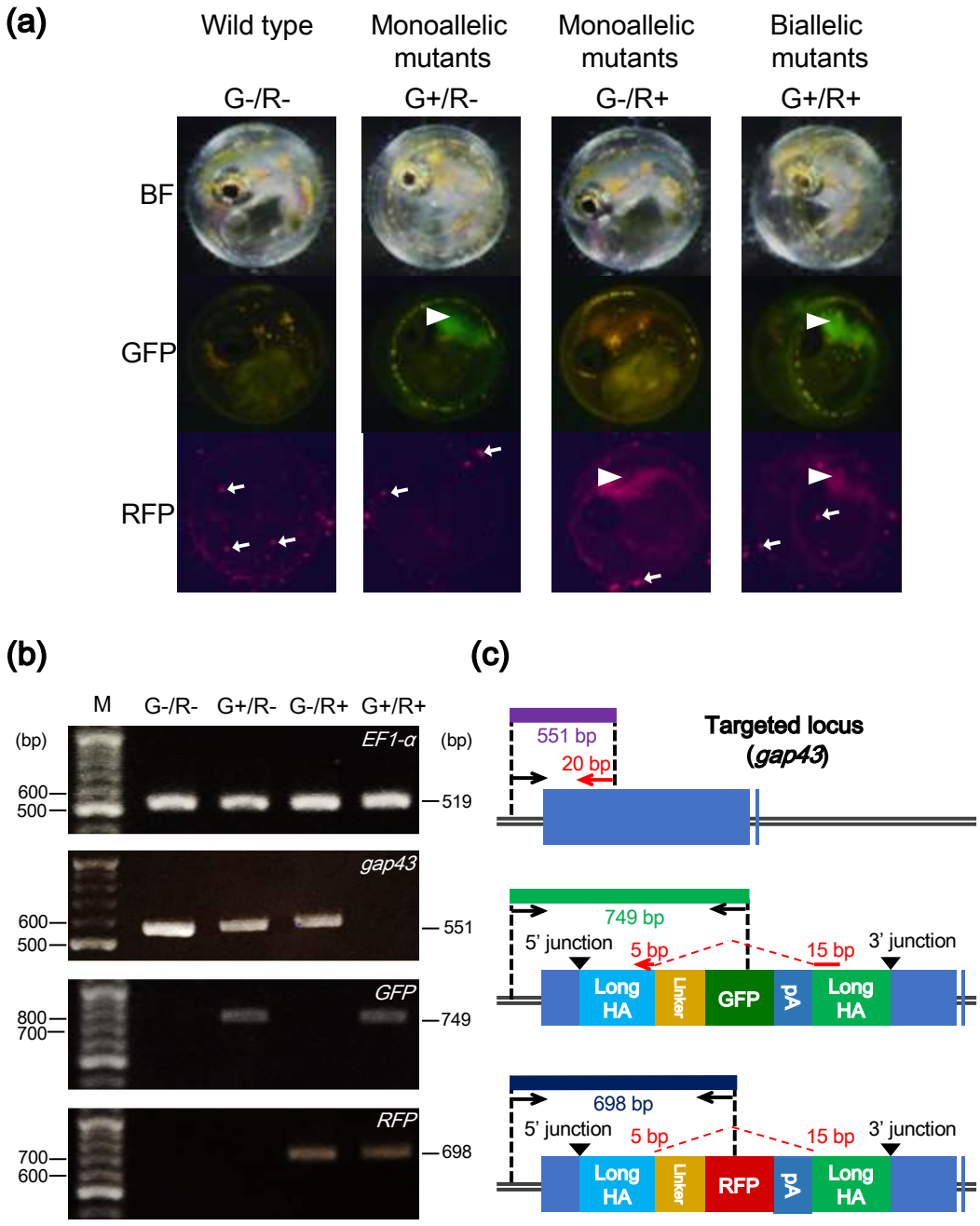


Fig. 2-7. Selection of biallelic mutants using two different colors of fluorescence. (a) Fluorescence of F2 embryos derived from mating with each F1 monoallelic mutants harboring either the GFP or the RFP gene in the gap43 locus. Each of the F2 embryos were categorized into the following four groups; no fluorescence (G－/R–), green fluorescence (G+/R–), red fluorescence (G－/R+), and both green and red fluorescence (G+/R+). The images were captured at 4 days post fertilization (dpf). White arrowheads show fluorescence in the central nervous system (CNS), while white arrows show autofluorescence of bacteria on the surface of the egg membrane. (b, c) PCR genotyping of the F2 embryos derived from mating between the F1 monoallelic mutants. (b) The EF1-α fragment (519 bp) was used to confirm genomic DNA extraction, and was detected in all samples. The gap43 fragment (551 bp) was detected from the groups (G－/R–), (G+/R–), and (G－/R+). The GFP fragment (749 bp) was detected from (G+/R–) and (G+/R+) groups, while the RFP fragment (698 bp) was detected from (G－/R+) and (G+/R+) groups. In (G+/R+), no PCR product of gap43 was detected. (c) Design of PCR primers to detect the targeted gene integration into the gap43 locus. (Upper): Genomic structure of the intact gap43 locus. The size of the PCR product amplified using primer pair (GAP43-for-Seq-Fw and GAP43-Rv) is 551 bp. The latter primer is 20 bp in size and is shown as a red arrow. The reverse primer is designed on the boundary between upstream and downstream long homology arms, so that this primer pairs can amplify the intact locus but not the GFP- or RFP-integrated loci. (Middle): Genomic structure of gap43 locus at which the donor plasmid harboring the GFP gene is integrated. The size of the PCR product amplified using the primer pair (GAP43-for-Seq-Fw and mAG-Rv) is 749 bp. (Lower): Genomic structure of the gap43 gene at which the donor plasmid harboring the RFP gene is integrated. The size of the PCR product amplified using the primer pair (GAP43-for-Seq-Fw and tdTomato-Rv) is 698 bp.

2.4 Discussion

Previous studies suggested that the length of homologous sequences of donor plasmids is important to determine which DSB repair pathways can be induced by the targetable nuclease systems [7]. Targeted gene integration mediated by MMEJ using plasmids with short homologous arms (10-40 bp) has been applied in cultured cells, frogs, mice, and zebrafish because those short arms can be inserted into the donor plasmids easily by PCR or oligonucleotide annealing [8,9,11,25]. Although our results showed that the donor plasmid with 40 bp of homologous arms slightly improved the integration efficiency than that with 20 bp of arms, the efficiencies were much lower than the plasmid with longer homology arms (~500 bp each) in medaka (Table 2-3). These results indicate that donor plasmids with longer homology arms (ca. 500 bp) become a good system for targeted gene integration in medaka. A previous study showed that MMEJ is active during G1 and early S phases when HDR is inactive [25]. Also, it is known that some molecules are involved in MMEJ but not in HDR [7]. These evidences indicate that the activity of each DSB repair pathway induced by the nuclease systems may vary among species and the developmental stages. Thus, specifying the highly active pathways in species and/or developmental stages of interest will be required to establish a highly efficient system for targeted gene integration in each model system.

Previous studies demonstrated that simultaneous cleavage of the bait sequences of a donor plasmid and a genomic target site by targetable nucleases can improve targeted integration efficiency of the plasmid [10,13]. In my experiments, donor plasmids that were cleaved by the bait systems also increased their integration efficiencies compared to that of a donor plasmid without any targeted cleavages (Fig. 2-3c). This result indicates

that linearization of donor plasmids using the bait systems is effective to enhance integration efficiencies by HDR in medaka. In addition, I designed a new bait sequence BaitD and demonstrated successful targeted gene integration using donor plasmids harboring the bait sequences. My data indicate that the BaitD system become a good system for targeted gene integration with high efficiency in medaka and other fish species because of the following reasons. 1) In a *in vivo* screening by gene knockin in medaka embryos, the BaitD system showed the highest integration efficiency among 8 bait sequences tested in this study (Fig. 2-3c). 2) Unlike Gbait, the BaitD system can be used in previously established GFP-transgenic strains because there is no potential target site of the sgRNA for the BaitD in *EGFP* gene. 3) *In silico* screening in the reference genomes of 12 fish species showed that the BaitD system has less potential off-target effects in the fish genomes (Table 2-4 and 5).

So far, to know the genotype of the targeted locus in the genome-edited animals, the genomic DNA extraction from a piece of their tissue and the subsequent PCR-based genotyping works are required. However, these processes injure the tissues of individuals and are laborious and time-consuming. In this study, I demonstrated a simple method to genotype the genome-edited fish with two different colors of fluorescent protein genes that are inserted at the target locus (Fig. 2-7a). With this method, I can determine the genotype of each individual simply only by observing the fluorescence without invasion and sacrifice of individuals and laborious works such as genomic DNA preparation. This is especially advantageous for embryos and larva which are difficult to prepare genomic DNA alive, for example, the genotyping of mutants with embryonic lethal phenotypes before their death and the selection of fish harboring desired mutations at the early stages to reduce the costs for breeding of fish. Thus, generations of the gene knockin strains

harboring the two different colors of fluorescent protein genes will also become an effective approach for targeted mutagenesis in medaka and others.

Although some G_0 fish injected with components to target the *gap43* locus transmitted the genes integrated at the target site to their progenies, the transmission rates are lower than those of other methods to mimic the endogenous gene expressions by reporter genes. One example is the homology-independent gene knockin using donor plasmids with a hsp70 promoter and a reporter gene in zebrafish [27]. The study demonstrated that 5.0-10.0 % of injected embryos showed broad reporter gene expression and 30.0-40.0% of them transmitted the gene to next generation; that is, 1.5-4.0% of injected embryos became transgenic founders. In my study, targeted integration of reporter genes into the *gap43* locus showed that only 0.6-1.0% of injected embryos transmitted to their progenies (Table 2-6). The lower efficiency of this event would be caused by the usage of promoter-less constructs in this study. While the constructs harboring a promoter can express when they are integrated in any direction and frame, in-flame integrations into the genomic target site are required for expression of the promoter-less constructs. Therefore, even if the constructs with or without a promoter are integrated at the similar efficiencies, the integration efficiency of the promoter-less constructs can be lower than that of constructs with promoter. In another example, bacterial artificial chromosome (BAC)-based transgenes were transmitted with high germline transmittion rate (~15%) mediated by Tol2 transposon in zebrafish [28]. Although the integration rate of my method is lower than that of the BAC transgenesis, my gene knockin system has an advantage that the transgenes can be precisely integrated into the target site without any positional effects. However, considering these lower rates of the germline transmission events in my system, further studies will be needed to

establish a method for improving the efficiency of targeted gene integration in medaka.

Here, I have established an efficient system for targeted gene integration in medaka. This system can be applicable to the generation of medaka strains with a wide range of genetic modifications. In addition to the gene tagging and knockout by fluorescent reporter genes as shown in this study, this system will allow for generation of conditional knockout strains by inserting target sites of site-specific recombinase (e.g. LoxP sites for Cre recombinase or FRT sites for Flp) at the genomic targeted positions and for precise introduction of site-specific point mutations in the genes of interest [10,12]. These types of advanced genome editing by HDR-mediated targeted gene integration will be useful to accelerate detailed analysis of the gene functions in medaka. Recently, targeted mutagenesis with small indels by the targetable nucleases was also demonstrated in a wide range of fish species in some similar ways to medaka and zebrafish [26–35]. These indicates that my findings in improving efficiency of the targeted gene integration can be effective for establishing the targeted gene integration system in those fish species.

Chapter2 References

1. Peng Y, Clark KJ, Campbell JM, Panetta MR, Guo Y, Ekker SC. (2014) Making designer mutants in model organisms. Development. 141:4042–54.
2. Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD. (2010) Genome editing with engineered zinc finger nucleases. Nat Rev Genet. 11:636–46.
3. Ansai S, Sakuma T, Yamamoto T, Ariga H, Uemura N, Takahashi R, et al. (2013) Efficient targeted mutagenesis in medaka using custom-designed transcription activator-like effector nucleases. Genetics. 193:739–49.

4. Ansai S, Inohaya K, Yoshiura Y, Schartl M, Uemura N, Takahashi R, et al. (2014) Design, evaluation, and screening methods for efficient targeted mutagenesis with transcription activator-like effector nucleases in medaka. Develop Growth Differ. 56:98–107.
5. Ansai S, Kinoshita M. (2014) Targeted mutagenesis using CRISPR/Cas system in medaka. Biol Open. 3:362–71.
6. Stemmer M, Thumberger T, Del Sol KM, Wittbrodt J, Mateo JL. (2015) CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. PLoS One. 10:e0124633.
7. McVey M, Lee SE. (2008) MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. Trends Genet. 24:529–38.
8. Nakade S, Tsubota T, Sakane Y, Kume S, Sakamoto N, Obara M, et al. (2014) Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. Nat Commun. 5:5560.
9. Sakuma T, Nakade S, Sakane Y, Suzuki K-IT, Yamamoto T. (2016) MMEJ-assisted gene knock-in using TALENs and CRISPR-Cas9 with the PITCh systems. Nat Protoc. 11:118–33.
10. Irion U, Krauss J, Nusslein-Volhard C. (2014) Precise and efficient genome editing in zebrafish using the CRISPR/Cas9 system. Development. 141:4827–30.
11. Hisano Y, Sakuma T, Nakade S, Ohga R, Ota S, Okamoto H, et al. (2015) Precise inframe integration of exogenous DNA mediated by CRISPR/Cas9 system in zebrafish. Sci Rep. 5:8841.
12. Hoshijima K, Jurynec MJ, Grunwald DJ. (2016) Precise Editing of the Zebrafish Genome Made Simple and Efficient. Dev Cell. 36:654–67.

13. Ochiai H, Sakamoto N, Fujita K, Nishikawa M, Suzuki K, Matsuura S, et al. (2012) Zinc-finger nuclease-mediated targeted insertion of reporter genes for quantitative imaging of gene expression in sea urchin embryos. Proc Natl Acad Sci U S A. 109:10915–20.
14. Xu H, Xiao T, Chen CH, Li W, Meyer CA, Wu Q, et al. (2015) Sequence determinants of improved CRISPR sgRNA design. Genome Res. 25:1147–57.
15. Bae S, Park J, Kim JS. (2014) Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics. 30:1473–5.
16. Auer TO, Duroure K, De Cian A, Concordet JP, Del Bene F. (2014) Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. Genome Res. 24:142–53.
17. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, et al. (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol. 31:827–32.
18. Okuyama T, Isoe Y, Hoki M, Suehiro Y, Yamagishi G, Naruse K, et al. (2013) Controlled Cre/loxP site-specific recombination in the developing brain in medaka fish, Oryzias latipes. PLoS One. 8:e66597.
19. Moreno-Mateos MA, Vejnar CE, Beaudoin J, Fernandez JP, Mis EK, Khokha MK, et al. (2015) CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. Nat Methods. 12:982–8.
20. Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, et al. (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat Biotechnol. 31:227–9.
21. Kinoshita M, Kani S, Ozato K, Wakamatsu Y. (2000) Activity of the medaka

translation elongation factor 1alpha-A promoter examined using the GFP gene as a reporter. Develop Growth Differ. 42:469–78.

22. Kusakabe R, Kusakabe T, Suzuki N. (1999) In vivo analysis of two striated muscle actin promoters reveals combinations of multiple regulatory modules required for skeletal and cardiac muscle-specific gene expression. Int J Dev Biol. 43:541–54.
23. Kinoshita M. (2004) Transgenic medaka with brilliant fluorescence in skeletal muscle under normal light. Fish Sci. 70:645–9.
24. Fujimori KE, Kawasaki T, Deguchi T, Yuba S. (2008) Characterization of a nervous system-specific promoter for growth-associated protein 43 gene in Medaka (*Oryzias latipes*). Brain Res. 1245:1–15.
25. Aida T, Nakade S, Sakuma T, Izu Y, Oishi A, Mochida K, et al. (2016) Gene cassette knock-in in mammalian cells and zygotes by enhanced MMEJ. BMC Genomics. 17:979.
26. Taleei R, Nikjoo H. Biochemical DSB-repair model for mammalian cells in G1 and early S phases of the cell cycle. (2013) Mutat. Res. - Genet. Toxicol. Environ. Mutagenesis. 756:206–12.
27. Kimura Y, Hisano Y, Kawahara A, Higashijima S-I. (2014) Efficient generation of knock-in transgenic zebrafish carrying reporter/driver genes by CRISPR/Cas9-mediated genome engineering. Sci Rep. 4:6545.
28. Suster ML, Abe G, Schouw A, Kawakami K. (2011) Transposon-mediated BAC transgenesis in zebrafish. Nat Protoc. 6:1998–2021.
29. Yano A, Guyomard R, Nicol B, Jouanno E, Quillet E, Klopp C, et al. (2012) An immune-related gene evolved into the master sex-determining gene in rainbow trout, *Oncorhynchus mykiss*. Curr Biol. 22:1423–8.

30. Edvardsen RB, Leininger S, Kleppe L, Skaftnesmo KO, Wargelius A. (2014) Targeted mutagenesis in Atlantic salmon (*Salmo salar* L.) using the CRISPR/Cas9 system induces complete knockout individuals in the F0 generation. PLoS One. 9:e108622.
31. Li M, Yang H, Zhao J, Fang L, Shi H, Li M, et al. (2014) Efficient and heritable gene targeting in tilapia by CRISPR/Cas9. Genetics. 197:591–9.
32. Harel I, Benayoun BA, Machado B, Singh PP, Hu CK, Pech MF, et al. (2015) A platform for rapid exploration of aging and diseases in a naturally shortlived vertebrate. Cell. 160:1013–26.
33. Ma L, Jeffery WR, Essner JJ, Kowalko JE. (2015) Genome editing using TALENs in blind Mexican Cavefish, *Astyanax mexicanus*. PLoS One. 10:e0119370.
34. Erickson PA, Ellis NA, Miller CT. (2016) Microinjection for Transgenesis and Genome Editing in Threespine Sticklebacks. J Vis Exp. 111:e54055.
35. Cui Z, Liu Y, Wang W, Wang Q, Zhang N, Lin F, et al. (2017) Genome editing reveals dmrt1 as an essential male sex-determining gene in Chinese tongue sole (*Cynoglossus semilaevis*). Sci Rep. 7:42213.

Chapter3

Targeted gene integration into a lethal gene using Cas9 nickase

3.1 Introduction

The CRISPR/Cas9 system has become a powerful technology for reverse genetics approaches, allowing to modify targeted genes simply and efficiently in a wide range of organisms. In medaka, the previous studies have demonstrated that this system enabled the efficient gene modifications including targeted gene disruptions and targeted gene integrations [1,2]. However, despite of its efficiency and robustness, it still difficult to integrate foreign genes into the target genes essential for survival. Indeed, all of the F_0 knock-in medaka expressing green fluorescence in skeletal muscle died before or soon after hatching as described in chapter2 of this study (Fig. 2-3). The cause of the death is presumed to be muscular dysfunction resulting from the biallelic mutations on *acta1* locus. Excess knockout of such genes essential for maintaining life makes lethal effects. In my previous experiments for the targeted gene integration into *acta1* locus with the CRISPR/Cas9 system, the generated genetic mutations on each allele are largely classified into small insertions or deletions (indels) of several nucleotides mediated by NHEJ or integration of the foreign gene mediated by HDR. Since NHEJ is highly active throughout the cell cycle comparing to HDR [3], these indels are frequently occurred leading to make undesirable side effects such as embryonic or larval lethality as shown in the study. Thus, to perform the HDR-mediated knock-in targeting lethal genes without killing individuals, a new knock-in method that reduces unwanted NHEJ-mediate small

indels at the target genes is needed.

As a method to selectively induce HR without stimulating NHEJ, the use of targeted single strand breaks (nicks) attracts attention. The nicks can be generated by the Cas9 nickases (Cas9n D10A and Cas9n H840A), which are the variants of Cas9 containing a single amino acid substitution [4]. The Cas9n D10A or Cas9n H840A allows to generate the site-specific nick on the gRNA complementary or non-complementary DNA single strand, respectively [5]. The generated nicks on the targeted site have been reported to be repaired mainly by HR and seldom by NHEJ in human cells [6]. This fact prompted me to use the Cas9n for the HR-mediated gene integration without inducing excess gene disruption by NHEJ. However, no studies have achieved the integration of the exogeneous genes into the genomic targeted locus with the Cas9n in teleost fish. Thus, the detailed knowledge is required to establish efficient protocols for this technology.

The simultaneous cleavage of the targeted gene and the donor plasmid by targetable nucleases enables to enhance the HR-mediated gene integration efficiency in sea urchin and zebrafish embryos [7,8], although the detailed mechanism remains unclear. It was also confirmed in medaka that linearization of circular donor plasmids by the cleavage of bait sequences on the plasmids is essential for the efficient HR-mediated knock-in in medaka as shown in chapter2. However, Cas9 can induce DNA double strand breaks at the target site, allowing to linearize the donor plasmid containing one pair of BaitD sequence, whereas Cas9n, which can cleave only the single strand of DNA, cannot linearize the conventional donor plasmid. To overcome this problem, it is necessary to develop a new donor plasmid that can be linearized by Cas9n-mediated nicking.

Here, I aimed to show the capability of the Cas9n (Cas9 D10A) to establish gene knock-in strains targeting lethal genes in medaka. Additionally, I sought an efficient

method for the HR-mediated targeted gene integration with Cas9n by constructing a novel donor plasmid (p2BaitD) containing two pairs of BaitD sequences. Because BaitD has been confirmed to achieve the efficient gene knock-in in medaka and minimize the off-target effects in 12 fish species in chapter2, a new knock-in method using Cas9n and the p2BaitD plasmid can contribute to achieve a flexible and efficient gene modification targeting lethal genes in medaka and other fish species.

3.2 Materials and Methods

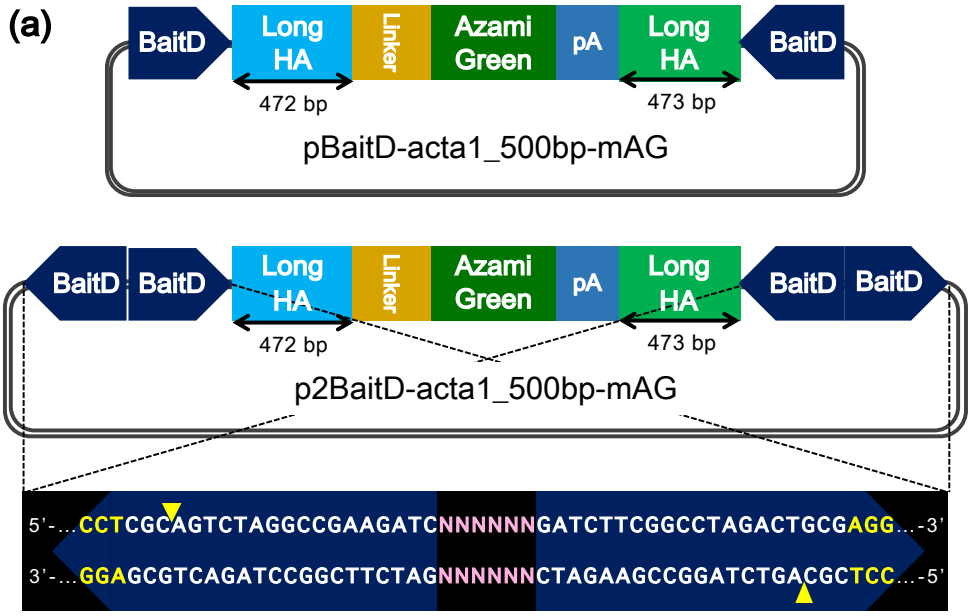
Fish

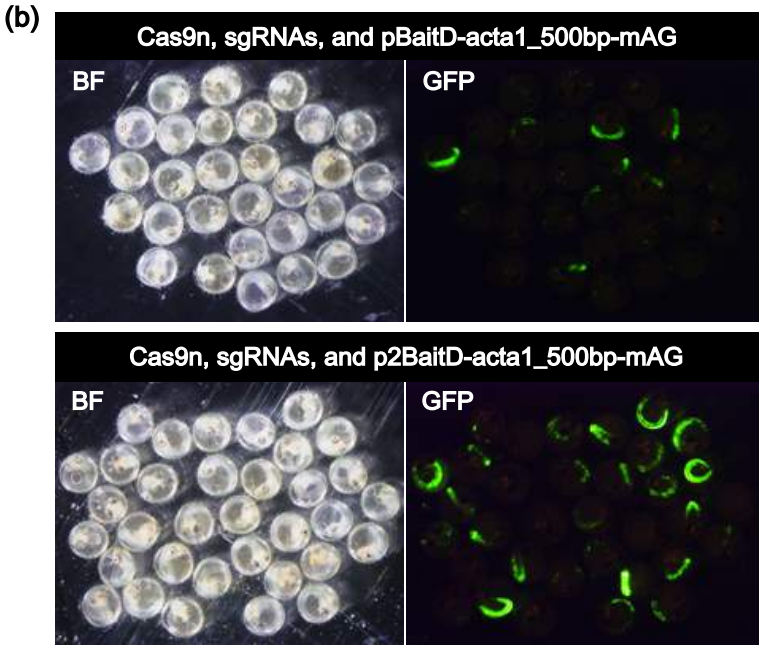
The Cab strain was used in this study. Fish were maintained in an aquarium with recirculating water in a 14/10-h light/dark cycle at 26 °C. The use and care of fish in this study were in accordance with the guidelines of the Animal Experimentation Committee of Kyoto University.

Construction of donor plasmids

The donor plasmid pBaitD-acta1_500 bp-mAG containing one pair of BaitD sequence as described in chapter2 (Fig. 3-3a) was used a control to evaluate the effect of the linearization of a novel donor plasmid p2BaitD by Cas9n on the gene knock-in efficiency. To construct the novel plasmid (p2BaitD-acta1_500 bp-mAG: Fig. 3-3a) containing two pairs of BaitD sequence, backbone fragments containing a pUC replication origin, an ampicillin resistance, and BaitD sequences were amplified from the plasmid pBaitD-acta1_500 bp-mAG by PCR using a primer pair (Vec-PstI-acta1-FW/Vec-Asp718I-acta1-RV) (Table 3-1). An insert fragment containing two BaitD sequences, homology

arms, mAG with a N-terminal linker, and a SV40 polyA, was amplified from the plasmid pBaitD-acta1_500 bp-mAG by PCR using a primer pair (In-Asp718I-acta1-FW/ In-PstI-acta1-RV) (Table 3-1). These fragments were digested with Asp718I and PstI, and ligated using a Ligation high Ver.2 Kit (Toyobo). Both plasmids were purified by NucleoSpin Plasmid QuickPure kit (MACHEREY-NAGEL).





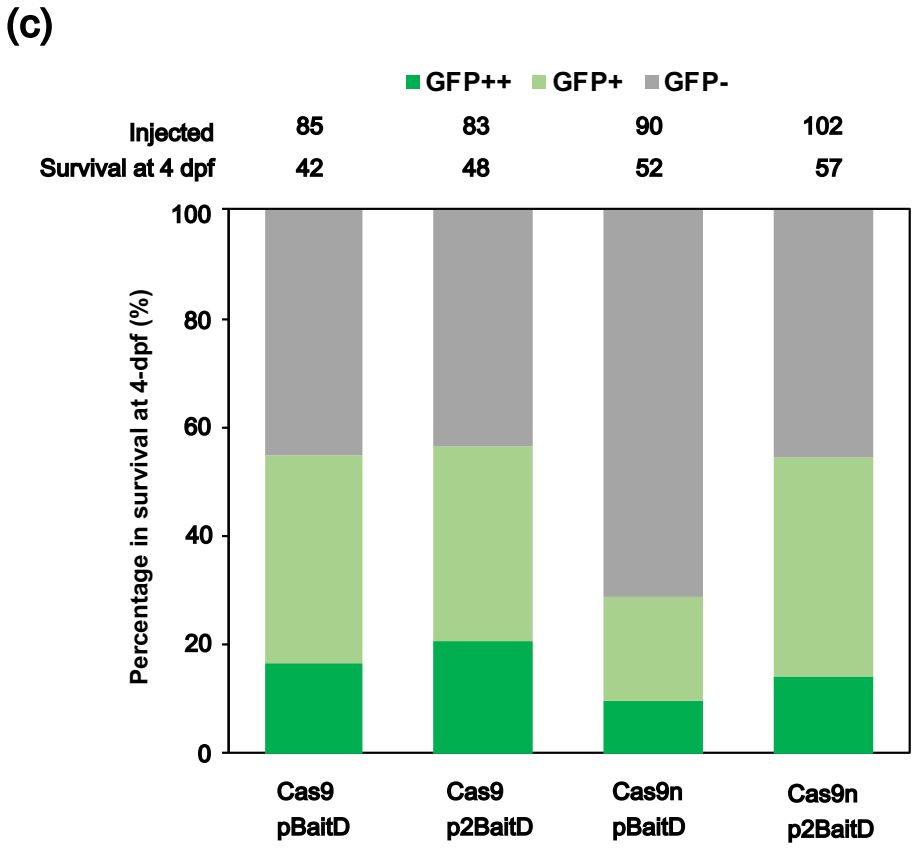


Fig. 3-3. Targeted gene integration of GFP into *acta1* locus. **(a)** Schematics of each donor plasmid for evaluating the effects of two pairs of BaitD sequence. The pBaitD-acta1_500bp-mAG plasmid contains long homology arms (Long HAs) and one pair of BaitD. The p2BaitD-acta1_500bp-mAG plasmid contains two pairs of BaitD, which are shown the DNA sequences at the lower side. White and yellow letters show the target sequence and AGG protospacer adjacent motif (PAM), respectively. Yellow triangles show the nicked sites by Cas9n. Pink letters between BaitD sequences (NNNNNN) show each of restriction enzyme sites, Asp718I (GGTACC) and PstI (CTGCAG), used for the ligation reaction of two DNA fragments, respectively. **(b)** Observation of green fluorescence in injected embryos at 4 days post fertilization (4 dpf). The p2BaitD-acta1_500bp-mAG plasmid (lower) is more efficiently integrated at *acta1* locus than the pBaitD-acta1_500bp-mAG plasmid (upper). BF: bright field, GFP: GFP fluorescent image. **(c)** Comparison of the knock-in efficiencies among each of donor plasmids combined Cas9 or Cas9n

at 4 dpf. The p2BaitD-acta1_500bp-mAG plasmid or the pBaitD-acta1_500bp-mAG plasmid is shown as “p2BaitD” or “pBaitD”, respectively. The number of injected eggs is shown as “Injected”, and the number of survived embryos at 4 days post fertilization (dpf) among the injected eggs is shown as “Survival at 4 dpf”.

Table 3-1. Oligonucleotide sequences used in this study.

Name	Sequence (5'-3')	Usage
Vec-PstI-acta1-FW	GCTCTGCAGGATCTTCGGCCTAGACTGCGAGGAGCAAAAGGCCAGCAAAAGGC	Construction of plasmid
Vec-Asp718I-acta1-RV	GCAGGTACCGATCTTCGGCCTAGACTGCGAGGCACTACGTGTATCCGCTCATG	Construction of plasmid
In-Asp718I-acta1-FW	GCTGGTACCGATCTTCGGCCTAGACTGCGAGGCTCGAGCGTGTT	Construction of plasmid
In-PstI-acta1-RV	GCACTGCAGGATCTTCGGCCTAGACTGCGAGGACTAGTTCACAC	Construction of plasmid

Preparation of donor plasmids and sgRNA

To prevent from degrading sgRNAs by residual RNase in the plasmid solutions, the donor plasmids were dissolved with 50 μL of 5 mM Tris-HCl buffer (pH 8.5) were incubated with 5 μL of 10% sodium dodecyl sulfate (SDS) and 2 μL of Proteinase K (20 mg/mL) at 55 °C for 30 min, and then purified using NucleoSpin Gel and PCR Clean-up kit (MACHEREY-NAGEL) with the Buffer NTB. The sgRNA targeting *acta1* locus, sgRNA-acta1 #1 (Fig. 2-1), was synthesized as described previously [1]. Briefly, a pair of oligonucleotides (Table 2-2) was annealed and ligated into the pDR274 vector (Addgene Plasmid #42250) digested with BsaI-HF (New England Biolabs). The sgRNA expression vector was digested by DraI, and used for the DNA template to synthesis sgRNA with the AmpliScribe T7-Flash Transcription Kit (Epicentre). The synthesized sgRNA was purified with RNeasy Mini Kit (Qiagen) for microinjection.

Microinjection

To evaluate the DNA double strand breaks activity of Cas9 or Cas9n at *acta1* locus, an

injection mixture containing 500 ng/μL of Cas9 or Cas9n protein (Integrated DNA Technology) and 100 ng/μL of sgRNA-acta1 #1 was prepared. To examine the efficiency of gene knock-in at *acta1* locus, an injection mixture containing 500 ng/μL of Cas9 or Cas9n protein (Integrated DNA Technology), 100 ng/μL of sgRNA-acta1 #1 and sgRNA targeting BaitD sequences, and 2.5 ng/μL of the donor plasmid p2BaitD-acta1_500 bp-mAG or pBaitD-acta1_500 bp-mAG was prepared. These mixtures were injected into the cytosol of each egg at the one-cell stage, as described previously [9].

Genomic DNA extraction

Embryos at 4 days post fertilization (dpf) were lysed individually in 25 L of alkaline lysis buffer containing 25 mM NaOH and 0.2 mM EDTA and incubated at 95 °C for 15 min after breaking the egg envelope with forceps. Samples were neutralized with 25 μL of 40 mM Tris-HCl (pH 8.0) and used as genomic DNA samples.

Heteroduplex mobility assay

Heteroduplex mobility assay (HMA) was conducted to detect the genomic mutation on the target site [10]. The DNA fragment containing the target site (235 bp) was amplified with KOD-FX (Toyobo) and a primer set acta1-HMA-FW and acta1-HMA-RV (Table 2-2). The reaction mixture was prepared as follows: 1 μL of genomic DNA as template, 1× PCR buffer for KOD FW, 0.4 mM of each dNTP, 0.2 μM of each primer, and 0.05 unit of KOD FX (Toyobo) in a total volume of 10 μL. The cycle of the reaction was as follows: one cycle at 94 °C for 2 min followed by 35 PCR cycles of 98 °C for 10 s, 58 °C for 30 s and 68 °C for 5 s. The resulting amplicons were analyzed using a microchip electrophoresis system (MCE-202 MultiNA; Shimadzu, Kyoto, Japan) with DNA-500

reagent kit.

Observation of GFP fluorescence

The fluorescence of the skeletal muscle was observed using a fluorescence stereomicroscope MZFLIII (Leica Microsystems) with a GFP2 filter set. Microscopic images were captured using a digital color-cooled charge-coupled camera and the VB-7010 image control system (Keyence).

Sequence analysis

To evaluate the precise targeted integration, the genomic region including the integration site on *acta1* locus was investigated. The junction regions of the target site on the host genome and introduced gene were amplified by PCR using the following primer pairs: acta1-for-Seq-Fw and mAG-Rv, mAG-Fw and acta1-for-Seq-Rv (Table 2-2), and KOD FX DNA polymerase (Toyobo). The PCR was run as follows: one cycle at 94 °C for 2 min followed by 35 PCR cycles of 98 °C for 10 s, 58 °C for 30 s and 68 °C for 30 s. The resulting PCR products were subjected to electrophoresis with a 1% agarose gel. PCR fragments predicted to contain the introduced gene were excised from the gel and purified using NucleoSpin Gel and PCR Clean-up (MACHEREY-NAGEL). The purified fragments were sequenced using the primers acta1-for-Seq-Fw and acta1-for-Seq-Rv (Table 2-2).

3.3 Results

Lethal effects induced by targeted gene integration using Cas9

To evaluate the lethal effects generated by gene knock-in targeting *acta1* locus, microinjection of mixtures containing Cas9, sgRNA-acta1 #1 (Fig. 2-1), sgRNA targeting BaitD, and the donor plasmid pBaitD-acta1_500 bp-mAG into 85 eggs was performed. Green fluorescence was clearly observed in the skeletal muscle of 23 of 85 injected eggs at 4 days post fertilization (dpf), however, no embryos expressed GFP in the skeletal muscle could be raised to adulthood (Fig. 3-1 and 3c, Table 3-2). The result indicates that excess biallelic mutations on *acta1* locus led to death of the injected embryos. The genetic mutations on each allele are classified into two groups including NHEJ-mediated small indels or HR-mediated gene integration. Then, to alleviate the lethal effects by the biallelic mutations, I focused on Cas9n that may largely reduce the undesirable mutations mediated by NHEJ.

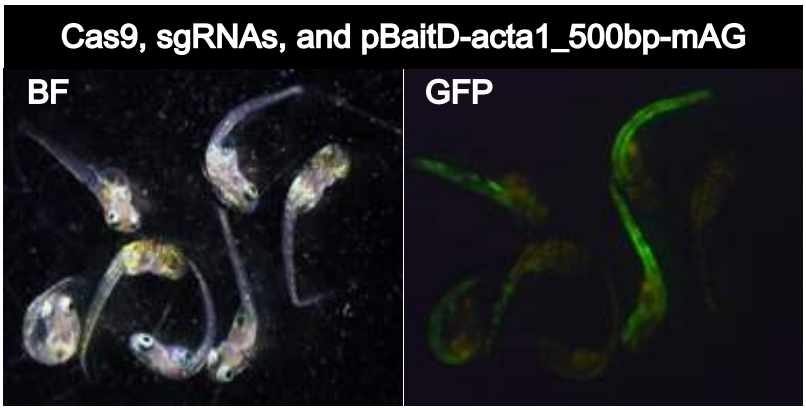


Fig. 3-1. Malformation in F_0 knock-in larvae at 9 days post fertilization (dpf). These larvae were injected with the pBaitD-acta1_500bp-mAG plasmid, Cas9 protein, sgRNAs targeting BaitD and *acta1*. BF: bright field, GFP: GFP fluorescent image.

Table 3-2. Results of microinjection targeting the *gap43* locus with GFP or RFP

Donor plasmid	Cas9	Cas9n	Injected	Sexually matured	Germ-line transmission rate
pBaitD	+	-	85	0	-
p2BaitD	+	-	83	0	-
pBaitD	-	+	90	8	12.5% (1/8)
p2BaitD	-	+	102	14	14.3% (2/14)

Single nicking by Cas9n-sgRNA does not generate any mutations on the genomic targeted site

To confirm that Cas9n and sgRNA targeting *acta1* gene do not activate NHEJ at the locus in medaka, the indel inducing efficiency with the injection of Cas9n mixture was compared with the injection of Cas9 mixture. HMA analysis using the injected eggs revealed that Cas9 and the sgRNA can efficiently generate small indels on *acta1* locus, whereas Cas9n and the sgRNA cannot induce any mutation on the locus corresponding to the result of a negative control using an un-injected egg (Fig. 3-2). The results show the inability of the sgRNA-Cas9n complex to stimulate NHEJ on the genomic targeted locus. Thus, Cas9n has the potential to use for the targeted gene integration mediated by HR without generating unnecessary small indels mediated by NHEJ on the genomic targeted site in medaka.

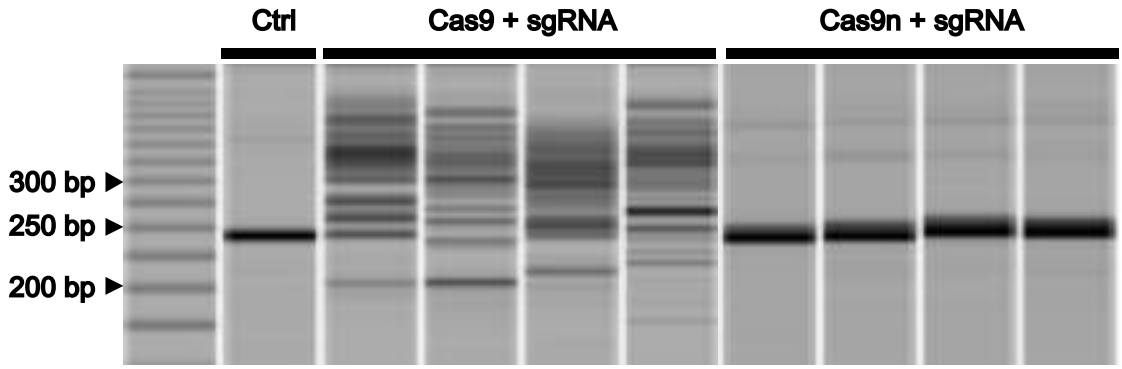


Fig. 3-2. Evaluation of the genomic mutations on *acta1* using HMA analysis. The embryos injected with a mixture containing sgRNA targeting *acta1* and Cas9 or Cas9n were subjected. Control (Ctrl) shows the result from an embryo without injection.

Establishment of a gene knock-in strain harboring a transgene on the lethal locus by Cas9n

To show that Cas9n allows to establish a gene knock-in strain targeting the lethal locus, microinjection of mixtures including Cas9n, sgRNA-acta1 #1, sgRNA recognizing BaitD, and the donor plasmid pBaitD-acta1_500 bp-mAG into 90 eggs was performed. Expression of green fluorescence in the skeletal muscle was detected at 4 dpf (Fig. 3-3b), showing the successful gene knock-in at *acta1* locus using Cas9n. Interestingly, of 90 injected eggs, 8 individuals with GFP expression in the skeletal muscle could be raised to adulthood, demonstrating that Cas9n enables to survival of the individuals that could not be achieved with Cas9 (Fig. 3-3c, Table 3-2). Mating the GFP-expressing 8 founders with wild-type fish, one individual transmitted the insert gene to the next generation (Table 3-2 and 3). To investigate the precision of the gene knock-in events at *acta1* locus, I sequenced the DNA surrounding the inserted fragment after PCR amplification using genomic DNA extracted from five F1 embryos with GFP fluorescence in the skeletal muscle (Fig. 3-4a and c). In all five embryos derived from one F_0 founder, the insert fragment was precisely integrated on the target site without any indels (Fig. 3-4b).

However, the gene integration rate of Cas9n at 4 dpf (28.9%) is lower than that of Cas9 (54.8%) (Fig. 3-3c: first and third column from left), thus, a new method that can achieve higher integration efficiencies using Cas9n was needed.

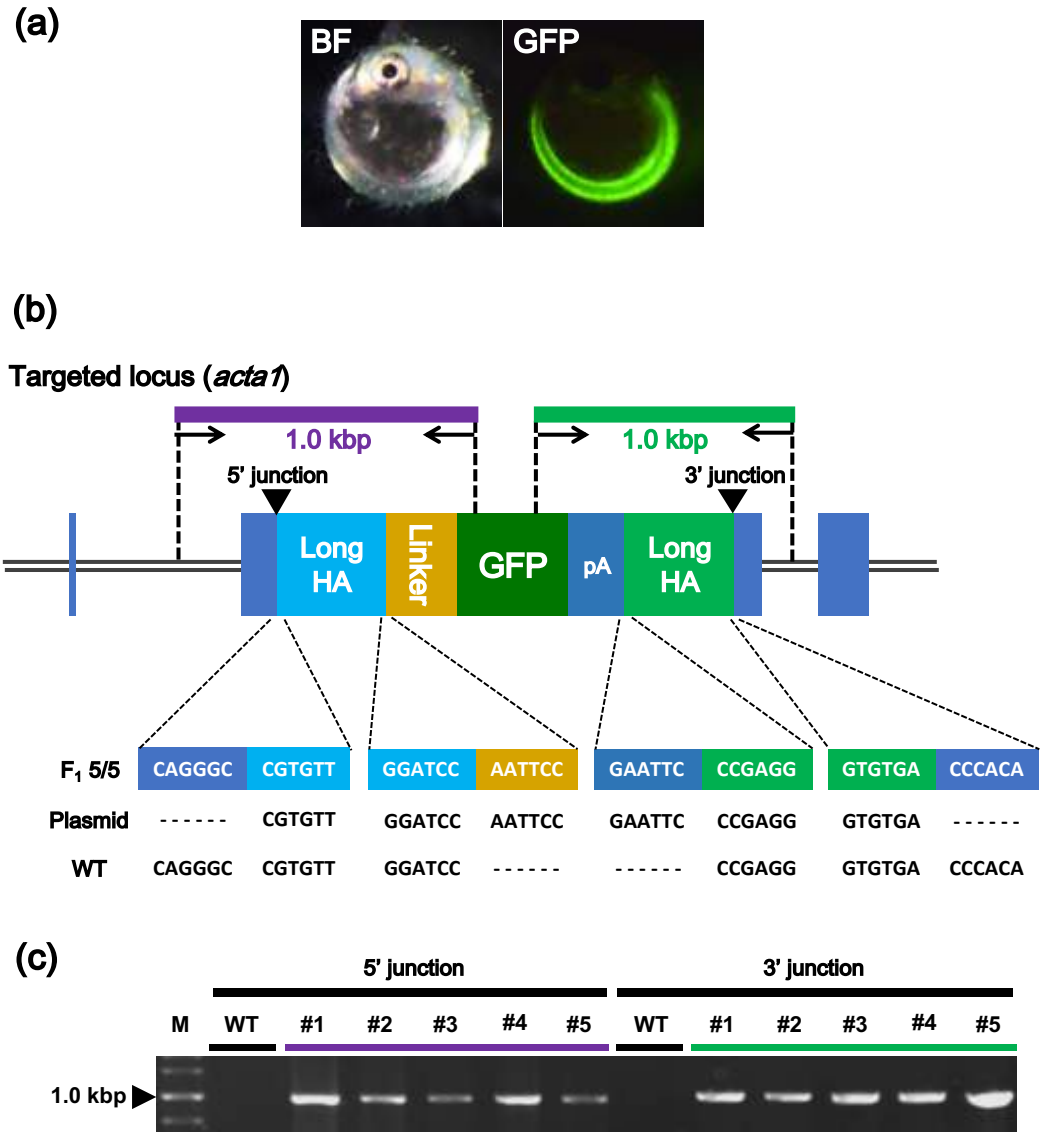


Fig. 3-4. Precise gene integration of a donor vector containing the GFP gene into the acta1 locus. **(a)** Green fluorescence of the F_1 embryo at 4 days post fertilization (dpf) derived from mating F_0 with and wild-type. **(b)** Schematic illustration of the locus. The primer pair (acta1-for-Seq-Fw and mAG-Rv or mAG-Fw and acta1-for-Seq-Rv) was used for investigating the precise integration into the acta1 locus. The amplicons containing 5′ junction or 3′ junction are 1.0 kbp, respectively, which were subjected to the sequence analysis. (Upper): The sequence observed in the F_1 fish with the inserted gene (#1–#5). (Middle): The sequence in the donor plasmid, pBaitD-acta1_500 bp-mAG (Plasmid). (Lower): The sequence in the wild type fish without the insertion

(WT). **(c)** Electrophoresis image of the PCR amplicons of each junction region on the host genome and introduced gene. The PCR amplicons (1.0 kbp) were detected from the F_1 embryos with GFP expression (#1–#5), but not from those of wild-type (WT).

Table 3-3. Germ-line transmition efficiency of F_0 individuals harboring GFP in acta1 locus

	Collected eggs	Expressed embryos	Transmit efficiency (%)
pBaitD#1	46	11	23.9
p2BaitD#1	42	9	21.4
p2BaitD#2	47	15	31.9

Transmit efficiency (%) = Expressed embryos / Collected eggs

Cas9n-mediated efficient gene knock-in with a donor plasmid containing two pairs of BaitD sequences

Since the conventional donor plasmid pBaitD-acta1_500 bp-mAG containing one pair of BaitD cannot be linearized by using Cas9n, the novel donor plasmid p2BaitD-acta1_500 bp-mAG containing two pairs of BaitD was constructed to be linearized by Cas9n (Fig. 3-3a). Mixtures containing the constructed donor plasmid p2BaitD-acta1_500 bp-mAG, Cas9n, sgRNA-acta1 #1, and sgRNA targeting BaitD was injected into 102 eggs. Surprisingly, the novel plasmid p2BaitD-acta1_500 bp-mAG can significantly increase the rate of GFP-expressing individuals in the skeletal muscle at 4 dpf (54.4%) compared to those with conventional donor plasmid pBaitD-acta1_500 bp-mAG (28.9%) (Fig. 3-3c: first and second column from right). Of 14 adult fish, 2 individuals transmitted the insert gene on the novel donor plasmid to the next generation (Table 3-2).

On the other hand, no significant difference of the GFP-expressing rate at 4 dpf

was observed between both plasmids, when combining Cas9 and each plasmid, pBaitD-acta1_500 bp-mAG (54.8%) or p2BaitD-acta1_500 bp-mAG (56.3%) (Fig. 3-3c: first and second column from left).

3.4 Discussion

In this study, I successfully established a novel gene knock-in system with Cas9n for lethal genes in medaka. Because Cas9n-gRNA complex didn't cause NHEJ (Fig. 3-2), this system could largely reduce the frequency of the unwanted small indels in each cell, resulting in the successful survival of the knock-in individuals (Table. 3-2). In addition to the gene tagging by the fluorescent gene as shown in the present study, this system enables other types of the HDR-mediated gene modifications including conditional knockout using recombinases such as Cre and Flp or introduction of point mutation in genes of interest. These flexible modifications for lethal genes by this system using Cas9n can broaden the possibilities for analyzing the gene functions even if the desired gene modified strains cannot be obtained using Cas9. It should be noted that, however, this system can induce biallelic mutations on targeted loci, because exogenous genes may be integrated into biallele mediated by HR. The biallelic mutations would make lethal or serious effects as shown in this study when the target gene is lethal gene. This problem indicates the need for further study to develop a new gene modified technology allowing to integrate foreign genes into the only monoallele.

I also found that linearization of the circular donor plasmid increased the efficiency of HDR-mediated gene integration. The new method using Cas9n and the p2BaitD plasmid can doubly enhance the efficiency of the gene knock-in comparing to

the method using Cas9n and the pBaitD plasmid (Fig. 3-3c: first and second column from right). On the other hand, the efficiencies of the gene knock-in were not different between two groups in the use of p2BaitD or pBaitD with Cas9 (Fig. 3-3c: first and second column from left), suggesting that p2BaitD can efficiently induce the targeted gene integration events when combined with Cas9n.

As shown in Table 3-2, the targeted integration of *GFP* gene into *acta1* locus using Cas9n resulted in transmission by only 1.1–2.0% of injected embryos to their progeny. This is no different than the transmission rate of reporter genes integrated into *gap43* locus by using Cas9 (0.6–1.0%) as described in chapter2. These efficiencies of the HDR-mediated gene integration are still low, comparing to other types of gene integration such as the NHEJ-mediated gene knock-in in zebrafish (1.5–4.0%) and the conventional transgenesis with the transposase Tol2 in zebrafish (~15%) [11,12]. Thus, a new technology allowing to enhance the efficiency of HDR-mediated gene knock-in is needed. Recently, successful trials have been reported for switching the DSB repair pathway from NHEJ to HDR by suppressing NHEJ key molecules such as KU70, KU80, and DNA ligase IV in rats and cell lines [13,14]. Combination of these molecules with the HDR-mediated gene knock-in system in the present study may be effective for enhancing the efficiency of germline transmission in medaka.

In conclusion, I established a universal and efficient strategy in the present study for increasing the efficiency of HDR-mediated gene integration in medaka through combination of Cas9n and the p2BaitD plasmid. This is the first report in teleost fish demonstrating that Cas9n is helpful for the HR-mediated gene knock-in without inducing the NHEJ-mediated gene disruption at the genomic targeted site. The established knock-in system in medaka will expand the possibility for functional study of genes at molecular

level, modeling of human genetic diseases, and molecular breeding of farmed fish in the future.

Chapter3 References

1. Ansai, S. & Kinoshita, M. (2014) Targeted mutagenesis using CRISPR/Cas system in medaka. Biol Open. 3: 362-71.
2. Murakami, Y., Ansai, S., Yonemura, A. & Kinoshita, M. (2017) An efficient system for homology-dependent targeted gene integration in medaka (*Oryzias latipes*) Zoological Lett. 3: 10.
3. Mao, Z., Bozzella, M., Seluanov, A. & Gorbunova, V. (2008) DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells. Cell Cycle. 7: 2902-6.
4. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A. & Zhang, F. (2013) Multiplex genome engineering using CRISPR/Cas systems. Science. 339: 819-23.
5. Mcconnell Smith, A., Takeuchi, R., Pellenz, S., Davis, L., Maizels, N., Monnat, R.J. & Stoddard, B.L. (2009) Generation of a nicking enzyme that stimulates site-specific gene conversion from the I-AniI LAGLIDADG homing endonuclease. Proc Natl Acad Sci U S A. 106: 5099-104.
6. Metzger, M.J., Mcconnell-Smith, A., Stoddard, B.L. & Miller, A.D. (2011) Single-strand nicks induce homologous recombination with less toxicity than double-strand breaks using an AAV vector template. Nucleic Acids Res. 39: 926-35.
7. Irion, U., Krauss, J. & Nusslein-Volhard, C. (2014) Precise and efficient genome

editing in zebrafish using the CRISPR/Cas9 system. Development. 141: 4827-30.

8. Ochiai, H., Sakamoto, N., Fujita, K., Nishikawa, M., Suzuki, K., Matsuura, S., Miyamoto, T., Sakuma, T., Shibata, T. & Yamamoto, T. (2012) Zinc-finger nuclease-mediated targeted insertion of reporter genes for quantitative imaging of gene expression in sea urchin embryos. Proc Natl Acad Sci U S A. 109: 10915-20.
9. Kinoshita, M., Kani, S., Ozato, K. & Wakamatsu, Y. (2000) Activity of the medaka translation elongation factor 1alpha-A promoter examined using the GFP gene as a reporter. Dev Growth Differ. 42: 469-78.
10. Ansai, S., Inohaya, K., Yoshiura, Y., Schartl, M., Uemura, N., Takahashi, R. & Kinoshita, M. (2014) Design, evaluation, and screening methods for efficient targeted mutagenesis with transcription activator-like effector nucleases in medaka. Dev Growth Differ. 56: 98-107.
11. Chu, V.T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K. & Kühn, R. (2015) Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nat Biotechnol. 33: 543-8.
12. Ma, Y., Chen, W., Zhang, X., Yu, L., Dong, W., Pan, S., Gao, S., Huang, X. & Zhang, L. (2016) Increasing the efficiency of CRISPR/Cas9-mediated precise genome editing in rats by inhibiting NHEJ and using Cas9 protein. RNA Biology. 13: 605-612.
13. Kimura, Y., Hisano, Y., Kawahara, A. & Higashijima, S. (2014) Efficient generation of knock-in transgenic zebrafish carrying reporter/driver genes by CRISPR/Cas9-mediated genome engineering. Sci Rep. 4: 6545.

14. Suster, M.L., Abe, G., Schouw, A. & Kawakami, K. (2011) Transposon-mediated BAC transgenesis in zebrafish. Nat Protoc. 6: 1998-2021.

Chapter4

A useful system for transporting foreign proteins into eggs with vitellogenin signal

4.1 Introduction

To date, various bioreactor systems have been established for production of recombinant proteins. The use of bacteria is still a popular choice, because they can be cultured in larger scale and at lower cost than other expression system. However, they cannot produce glycosylated proteins of vertebrates, and also often form misfolded recombinant proteins as the inclusion body [1]. Although yeasts and insect cell lines such as Sf9 can perform post-translational modifications (PTMs) to enhance the function of proteins, it is difficult to completely reproduce the precise glycosylation pattern of vertebrates; the bioactivities of the produced recombinant proteins may not fully correspond to those of the native forms in vertebrates [2]. Mammalian cultured cell lines such as 293T can achieve complex PTMs, however, culture of the cell lines is expensive so that the recombinant protein cannot be produced cost-effectively. As each technique has both advantages and disadvantages, it is necessary to select the appropriate system depending on the properties of the recombinant proteins.

On the other hand, fish also have the following desirable features for use as a bioreactor system; (i) low cost for maintenance, (ii) ease of collecting eggs, and (iii) availability of the technology for gene modification. Indeed, two studies demonstrated that recombinant proteins could be purified from fish eggs into which the expression

plasmid was injected. One is human coagulation factor VII (hVFII) expressed by the cytomegalovirus (CMV) promoter in nile tilapia *Oreochromis niloticus* embryos, and the other is the recombinant goldfish luteinizing hormone (LH) expressed by the medaka β-actin promoter in rainbow trout *Oncorhynchus mykiss* embryos [3,4]. Both recombinant proteins were successfully expressed in fish embryos, however these studies did not investigate whether the eggs containing each respective recombinant protein can be stably produced from the female transgenic founders. Moreover, Hu et al. (2011) reported that the F_2 fertilized eggs from intercrossed F_1 transgenic zebrafish could produce mature tilapia insulin-like growth factors (IGFs) proteins using the zona pellucida (*zp3*) promoter, which promotes oocyte-specific expression. This report indicated that transgenic fish can stably produce and strore exogenous proteins in its eggs. Most teleost fish, however, can generate the precursors of the egg envelope and/or yolk proteins in the liver, and then deliver them into eggs for the formation of matured eggs [5,6]. These results prompted me to examine the possibility of a novel bioreactor system; liver-expressed recombinant proteins may be efficiently transported into eggs, and thus allowing for the mass accumulation of the proteins in the eggs. The liver generally plays a role in producing the various proteins in quantity, and thus has the potential to stably and abundantly generate recombinant proteins for delivering into eggs as well.

Vitellogenin (Vtg) is a major precursor of egg yolk protein for embryonic development in oviparous organisms [7]. In general, teleost fish produce Vtg in the liver with the stimulation of estradiol-17β (E_2), and then it is secreted into the bloodstream [6]. The secreted Vtg is incorporated into the growing oocytes via specific membrane receptors during vitellogenesis [8-10], and then it is proteolytically cleaved by cathepsin D to generate multiple egg yolk proteins such as lipovitellin (Lv) and phosvitin (Pv) [11].

Functional analyses of domains of Vtg have depended mainly upon *in vitro* experiments and biochemical approaches. For example, a previous study using the yeast two-hybrid system revealed that a peptide sequence (HLTKTKDL) in the Lv domain of the Vtg of blue tilapia *Oreochromis aureus* played a critical role in binding to the Vtg receptor [12]. However, the critical sequence of Vtg that enables the transportation of proteins secreted from the liver into the eggs *in vivo* remains unknown. For easier construct design and expression, it is necessary to identify more precisely the effective region in the full length of Vtg, which is designated as the ‘‘Vtg signal’’ in this study.

Here, I aimed to establish a novel system for transporting foreign proteins into eggs by using the Vtg signal in medaka. First, I generated transgenic medaka lines that express enhanced green florescent protein (EGFP) with or without the Vtg signal (N-terminal part of Vtg protein) in their liver using the liver-specific *chg*H promoter [13]. Second, I demonstrated that the EGFP with Vtg signal could be accumulated in eggs, but not without Vtg signal. I also confirmed that the larger size of exogenous protein (EGFP-fused luciferase: 798 amino acid, 88 kDa) than EGFP (240 aa, 27 kDa) could be transport into eggs with the Vtg signal. My results showed that the Vtg signal is a valuable tool for bioreactor.

4.2 Materials and Methods

Fish

The Cab inbred closed colony of medaka *Oryzias latipes* was used in this study. The fish were kept under a 14/10-h light/dark cycle at 26 °C. Fish handling and sampling methods

were approved by Kyoto University (No.26–71). All efforts were made to minimize suffering.

Identification and isolation of the vtg signal in medaka

To identify the Vtg signal from the full-length amino acid sequence, the secretory signal of the medaka vitellogenin (NCBI, accession number; AY074891) was analyzed by the Signal Peptide Predictor software (SignalP, http://www.cbs.dtu.dk/services/SignalP/). In addition, the amino acid sequence of the medaka Vtg was aligned with the Vtg receptor-binding region (HLTKTKDL) of the blue tilapia *Oreochromis aureus* Vtg to investigate the presence of conserved regions using the sequence alignment tool (ClustalW, http://clustalw.ddbj.nig.ac.jp/). The resulting 300 amino acid of N-terminal region containing the predicted secretory signal and the receptor-binding region was designated as Vtg signal and the coding sequence (*vtg* signal) of Vtg singal was isolated by reverse transcription-polymerase chain reaction (RT-PCR). Total RNA was extracted from the liver of a sexually matured female with RNeasy Plus Mini kit (Qiagen). One μg of the total RNA was used for synthesizing first strand cDNA with random hexamers and SuperScript III Reverse Transcriptase (Invitrogen) according to the manufacturer's instruction. The cDNA was used for PCR amplification of the *vtg* signal with KOD-FX DNA polymerase (Toyobo) using primers Asp718I-vtg-Fw and vtg-NcoI-Rv (Table 4-1). The PCR condition was as follows: incubation at 94 °C for 2 min, followed by 35 cycles of 98 °C for 10 min, 58 °C for 30 s and 68 °C for 30 s. The PCR products purified with 1% agarose gel electrophoresis and the DNA sequence was confirmed by direct sequencing analysis (data not shown).

Plasmid construction

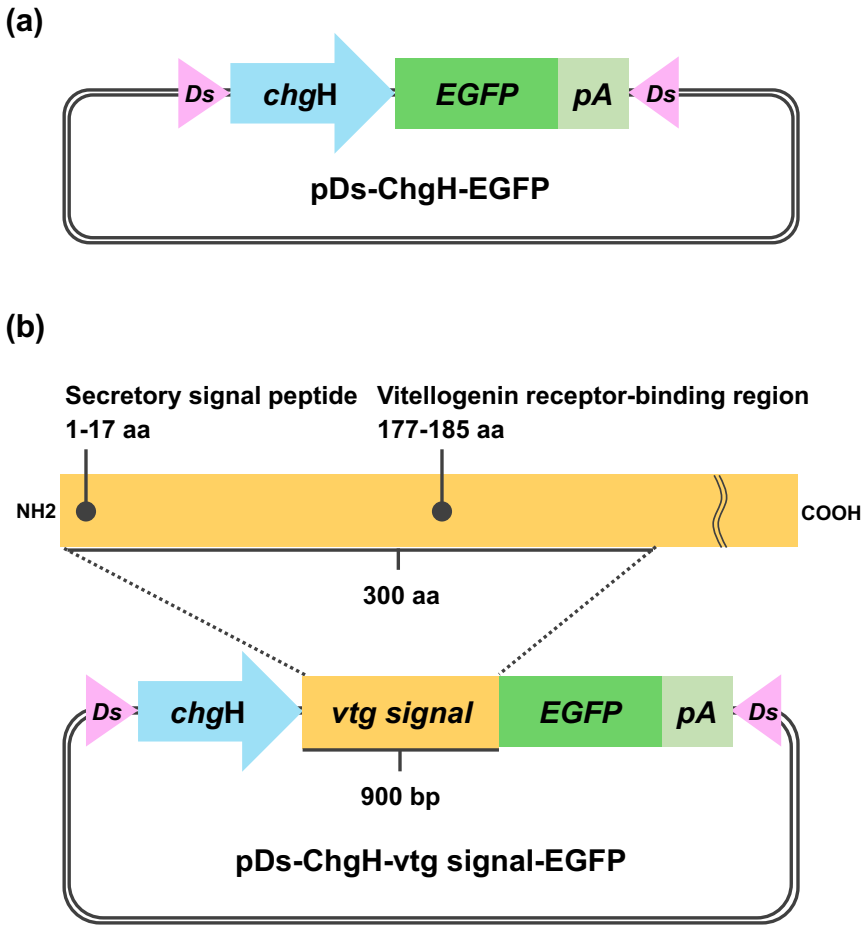
All DNA fragments for constructing each plasmid were amplified by PCR using KOD-FX DNA polymerase (Toyobo). The PCR programs were as follows: one cycle at 94 °C for 2 min followed by 35 PCR cycles of 98 °C for 10 s, 58 °C for 30 s and 68 °C for 30 s per 1 kbp.

pDs-ChgH-EGFP (Fig. 4-1a): A backbone fragment containing 5′- and 3′-*Ds* sequences, a poly (A) signal, a pUC replication origin, and an ampicillin resistance gene was amplified from a plasmid pDs-actb2k-EGFP [14] by PCR using a primer pair (NotI-Backbone-Fw and Backbone-XhoI-Rv) (Table 4-1). An insert fragment containing a *chg*H promoter (approximately 2 kbp) and *EGFP* was amplified from a plasmid pChgH-EGFP by PCR using a primer pair (XhoI-chgH-Fw and EGFP-NotI-Rv) (Table 4-1) [13]. These two fragments were digested with XhoI/NotI, and then ligated to obtain pDs-ChgH-EGFP. The c*hg*H promoter (approximately 2 kbp), which was isolated from the 5'-upstream sequence of the medaka *choriogeninH* gene, induces liver-specific gene expression in response to endogenous and exogenous 17-β estrogen (E2) [13].

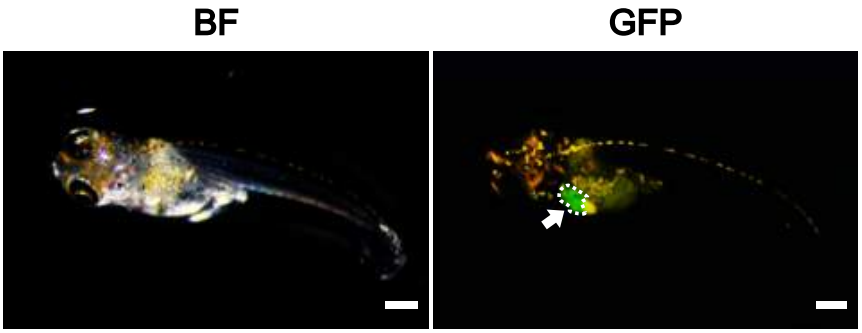
pDs-ChgH-vtg signal-EGFP (Fig. 4-1b): To generate the plasmid of pDs-ChgH-vtg signal-EGFP, the PCR-amplified *vtg* signal which contains Asp718I and NcoI recognition site at 5' and 3' respectively was digested and inserted into the Asp718I/NcoI site of pDs-ChgH-EGFP construct.

pDs-ChgH-LUC-EGFP (Fig. 4-2a): The fragment of firefly *Photinus pyralis* luciferase (LUC) was amplified from the pGL 4.13 plasmid (Promega) by PCR using a primer pair (NcoI-LUC-Fw and LUC-Linker-SpeI-Rv) (Table 4-1), and then digested with NcoI/SpeI. The resultant LUC fragment was ligated into the NcoI/SpeI site of pDs-ChgH-EGFP

construct to generate the pDs-ChgH-LUC-EGFP construct. LUC and EGFP were produced as a fusion reporter protein by this construct.



(c)



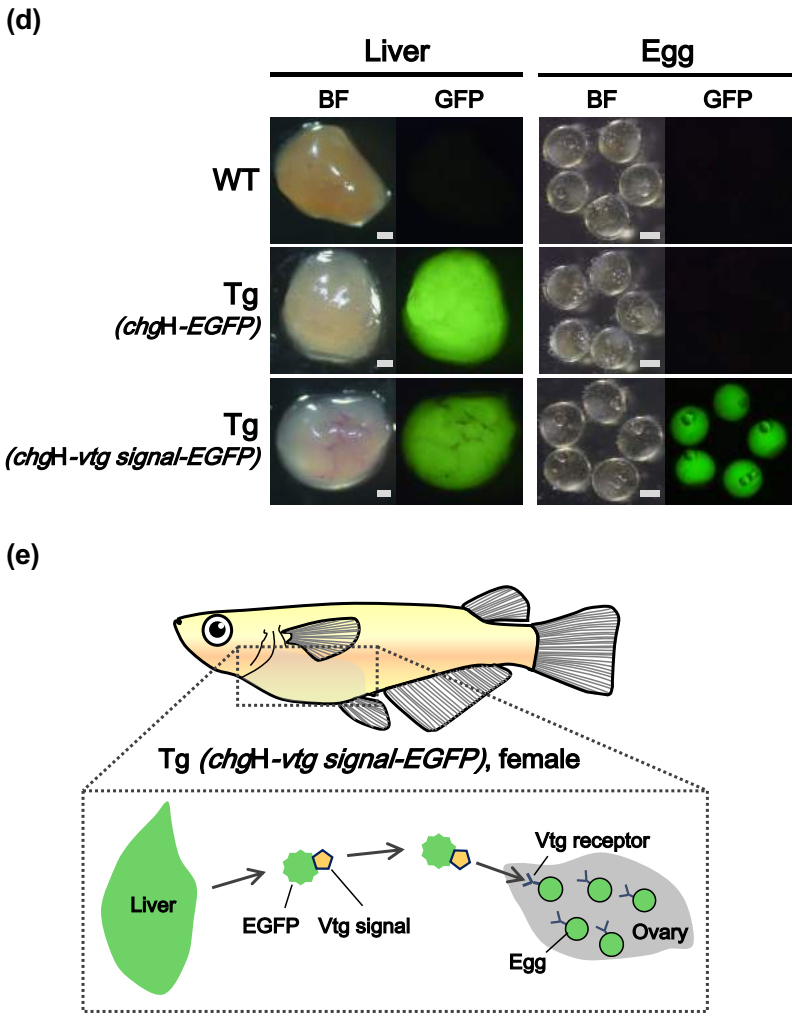
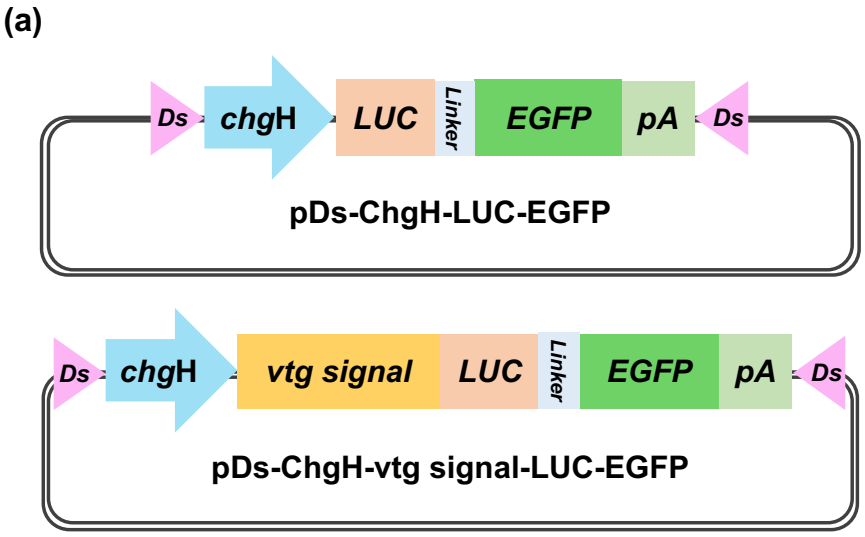


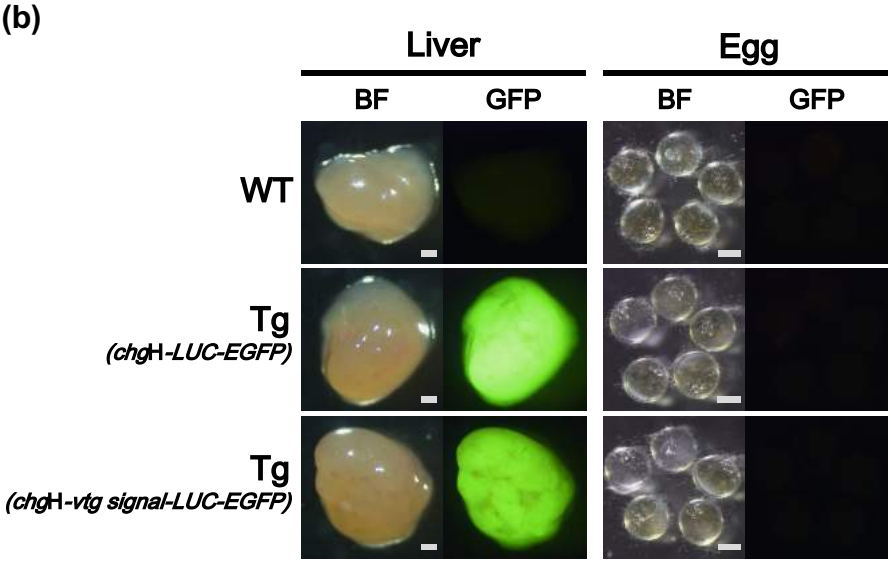
Fig. 4-1. The Vtg signal-mediated delivery of EGFP secreted in the liver into eggs. **(a)** DNA construct of the pDs-ChgH-EGFP for establishing Tg (chgH-EGFP). The Ds transposon sequences flanking the plasmid were used to enhance the efficiency of genomic integration. The chgH promoter can drive the downstream EGFP in liver in response to stimulation by 17-β estrogen. **(b)** Schematics of the full-length amino acid sequence of the medaka Vtg, and the pDs-ChgH-vtg signal-EGFP for establishing Tg (chgH-vtg signal-EGFP). The vtg signal is 900 bp in length encoding 300 aa containing the secretory signal peptide and the receptor-binding region of the Vtg. **(c)** GFP expression in the liver of Tg (chgH-EGFP) at 2 days post hatching (dph). The yolk-sac larva was exposed to 100 ng/μL of E2 for overnight to drive the chgH promoter. A white arrow and a site surrounded by a dotted white line show the green fluorescence in the liver. BF:

bright field, GFP: GFP fluorescent image, Scale bar: 200 μm. **(d)** Fluorescent images of livers and eggs of sexually mature females in the Cab (WT), Tg (chgH-EGFP), and Tg (chgH-vtg signal-EGFP). Each female was mated with a wild-type male to observe the green fluorescence of the fertilized eggs on the day of spawning. There was no GFP signal in the liver or eggs of the Cab (WT). In Tg (chgH-EGFP), a GFP signal was detected from the liver but not from the eggs, though Tg (chgH-vtg signal-EGFP) exhibited GFP signals in both the liver and eggs. The transgenic lines in the Figure carried each transgene heterozygously. BF: bright field, GFP: GFP fluorescent image, scale bar: 200 μm. **(e)** Schematics of delivery of EGFP secreted in the liver into the eggs in Tg (chgH-vtg signal-EGFP) females. The Vtg signal enabled transportation of EGFP secreted in the liver into the eggs.

pDs-ChgH-vtg signal-LUC-EGFP (Fig. 4-2a): The *vtg* signal fragment was obtained by digesting the pChgH-vtg signal-EGFP with Asp718I/NcoI, and then ligated into the the pDs-ChgH-LUC-EGFP construct at Asp718I/NcoI site.



(b)


Liver
Egg
BF
GFP
BF
GFP
WT
Tg
(chgH-LUC-EGFP)
Tg
(chgH-vtg signal-LUC-EGFP)

(c)

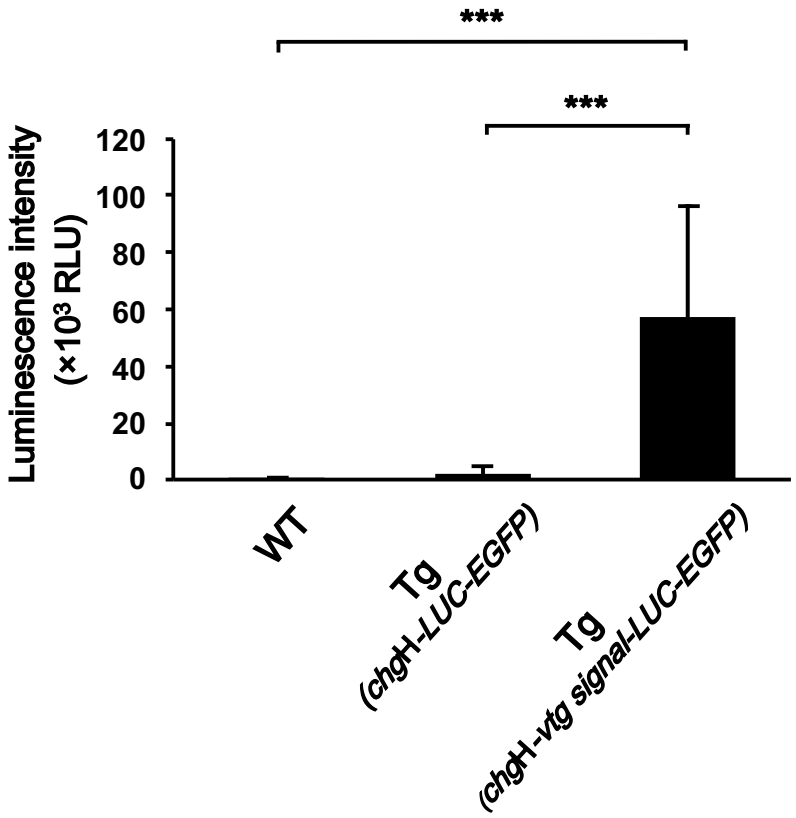

***
***
Luminescence intensity
(×10³ RLU)
120
100
80
60
40
20
0
WT
Tg
(chgH-LUC-EGFP)
Tg
(chgH-vtg signal-LUC-EGFP)

Fig. 4-2. The transportation of the EGFP-fused LUC expressed in the liver into eggs using the Vtg signal. **(a)** DNA construct of the pDs-ChgH-LUC-EGFP for establishing Tg (chgH-LUC-EGFP), and the pDs-ChgH-vtg signal-LUC-EGFP for establishing Tg (chgH-vtg signal-LUC-EGFP). A linker sequence (GGSGGT) was located between LUC and EGFP to increase the flexibility of the fusion protein, LUC and EGFP. **(b)** Fluorescent images of livers and eggs of sexually mature females in the Cab (WT), Tg (chgH-LUC-EGFP), and Tg (chgH-vtg signal-LUC-EGFP). The females were crossed with wild-type males to observe the green fluorescence of the fertilized eggs on the day of spawning. No GFP signal was detected in the liver or eggs of the Cab (WT). GFP fluorescence was detected from the livers of Tg (chgH-LUC-EGFP) and Tg (chgH-vtg signal-LUC-EGFP), but not from the eggs of either transgenic lines. BF: bright field, GFP: GFP fluorescent image, scale bar: 200 μm. **(c)** Detection of LUC activity in the eggs derived from females of the Cab (WT), Tg (chgH-LUC-EGFP), and Tg (chgH-vtg signal-LUC-EGFP). The luciferase activity in the eggs was normalized by the concentration of total protein extracted from egg samples, and expressed as relative luminescence units (RLU). Data are presented as mean ± SD of triplicate determinations from a single experiment ($n = 9$/group). Stars (***) indicate that the values are significantly different between each group by one-way ANOVA followed by Tukey's multiple comparisons ($P < 0.001$). All transgenic fish in the Figure carried each transgene heterozygously.

pBaitD-vtg-Cas9-mAG-pA (Fig. 4-4a): To obtain homology arms, an upstream (522 bp) or downstream (550 bp) genomic region at the target site of sgRNA-vtg was amplified by PCR with primer pairs In-vtg-LHA-FW and In-vtg-LHA-Linker-RV or In-vtg-RHA-FW

Table 4-1. Oligonucleotide sequences used in this study.

Name	Sequence (5'-3')	Usage
Asp718I-vtg-Fw	AGGTGGTACCATGAGGGGGCTGATTCTTGC	RT-PCR
vtg-NcoI-Rv	ATCGCCATGGTCTCAACTGGCGTTTGGAGGAGC	RT-PCR
vtg-OligoA	CTAATACGACTCACTATAGTTGAGAAGAGCATTGATTTGGTTTTAGAGCTAG AAATAGCA	Thynsthesis of sgRNA
vtg-OligoB	AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTAT TTTAACTTGCTATTTCTAGCTCTAAAAC	Thynsthesis of sgRNA
vtg-OligoC	AAAAGCACCGACTCGGTGCC	Thynsthesis of sgRNA
NotI-Backbone-Fw	CGTGCGGCCGCGACTCTAGATCATAATC	Construction of plasmid
Backbone-XhoI-Rv	GCACTCGAGTTTCCCCGTATCCCCCCAGG	Construction of plasmid
XhoI-chgH-Fw	CTGCTCGAGACACGTTATTTTGTGATAAC	Construction of plasmid
EGFP-NotI-Rv	GCAGCGGCCGCTTTACTTGTACAGCTCGT	Construction of plasmid
NcoI-LUC-Fw	AGCTCCATGGAAGACGCCAAAAACAT	Construction of plasmid
LUC-Linker-SpeI-Rv	AGCACTAGTAGTTCCACCGCTACCTCCCACGGCGATCTTGCC	Construction of plasmid
In-vtg-LHA-FW	ACCTTTGCTCATTCCTGGGTTCTGTCGGGA	Construction of plasmid
In-vtg-LHA-Linker-RV	TCTTTGAGAAGAGCATTGATGGAGGTAGCGGTGGA	Construction of plasmid
In-vtg-RHA-FW	CATCAATGTATCTTATTGAGGGAGACCGCAGAGGC	Construction of plasmid
In-vtg-RHA-RV	GAGCAGATGACCTGACAAGGATGTACAGAAACCA	Construction of plasmid
In-Linker-Cas9-FW	GGAGGTAGCGGTGGAACTATGGACTATAAGGACCACGACGGAGAC	Construction of plasmid
In-pA-RV	AGTTGTGGTTTGTCCAAACTCATCAATGTATCTTA	Construction of plasmid
Vec-vtg-FW	CCTCGCAGTCTAGGCCGAAGATCAGCAAAAG	Construction of plasmid
Vec-vtg-RV	CCTCGCAGTCTAGGCCGAAGATCCACTACGTGTATCCGCTCAT	Construction of plasmid
EGFP-Fw	TGCCACCTACGGCAAGCTGA	RT-PCR
EGFP-Rv	TGTTGCCGTCCTCCTTGAAG	RT-PCR
ef1α-Fw	CAGGACGTCTACAAAATCGG	RT-PCR
ef1α-Rv	AGCTCGTTGAACTTGCAGGC	RT-PCR

and In-vtg-RHA-RV (Table 4-1). Insert fragments containing Cas9, mAG, and SV40 polyA were amplified from a plasmid pOlhsp70-Cas9-mAG-pA (kindly gifted by Dr. Ansai) by PCR using a primer pair In-Linker-Cas9-FW and In-pA-RV (Table 4-1). Backbone fragments containing a pUC replication origin, an ampicillin resistance gene, and BaitD sequences were amplified from plasmid pBaitD-acta1_500 bp-mAG by PCR using a primer pair Vec-vtg-FW and Vec-vtg-RV (Table 4-1). These four fragments were digested by DpnI at 37 ºC for 1 hour to degrade the DNA templates in the PCR reaction mixtures, and then subjected to electrophoresis with a 1% agarose gel. The desired bands were extracted from the gel and purified using NucleoSpin Gel and PCR Clean-up

(MACHEREY-NAGEL). The purified fragments were ligated using an In-Fusion Cloning kit (Takara).

All constructed plasmids were extracted using NucleoSpin Plasmid QuickPure kit (MACHEREY-NAGEL) according to the manufacturer's direction. To eliminate residual RNase activity of the extracted plasmids, the plasmids dissolved with 50 μL of 5 mM Tris-HCl buffer (pH 8.5) were incubated with 2 μL of Proteinase K (20 mg/mL) and 5 μL of 10% sodium dodecyl sulfate (SDS) at 55 ºC for 30 min, and then purified using NucleoSpin Gel and PCR Clean-up kit (MACHEREY-NAGEL) with the buffer NTB.

Ac RNA and sgRNA preparation for microinjection

The *Ds* sequences in the above plasmids contain the cis-required terminal repeats that can be transactivated by an *A*c transposase [15]. Introduction of plasmids with the *Ds* elements and the *Ac* RNA in medaka embryos enables to enhance the efficiency of germline transmission in transgenic medaka [16]. The *Ac* RNA for microinjection was synthesized and purified as described previously [17]. Briefly, the linearized pAcII plasmid with BamHI was purified using NucleoSpin Gel and PCR Clean-up kit (MACHEREY-NAGEL) for *in vitro* transcription. The capped *Ac* RNA was transcribed from the purified DNA template using the Message mMachine SP6Kit (Life Technologies, Gaithersburg, MD).

The sgRNA targeting the *vtg* locus was designed using a track for the UCSC genome browser of CRISPRscan [18]. The template DNA for the designed sgRNA was amplified by PCR using KOD-FX DNA polymerase (Toyobo) and three oligonucleotides, vtg-OligoA, vtg-OligoB, and vtg-OligoC (Table 4-1). The PCR programs were as

follows: one cycle at 94 °C for 2 min followed by 35 PCR cycles of 98 °C for 10 s, 58 °C for 30 s and 68 °C for 5 s. The resulting PCR products were purified using NucleoSpin Gel and PCR Clean-up kit (MACHEREY-NAGEL) and then transcribed *in vitro* with CUGA®7 gRNA Synthesis Kit (Nippon Gene). All synthesized RNA was purified using RNeasy Mini Kit (Qiagen) for microinjection.

Generation of transgenic and knock-in medaka line

In the present study, 4 different transgenic lines that harbor pDs-ChgH-EGFP, pDs-ChgH-vtg signal-EGFP, pDs-ChgH-LUC-EGFP and pDs-ChgH-vtg signal-LUC-EGFP are designated as Tg (*chg*H-*EGFP*), Tg (*chg*H-*vtg signal-EGFP*), Tg (*chg*H-*LUC-EGFP*), and Tg (*chg*H-*vtg signal-LUC-EGFP*), respectively. To establish the 4 transgenic lines, injection mixtures containing 10 ng/μL of each plasmid and 100 ng/μL of *Ac* RNA were injected into fertilized eggs at the one-cell stage as previously described [17]. Additionally, to establish a knock-in strain KI (*Cas9-mAG*), injection mixtures containing 2.5 ng/μL of pBaitD-vtg-Cas9-mAG-pA, 500 ng/μL of Cas9 Protein (Integrated DNA Technologies), and two sgRNAs for cleavage of the targeted site on the *vtg* locus and BaitD sequences on the donor plasmid were injected into fertilized eggs as described above. The hatched yolk-sac larvae were screened by observing EGFP expression in the liver after the overnight exposure to 100 ng/μL of E2. The screened larvae were reared to adulthood as founders (F_0s), and then mated with wild-type fish to obtain the F_1 transgenic or knock-in fish. In this study, $F_{1\text{-}8}$ heterozygous females were used for verifying the accumulation of foreign proteins in the fertilized eggs.

Observation of GFP fluorescence

The fluorescence of the liver and fertilized eggs was observed using a fluorescence stereomicroscope MZFLIII (Leica Microsystems) with a GFP2 filter set. Microscopic images were captured using a digital color-cooled charge-coupled camera and the VB-7010 image control system (Keyence).

Reporter assay for the detection of LUC activity

To investigate Vtg signal-mediated accumulation of LUC in eggs, bioluminescence was detected with Luciferase assay system kit (Promega) and GloMax 96 microplate luminometer (Promega) as described by Saito et al. (2009). Egg samples for this analysis were collected from three groups, wild-type, Tg (*chgH-LUC-EGFP*), and Tg (*chgH-vtg signal-LUC-EGFP*). In each group, 9 females were mated with wild-type counterparts to obtain 5 fertilized eggs per 1 female. The eggs of each transgenic line were prepared as follows: (i) 3 founders (F_0s) of Tg (*chgH-LUC-EGFP*) or Tg (*chgH-vtg signal-LUC-EGFP*) were mated with wild-type counterparts to establish 3 independent heterozygous lines (F_1s) of each transgene. (ii) 3 heterozygotes (F_1s) were mated with wild-type counterparts to generate 3 heterozygous females (F_2s) in each line. (iii) the resulting 9 females (F_2s) were mated with wild-type males to collect 5 fertilized eggs in each pair. The collected 5 fertilized eggs were pooled in a 1.5-mL microtube containing 300 μL of the Reporter lysis buffer (Promega), and then homogenized with a pestle. After freezing (-80 ^{o}C) and thawing (room temperature), 80 μL of each homogenate was dispensed into 3 wells on a 96-well microplate for triplicate assays. Then, 20 μL of the luciferase assay substrate (Promega) was added in each well to measure the bioluminescence with

GloMax 96 microplate luminometer (Promega). The measured bioluminescence intensity was normalized by the concentration of total protein extracted from each homogenate, and expressed as relative luminescence units (RLU). The protein concentrations were determined with the Pierce BCA protein assay kit (Thermo Scientific) according to the manufacturer's protocol.

Evaluation of transcripts of foreign genes in ovary

RT-PCR was performed as described in the above "Identification and isolation of the *vtg* signal in medaka" section to evaluate the transcripts of the elongation factor-1α (*ef-1α*) gene and *EGFP* gene in the ovary. Briefly, to evaluate possible contamination of genomic DNA (gDNA), *ef-1α* was amplified by PCR using KOD-FX DNA polymerase (Toyobo) with primers ef1α-Fw and ef1α-Rv (Table 4-1). This primer set gives different length of PCR products from gDNA (519 bp) and cDNA (374 bp). The expression analysis of *EGFP* was performed by PCR with KOD-FX DNA polymerase (Toyobo) and primers EGFP-Fw and EGFP-Rv (Table 4-1) to evaluate the presence of transcripts of each foreign gene, GFP or GFP-LUC, in the ovary. The ovary of a transgenic female (*42Sp50-EGFP*) expressing EGFP in the oocyte was used as a positive control to confirm the detection of *EGFP* in the RT-PCR analysis [20]. The PCR conditions were as follows: one cycle at 94 °C for 2 min followed by 35 PCR cycles of 98 °C for 10 min, 58 °C for 30 s and 68 °C for 30 s. The resulting PCR products were analyzed by 1% agarose gel electrophoresis.

Statistical analysis

Statistical analysis was performed by one-way ANOVA followed by Tukey's multiple comparisons using GraphPad Prism (MDF). P values less than 0.05 were considered statistically significant.

4.3 Results

No accumulation of EGFP without the *vtg* signal in eggs spawned by Tg (*chg*H-*EGFP*)

To evaluate whether EGFP produced in liver can be delivered into eggs, I established a Tg (*chg*H-*EGFP*) line, expressing EGFP in the liver under the regulation of *chg*H promoter (Fig. 4-1a and c). The sexually matured transgenic fish were mated with wild-type counterparts to generate the heterozygous females. The resultant females exhibited intense green fluorescence in the liver, whereas no GFP signal was detected from the fertilized eggs obtained by mating with wild-type males (Fig. 4-1d), showing no accumulation of EGFP in the eggs. Thus, to achieve the transport of EGFP from the liver into the eggs, I focused on Vtg that is synthesized in the liver and delivered into the eggs.

Identification and cloning of the *vtg* signal in medaka

I tried to identify a region of Vtg sufficient for uptake of maternal factors into medaka. The *in silico* analysis using SignalP (http://www.cbs.dtu.dk/services/SignalP/) nominated the possible secretion signal peptides in the first 17 aa (^{1}MRGLILALSLALVAANQ^{17}) of Vtg. In addition, the alignment analysis of Vtg amino acid sequence of various fish species revealed that a similar amino acid sequence (identical in 7 of the 8 amino acids) to receptor -binding region (HLTKTKDL) identified in tilapia Vtg [12] was located

between amino acids 178 and 185 (^{178}HLSKTKDL^{185}) of the medaka Vtg. From these results, I hypothesized that the 300 aa N-terminal portion of Vtg ("Vtg signal") was sufficient to deliver the foreign proteins produced in the liver into eggs. The 900 bp of *vtg* signal encoding the Vtg signal (300 aa) was isolated from cDNA of the female liver by RT-PCR, and then cloned into reporter gene plasmids (Fig. 4-1b and 2a) to generate transgenic medaka.

Transportation of EGFP expressed in liver into eggs by the Vtg signal

To investigate whether the Vtg signal could be effective for the transportation of secreted EGFP from the liver into eggs, I observed the green fluorescence in the liver and fertilized eggs of Tg (*chg*H-*vtg signal-EGFP*) female fish. The liver sample, which was dissected from a sexually matured female of the transgenic line, showed the intense green fluorescence of introduced *EGFP* driven by the *chg*H promoter similar to the liver of Tg (*chg*H-*EGFP*) (Fig. 4-1d). In contrast, there was a significant difference of EGFP fluorescence in fertilized eggs between the transgenic lines with and without the *vtg* signal. Fertilized eggs of Tg (*chg*H-*vtg signal-EGFP*) fish exhibited intense green fluorescence, whereas those of Tg (*chg*H-*EGFP*) had no green fluorescence (Fig. 4-1d). These findings show that the Vtg signal can promote the transportation of EGFP expressed in the liver into eggs.

Transportation of LUC-EGFP expressed in liver into eggs by the Vtg signal

To test whether a larger molecule than EGFP (240 aa) can be delivered into eggs by the Vtg signal, I compared females of two transgenic lines, Tg (*chg*H-*LUC-EGFP*) and Tg (*chg*H-*vtg signal-LUC-EGFP*) that stably expressed the fused protein (798 aa) of LUC

and EGFP in the liver. The transgenic line without the *vtg* signal, Tg (*chg*H-*LUC-EGFP*), was used as a negative control to assess the effect of the *vtg* signal. The dissected livers of both heterozygous transgenic females appeared bright green under the fluorescent microscope (Fig. 4-2b). In contrast, there is no EGFP fluorescence in fertilized eggs from both transgenic lines of the heterozygous females, showing that the transportation efficiency of Tg (*chg*H-*vtg signal-LUC-EGFP*) expressing a fusion protein LUC and EGFP (798 aa) significantly decreased in comparison to that of Tg (*chgH-vtg signal-EGFP*) producing a single protein EGFP (240 aa) (Fig. 4-1d and 2b). Additionally, I measured the LUC activity in the fertilized eggs using the luciferase reporter assay system because luminescence of luciferase and luciferin is more sensitive than fluorescence of EGFP. As a result, the eggs of Tg (*chg*H-*vtg signal-LUC-EGFP*) showed a significantly higher value of the luminescence than those of wild-type and Tg (*chg*H-*LUC-EGFP*) lacking the Vtg signal (Fig. 4-2c). My data show that the Vtg signal enables the delivery of not only EGFP (240 aa) but also the fusion protein of EGFP and LUC (798 aa) into eggs.

Evaluation of transcripts of the transgene, *EGFP*, in the ovary by RT-PCR

To evaluate the quality of RNA extracted from ovary, RT-PCR with specific primers of e*f-1α* was performed. As shown in Fig. 4-3a, smaller PCR products (374 bp), which are derived from e*f-1α* RNA, were observed in WT and transgenic lines, whereas larger PCR products (519 bp), which are amplified from genomic DNA, were not detected. These results showed that the contamination of gDNA was eliminated in our RNA extract.

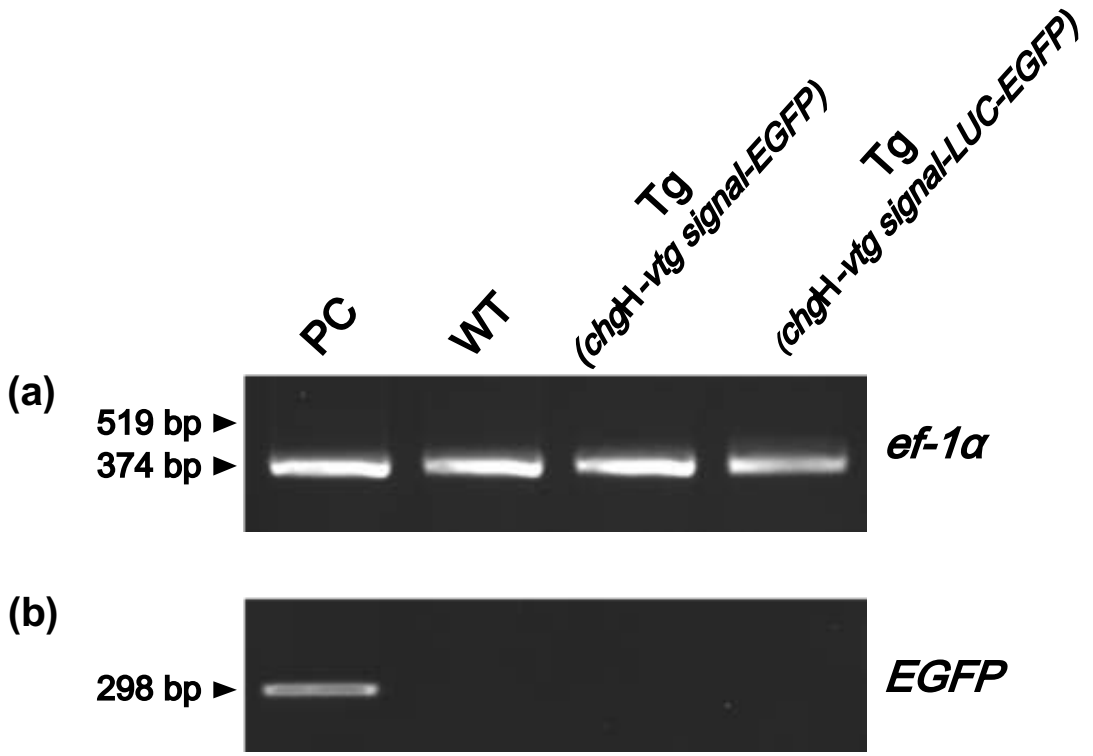


Fig. 4-3. Expression analysis of EGFP in the ovary using RT-PCR. **(a)** Electrophoresis image of the PCR amplicons of ef-1α gene. In all samples, the smaller size of PCR products (374 bp) derived from the exon sequence were observed, whereas the longer size of PCR products (519 bp) derived from the exon and intron sequence were not detected. **(b)** Electrophoresis image of the PCR amplicons of EGFP gene. The PCR amplicons (298 bp) containing the EGFP gene were detected only from the ovary of a transgenic female which expresses EGFP in the oocytes and was used as a positive control (PC), but not from that of WT (Cab), Tg (chgH-vtg signal-EGFP) and Tg (chgH-vtg signal-LUC-EGFP).

To confirm that EGFP or LUC-EGFP accumulated in the eggs of Tg (*chg*H-*vtg signal-EGFP*) or Tg (*chg*H-*vtg signal-LUC-EGFP*), respectively, was not derived from the transcript of each transgene in the ovary, I evaluated the presence of the transcript of

EGFP in the ovary, which is an indicator of both GFP and LUC-EGFP RNA. As shown in Fig. 4-3b, the PCR product derived from *EGFP* (298 bp) was clearly detected in the ovary of Tg (*42Sp50-EGFP*) which expresses EGFP in the oocyte and was used as an EGFP positive sample [20]. On the other hand, the expression of *EGFP* was not detected in the ovary of wild-type (WT), Tg (*chg*H-*vtg signal-EGFP*) and Tg (*chg*H-*vtg signal-LUC-EGFP*) (Fig. 4-3b). These results show that the Vtg signal-fused EGFP or LUC-EGFP was delivered from the outside of the ovary into the eggs and was not produced by ectopic expression of each transgene in the ovary.

Transportation of Cas9 expressed in liver into eggs

To teste transportation of Cas9 into eggs with the Vtg signal, a transgenic fish expressing the Vtg signal-fused Cas9-GFP in the liver was created in the same way as described above. However, the green fluorescence of the liver was detected whereas that of eggs from the transgenic fish was not observed at all (data not shown). This result suggests that the Vtg signal was not long enough to deliver Cas9 secreted in the liver into eggs.

To achieve the transportation of Cas9 into eggs, gene knock-in was performed to integrate *Cas9-mAG* just before stop codon of *vtg* locus, allowing to express the full length of Vtg fused-Cas9-mAG in the liver (Fig. 4-4a). The fluorescent observation revealed that WT fish used as negative control expressed no green fluorescence in the liver and eggs, whereas the KI fish showed the bright green fluorescence of the Cas9-mAG fusion protein driven by the endogenous *vtg* promoter in the liver (Fig. 4-4b). Additionally, all eggs spawned by the KI fish fluoresced slight green, showing successful transportation of Cas9-mAG into eggs with the full length of Vtg (Fig. 4-4b). Although sgRNA targeting *slc45a2* locus was injected into the eggs to confirm the nuclease activity

of Cas9, targeted mutagenesis in the locus was not detected by HMA analysis (data not shown) [21].

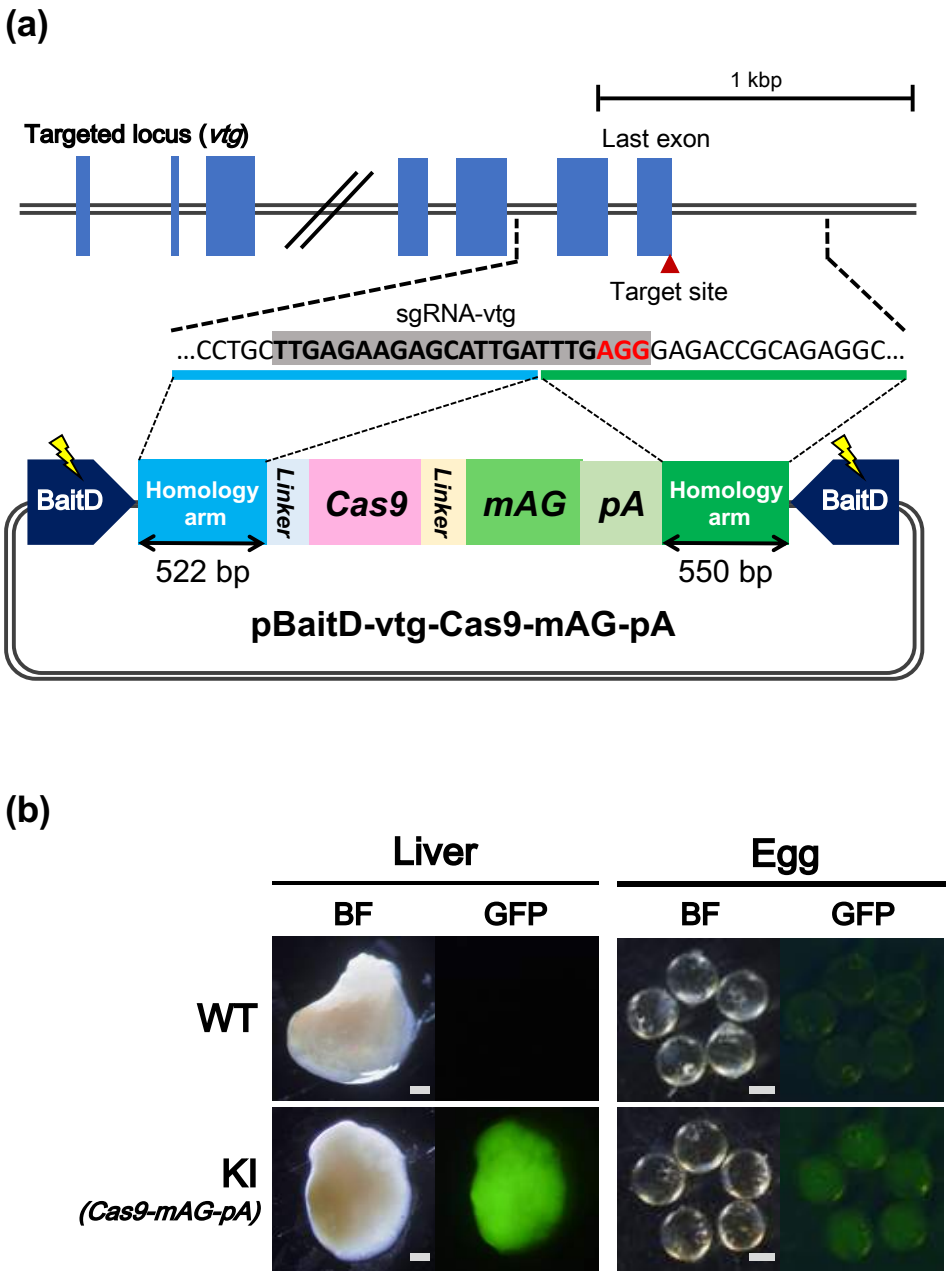


Fig. 4-4. The transportation of the mAG-fused Cas9 expressed in the liver into eggs using the full length of Vtg. **(a)** Schematic illustration of the genomic targeted locus, *vtg*, and the donor plasmid containing Cas9. To increase the flexibility of the fusion protein, Vtg and Cas9 and mAG, short

linker (GGSGGT) located between left homology arm and Cas9 or long linker (24 amino acids) located between Cas9 and mAG is shown in blue or yellow box, respectively. **(b)** Fluorescent images of livers and eggs of sexually mature females in the Cab (WT), KI (Cas9-mAG). Each female was mated with a wild-type male to observe the green fluorescence of the fertilized eggs on the day of spawning. There was no GFP signal in the liver or eggs of the Cab (WT). In KI (Cas9-mAG), a strong GFP signal was detected from the liver, whereas a weak GFP signal was observed in the eggs. This knock-in fish harbored transgene heterozygously. BF: bright field, GFP: GFP fluorescent image, scale bar: 200 μm.

4.4 Discussion

In this study, I successfully established a novel bioreactor system in medaka using the N-terminal 300 aa of Vtg, which is a major yolk precursor protein. Based on the results of *in silico* analysis showing that medaka Vtg contains the secretory signal peptide (^{1}MRGLILALSLALVAANQ^{17}) and the receptor-binding region (^{178}HLSKTKDL^{185}), I predicted that the 300 aa N-terminal portion of Vtg ("Vtg signal") was sufficient to transport the foreign proteins produced in the liver into eggs. To investigate whether the prediction is true or not, I generated transgenic medaka expressing the foreign proteins (Vtg signal-EGFP and Vtg signal-EGFP-LUC) in the liver. The fertilized eggs spawned by both transgenic lines showed the fluorescence of EGFP and the luminescence of LUC, respectively, indicating that the Vtg signal-fused proteins were incorporated into the eggs through the Vtg receptor on the surface of the oocytes (Fig. 4-1d, 1e and 2c). I therefore concluded that the identified Vtg signal could be effective for the *in vivo* delivery of secreted foreign proteins from the liver into the eggs.

Several studies have shown that fish eggs could be used for producing recombinant proteins with transgenesis. For example, Hwang et al. (2004) and Morita et al. (2004) succeeded in obtaining the recombinant protein (human coagulation factor VII or goldfish luteinizing hormone) from the first-generation (G_0) eggs of nile tilapia *Oreochromis niloticus* or rainbow trout *Oncorhynchus mykiss* into which the expression vector containing the CMV promoter or the medaka β-actin promoter was injected, respectively. This method is helpful to obtain the target proteins quickly, but it is necessary to microinject into a large number of eggs sufficient to reach the desired amount of the protein. In contrast, when producing the target proteins stably and continuously over a long period with less work, it is efficient to obtain the protein from the eggs spawned by established transgenic females of subsequent generations (F_1, F_2, and F_3), as shown in the present study. In this respect, small model teleost fish, such as zebrafish and medaka, are the promising models for bioreactors because of their frequent spawning and ease of generating and maintaining a large number of transgenic lines. Nevertheless, there have been only a few reports verifying the use of bioreactors for producing exogenous proteins in the eggs of these fish. A previous study using zebrafish reported that transgenic females could produce proteins of tilapia insulin-like growth factors (IGFs) in the eggs under the oocyte-specific zona pellucida (*zp3*) promoter [22]. Unlike the previous study, I demonstrated for the first time that the proteins expressed in the liver could be delivered into the eggs by using the Vtg signal in medaka. In general, liver synthesizes large quantities of diverse proteins, thus, liver is suitable for the mass production of recombinant proteins. Additionally, my transgenic medaka appeared normal and had no obvious health or growth problems (up to the F_8 generation), thus, I will be able to routinely obtain the targeted protein from eggs without sacrificing the matured females

unless the protein causes toxicity or infertility in the females.

The delivery system into eggs with the Vtg signal can potentially be applied for the transportation of other valuable substances into eggs. One example of delivery of substances with the Vtg signal is an antimicrobial peptide against a broad spectrum of bacteria. Recently, due to human population growth, health trends, and economic development, the demand for aquaculture products is increasing. Supplying high quality fish seeds is a key factor to meet this demand, however, bacterial pathogens are still significant threats for sustainable intensive aquaculture [23]. Cecropins, which are antimicrobial peptides isolated from the hemolymph of silk moth *Hyalophora cecropia* pupae, have been proven to increase the resistance to infection by bacterial pathogens in channel catfish *Ictalurus punctatus* [24,25]. The Vtg signal-mediated delivery of antimicrobial peptides such as cecropins into eggs is potentially effective for producing specific pathogen-free fish seeds. Furthermore, in the medical and pharmaceutical field, because of the progress in research on peptides for human health, it is possible to synthesize novel artificial peptides with antitumor activities, and discover the potential effects of existing antimicrobial peptides such as defensins [26,27]. I thus anticipate that these functional peptides can be efficiently generated by the strategies described in this study in the future. To expand the application possibilities of the delivery system with the Vtg signal, I am planning to investigate the versatility of the system for these valuable substances and other fish species including commercially cultured fish.

The present study also found an issue related to the delivery of substances. As shown Fig. 4-1d and 2b in this study, I observed GFP fluorescence in the eggs of Tg (*chgH-vtg signal-EGFP*), whereas I could not detect any GFP signal in those of Tg (*chg*H-*vtg signal-LUC-EGFP*). These results show that the transportation efficiency of Tg

(*chg*H-*vtg signal-LUC-EGFP*) expressing a fusion protein LUC and EGFP (798 aa) significantly decreased in comparison to that of Tg (*chgH-vtg signal-EGFP*) producing a single protein EGFP (240 aa). This lower efficiency may be attributable to the weakened dimerization caused by fusing LUC to the Vtg signal and EGFP. The circulating Vtg generally exists in the blood as a dimer, thus, the Vtg conformation is considered to play an important role in incorporating itself into eggs [28]. On the contrary, in Tg (*chg*H-*vtg signal-LUC-EGFP*), the transduced protein, LUC-EGFP may lead to a weakened dimerization of the fusion protein, because LUC is a monomer protein unlike the weakly dimeric EGFP [29]. The difference of the delivery efficiency between these two transgenic lines indicates the need for further study to identify what protein features, including the structure, electrical charge, polar characteristics, and molecular weight, are critical for being taken up into eggs.

Recently, Chaverra-Rodriguez et al. (2018) have developed a new technology for gene knockout of progenies (F_1) by transporting the Cas9-guide RNA (gRNA) ribonucleoprotein (RNP) injected in the hemolymph of female (G_0) into the eggs of the mosquito *Aedes aegypti*. The delivery was achieved by injecting a fused protein Cas9 and *Drosophila melanogaster* yolk protein 1 (DmYP1) possessing a similar function of Vtg; the DmYP1 is produced in fat bodies, secreted into the bloodstream, and then transported into the eggs [31]. To simplify the gene modification in a wide range of organisms, it is essential to develop a delivery system of genome editing tools such as the CRISPR/Cas9 system without conventional embryonic microinjection. Then, I tested to deliver Cas9 secreted in the liver into eggs by using the Vtg signal or the full length of Vtg, but targeted mutagenesis of the eggs could not be achieved when injecting sgRNA into the eggs spawned by each gene modified female (Fig. 4-4b). Because Cas9 is a monomer protein

like LUC, Cas9 may make the dimerization of the fusion proteins weak resulting in the significant decrease of the transportation efficiency as described above. The result indicates the need for further studies to open the way for enhancing the transportation efficiency. It may be effective to use other dimer nucleases such as ZFN and TALEN [32] or to vary the integration position of transgenes at *vtg* locus.

In summary, I demonstrated that the Vtg signal was beneficial for accumulating exogenous proteins (EGFP and LUC-fused EGFP) into the eggs in medaka. This finding would offer a platform for other fish species for establishment of the novel bioreactor generating recombinant proteins, production of healthy seed without pathogens, and comprehension of the uptake mechanism of yolk proteins into eggs. Vtg is a highly conserved protein among almost all oviparous organisms including fish, amphibians, reptiles, birds, and insects, thus, the Vtg signal has great potential for applications as described above in various organisms [7,33]. Further examinations on the Vtg-mediated transportation system into eggs would lead to the technological innovations in the future.

Chapter4 References

1. Villaverde A, Carrió MM (2003) Protein aggregation in recombinant bacteria: biological role of inclusion bodies. Biotechnol Lett 25:1385–1395.
2. Brooks SA (2004) Appropriate glycosylation of recombinant proteins for human use: implications of choice of expression system. Mol Biotechnol 28:241–255.
3. Hwang G, Müller F, Rahman MA, Williams DW, Murdock PJ, Pasi KJ, Goldspink G, Farahmand H, Maclean N (2004) Fish as bioreactors: transgene expression of human coagulation factor VII in fsh embryos. Mar Biotechnol (NY) 6:485–492.

4. Morita T, Yoshizaki G, Kobayashi M, Watabe S, Takeuchi T (2004) Fish eggs as bioreactors: the production of bioactive luteinizing hormone in transgenic trout embryos. Transgenic Res 13:551–557.
5. Liu X, Wang H, Gong Z (2006) Tandem-repeated Zebrafsh zp3 genes possess oocyte-specifc promoters and are insensitive to estrogen induction. Biol Reprod 74:1016–1025.
6. Hara A, Hiramatsu N, Fujita T (2016) Vitellogenesis and choriogenesis in fshes. Fish Sci 82:187–202.
7. Romano M, Rosanova P, Anteo C, Limatola E (2004) Vertebrate yolk proteins: a review. Mol Reprod Dev 69:109–116.
8. Hiramatsu N, Chapman RW, Lindzey JK, Haynes MR, Sullivan CV (2004) Molecular characterization and expression of vitellogenin receptor from white perch (*Morone americana*). Biol Reprod 70:1720–1730.
9. Mizuta H, Luo W, Ito Y, Mushirobira Y, Todo T, Hara A, Reading BJ, Sullivan CV, Hiramatsu N (2013) Ovarian expression and localization of a vitellogenin receptor with eight ligand binding repeats in the cutthroat trout (*Oncorhynchus clarki*). Comp Biochem Physiol B: Biochem Mol Biol 166:81–90.
10. Mushirobira Y, Mizuta H, Luo W, Todo T, Hara A, Reading BJ, Sullivan CV, Hiramatsu N (2015) Molecular cloning and partial characterization of a low-density lipoprotein receptor-related protein 13 (Lrp13) involved in vitellogenin uptake in the cutthroat trout (*Oncorhynchus clarki*). Mol Reprod Dev 82:986–1000.
11. Hiramatsu N, Ichikawa N, Fukada H, Fujita T, Sullivan CV, Hara A (2002) Identifcation and characterization of proteases involved in specifc proteolysis of vitellogenin and yolk proteins in salmonids. J Exp Zool 292:11–25.

12. Li A, Sadasivam M, Ding JL (2003) Receptor-ligand interaction between vitellogenin receptor (VtgR) and vitellogenin (Vtg), implications on low density lipoprotein receptor and apolipoprotein B/E. The frst three ligand-binding repeats of VtgR interact with the amino-terminal region of Vtg. J Biol Chem 278:2799–2806.
13. Kurauchi K, Nakaguchi Y, Tsutsumi M, Hori H, Kurihara R, Hashimoto S, Ohnuma R, Yamamoto Y, Matsuoka S, Kawai S, Hirata T, Kinoshita M (2005) In vivo visual reporter system for detection of estrogen-like substances by transgenic medaka. Environ Sci Technol 39:2762–2768.
14. Uemura N, Koike M, Ansai S, Kinoshita M, Ishikawa-Fujiwara T, Matsui H, Naruse K, Sakamoto N, Uchiyama Y, Todo T, Takeda S, Yamakado H, Takahashi R (2015) Viable neuronopathic Gaucher disease model in Medaka (*Oryzias latipes*) displays axonal accumulation of alpha-synuclein. PLoS Genet 11:e1005065.
15. Lazarow K, Doll ML, Kunze R (2013) Molecular biology of maize Ac/ Ds elements: an overview. Methods Mol Biol 1057:59–82.
16. Boon Ng GH, Gong Z (2011) Maize Ac/Ds transposon system leads to highly efcient germline transmission of transgenes in medaka (*Oryzias latipes*). Biochimie 93:1858–1864.
17. Inoue T, Iida A, Maegawa S, Sehara-Fujisawa A, Kinoshita M (2016) Generation of a transgenic medaka (*Oryzias latipes*) strain for visualization of nuclear dynamics in early developmental stages. Dev Growth Difer 58:679–687.
18. Moreno-Mateos MA, Vejnar CE, Beaudoin J, Fernandez JP, Mis EK, Khokha MK, et al. (2015) CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. Nat Methods. 12:982–8.
19. Saito S, Furuno A, Sakurai J, Sakamoto A, Park HR, Shin-Ya K, Tsuruo T, Tomida A

(2009) Chemical genomics identifies the unfolded protein response as a target for selective cancer cell killing during glucose deprivation. Cancer Res 69:4225–4234.

20. Kinoshita M, Okamoto G, Hirata T, Shinomiya A, Kobayashi T, Kubo Y, Hori H, Kanamori A (2009) Transgenic medaka enables easy oocytes detection in live fsh. Mol Reprod Dev 76:202–207.
21. Ansai S, Inohaya K, Yoshiura Y, Schartl M, Uemura N, Takahashi R, et al. (2014) Design, evaluation, and screening methods for efficient targeted mutagenesis with transcription activator-like effector nucleases in medaka. Develop Growth Differ. 56:98–107.
22. Hu SY, Liao CH, Lin YP, Li YH, Gong HY, Lin GH, Kawakami K, Yang TH, Wu JL (2011) Zebrafsh eggs used as bioreactors for the production of bioactive tilapia insulin-like growth factors. Transgenic Res 20:73–83.
23. Toranzo AE, Magariños B, Romalde JL (2005) A review of the main bacterial fish diseases in mariculture systems. Aquaculture 246:37–61.
24. Dunham RA, Warr GW, Nichols A, Duncan PL, Argue B, Middleton D, Kucuktas H (2002) Enhanced bacterial disease resistance of transgenic channel catfsh *Ictalurus punctatus* possessing cecropin genes. Mar Biotechnol (NY) 4:338–344.
25. Sarmasik A, Warr G, Chen TT (2002) Production of transgenic medaka with increased resistance to bacterial pathogens. Mar Biotechnol (NY) 4:310–322.
26. Papo N, Braunstein A, Eshhar Z, Shai Y (2004) Suppression of human prostate tumor growth in mice by a cytolytic d-, l-amino acid peptide: membrane lysis, increased necrosis, and inhibition of prostate-specifc antigen secretion. Cancer Res 64:5779–5786.
27. Suarez-Carmona M, Hubert P, Delvenne P, Herfs M (2015) Defensins: “simple”

antimicrobial peptides or broad-spectrum molecules? Cytokine Growth Factor Rev 26:361–370.

28. Reading BJ, Hiramatsu N, Sawaguchi S, Matsubara T, Hara A, Lively MO, Sullivan CV (2009) Conserved and variant molecular and functional features of multiple egg yolk precursor proteins (vitellogenins) in white perch (*Morone americana*) and other teleosts. Mar Biotechnol (NY) 11:169–187.
29. Shaner NC, Steinbach PA, Tsien RY (2005) A guide to choosing fuorescent proteins. Nat Methods 2:905–909
30. Chaverra-Rodriguez D, Macias VM, Hughes GL, Pujhari S, Suzuki Y, Peterson DR, Kim D, McKeand S, Rasgon JL (2018) Targeted delivery of CRISPR-Cas9 ribonucleoprotein into arthropod ovaries for heritable germline gene editing. Nat Commun 9:3008.
31. Bownes M, Hurd H, Büsgen T, Servay D, Alvis S, Popovic B, Bruce S, Burns I, Rothwell K, Walkinshaw M (2002) Drosophila yolk protein produced in E. coli is accumulated by mosquito ovaries. Insect Mol Biol 11:487–496.
32. Thomas G, Charles AG, Carlos FBIII. (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends in Biotech. 31:397-405.
33. Sappington TW, Raikhel AS (1998) Molecular characteristics of insect vitellogenins and vitellogenin receptors. Insect Biochem Mol Biol 28:277–300.

Chapter5

General discussion

The genome editing technology is rapidly evolving because of its simple customization, flexible design, and high success rate. In the field of aquaculture, this technology has enabled the targeted mutagenesis in multiple aquatic organisms including pacific oyster (*Crassostrea gigas*), ridgetail white prawn (*Exopalaemon carinicauda*), and sea urchin (*Strongylocentrotus purpuratus*) [1,2,3]. In popular aquaculture fish in Japan such as red sea bream (*Pagrus major*), a new breed, which shows increased muscle mass, was established by the NHEJ-mediated gene disruption using the CRISPR/Cas9 system [4]. On the two grounds that the established fish never harbor any transgenes, and that the genetic mutation can be naturally occurred by the classical selective breeding, the genome edited fish may soon be sold as a new fishery product in Japan. In contrast, any products that are created by the gene knock-in method are difficult to be accepted by general society at present, because they harbor transgenes on genome just like the conventional Tg or gene modified organisms (GMO). Under these circumstances, the HDR-mediated knock-in system as shown in chapter2 and 3 can be a powerful approach for resolving the global food crisis in the future. To realize sustainable and productive aquaculture, it is necessary to accelerate the molecular breeding, and obtain the economic traits such as disease resistance, rapid growth, cold tolerance, and enhanced food digestibility. The gene knock-in system can achieve more advanced and complex gene modifications than the NHEJ-based genome editing or Tg technology, allowing to broaden the possibilities for gaining various economic traits. Indeed, our research group applies the established HDR-

mediated gene knock-in system in this study for red sea bream (*Pagrus major*) to produce a new breed that can self-synthesize the omega-3 fatty acid docosahexaenoic acid (DHA), which is an essential functional ingredient for growth and development in most of marine cultured fish. Since the cost of fish oil, which is a crude material for DHA, continues to escalate, the fish generating DHA on their own can contribute to cost reduction for farm producers. Additionally, because DHA has a positive effect on various human diseases such as hypertension, arthritis, and atherosclerosis, the new farmed fish may be acceptable for consumers [5]. To alleviate antagonistic attitudes towards GM fish, such farmed fish that can bring great value to both consumers and producers are required for the future. The HDR-mediated gene knock-in system in the present study has strong potential to overcome the current and future challenges in aquaculture, thus, further research on development of the efficient and simplified technical procedures is needed. On that point, small model teleost fish such as medaka and zebrafish are desirable models for cultured fish because of their short generation time, almost daily spawning, and ease of keeping and creating genome edited strains.

The gene knock-in system in medaka is also useful in the medical and pharmaceutical field. Medaka has been one of models for studying human diseases, and used for analyzing their causative genes. Indeed, medaka strains harboring small mutations in genes responsible for human diseases have already been established using random mutagenesis by the Targeting Induced Local Lesions in Genomes (TILLING) method or the NHEJ-mediated gene disruption by TALEN [6,7,8]. On the contrary, the HDR-mediated gene knock-in system in this study can more flexibly modify the targeted genes than the conventional mutagenesis technologies, allowing to obtain the various types of the genetic mutants. For example, this system can visualize expression of

disease-causing genes by reporter genes, or spatiotemporally regulate their expression by using tissue-specific or thermos-responsive promoters such as heat shock promoters [9,10]. Even if the desired mutants cannot be created because of the lethality caused by the excess NHEJ-mediated gene disruption, the knock-in system with Cas9n as described in chapter3 may be effective for establishing the strains without killing individuals. These genetic analytical approaches using the HDR-mediated gene integration method in this study can contribute to help comprehensive understanding of mechanisms of human diseases. Furthermore, because medaka is suitable for high-throughput chemical screening owing to its high reproductivity, the embryos harboring genetic mutations can be used for simple and rapid assay for the effects of chemical compounds in *in vivo*, which be useful in developing novel medicines for treatment of diseases.

Although the genome editing technology has been applied to multiple fish species, embryonic injection of the genome editing tools is still essential for success of the targeted gene modification in fish [11]. To achieve the gene modification with high efficiency, researchers need to obtain eggs at the one-cell stage just after fertilization, which chorions are enough soft for injection. In small experimental fish such as medaka and zebrafish, we can simply collect the early embryos for microinjection every day by controlling several environmental factors including day length and water temperature. In contrast, it is more difficult or impossible to prepare the one-cell eggs in large cultured fish or viviparous fish than such experimental fish. Thus, the development of a simple alternative method that can extend the range of application of genome editing for various fish is strongly desired at present. One of prospective candidates for the technological innovation is the use of Vtg, allowing to deliver foreign proteins secreted from liver into eggs as shown in chapter4. The results prompted me to hypothesize about the

development of a novel genome editing method: genome editing tools expressed in liver can be transported into eggs with aid of Vtg, resulting in the successful targeted gene modification of all eggs in maternal body before spawning. To complete the novel method using the transport capability of Vtg, the following 3 tasks should be addressed in the future. 1) It is necessary to specify the length of Vtg that can deliver genome editing tools into eggs without deactivating their nuclease activities as described in chapter4. 2) Development of a method for introducing DNA construct into adult liver with high efficiency is needed. In terms of cost especially for large farmed fish, the use of DNA is more suitable for the novel genome editing method than that of RNA or Protein, because DNA can be more simply and inexpensively prepared at large scale comparing to RNA or Protein. Although I have already tested DNA introduction in matured medaka using several techniques such as electroporation using electric pulses and sonoporation using ultrasounds, the introduction efficiencies of transgenes into the liver were significant low (data not shown), suggesting the need for further studies to explore different approaches [12,13]. 3) Some devices may be required for transport of the Vtg-fused genome editing tools from egg yolk into nucleus at cytoplasm. Because Vtg is just a precursor of egg yolk proteins, the Vtg-fused genome editing tools may stay inside egg yolks, which makes it impossible for the tools to access genome. Indeed, in Tg (*chg*H-*vtg signal-EGFP*) as shown in Fig. 4-1d, green fluorescence was detected from only the egg yolks but not from the cytoplasm. To overcome this problem, chemical membrane destabilizers such as chloroquine and saponin, or cell-penetrating peptides (CPPs) such as transactivator of transcription (TAT) and polyarginine (R9) may be effective [14,15]. The resolution of the above-mentioned 3 issues would lead to produce technical breakthrough for the simple genome editing without the conventional embryonic injection.

Chapter5 References

1. Yu, H., Li, H., Li, Q., Xu, R., Yue, C. & Du, S. (2019) Targeted Gene Disruption in Pacific Oyster Based on CRISPR/Cas9 Ribonucleoprotein Complexes. Mar Biotechnol (NY). 21, 301-309.
2. Gui, T., Zhang, J., Song, F. et al. (2016) CRISPR/Cas9-Mediated Genome Editing and Mutagenesis of. G3 (Bethesda). 6, 3757-3764.
3. Lin, C. Y. & Su, Y. H. (2016) Genome editing in sea urchin embryos by using a CRISPR/Cas9 system. Dev Biol. 409, 420-428.
4. Kishimoto, K., Washio, Y., Yoshiura, Y. et al. (2018) Production of a breed of red sea bream *Pagrus major* with an increase of skeletal muscle mass and reduced body length by genome editing with CRISPR/Cas9. Aquaculture. 495, 415-427.
5. Horrocks, L.A. & Yeo, Y.K. (1999). HEALTH BENEFITS OF DOCOSAHEXAENOIC ACID (DHA). Pharmacological Research. 40: 211-225.
6. Taniguchi, Y., Takeda, S., Furutani-Seiki, M. et al. (2006) Generation of medaka gene knockout models by target-selected mutagenesis. Genome Biol. 7, R116.
7. Uemura, N., Koike, M., Ansai, S. et al. (2015) Viable neuronopathic Gaucher disease model in Medaka (*Oryzias latipes*) displays axonal accumulation of alpha-synuclein. PLoS Genet. 11, e1005065.
8. Morita, A., Nakahira, K., Hasegawa, T. et al. (2012) Establishment and characterization of Roberts syndrome and SC phocomelia model medaka (*Oryzias latipes*). Dev Growth Differ. 54, 588-604.
9. Oda, S., Mikami, S., Urushihara, Y. et al. (2010) Identification of a functional medaka heat shock promoter and characterization of its ability to induce exogenous gene

expression in medaka in vitro and in vivo. Zoolog Sci. 27, 410-415.

10. Deguchi, T., Itoh, M., Urawa, H. et al. (2009) Infrared laser-mediated local gene induction in medaka, zebrafish and Arabidopsis thaliana. Development, Growth & Differentiation, 51, 769-775.
11. Goto, R., Saito, T., Matsubara, T. & Yamaha, E. (2019). Microinjection of Marine Fish Eggs. Methods Mol Biol, 1874: 475-487.
12. Hostetler, H. A., Peck, S. L. & Muir, W. M. (2003) High efficiency production of germ-line transgenic Japanese medaka (*Oryzias latipes*) by electroporation with direct current-shifted radio frequency pulses. Transgenic Res. 12, 413-424.
13. Ohta, S., Suzuki, K., Ogino, Y. et al. (2008) Gene transduction by sonoporation. Dev Growth Differ. 50, 517-520.
14. Chaverra-Rodriguez, D., Macias, V. M., Hughes, G. L. et al. (2018) Targeted delivery of CRISPR-Cas9 ribonucleoprotein into arthropod ovaries for heritable germline gene editing. Nat Commun. 9, 3008.
15. Derakhshankhah, H. & Jafari, S. (2018) Cell penetrating peptides: A concise review with emphasis on biomedical applications. Biomed Pharmacother. 108, 1090-1096.

Acknowledgement

I would like to express the deepest appreciation to Dr. Masato Kinoshita, Division of Applied Bio-sciences, Kyoto University, for his insightful comments and suggestions. Without his gentle guidance and worm encouragement, this thesis would not have been possible. I would like to thank Dr. Kenji Sato, Division of Applied Bio-sciences, Kyoto University, for supporting my laboratory life and giving lessons about food science. I am also grateful to Dr. Haruhiko Toyohara, Division of Applied Bio-sciences, Kyoto University, for suggestive comments in the laboratory seminar. I would like to thank Dr. Tomohisa Horibe, Department of Pharmacoepidemiology, Graduate School of Medicine and Public Health, Kyoto University, for helpful discussions on detection of luciferase.

I would like to express my gratitude to all members including graduates in Laboratory of Marine Biological Function. Thanks to them, I could live a creative and enjoyable life leading to my personal growth.

Finally, I would particularly like to thank my family who have helped me since I was a child. Without their support, I might not have such a fulfilling time. Thank you.