# An efficient strategy for TALEN-mediated genome engineering in *Drosophila*

Tomonori Katsuyama[1], Arslan Akmammedov[1], Makiko Seimiya[1], Samuel C. Hess[1], Cem Sievers[1] and Renato Paro[1,2,*]

[1]Department of Biosystems Science and Engineering, Eidgenössische Technische Hochschule Zürich, Mattenstrasse 26, 4058 Basel, Switzerland and [2]Faculty of Science, University of Basel, Klingelbergstrasse 50, 4056 Basel, Switzerland

## ABSTRACT

**In reverse genetics, a gene's function is elucidated through targeted modifications in the coding region or associated DNA *cis*-regulatory elements. To this purpose, recently developed customizable transcription activator-like effector nucleases (TALENs) have proven an invaluable tool, allowing introduction of double-strand breaks at predetermined sites in the genome. Here we describe a practical and efficient method for the targeted genome engineering in *Drosophila*. We demonstrate TALEN-mediated targeted gene integration and efficient identification of mutant flies using a traceable marker phenotype. Furthermore, we developed an easy TALEN assembly (easyT) method relying on simultaneous reactions of DNA Bae I digestion and ligation, enabling construction of complete TALENs from a monomer unit library in a single day. Taken together, our strategy with easyT and TALEN-plasmid microinjection simplifies mutant generation and enables isolation of desired mutant fly lines in the $F_1$ generation.**

## INTRODUCTION

The generation of mutant organisms with defined genomic modifications is the most efficient way in reverse genetics to elucidate the functions of gene products or unravel gene regulatory mechanisms. This has been successfully exploited in yeast owing to the ability to exchange DNA via homologous recombination (HR) between endogenous target locus and an exogenously introduced DNA having desired genetic alteration (1,2). In the cases of multicellular organisms, however, such targeted HR is rarely induced, when homologous foreign DNAs is provided, with a few exceptions including mouse embryonic stem cells or the moss *Physcomitrella patens* (3,4).

A breakthrough in targeted genome manipulation was achieved by the development of the engineered nucleases: zinc-finger nucleases (ZFNs) (5) and, more recently, transcription activator-like effector nucleases (TALENs) (6), which consist of repeats of DNA-binding domains and a Fok I nuclease domain. Because Fok I is only active as a dimer (7), these enzymes can cut genomic DNA only where a pair of two engineered nucleases binds nearby in a head-to-head manner. In eukaryotic cells, double-strand breaks (DSBs) stimulate intrinsic DNA repair mechanisms, either error-prone non-homologous end-joining (NHEJ) or error-free HR. In addition to unspecific insertions and deletions mediated by NHEJ, targeted integration of exogenously provided DNA sequences by HR is of substantial value for genome engineering.

TALENs are lately getting more attention over ZFNs because of their unique and simple modularity of DNA-binding domains conferring target specificity (8). Essential to the sequence-specific DNA recognition by TALENs is the contiguous repeat of DNA-binding modules of transcription activator-like effectors (TALE-repeats), which varies in number from 15.5 to 19.5 repeats among most of naturally occurring TALEs (9). Each module consists of 34 amino acids recognizing a single base on DNA. The base preference of each module is specified by the two residues at 12th and 13th positions, called repeat variable di-residues (RVDs), and a set of RVD–base codes has been deciphered (10–13). Because the modules that preferentially bind to a unique base have been identified (e.g. NI–adenine, NG–thymine, NK–guanine and HD–cytosine), TALENs with desired sequence specificity can be constructed by arranging these distinctive modules in a specific order. There is, however, a constraint in selection of TALEN target sites. A thymine is the base that precedes the sequences bound by TALE-repeats. The thymine, often called T at

---

position 0, is essential for the targeting of designer TALEs or TALENs and recognized by two degenerate modules at the N-terminus (14,15). Nonetheless, the customizability of TALENs for the design and construction of tailored molecular scissors to target any sequence of interest is simple and straightforward.

The utility of TALENs for targeted germ line mutagenesis has been successfully proven in a wide range of model organisms, including nematodes (16), insects (17–21), zebrafish (22–26), *Xenopus* (27) and some mammals (28,29). In the fruit fly *Drosophila melanogaster*, based on studies with ZFNs (30), the microinjection of TALEN mRNA into the syncytial embryo has been used to generate targeted mutations in the germ line (18,21). However, the efficiency of TALEN-mediated targeted mutagenesis in *Drosophila* is generally lower than in other model organisms such as zebrafish or *Xenopus*, where TALEN mRNAs are delivered at a single-cell stage. Therefore, identification of mutant alleles without a visible phenotype is laborious and requires interrogation of the genomic TALEN target loci. The use of genetic markers has made identification of transgenic mutants and tracking of the transgenes an uncomplicated matter. However, to date, no efforts have been made to use a traceable marker as an indicator of TALEN-induced mutations. TALEN-mediated direct in-frame introduction of fluorescence genes into endogenous genes has been attempted in cultured cells (31,32) and zebrafish (26). Unfortunately, the utility of these selection markers depends on the expression levels of the fusion proteins.

Here we develop a simple and efficient method of assembling TALENs into an expression vector that can subsequently be used for *Drosophila* embryo microinjections. We demonstrate that we can induce targeted locus-specific mutations via NHEJ as well as marker gene integration via HR in the *Drosophila* genome. This traceable marker allows efficient identification of the engineered flies.

## MATERIALS AND METHODS

### Construction of TALEN backbone plasmid

The TALEN backbone created in this work is a modification of the AvrXa7-FN from B. Yang's lab (33). Owing to a high homology and related origin of AvrXa7 and AvrXa10 proteins, the truncation produced in this study was modelled after the AvrXa10 truncation of Sun *et al.* (34). The N3-C1 truncation produced in this study retained 207 amino acids on the N-terminus and 63 amino acids on the C-terminus (Supplementary Figure S1). The nuclear localization signal (NLS) from SV40 T-antigen was added to the N-terminus. The CAT and *ccdB* selection markers were inserted between N3 and C1 fragments with Mlu I and Aat II. To allow insertion of a tailored TALE-repeat in easy TALEN assembly (easyT) protocol, Bae I sites were created on either side of the CAT–*ccdB* cassette. The TALEN.N3-C1 backbone was placed under *Copia* promoter for subsequent expression in *Drosophila* embryos.

### Construction of unit template plasmids

The template DNA sequences for individual monomer units were created by dimerizing each pair of DNA-oligos shown in Supplementary Table S1, and cloned into pGEM-T Easy vector (Promega). The RVDs used were NI for A, NG for T, NK for G and HD for C. Because polymerase chain reaction (PCR) amplification of repetitive sequences is inefficient and 4-mer repeats amplification is required in the easyT protocol, four types of the template DNA sequences (type a, b, c and d) were created for unit 2 to 19 without changing the amino acid sequences (Supplementary Figure S2).

### TALEN construction

The TALEN–*y* pair targeting *y* (chrX: 253 627–263 693) was constructed with a modular assembly protocol (33), and subsequently cloned into the TALEN N3-C1 backbone. TALEN–*w* pair (chrX: 2 686 269–2 686 214), TALEN–*upd* pair (chrX: 18 207 553–18 207 606) and TALEN–*wg* pair (chr2L: 7 325 600–7 325 658) were created with the easyT protocol. See Supplementary Methods for step-by-step easyT protocols and backup troubleshooting protocols. The TALE-repeats and target DNA sequences for each TALEN pair were shown in Table 1.

### Construction of Donor Plasmids

The right and left homology arms of donor DNA were separately amplified by PCR from a wild-type fly genome as a template. Both arms and 3xP3-EGFP-SV40 cassette (35) were cloned into pGEM-T Easy vector. The regions used for homologous donor sequences were as follows: Donor-*y*, chrX: 251 198–255 994; Donor-*upd*, chrX: 18 205 583–18 209 591; and Donor-*wg*, chr2L: 7 323 606–7 327 640. On Donor-*upd* and Donor-*wg*, the putative activator protein-1 (AP-1)–binding sequences TGA[C/G]TCA at TALEN-*upd* or TALEN-*wg* target sites were deleted for further study (TK, unpublished data). The primers used are available on request.

### *Drosophila* embryo microinjection

The concentrations of TALEN and donor plasmids used were shown in Tables 2 and 3. The syncytial blastodermal embryos of wild type or $w^{1118}$ $lig4^{169}$ (36) were used for microinjection. To detect TALEN-induced germ line *y* or *w* mutants, the eclosed flies ($F_0$) were individually crossed with lesion-known $y^1$ $w^1$ or $w^{1118}$ mutants, respectively. Because the levels of 3xP3-EGFP expression are influenced by the chromosomal location, the identification of the mutant ($F_1$) expressing EGFP in its eyes was performed in $w^{1118}$ background. Frequencies of targeted mutagenesis or gene integration were calculated as the frequency of yielders: the number of $F_0$ yielding mutant offspring in $F_1$ per the total number of fertile $F_0$.

### Validation of targeted mutagenesis

Genomic DNA of TALEN-mediated mutants were isolated from single fly with QIAamp DNA Micro Kit (QIAGEN). The TALEN-*y* and TALEN-*w* target sites

**Table 1.** TALENs constructed in this work

| TALEN | RVDs and target sequence | | Repeats | Assembly protocol (reference) |
|---|---|---|---|---|
| TALEN-*y*-L | NI NI NG NG HD NG HD HD NG NI NG NG NG NG NI NG NG NG NG NG NG NI NK NK<br>a a t t c t c c t a t t t t a t t t t t a g G | | 23.5 | Modular assembly (33) |
| TALEN-*y*-R | HD NI NI NI HD NG NK HD NG NK NG NG NI NG NK NG NG NG NI NG NI NG NI<br>C A A A C T G C G G T C C A T G T T T A T A T A | | 23.5 | Modular assembly (33) |
| TALEN-*w*-L | HD NK NK HD HD NI NG NG NI NK NI NI NK NK NI NG HD NG NG<br>C G G C C A T C A G A A G G A T C T T | | 18.5 | easyT (this study) |
| TALEN-*w*-R | NG HD NI NG HD NI NK HD HD NK NG HD NG NG HD HD NK NI NK<br>T C A T C A G C C G T C T T C C G A G | | 18.5 | easyT (this study) |
| TALEN-*upd*-L | NK HD HD HD NK NI HD NI NG NG NG NG NK HD HD<br>g c c g a c a t t t t g c c | | 14.5 | easyT (this study) |
| TALEN-*upd*-R | NI HD HD NK NG NI NI NK HD HD NK NG HD NG<br>a c c g t a c a g c c g t c t | | 14.5 | easyT (this study) |
| TALEN-*wg*-L | HD NI NG HD NG NK NI NG NK HD NG NG HD NI HD NI NK NI NI<br>c a t c t g a t g c t t c a c a g a a | | 18.5 | easyT (this study) |
| TALEN-*wg*-R | NK NI NI NG NG HD HD NG NK NI NI NI NI HD NG NK NI NI NG HD<br>g a a t t c c t g a a a c t g a a t c | | 18.5 | easyT (this study) |

In each TALEN target sequence, protein coding sequences are shown in upper cases, and intron or intergenic region are written in lower cases.

**Table 2.** TALEN-mediated targeted gene mutagenesis

| Target | | Recipient | Concentration (ng/μl each) | Frequency of targeted gene integration[a] | | |
|---|---|---|---|---|---|---|
| Gene | Chr. | | | Male | Female | Total |
| *y* | X | Wild type | 25 | 27.6% (8/29) | 8.3% (1/12) | 22.0% (9/41) |
| | | | 50 | 43.2% (19/44) | 11.1% (5/45) | 27.0% (24/89) |
| | | | 100 | 22.4% (11/49) | 6.8% (3/44) | 15.1% (14/93) |
| | | | 250 | 0.0% (0/39) | 0.0% (0/42) | 0.0% (0/81) |
| *w* | X | Wild type | 50 | 12.5% (3/24) | 0.0% (0/14) | 7.9% (3/38) |
| | | | 100 | 0.0% (0/68) | 0.0% (0/50) | 0.0% (0/118) |

[a]The frequencies were calculated as frequency of yielder.

**Table 3.** TALEN-mediated targeted gene integration

| Target | | Homology length of donor DNA (kb) | Concentration (ng/μl) | | Frequency of targeted gene integration[a] | | |
|---|---|---|---|---|---|---|---|
| Gene | Chr. | | TALEN | Donor | Male | Female | Total |
| *y* | X | 5.0 | 50 | 100 | 1.6% (1/63) | 1.5% (1/67) | 1.5% (2/130) |
| *upd* | X | 4.0 | 50 | 500 | 8.3% (9/108) | 1.6% (3/189) | 4.0% (12/297) |
| | | 4.0 | 50 | 100 | 14.3% (11/77) | 6.7% (5/75) | 10.5% (16/152) |
| *wg* | II | 4.0 | 50 | 100 | 2.6% (3/114) | 1.4% (1/72) | 2.2% (4/186) |

[a]The frequencies were calculated as frequency of yielder.

at *y* or *w* loci were amplified and sequenced. TALEN-mediated targeted 3xP3-EGFP integrations were also determined by PCR amplification and sequencing. Primer pairs used in genomic DNA amplification are shown in Supplementary Table 3.

### *In silico* estimation of TALEN specificity as a function TALEN-binding sequence length

To assess the TALEN target specificity, we used a custom-made R script using different bioconductor packages (37) and multicore (parallel processing of R code on machines with multiple cores or CPUs, Simon Urbanek, version 0.1-7). To estimate target specificity of a TALEN-L/R pair of a given TALE-repeat length, 500 TALEN target sites, which consist of TALEN-L– and -R–binding sequences separated by a spacer sequence ranging from 12 to 32 bases (randomly assigned in each sample), were sampled uniformly from *D. melanogaster* genome. Here, TALE-repeat lengths ranging from 5 to 20 were considered. In total, 500 samples were independently generated from each of the following chromosomal arms: chr2L, chr2R, chr3L, chr3R, chrX, chr2LHet, chr2RHet, chr3LHet, chr3RHet and chrXHet. The TALEN-L– and -R–binding sequences were subsequently used to generate a set of sequences representing

all potential targeting sites. For this purpose, the TALEN-L– and -R–binding sequences are joined to either end of spacer sequences of Ns, which vary in length from 12 to 32; [TALEN-L–binding sequence]+[spacer: $N_{12-32}$]+[TALEN-R–binding sequence]. In addition, and to account for the targeting sites caused by TALEN-L and TALEN-R self-pairing, [TALEN-L–binding sequence]+[spacer: $N_{12-32}$]+[TALEN-L–binding sequence] and [TALEN-R–binding sequence]+[spacer: $N_{12-32}$]+[TALEN-R–binding sequence] were generated. All sequences (63 in total) were aligned to the *D. melanogaster* reference genome (BDGP Release 5 dm3) requiring a perfect match for both TALEN-binding sequences while ignoring the spacer sequence of Ns. A TALEN target site was considered to be specific if only the initial sequence uniquely aligns to its origin and all other constructed sequences fail to align. Hence, sequence sets exhibiting more than one alignment are considered to be unspecific. Using all 500 samples, the TALEN target specificity for a given length and chromosomal arm was calculated as the number of specific TALENs divided by the number of samples. To assess the contribution of genomic repeat regions to unspecific TALEN-binding sites, we subtracted all samples derived from repeat regions of the genome. Repeat regions were identified by Tandem Repeats Finder and RepeatMasker masks, including, for example, simple repeats and transposable elements, contained in the *D. melanogaster* object provided by the BSgenome bioconductor package.

## RESULTS

### The length of TALE-repeats required for specific targeting in *Drosophila* genome

TALENs with desired sequence specificities are constructed by arranging the order of TALE-repeat modules. To gain the target specificity, TALE-repeats have to contain a certain number of modules. Additionally, highly site-specific DSB introduced by a TALEN pair is in part supported by the requirement that Fok I nuclease domain acts as dimmer (7). The genome of *D. melanogaster* is ~180 Mb in length, consisting of ~120 Mb of euchromatic sequences and 60 Mb of heterochromatic regions (38). Thus, to determine the length of the TALE-repeats required for minimizing unfavourable off-target DNA cleavages in *D. melanogaster* genome, we computationally assessed TALEN-pair–binding specificity as a function of the TALE-repeat–binding sequence length. To this end, we sampled 500 TALEN-L/R pairs (TALEN target sites) of a given TALE-repeat length uniformly from the euchromatin or heterochromatin of each chromosomal arm of *D. melanogaster*. Subsequently, a set of possible target sites by TALEN-L/R pair and TALEN-L or -R self-pairing were constructed and aligned to *D. melanogaster* reference genome (Figure 1A). A TALEN target site was considered to be specific if only the initial sequence uniquely aligns to its origin and all other constructed sequences fail to align. The *in silico* analysis revealed that TALE-repeats longer than 12 bp reached a plateau
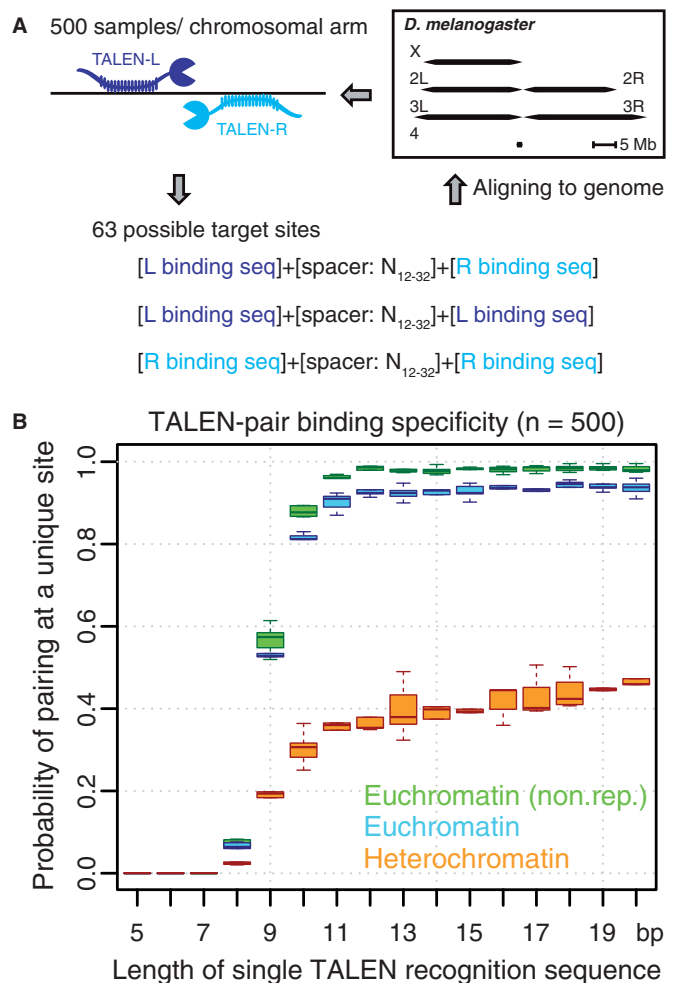


**Figure 1.** Target specificity of TALEN-pairs as a function of the binding sequence length. (**A**) The overview of *in silico* analysis. Five hundred TALEN-pair target sites were sampled uniformly from euchromatic or heterochromatic regions of the different chromosomal arms. The specificity of a TALEN-L/R pair in each sample was determined by aligning all the potential targeting sites to *D. melanogaster* reference genome (BDGP Release 5 dm3). Specific TALEN-pairs contain exactly one targeting site within the genome, and hence, TALEN-pairs exhibiting multiple targeting sequences are considered to be unspecific. (**B**) The probability of specific targeting was calculated as the number of specific TALEN pairs divided by the number of samples. Results obtained for different chromosome arms were summarized using a boxplot representation. To assess the contribution of genomic repeat regions (e.g. simple repeats or transposable elements) to the fraction of unspecific TALEN pairs, the specificity was estimated using only samples localizing outside of annotated repeat regions within the euchromatin.

in TALEN-pair–binding specificity in the euchromatic regions of the genome (Euchromatin in Figure 1B). To assess the contribution of genomic repeat regions to unspecific TALEN-binding sites, we subtracted the samples derived from repeat regions such as simple repeats or transposable elements on the euchromatin [Euchromatin (non-repetitive) in Figure 1B]. On the other hand, more than half of TALEN-pairs sampled from heterochromatic regions still exhibited off-targets even with TALE-repeat length of 20 bp (Heterochromatin in Figure 1B). These results suggested that TALEN-pairs of TALE-repeat
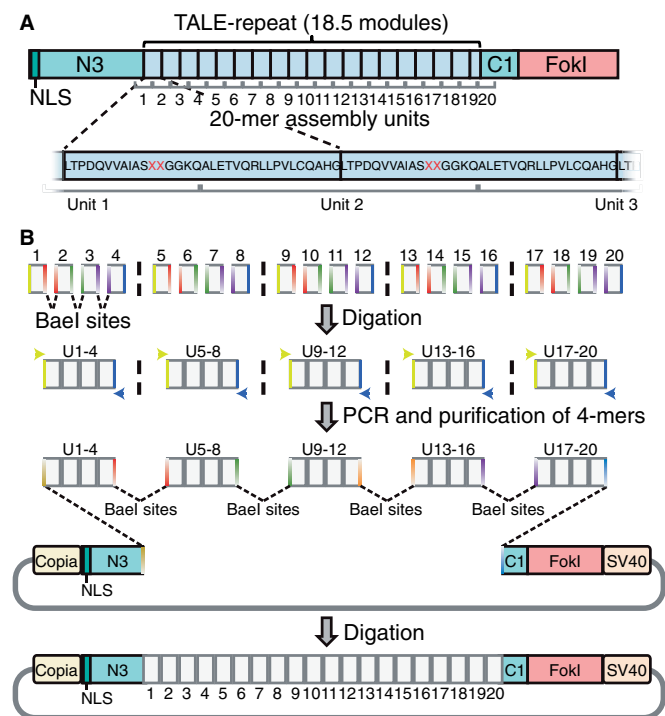
**Figure 2.** Construction of TALENs with the easyT protocol. (**A**) A schematic representation of a TALEN with a TALE-repeat length of 18.5 modules. The TALE-repeat is assembled from 20 monomer units. The boundaries of monomer units were shifted from those of the TALE-repeat modules. (**B**) Overview of TALEN cloning. In the first step, four units are assembled into 4-mers in a 'digation' reaction. In the second step, 4-mers are PCR-amplified, run on an agarose gel, gel-extracted and concentrated. Finally, 4-mers were assembled into the TALEN backbone plasmid in the second digation reaction. Yellow and blue arrows indicate primers used for 4-mer amplification.

longer than 12 bp are required for specific targeting within non-repetitive euchromatic regions of *D. melanogaster* genome. Accordingly, we constructed the TALEN-pairs recognizing sequences >15 bp (Table 1).

## Development of the easyT protocol

A variety of TALEN assembly protocols are now available (39). However, these protocols rely on extensive plasmid libraries, special reagents or extended cloning schemes requiring many days of work (22,33,40–44). We developed a new easyT protocol that enables us to construct TALENs of up to 18.5 TAL-repeat modules in a single cloning process (Figure 2, Supplementary Figures S1–S5 and Supplementary Methods). Importantly, the activity of the restriction enzyme Bae I at 25°C allowed combining DNA digestion and ligation into a single 1-h 'digation' reaction, thereby saving time and effort (Figure 2B). In a first step, individual units were assembled into 4-mers in a first digation reaction. DNA sequence of each unit within a 4-mer has been altered to reduce nucleotide sequence homology without changing amino acid sequence. This enabled efficient PCR-amplification of 4-mers regardless of combination of four units (Supplementary Figure S2). In a subsequent second digation, a TALE-repeat was assembled into the pre-digested TALEN backbone plasmid and followed by

conventional cloning procedure. Thus, construction of custom TALENs can be achieved in a single day. Besides the simplicity of the cloning procedures, the easyT prepares a unique troubleshooting protocol. In the regular Golden Gate based TALEN assembly protocols (39), the assembled units are no longer replaceable. By shifting the boundaries of assembly units from those of TALE-repeat modules, however, we made it possible to embed unique restriction sites at the borders of every 4-mer units, enabling users to replace any 4-mer in case there is a mutation by PCR-error or misligation (Supplementary Figure S3).

## Targeted mutagenesis in *Drosophila* germ line by microinjection of TALEN expression plasmids

The microinjection of plasmid DNA into *Drosophila* embryos has been routinely used to generate transgenic flies. Co-injection of P-element–based transformation vector with a transposase-expressing helper plasmid can efficiently integrate the transgenes into the genome of germ cell lineage (45). Additionally, the use of plasmid DNA for TALEN introduction eliminates an mRNA synthesis step. Hence, the TALEN plasmids synthesized via easyT were used directly in embryo microinjections. We first tested whether TALEN plasmid DNA microinjection can be used to introduce DSB in *Drosophila* embryos by targeting two genes, *yellow* (*y*) and *white* (*w*), mutations of which result in a visible phenotype. Flies carrying a knockout of *y* display a characteristic yellow body phenotype, while mutation of *w* results in reduced or completely lost red eye pigmentation. As shown in Table 2, a series of different TALEN plasmid concentrations were tested for *y* targeting, with a peak [27.0% (24/89)] in mutation frequency observed at 50 ng/µl concentration for each TALEN plasmid. Interestingly, the mutation frequency was consistently higher in males than in females at all concentrations tested. For *w* targeting, 50 ng/µl and 100 g/µl concentrations were tested, yielding germ line *w* mutants with 7.9% (3/38) and 0.0% (0/118) frequencies, respectively (Table 2). The optimal TALEN concentration is likely to be a function of the TALEN pair affinity to its recognition sequence, the chromatin status of the target locus, and potential off-target sites in the genome and thus might need to be adjusted for individual TALEN pairs.

The targeted mutations were determined by sequencing genomic TALEN target region. For the *y* mutants generated, randomly selected 10 independent mutant lines had a mutation by NHEJ-induced indels at the target site (Figure 3A, C and D). Interestingly, targeting of the *w* gene resulted in a distinctive genomic lesion in all three mutant lines derived from individual microinjected embryos (Figure 3B). The deletion of nine nucleotides from the coding sequence caused reduced red eye pigmentation in mutants (Figure 3E and F) and could be due to an accidental 5-bp microhomology in the spacer sequence between two TALEN-binding sites. In summary, TALEN plasmid microinjection is a convenient alternative to mRNA microinjection for germ line mutagenesis in *Drosophila* embryos.
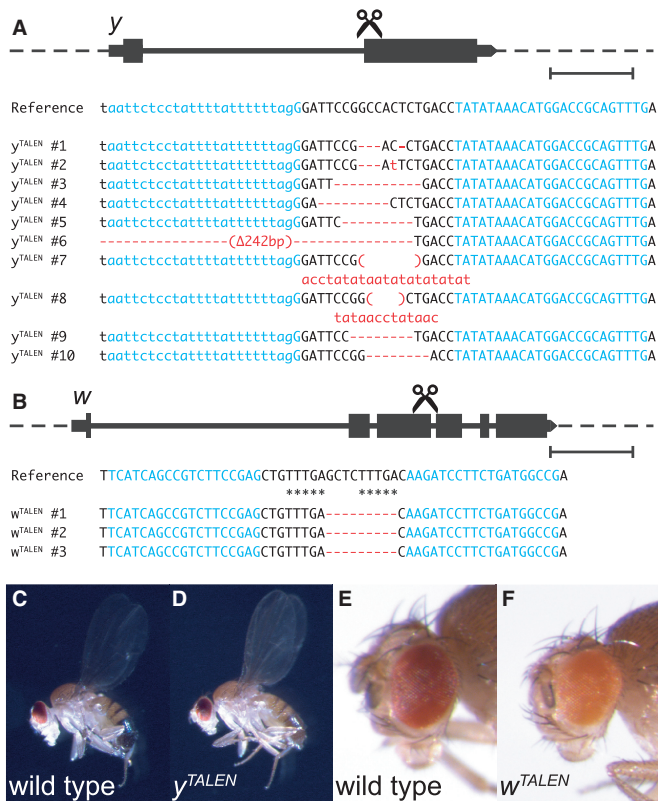
**Figure 3.** Targeted germ line gene mutagenesis. (**A**) TALEN-induced mutations in *y*. The target site was chosen at the junction of exon 2 and intron, as indicated with the scissors mark. Ten *y* mutants ($y^{TALEN}$) derived individual $F_0$ were sequenced. (**B**) TALEN-induced mutations in *w*. The target site encodes the end part of the ATP-binding cassette transporter domain. Three independently generated $F_1$ mutants have the same genomic lesions of a 9-bp deletion. The sequence microhomology shown with asterisk at the target site indicates the probable DSB repair through microhomology-mediated end-joining. (**C** and **D**) The body colour phenotypes of wild type and $y^{TALEN}$ flies, both male, are shown. (**E** and **F**) The eye colour phenotypes of wild type and $w^{TALEN}$ flies are shown. $w^{TALEN}$ is a hypomorphic allele, showing an orange eye colour. In the TALEN target sequences, the DNA aberrations are shown in red and TALEN-binding sites in blue. The intronic and exonic sequences are shown in lowercase and uppercase, respectively. Scale bar represents 1 kb length.

### TALEN-mediated targeted gene integration by HR in *Drosophila* germ line

Next, we investigated TALEN-mediated targeted gene integration via HR between the target locus and an exogenously provided donor DNA. To expand the TALEN application to any region in the genome, we constructed two TALEN-pairs targeting the intergenic region of *unpaired* (*upd*) and the *wingless* (*wg*) locus, respectively (Table 1). To that end, three donor DNA plasmids were constructed to contain an eye-specific enhanced green fluorescent protein marker gene [3xP3-EGFP (35)] flanked on either side by homology arms (Figure 4A–C). Insertion of the marker gene in *y* donor DNA disrupts the coding sequence of the *y* gene. In the other two cases, the insertion results in a deletion of a putative AP-1–binding sequence in the *upd* and *wg* donor DNA. Co-injection of donor DNA plasmids (100 ng/µl) along with TALEN plasmids (50 ng/µl each) was carried out in a DNA

ligase IV (*lig4*) mutant background (36), shown previously to have higher efficiency of donor DNA integration via HR (30).

The flies carrying germ line integrations of the marker gene were identified by green fluorescence in eyes of the progeny of microinjected embryos (Figure 4D–F). The fluorescence level and pattern were different for the three genomic locations tested, probably reflecting the influence of surrounding DNA regulatory elements on the transcriptional activity of the marker gene. The triple-plasmid microinjection resulted in a successful introduction of germ line transmitting mutations at each locus: 1.5% (2/130) for *y*, 10.5% (16/152) for *upd* and 2.2% (4/186) for *wg* (Table 3). Using the best TALEN-pair targeting *upd*, we examined the frequency of HR event with higher concentrations (500 ng/µl) of donor DNA plasmids (Table 3). Unexpectedly, increasing donor DNA concentration resulted in a decrease of frequency of HR-mediated gene integration (4.0%, 12/297). Genomic DNA of EGFP-positive flies was amplified to confirm targeted gene integration (Figure 4G–J). Amplicons of $upd^{TALEN-EGFP}$ and $wg^{TALEN-EGFP}$ were also sequenced to verify that they indeed represent TALEN-mediated targeted gene integration (data not shown). These results demonstrate that targeted mutations via gene integration can be efficiently induced in *Drosophila* and easily scored with a traceable recombination marker.

## DISCUSSION

Introduction of site-specific genetic modifications in a controlled manner has been recognized as an ultimate approach in biosciences to understand the function of a gene product or gene regulation. The development of the customizable engineer nucleases such as ZFNs and, more recently, TALENs made a leap forward for targeted genome editing in a variety of organisms. In this study, we developed a new TALEN synthesis method named easyT. This protocol allows construction of custom TALENs from monomer units in a single day and relies only on standard techniques commonly used in a molecular biology lab. Other assembly methods that permit construction of TALENs in a single day require either solid phase high-throughput ligation technology or a library of preassembled 4- or 5-mer unit plasmids (40,43,44). Furthermore, an additional feature of easyT is the capability of replacing a part of assembled units every 4-mers. The time to synthesize a TALEN by a given protocol relies on the successful cloning. Many of the currently available TALEN assembly protocols are based on either plasmid- or PCR-based Golden Gate cloning (39), where the correct assembly is confirmed at the end by sequencing. At this point, mutations introduced in TALE-repeat cannot be fixed, and either more clones are sequenced or the assembly procedure has to be repeated. Therefore, the ability to replace a part of TALE-repeats is a valuable troubleshooting feature of easyT.

TALEN technology has been quickly and successfully applied to a wide variety of cells and organisms in the past years. To evaluate the utilities of TALENs in each model
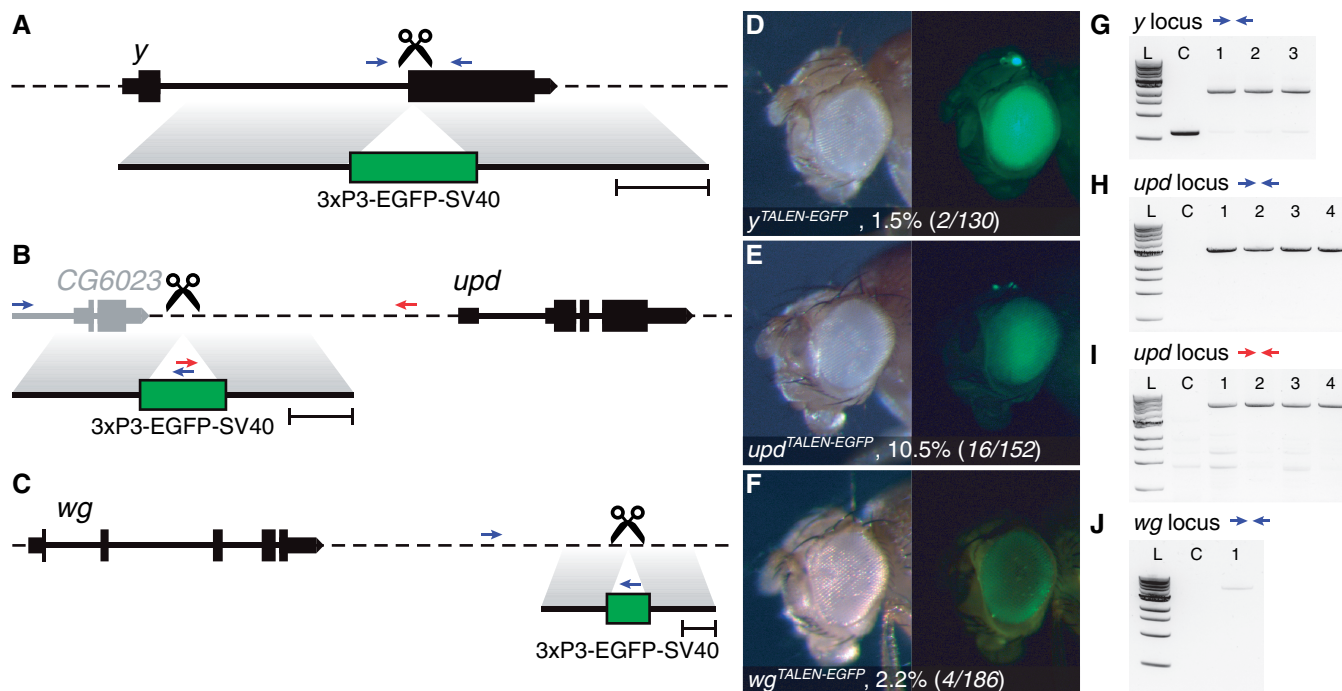
**Figure 4.** TALEN-mediated targeted marker gene integration. (**A–C**) Schematics of TALEN target loci and corresponding donor DNA. TALEN target sites are indicated with scissors. Scale bar in each panel represents 1 kb length. (**D–F**) Mutant progeny of microinjected embryos with TALEN-mediated targeted 3xP3-EGFP integration. $y^{TALEN-EGFP}$ fly also show the expected yellow body colour (**D**). $upd^{TALEN-EGFP}$ and $wg^{TALEN-EGFP}$ mutants show weaker levels and specific patterns of green fluorescence in the fly eyes (E and F). (**G–J**) Genomic DNA amplification of $y^{TALEN-EGFP}$, $upd^{TALEN-EGFP}$ and $wg^{TALEN-EGFP}$. Red and blue arrows indicate primers used for genomic DNA amplification.

organism, most of the pioneering work has scored the frequency of NHEJ-mediated targeted mutations using the genes that express known visible phenotypes. In the practical use of TALEN technology, however, it is often required to target loci in which resulting phenotypes are not known. Because the frequencies of mutant production are generally not high enough in *Drosophila* (18,21,30), the identification of mutants could be the key step in *Drosophila* genome engineering. In this work, we demonstrated that TALEN-directed targeted genome alteration through HR with exogenous donor DNA could be an efficient and straight approach for mutant generation and identification. In this regard, we demonstrate the utility of a traceable recombination marker introduced by the donor DNA. First, it allows identification of germ line transmitting mutants as early as the progeny of injected embryos start to hatch. Second, it enables tracking of the mutated alleles for generating hetero- or homozygous stable lines for subsequent experiments. And finally, the easy identification allows more flexibility in the mode of TALEN introduction into *Drosophila* embryos, and we showed that TALEN plasmid microinjection is a convenient alternative to the mRNA microinjection in *Drosophila*. As we speculate from the limited number of examples in this work and previous studies (18,21), the simple identification of mutants by traceable markers will not be substantially improved by the potential of a slight increase in mutant frequencies, thus shifting the focus from TALEN efficiency to the overall convenience of the method. For example, the use of plasmid DNA for TALEN introduction eliminates additional mRNA synthesis steps and possible concerns on the quality of TALEN mRNA during microinjection. Taken together, HR-mediated gene integration with donor DNA bearing a traceable marker appears as a more favourable and practical strategy.

The utility of TALEN-mediated targeted gene integration can be extended to precise genome engineering by combining TALEN-mediated targeted gene integration with site-specific recombinase/integrase systems such as flippase (Flp)-FRT, Cre-*loxP* or PhiC31 (46), and the *piggyBac* transposon system (47). For example, by using the 3xP3-EGFP cassette having flp/Cre recognition sites besides the 3xP3 promoter, the 3xP3 can be removed by appropriate recombinases, while leaving EGFP at the target site (Supplementary Figure S6). In this way, it is now possible to tag endogenous gene products with fluorescence protein regardless of their expression levels.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online, including [48].

## ACKNOWLEDGEMENTS

## REFERENCES

1. Baudin,A., Ozier-Kalogeropoulos,O., Denouel,A., Lacroute,F. and Cullin,C. (1993) A simple and efficient method for direct gene deletion in Saccharomyces cerevisiae. *Nucleic Acids Res.*, **21**, 3329–3330.
2. Hinnen,A., Hicks,J.B. and Fink,G.R. (1978) Transformation of yeast. *Proc. Natl Acad. Sci. USA*, **75**, 1929–1933.
3. Capecchi,M.R. (1989) Altering the genome by homologous recombination. *Science*, **244**, 1288–1292.
4. Schaefer,D.G. (2002) A new moss genetics: targeted mutagenesis in Physcomitrella patens. *Annu. Rev. Plant Biol.*, **53**, 477–501.
5. Kim,Y.G., Cha,J. and Chandrasegaran,S. (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc. Natl Acad. Sci. USA*, **93**, 1156–1160.
6. Christian,M., Cermak,T., Doyle,E.L., Schmidt,C., Zhang,F., Hummel,A., Bogdanove,A.J. and Voytas,D.F. (2010) Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*, **186**, 757–761.
7. Bitinaite,J., Wah,D.A., Aggarwal,A.K. and Schildkraut,I. (1998) FokI dimerization is required for DNA cleavage. *Proc. Natl Acad. Sci. USA*, **95**, 10570–10575.
8. Mussolino,C. and Cathomen,T. (2012) TALE nucleases: tailored genome engineering made easy. *Curr. Opin. Biotechnol.*, **23**, 644–650.
9. Boch,J. and Bonas,U. (2010) Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annu. Rev. Phytopathol.*, **48**, 419–436.
10. Boch,J., Scholze,H., Schornack,S., Landgraf,A., Hahn,S., Kay,S., Lahaye,T., Nickstadt,A. and Bonas,U. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, **326**, 1509–1512.
11. Moscou,M.J. and Bogdanove,A.J. (2009) A simple cipher governs DNA recognition by TAL effectors. *Science*, **326**, 1501.
12. Morbitzer,R., Romer,P., Boch,J. and Lahaye,T. (2010) Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. *Proc. Natl Acad. Sci. USA*, **107**, 21617–21622.
13. Miller,J.C., Tan,S., Qiao,G., Barlow,K.A., Wang,J., Xia,D.F., Meng,X., Paschon,D.E., Leung,E., Hinkley,S.J. et al. (2011) A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.*, **29**, 143–148.
14. Mak,A.N., Bradley,P., Cernadas,R.A., Bogdanove,A.J. and Stoddard,B.L. (2012) The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science*, **335**, 716–719.
15. Bochtler,M. (2012) Structural basis of the TAL effector-DNA interaction. *Biol. Chem.*, **393**, 1055–1066.
16. Wood,A.J., Lo,T.W., Zeitler,B., Pickle,C.S., Ralston,E.J., Lee,A.H., Amora,R., Miller,J.C., Leung,E., Meng,X. et al. (2011) Targeted genome editing across species using ZFNs and TALENs. *Science*, **333**, 307.
17. Sajwan,S., Takasu,Y., Tamura,T., Uchino,K., Sezutsu,H. and Zurovec,M. (2013) Efficient disruption of endogenous Bombyx gene by TAL effector nucleases. *Insect Biochem. Mol. Biol.*, **43**, 17–23.
18. Liu,J., Li,C., Yu,Z., Huang,P., Wu,H., Wei,C., Zhu,N., Shen,Y., Chen,Y., Zhang,B. et al. (2012) Efficient and specific modifications of the *Drosophila* genome by means of an easy TALEN strategy. *J. Genet. Genomics*, **39**, 209–215.
19. Watanabe,T., Ochiai,H., Sakuma,T., Horch,H.W., Hamaguchi,N., Nakamura,T., Bando,T., Ohuchi,H., Yamamoto,T., Noji,S. et al. (2012) Non-transgenic genome modifications in a hemimetabolous insect using zinc-finger and TAL effector nucleases. *Nat. Commun.*, **3**, 1017.
20. Ma,S., Zhang,S., Wang,F., Liu,Y., Xu,H., Liu,C., Lin,Y., Zhao,P. and Xia,Q. (2012) Highly efficient and specific genome editing in silkworm using custom TALENS. *PLoS One*, **7**, e45035.
21. Sakuma,T., Hosoi,S., Woltjen,K., Suzuki,K., Kashiwagi,K., Wada,H., Ochiai,H., Miyamoto,T., Kawai,N., Sasakura,Y. et al. (2013) Efficient TALEN construction and evaluation methods for human cell and animal applications. *Genes Cells*, **18**, 315–326.
22. Huang,P., Xiao,A., Zhou,M., Zhu,Z., Lin,S. and Zhang,B. (2011) Heritable gene targeting in zebrafish using customized TALENs. *Nat. Biotechnol.*, **29**, 699–700.
23. Cade,L., Reyon,D., Hwang,W.Y., Tsai,S.Q., Patel,S., Khayter,C., Joung,J.K., Sander,J.D., Peterson,R.T. and Yeh,J.R. (2012) Highly efficient generation of heritable zebrafish gene mutations using homo- and heterodimeric TALENS. *Nucleic Acids Res.*, **40**, 8001–8010.
24. Dahlem,T.J., Hoshijima,K., Jurynec,M.J., Gunther,D., Starker,C.G., Locke,A.S., Weis,A.M., Voytas,D.F. and Grunwald,D.J. (2012) Simple methods for generating and detecting locus-specific mutations induced with TALENs in the zebrafish genome. *PLoS Genet.*, **8**, e1002861.
25. Bedell,V.M., Wang,Y., Campbell,J.M., Poshusta,T.L., Starker,C.G., Krug,R.G. II, Tan,W., Penheiter,S.G., Ma,A.C., Leung,A.Y. et al. (2012) *In vivo* genome editing using a high-efficiency TALEN system. *Nature*, **491**, 114–118.
26. Zu,Y., Tong,X., Wang,Z., Liu,D., Pan,R., Li,Z., Hu,Y., Luo,Z., Huang,P., Wu,Q. et al. (2013) TALEN-mediated precise genome modification by homologous recombination in zebrafish. *Nat. Methods*, **10**, 329–331.
27. Lei,Y., Guo,X., Liu,Y., Cao,Y., Deng,Y., Chen,X., Cheng,C.H., Dawid,I.B., Chen,Y. and Zhao,H. (2012) Efficient targeted gene disruption in Xenopus embryos using engineered transcription activator-like effector nucleases (TALENs). *Proc. Natl Acad. Sci. USA*, **109**, 17484–17489.
28. Tesson,L., Usal,C., Menoret,S., Leung,E., Niles,B.J., Remy,S., Santiago,Y., Vincent,A.I., Meng,X., Zhang,L. et al. (2011) Knockout rats generated by embryo microinjection of TALENs. *Nat. Biotechnol.*, **29**, 695–696.
29. Carlson,D.F., Tan,W., Lillico,S.G., Stverakova,D., Proudfoot,C., Christian,M., Voytas,D.F., Long,C.R., Whitelaw,C.B. and Fahrenkrug,S.C. (2012) Efficient TALEN-mediated gene knockout in livestock. *Proc. Natl Acad. Sci. USA*, **109**, 17382–17387.
30. Beumer,K.J., Trautman,J.K., Bozas,A., Liu,J.L., Rutter,J., Gall,J.G. and Carroll,D. (2008) Efficient gene targeting in *Drosophila* by direct embryo injection with zinc-finger nucleases. *Proc. Natl Acad. Sci. USA*, **105**, 19821–19826.
31. Hockemeyer,D., Wang,H., Kiani,S., Lai,C.S., Gao,Q., Cassady,J.P., Cost,G.J., Zhang,L., Santiago,Y., Miller,J.C. et al. (2011) Genetic engineering of human pluripotent cells using TALE nucleases. *Nat. Biotechnol.*, **29**, 731–734.
32. Zhang,Y., Zhang,F., Li,X., Baller,J.A., Qi,Y., Starker,C.G., Bogdanove,A.J. and Voytas,D.F. (2013) Transcription activator-like effector nucleases enable efficient plant genome engineering. *Plant Physiol.*, **161**, 20–27.
33. Li,T., Huang,S., Zhao,X., Wright,D.A., Carpenter,S., Spalding,M.H., Weeks,D.P. and Yang,B. (2011) Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. *Nucleic Acids Res.*, **39**, 6315–6325.
34. Sun,N., Liang,J., Abil,Z. and Zhao,H. (2012) Optimized TAL effector nucleases (TALENs) for use in treatment of sickle cell disease. *Mol. Biosyst.*, **8**, 1255–1263.
35. Horn,C., Jaunich,B. and Wimmer,E.A. (2000) Highly sensitive, fluorescent transformation marker for *Drosophila* transgenesis. *Dev. Genes Evol.*, **210**, 623–629.
36. McVey,M., Radut,D. and Sekelsky,J.J. (2004) End-joining repair of double-strand breaks in *Drosophila* melanogaster is largely DNA ligase IV independent. *Genetics*, **168**, 2067–2076.
37. Gentleman,R.C., Carey,V.J., Bates,D.M., Bolstad,B., Dettling,M., Dudoit,S., Ellis,B., Gautier,L., Ge,Y., Gentry,J. et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.*, **5**, R80.
38. Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Scherer,S.E., Li,P.W.,

Hoskins,R.A., Galle,R.F. *et al.* (2000) The genome sequence of *Drosophila* melanogaster. *Science*, **287**, 2185–2195.

39. Xiao,A., Wu,Y., Yang,Z., Hu,Y., Wang,W., Zhang,Y., Kong,L., Gao,G., Zhu,Z., Lin,S. *et al.* (2013) EENdb: a database and knowledge base of ZFNs and TALENs for endonuclease engineering. *Nucleic Acids Res.*, **41**, D415–D422.

40. Reyon,D., Tsai,S.Q., Khayter,C., Foden,J.A., Sander,J.D. and Joung,J.K. (2012) FLASH assembly of TALENs for high-throughput genome editing. *Nat. Biotechnol.*, **30**, 460–465.

41. Cermak,T., Doyle,E.L., Christian,M., Wang,L., Zhang,Y., Schmidt,C., Baller,J.A., Somia,N.V., Bogdanove,A.J. and Voytas,D.F. (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.*, **39**, e82.

42. Reyon,D., Khayter,C., Regan,M.R., Joung,J.K. and Sander,J.D. (2012) Engineering designer transcription activator-like effector nucleases (TALENs) by REAL or REAL-Fast assembly. *Curr. Protoc. Mol. Biol.*, **Chapter 12**, Unit 12.15.

43. Briggs,A.W., Rios,X., Chari,R., Yang,L., Zhang,F., Mali,P. and Church,G.M. (2012) Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. *Nucleic Acids Res.*, **40**, e117.

44. Schmid-Burgk,J.L., Schmidt,T., Kaiser,V., Honing,K. and Hornung,V. (2013) A ligation-independent cloning technique for high-throughput assembly of transcription activator-like effector genes. *Nat. Biotechnol.*, **31**, 76–81.

45. Rubin,G.M. and Spradling,A.C. (1982) Genetic transformation of *Drosophila* with transposable element vectors. *Science*, **218**, 348–353.

46. Hirano,N., Muroi,T., Takahashi,H. and Haruki,M. (2011) Site-specific recombinases as tools for heterologous gene integration. *Appl. Microbiol. Biotechnol.*, **92**, 227–239.

47. Kim,A. and Pyykko,I. (2011) Size matters: versatile use of PiggyBac transposons as a genetic manipulation tool. *Mol. Cell Biochem.*, **354**, 301–309.

48. Sievers,F., Wilm,A., Dineen,D., Gibson,T.J., Karplus,K., Li,W., Lopez,R., McWilliam,H., Remmert,M., Soding,J. *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.*, **7**, 539.