1 **Multi-tissue GAL4-mediated gene expression in all *Anopheles gambiae* life**

2 **stages using an endogenous polyubiquitin promoter**

3

4 Adriana Adolfi[a1], Emilie Pondeville[b2], Amy Lynd[a], Catherine Bourgouin[b], and Gareth J Lycett[a*]

5

6 [a]Liverpool School of Tropical Medicine, Vector Biology Department, Liverpool, UK

7 [b]Institut Pasteur, Genetics and Genomics of Insect Vectors, CNRS unit URA3012, Paris, France

8 ***Corresponding Author:** Gareth.Lycett@lstmed.ac.uk, Tel. number: +44 (0)151 705 3137

9 Present addresses:

10 [1]Department of Microbiology and Molecular Genetics, School of Medicine, University of California,

11 Irvine, CA 92697-4500, Adriana.Adolfi@uci.edu

12 [2]MRC-University of Glasgow Centre for Virus Research, Henry Wellcome Building, 464 Bearsden

13 Road, Glasgow, UK, Emilie.Pondeville@glasgow.ac.uk

14

ABSTRACT

The ability to manipulate the *Anopheles gambiae* genome and alter gene expression effectively and reproducibly is a prerequisite for functional genetic analysis and for the development of novel control strategies in this important disease vector. However, *in vivo* transgenic analysis in mosquitoes is limited by the lack of promoters active ubiquitously. To address this, we used the GAL4/UAS system to investigate the promoter of the *An. gambiae Polyubiquitin-c* (PUBc) gene and demonstrated its ability to drive expression in mosquito cell culture before incorporation into *An. gambiae* transgenic driver lines. To generate such lines, piggyBac-mediated insertion was used to identify genomic regions able to sustain widespread expression and to create φC31 docking lines at these permissive sites. Patterns of expression induced by PUBc-GAL4 drivers carrying single intergenic insertions were assessed by crossing with a novel responder UAS-mCD8:mCherry line that was created by φC31-mediated integration. Amongst the drivers created at single, unique chromosomal integration loci, two were isolated that induced differential expression levels in a similar multiple-tissue spatial pattern throughout the mosquito life cycle. This work expands the tools available for *An. gambiae* functional analysis by providing a novel promoter for investigating phenotypes resulting from widespread multi-tissue expression, as well as identifying and tagging genomic sites that sustain broad transcriptional activity.

KEYWORDS

Transgenic mosquitoes; GAL4/UAS; *Polyubiquitin*; piggyBac transposon; φC31 docking lines

1   Introduction

Anopheles gambiae mosquitoes are the major vectors of malaria in the sub-Saharan Africa which accounts for most of the global malaria cases and related deaths (WHO, 2016). While a remarkable reduction in malaria incidence and mortality has been achieved in recent years (Bhatt et al., 2015), the spread of resistance to insecticides and antimalarial drugs poses a major challenge, and the latest report suggests that progress has stalled (WHO, 2017). Therefore, innovative methods for malaria control based on genetically-modified mosquitoes open a promising prospective (Hammond et al., 2016). Such strategies are enabled by the growing abundance of genomic and transcriptomic data combined with the ability to effectively reprogram gene expression in *An. gambiae.* Nevertheless, tools for *in vivo* functional analysis to examine the role of genes in affecting a phenotype are limited, and mosquito transgenic analysis remains a demanding task due to the labor-intensive husbandry of continuous lines.

Binary systems such as the GAL4/UAS (Brand and Perrimon, 1993) can be used in insects for ectopic gene expression or silencing in spatially and/or temporally controlled patterns. The GAL4/UAS system comprises two transgenic components, a driver or GAL4 and a responder or UAS, inserted in separate transgenic lines (Fig. 1A). The driver line carries the yeast transcription activator GAL4 under the control of a specific promoter; the responder line contains an Upstream Activating Sequence (UAS) that serves as GAL4 binding site and activates the transcription of a target gene located downstream. Expression occurs only in the progeny resulting from the cross of driver and responder lines in the transcription pattern dictated by the selected promoter. The modular nature of the system offers several advantages compared to linear transgene expression including amplification of gene expression, creation of lines bearing toxic genes, and repeated use of lines for combinational crosses. Since the first applications in *Drosophila melanogaster* (Brand and Perrimon, 1993; Piccin et al., 2001), the GAL4/UAS system has been applied to *Bombyx mori* (Imamura et al., 2003), *Tribolium castaneum* (Schinko et al., 2010), *Aedes aegypti* (Kokoza and Raikhel, 2011), *An. stephensi* (O'Brochta et al., 2012), as well as *An. gambiae* (Lynd and Lycett, 2011) for gene overexpression (Lynd and Lycett, 2012).

However, gene functional analysis using the GAL4/UAS system is restricted by the limited set of functionally-characterized promoters incorporated in *An. gambiae* driver lines. Indeed, while promoters specific for several mosquito tissues have been characterized [midgut (Moreira et al., 2000; Kim et al., 2004; Abraham et al., 2005; Nolan et al., 2011), testes (Catteruccia et al., 2005), salivary glands (Lombardo et al., 2000, 2005), germline (Papathanos et al., 2009), fat body (Volohonsky et al., 2015 and 2017), and olfactory receptors (Riabinina et al., 2016)], to date only the *An. gambiae* carboxypeptidase promoter has been used to establish a midgut-specific GAL4 line (Lynd and Lycett 2012). Most notable is the lack of a characterized endogenous ubiquitous promoter in *An. gambiae*, which greatly limits our understanding of phenotypes resulting from genes expressed in a multi-tissue manner. In an effort to obtain ubiquitous-like expression, promoters of highly conserved housekeeping genes such as *Polyubiquitin* (PUB) have been investigated in insects. Amongst these, the *D.*

74   *melanogaster* PUB promoter drives widespread expression in transgenic *D. melanogaster* (Handler

75   and Harrell, 1999), *Anastrepha suspensa* (Handler and Harrell, 2001), *Lucilia cuprina* (Heinrich et al.,

76   2002), and the mosquito *An. albimanus* (Perera et al., 2002). A similar phenotype was described for

77   the *Tribolium castaneum* PUB regulatory region in transgenic beetles (Lorenzen et al., 2002). Finally,

78   the promoter of the *Ae. aegypti* PUB gene drives widespread expression across all mosquito

79   developmental stages (Anderson et al., 2010).

80   Achieving the desired expression pattern of transgenes, however, does not solely depend on

81   identifying the regulatory region that will produce the intrinsic spatio-temporal pattern required, but it is

82   affected by the genomic environment into which it is inserted. This has been observed previously

83   examining potential 'ubiquitous promoters' in *An. gambiae*, most notably the *α-tubulin-1b* promoter

84   (Lycett et al., 2012) in which a core expression profile in a limited number of tissues was overlaid by

85   unique temporal and spatial expression patterns in each of the 9 isofemale lines examined. Positional

86   effect is a consequence of the essentially random integration mediated by transposons and results in

87   the laborious process of creation of multiple lines and their assessment for appropriate or 'desired'

88   gene expression and fitness (Handler, 2002). In general, defining universally permissive expression

89   sites is challenging as the properties of any DNA sequence are likely to be affected by its relocation in

90   the genome, yet it is desirable as these active sites would represent an established context for future

91   comparative experiments. By including docking sites in the constructs used to define these locations,

92   characterized lines can be established that bear reusable transcriptionally active sites. From such

93   docking strains it is possible to create a great variety of other lines by integrating the target DNA

94   alongside the existing construct or by replacing the existing cassette with a donor cargo via

95   recombinase-mediated cassette exchange (RMCE) (Bateman et al., 2006), the latter having the

96   advantage of integrating only the desired construct without the plasmid backbone. To date, while

97   φC31-mediated integration (Meredith et al., 2011; Pondeville et al., 2014, Volohonsky et al., 2015)

98   and RMCE (Hammond et al., 2016) have been applied to *An. gambiae*, a very limited number of

99   docking strains has been created.

100   Here, we use the GAL4/UAS system to characterize the regulatory regions of the *An. gambiae*

101   *Polyubiquitin-c* (PUBc) gene, orthologue of *Ae. aegypti* AAEL003877 (Anderson et al., 2010), as a

102   promising candidate for driving ubiquitous expression in transgenic mosquitoes. Driver lines carrying

103   a PUBc-GAL4 cassette were created using the piggyBac transposon and assessed to identify unique

104   genomic locations able to sustain multi-tissue expression to be exploited for subsequent φC31-

105   mediated integration and cassette exchange. To validate GAL4 expression patterns, a unique UAS-

106   mCD8:mCherry responder line was also created by site-specific integration and used to examine

107   expression patterns in the progeny of crosses to PUBc-GAL4 driver lines.

108

2   Materials and Methods

2.1   Plasmid construction

To obtain the PUBc-GAL4 driver plasmid, primers were designed from the annotation in the PEST genome. PUBc5 consists of 2005 bp upstream of the predicted start codon and includes the intron, the 5'UTR and the intergenic space that separates PUBc (AGAP001971) from the previous gene (PUBb, AGAP001970). PUBc3 includes the intergenic space (407 bp) between the stop codon of PUBc and the following gene (TSR4 protein, AGAP001972). Plasmid components were individually amplified and ligated into pSLfa1180fa by Gibson cloning (Gibson et al., 2009, 2010) to create pSL-Gyp:PUBc5:GAL4 and pSL-PUBc3:Gyp:attP. The PUBc3:Gyp:attP cassette was then inserted into the BglII/NotI sites of pSL-Gyp:PUBc5:GAL4 to generate pSL-Gyp:PUBc5:GAL4:PUBc3:Gyp:attP. The Gyp:PUBc5:GAL4:PUBc3:Gyp:attP fragment was finally cloned into the AscI sites of pBAC:attP:eCFP:LRIM to replace LRIM and obtain the final plasmid pBAC:attP:eCFP:Gyp:PUBc5:GAL4:PUBc3:Gyp:attP (pPUBc-GAL4). All primers and templates used are reported in Table S1.

The UAS-mCD8:mCherry responder plasmid was designed to carry an attB site for site-specific integration into an attP docking line, an enhanced yellow fluorescent protein (eYFP) driven by the 3xP3 promoter, and a fluorescent mCD8:mCherry membrane marker under the control of 10x upstream activating sequences (UAS). In a first step, attB-EcoRI from pTA-attB (gift from M. P. Calos, Addgene plasmid #18937 (Groth et al., 2000)) was cloned into pBAC[3xP3-eYFPafm] (obtained from E. Wimmer). The cassette 3xP3-eYFPafm-attB was further amplified and cloned into pGEM-T Easy Vector (Promega) to create a recipient plasmid pGEM-T[3xP3-eYFPafm-attB]. A second recipient plasmid pSLfa-MF3 was created by cloning the cassette NotI-10xUAS-MCS-ftzintron-SV40-BamHI (from pMF3, DGRC) into pSLfa (obtained from E. Wimmer). mCD8:mCherry was amplified from pUAS-mCD8:mCherry (obtained from F. Schnorrer) and further cloned into pSLfa-MF3. The AscI-10xUAS-mCD8:mCherry-ftzintron-SV40-FseI cassette was cloned into pGEM-T[3xP3-eYFPafm-attB] to obtain the final plasmid pGEM-T[3xP3-eYFPafm-10xUAS-mCD8:mCherry-ftzintron-SV40-attB] (pUAS-mCD8:mCherry). All primers and templates used are reported in Table S1.

2.2   An. gambiae mosquito cells transfection and luciferase assay

The hemocyte-like *An. gambiae* cell line Sua5.1 (Müller et al., 1995) was used as described in Lynd and Lycett, 2012 to perform cell transfection and luciferase assay using Effectene® Transfection Reagent (Qiagen) and Dual-Luciferase® Reporter Assay System (Promega) respectively. DNA samples consisted of 100 ng of pPUBc-GAL4 driver plasmid, 100 ng of a 14xUAS responder plasmid driving the expression of firefly luciferase (pUAS-Luc) (Lynd and Lycett, 2012), and 2 ng of Actin Renilla plasmid to normalize for efficiency of transfection. The pPUBc-eGFP plasmid (Lycett, unpublished), which expresses cytoplasmic eGFP under the control of a shorter version of the PUBc promoter sequence used in this study, was used to visually monitor transfection efficiency and measure background activity (blank) in the absence of a GAL4 driver. Relative light units (RLU)

146   measurements were obtained from 6 replicate wells, normalized by Renilla activity (RLU_{Luc} / RLU_{Ren})

147   and adjusted for the activity of the blank ($RLU_{sample} - RLU_{blank}$). One-way ANOVA with Tukey's

148   multiple comparison analysis was performed to determine statistical significant differences ($P < 0.05$).

149   *2.3   Transgenic lines*

150   The PUBc-GAL4 transgenic line was produced using the piggyBac transposon as previously

151   described (Lycett et al., 2012; Lombardo et al., 2009). G3 strain embryos were microinjected with 350

152   ng/µl of GAL4 driver plasmid and 150 ng/µl of transposase helper plasmid pHSP-pBac (Handler and

153   Harrell, 1999). F_0 L1 larvae expressing eCFP transiently in the anal papillae were reared separately

154   and backcrossed in sex-specific cages. F_1 progeny was assessed for eCFP stable inheritance with

155   positive females then backcrossed to wild type and allowed to lay in individual tubes. F_2 isofemale

156   lines were scored for inheritance of the eCFP marker. Lines yielding a percentage of transformants

157   not compatible with a single insertion (*i.e* >50% fluorescent progeny) were discarded, while isofemale

158   lines showing ~50% of transgenic progeny were interbred to create stable lines.

159   The transgenic line UAS-mCD8:mCherry was created by injecting embryos of the E φC31-docking

160   line (Meredith et al., 2011) with 250 ng/µl of UAS-mCD8:mCherry plasmid and 800 ng/µl of mRNA

161   encoding an insect codon optimized mutant φC31 integrase (obtained from A. A. James). Of the F_0

162   adults recovered, pools of 10 females were backcrossed to 50 males, while all males were crossed

163   with 5x females of the E line. Microinjections, screening and stable homozygous line generation were

164   carried out as described by Pondeville et al. (2014).

165   *2.4   Inverse-PCR*

166   Inverse PCR was conducted as described by Lycett et al., 2012. Genomic DNA was extracted from

167   pools of 20 F_2 transgenic adults from selected isofemale lines using Qiagen Genomic tips (Qiagen). 1

168   µg of gDNA was then digested with BfuCI or TaqαI, self-ligated, and PCR was performed to amplify

169   DNA regions flanking the piggyBac arms at the site of insertion (primers ITRL1F and ITRL1R for

170   piggyBac left arm, ITRR1F and ITRR1R for right arm, Table S1). PCR products were sequenced and

171   genomic location of insertions identified using the BLAST tool integrated in VectorBase (Giraldo-

172   Calderón et al., 2015).

173   *2.5   Dissections*

174   Dissections were performed on 2-5-day-old GAL4/UAS adult females in PBS supplemented with

175   EDTA-free protease inhibitor cocktail (Roche). Body parts were incubated in fixing solution (4%

176   paraformaldehyde, 1X PBS pH 7.4, 2 mM $MgSO_4$, 1 mM EGTA) for 30-45 minutes, washed in PBS

177   and mounted on a microscope slide using Vectashield mounting medium (Vector Laboratories). For

178   preparation of abdomens, after removing internal organs, they were incubated in methanol for 2

179   minutes before fixing.

180

3   Results

3.1   *An. gambiae PUBc regulatory regions drive expression in mosquito cells*

To examine the activity of the regulatory regions surrounding PUBc, the entire intergenic sequences upstream and downstream of the coding region were used to flank the GAL4 gene. This transcription unit was placed between gypsy insulators (Cai and Levine, 1995) to potentially reduce position effect variation, and flanked by inverted φC31 attP repeats to generate RMCE docking sites. Finally, the construct was inserted in a piggyBac transformation vector marked with 3xP3-eCFP (Fig. 1B).

Initially, the ability of PUBc to activate GAL4-mediated expression was investigated in *An. gambiae* SUA5.1 cells. When co-transfected with a responder plasmid expressing UAS-regulated luciferase (pUAS-Luc) (Lynd and Lycett 2012), pPUBc-GAL4 showed ~4000-fold higher activity than control transfections with pUAS-Luc only ($P < 0.003$) (Fig. 1D). Additionally, background activity from a PUBc-eGFP construct without the luciferase gene (blank) was not significantly different to the pUAS-Luc control ($P = 0.16$), indicating limited leakiness in UAS-regulated expression (Fig. 1D).

3.2   *Generation and characterization of transgenic lines*

3.2.1   PUBc-GAL4 driver lines

The piggyBac transposon was used to create a series of driver lines carrying the PUBc-GAL4 cassette in single, unique genomic locations. From 180 *An. gambiae* G3 strain embryos injected, 113 (63%) larvae hatched, 77 (68%) of which showed eCFP transient expression in the anal papillae. 97 F_0 adults (54%) were obtained and pooled into 6 sex-specific F_0 founder cages (A-F) (Table 1). Transgenic F_1 progeny was recovered from each F_0 cage established from founders showing transient expression (A-D), while no transformants were recovered from negative founders (cages E,F) (Table 1). 21 F_1 transgenic females were backcrossed with wild-type and F_2 progeny from individual females were assessed for inheritance of the fluorescent marker as a proxy for transgene copy number. We identified 14 isofemale lines showing 47-53% transgene inheritance (A2, A3, A6, A7, A8, A10, A11, A12, A15, B3, B8, C1, C2, D2), suggestive of single insertions (Table 1). Upon verification of copy number and location of integration sites by inverse-PCR we found several instances of identical integration sites in distinct isofemale lines from the same founder cage (Fig. S1). Replicate lines were discarded thus overall the final analysis was consistent with the isolation of 7 isofemale lines carrying separate single integration sites (Table 2). Four lines (A8=A15, A10=A12, B3, C1), carried intergenic insertions, one line (A3=A6=A7=A11) carried an insertion within the open reading frame of *Cyp6m4* (AGAP008214), and two (A2 and D2) carried insertions within highly repetitive regions that precluded their exact localization in the genome (Table 2). iPCR failed to amplify the region flanking the piggyBac right arm of lines A8 and A10 and location in these lines was confirmed using primers designed on the genomic DNA flanking the insertion site predicted from the sequencing data obtained from the left arm (Fig. S2). All sites of integration characterized occurred at

217   piggyBac canonical TTAA integration sites (Table 2). Overall, we achieved a minimum transformation
218   efficiency of 7% (*i.e.* number of independent insertions / total F_0 survivors). All 7 lines were interbred
219   and underwent preliminary assessment of *in vivo* promoter activity in larvae.

220   3.2.2   UAS-mCD8:mCherry line

221   The responder plasmid was designed to carry an attB site for φC31-mediated site-specific integration
222   into the E docking line (Meredith et al., 2011), an enhanced yellow fluorescent protein (eYFP) driven
223   by the 3xP3 promoter, and a fluorescent mCD8:mCherry membrane marker under the control of 10x
224   UAS (Fig. 1C). From the 1182 E line embryos injected, 374 F_0 adults (32%) were recovered, 185
225   females and 189 males, which were backcrossed to individuals of the E line. A total of 23 transgenic
226   F_1 larvae were recovered from the F_0 males only and from which the line was established. Crosses of
227   the UAS-mCD8:mCherry responder line with a carboxypeptidase-GAL4 *An. gambiae* line (Lynd and
228   Lycett 2012) showed the expected adult midgut specific expression of red fluorescence (Fig. S3) in
229   the progeny.

230   *3.3   PUBc drives widespread expression in transgenic An. gambiae mosquitoes*

231   To investigate the pattern of expression driven by PUBc and identify unique insertion sites able to
232   sustain ubiquitous expression, profiles generated by transgenic PUBc-GAL4 isofemales lines (A2, A3,
233   A8, A10, B3, C1 and D2) and mixed male pools were assessed and compared after crossing with the
234   responder line UAS-mCD8:mCherry. Preliminary analysis of the larval expression profiles driven by
235   individual PUBc-GAL4 lines differed in terms of signal intensity and distribution: A10 and D2 derived
236   progeny had similar intense and widespread mCherry signal throughout the whole body; A8 produced
237   symmetrically patterned fluorescent signal along the abdomen that was noticeably weaker than A10
238   and D2 but with intense reporter expression in the head and mouthparts; A3 and C1 progeny
239   displayed asymmetrical mCherry signal along the abdomen, in the head and mouthparts, whereas in
240   A2 and B3 derived progeny mCherry signal was largely limited to the terminal part of the abdomen,
241   and intense signal in the brain and mouthparts respectively. Unfortunately, these profiles were not
242   documented as images (data not shown). We were unable to detect distinct phenotypes from the
243   mixed male pool progeny (Fig. S4) which were not also observed from the single isofemales lines,
244   and therefore our further analysis focused on female derived lines.

245   A8 and A10 were chosen for further in depth analysis as they displayed a similar temporal and spatial
246   distribution, yet displayed distinct intensities of reporter gene expression, with A10 being generally
247   higher in most tissues. In larval progeny from both drivers, mCherry fluorescence was evident in the
248   mouthparts, brain, fat body, muscles surrounding the aorta and the heart (Fig. 2A,B), developing
249   thoracic imaginal discs, and central nerve cord (Fig. 2C,D). Signal in A10/UASmCherry (A10/ch)
250   larvae (Fig. 2A) was more intense compared to A8/ch larvae (Fig. 2B), with the exception of the
251   mouthparts. Furthermore, while A10/ch individuals displayed a very bright and widespread signal
252   throughout the larval body (Fig. 2A,C), in A8/ch larvae a distinct symmetrical expression pattern was
253   detected dorsally and ventrally in each abdominal segment (Fig. 2B,D), most likely derived from the

integument. In pupae, the differences in signal intensity and the distinct abdominal pattern of expression were maintained, and strong fluorescence was detected in the developing antennae in both A10/ch (Fig. 2E) and A18/ch pupae (Fig. 2F). At adult stage, mCherry was widespread in intact females (Fig. 2 G,H) and males (Fig. 2 I,J) of both A10 and A8 crosses with A10/ch displaying a brighter signal than A8/ch mosquitoes. Fluorescence was widely detected in the thorax and along the abdomen with bright signal visible through the non-sclerotized areas of the cuticle. In both crosses, fluorescence was detected in all of the appendages including legs, veins of the wings, antennae, palps, and proboscis/labium (Fig. 2G to J). Expression was also robust at the base of the antennae in the pedicel, which hosts the Johnston's organ, and in the brain (Fig. 2G to J). A10/ch adults showed expression in some but not all of the ommatidia, while in A8/ch individuals this was less evident and occurred in fewer ommatidia (Fig. 2G to J).

In dissected abdomens of both crosses expression was detectable in all of the tissues present in the integument including the epidermis, fat body, oenocytes, lateral muscles and nerve cord (Fig. 3A,C). Here, besides displaying a lower signal intensity, A8/ch abdomens showed a mosaic pattern of expression in muscular tissues (Fig. 3C). A comparable level of fluorescence was detected in A10/ch (Fig. 3E) and A8/ch (Fig. 3G) mosquitoes along the majority of trachea surrounding the internal organs including the foregut, midgut, hindgut and Malpighian tubules of unfed females; however, no expression was detectable in the midgut or Malpighian tubule epithelium. PUBc was active in the salivary glands of A10/ch (Fig. 3I) with mCherry fluorescence detected in all lobes, while signal was weaker in A8/ch salivary glands (Fig. 3K). Ovaries (Fig. 3 M,O) of sugar-fed A10/ch females showed bright oviducts and follicles, and fluorescence was also strongly detected 24-48 h after blood meal in the developing oocytes; while fluorescence in A8/ch mosquitoes was weaker (not shown). In male reproductive system, signal was moderately detected in the testes (not shown). Finally, hemocytes attached to tissues and circulating in the hemolymph had detectable mCherry expression (not shown).

4 Discussion

To expand the available tools for functional genetic analysis of *An. gambiae*, we have isolated and characterized the regulatory regions of the *An. gambiae Polyubiquitin-c* gene and validated their ability to drive robust expression in cultured mosquito cells and widespread multi-tissue expression in all life stages of transgenic mosquitoes. This activity was tested in the context of the GAL4/UAS system to allow flexible use of the generated lines in future analysis. As part of the validation, we also established and tested a new UAS reporter construct that yields bright mCherry expression with all of the GAL4 lines so far examined.

The generation of piggyBac GAL4 driver lines was achieved with an efficiency equal to or higher than reported elsewhere using a variety of promoter/effector combinations (Grossman et al., 2001; Kim et al., 2004; Lobo et al., 2006; Lombardo et al., 2009; Lycett et al., 2012; Meredith et al., 2013). This would suggest that any potential toxicity of widespread GAL4 expression does not manifest as

291 reduced transformation efficiency. Furthermore, as reported elsewhere (Pondeville et al., 2014),
292 transgenic progeny was only obtained from F_0 mosquito pools that had showed transient episomal
293 expression of the transformation marker in larval stages, not only suggesting that effective screening
294 in future could be confined to progeny from these pools, but that concurrent episomal expression of
295 GAL4 in the F_0 is readily tolerated. Homozygous A8 and A10 are viable during inbreeding of the stock
296 colonies, and we have created a true breeding homozygous A10 line that continues to be stable over
297 multiple generations. However, routine observation suggests reduced longevity in relation to the
298 parental G3 line, and although we haven't performed detailed life table analysis, there may be a
299 fitness cost associated with high levels of GAL4 expression achieved in the homozygous A10 line.
300 This will be explored further when the Gal4 driver is removed from these lines by RMCE.

301 PiggyBac-mediated integration occurs essentially randomly at TTAA genomic sites and can lead to
302 insertion of the transgenic construct in multiple sites of an individual genome (Lombardo et al., 2009).
303 The selection strategy used here aimed at maximizing the identification of lines carrying insertions at
304 single intergenic genomic sites, desirable for docking lines, as well as identifying insertion sites that
305 gave the broadest tissue expression of GAL4, whilst limiting workload. To do so, we performed
306 expression profiling and molecular characterization of insertion site on isofemale lines derived from F_1
307 females, while F_1 males were used as a backup, crossed in pools to UAS-mCD8:mCherry and
308 screened *en masse* to recover any distinct expression profiles (and thus different insertion sites) not
309 observed in female lines. However, no distinct phenotypes were recovered in the F_2 male progeny,
310 suggesting they originated from the same F_0 individuals as the females.

311 Although differences in the expression of the 3xP3-eCFP marker from the piggyBac transposon were
312 detected in different F_1 isofemale lines, these were minor and did not readily correlate with the
313 extensive variegation observed in the GAL4 driven mCherry signal, suggesting that not all genomic
314 sites supportive of robust neuronal expression from 3xP3 are capable of sustaining widespread
315 expression driven by other regulatory regions. This is supported by findings in *Ae. aegypti*, where
316 3xP3 was an unreliable promoter for inferring efficiency of gut specific transgene expression at
317 distinct genomic locations (Franz et al., 2011).

318 Position variegation derived from our PUB-GAL4 lines also suggests that the inclusion of gypsy
319 insulators flanking the GAL4 cassette did not overcome the influence of nearby effectors and/or
320 chromatin status on transgene expression. Gypsy insulators (Roseman et al., 1993) have been
321 previously reported to repress the action of nearby suppressors in *D. melanogaster* and *An. stephensi*
322 (Sarkar et al., 2006; Carballar-Lejarazú et al., 2013). Yet, no quantitative study has been conducted in
323 *An. gambiae* when the GAL4/UAS system is used. When insulating UAS cassettes in *D.
324 melanogaster*, gypsy insulators have proven useful to mitigate positional effect, but efficiency of
325 insulation was still dependent on genomic locus (Markstein et al., 2008). In the first analysis of
326 GAL4/UAS in *An. gambiae*, gypsy sequences were used to flank the UAS-reporter cassette, and all
327 combinations of carboxypeptidase-GAL4 driver lines crossed to 6 UAS reporter lines drove robust
328 midgut-specific expression regardless of location of insertion (Lynd and Lycett 2012). There may be,
329 thus, a difference in the insulation capacity of gypsy sequences when flanking UAS compared to

330   GAL4 cassettes or depending on the regulatory regions used. It should be noted though that RMCE

331   will remove the gypsy sequences from these loci, and so depending on the nature of the transgenics

332   being produced it may be necessary to include gypsy sequences in the exchanged cassette to

333   recreate a similar sequence context.

334   Overall, visual assessment of progeny derived from selected PUBc driver lines (A10 and A8) crosses

335   with UAS-mCD8:mCherry individuals, revealed that the regulatory regions identified for PUBc are

336   active in a variety of tissues and organs in all mosquito life stages examined and presumably in

337   embryonic stages too, as neonate larvae are fluorescent. In adults, these tissues include: eyes, brain,

338   and ventral nerve cord; muscles of the aorta and heart; trachea surrounding the digestive tract and

339   the Malpighian tubules; salivary glands; sugar-fed and blood-fed ovaries and developing oocytes; fat

340   body; appendages (legs, palps, antennae, proboscis); and hemocytes. Tissues where expression was

341   not readily visible include the epithelium of the digestive tract and Malpighian tubules, although it

342   cannot be excluded that expression occurs in these tissues at a level that is not detectable by

343   fluorescence. However, it should be noted that robust midgut epithelium expression of mCherry was

344   observed in crosses with carboxypeptidase-GAL4 lines.

345   The *An. gambiae* PUBc-derived expression pattern largely overlaps those described for promoter

346   fragments derived from other insect PUB genes (Handler and Harrell, 2001; Heinrich et al., 2002;

347   Perera et al., 2002) and, more specifically, with that of the *Ae. aegypti* PUB promoter described by

348   Anderson et al. (2010) where expression was robust and widespread throughout the mosquito body

349   from larvae to adults. Nevertheless, we found that the A10 line drives strong fluorescence in the

350   salivary glands, while in *Ae. aegypti* these tissues displayed little fluorescence and lacked mRNA

351   signal. Conversely, strong fluorescence was seen in *Ae. aegypti* midguts, whilst it was observed only

352   in the trachea surrounding the digestive tract but not the midgut epithelium in *An. gambiae*. These

353   differences may be due to innate variation in promoter activity from the region selected in the two

354   species or they may be the result of positional effect. Nonetheless, the widespread core expression

355   pattern described here for PUBc is unprecedented for an *An. gambiae* endogenous promoter.

356   Previous work analyzing a potential endogenous 'ubiquitous' promoter directly with an eGFP reporter

357   gene, had indicated that positional effects can be extremely frequent in *An. gambiae* transgenics

358   (Lycett et al., 2012). Each line produced had a core expression profile in the same limited tissues,

359   including nerve cord and chordotonal organs, but different patterns of expression in other tissues

360   were observed in each of the 9 isofemale lines analyzed. In the current study, we reasoned that by

361   using the GAL4/UAS system to generate and follow the expression pattern conferred by putative

362   regulatory regions, then selected GAL4 lines that gave a desired expression pattern could be archived

363   for future functional genetic studies. The two lines thus retained (A10 and A8) will be used to compare

364   different UAS-regulated transgenes and monitor resultant phenotypes based on a consistent, defined

365   and robust expression pattern. If the UAS transgenes are created through φC31 site-directed insertion

366   into the same genomic location (Pondeville et al., 2014), then such work can be performed in similar

367   genetic backgrounds to allow accurate phenotypic comparison.

368 Since the number of docking strains for φC31 integrase in *An. gambiae* is limited, we also included

369 attP sites into the PUB-GAL4 transformation construct so that these lines can also be used as

370 docking lines for RMCE or integration. Indeed, in the presence of two inverted attP sites, cassette

371 integration also occurs in some individuals as a result of a single crossover event, rather than the

372 double crossover that generates cassette exchange. As such, F_1 progeny can be selected by

373 screening for the presence of single or double markers respectively. RMCE allows complete

374 swapping of transgene cargoes for phenotypic analysis (Bateman et al., 2006). For example, the

375 activity of other promoters could be assessed and directly compared to that of PUBc; similarly,

376 different UAS cassettes could be exchanged into these sites, and then crossed to parental or other

377 GAL4 driver lines. In other instances, integration may be desirable, such as inserting a UAS construct

378 next to the PUBc-GAL4 in order to create stable GAL4/UAS lines without the need of crossing, and

379 thus to readily generate homozygous GAL4/UAS lines. Although the φC31 system may be the most

380 convenient way to exploit these genomic sites, they can also be targeted by CRISPR/Cas9

381 approaches in which transgenes can be targeted to defined transcriptionally permissive sites.

382 The A10 and A8 driver lines established as part of this work sustain widespread gene expression at

383 different levels and will prove valuable and versatile tools for gene functional analysis in *An. gambiae*

384 using the GAL4/UAS system. Their potential uses include gain- or loss-of-function of any gene

385 located in corresponding UAS lines. Owing to its multi-tissue expression, the PUBc promoter could

386 provide more reliable expression of transformation markers than 3xP3 and would be useful for

387 efficient mass screening of large numbers of transgenic individuals at lower magnification or by

388 automated screening systems (Marois et al., 2012; Volohonsky et al., 2015). It could also be used to

389 create a transgenic line that stably produces broad expression of Cas9, which can be transiently

390 modified by injection with guide RNAs or crossed with transgenic lines expressing guide RNAs.

391

392 Authors contributions

393 Conceived the work: AA, EP, CB, GJL

394 Participated in design of experiments AA, EP, AL, CB, GJL

395 Performed the experiments: AA, EP

396 Analyzed the data: AA, EP, CB, GJL

397 Drafted the manuscript: AA

398 Edited and reviewed manuscript: AA, EP, CB, AL, GJL

399

400 Funding

12

Conflicts of interest

Authors declare no conflict of interest.

Acknowledgments

References

Abraham, E.G., Donnelly-Doman, M., Fujioka, H., Ghosh, A., Moreira, L., Jacobs-Lorena, M., 2005. Driving midgut-specific expression and secretion of a foreign protein in transgenic mosquitoes with AgAper1 regulatory elements. Insect Mol. Biol. 14, 271–279.

Anderson, M.A.E., Gross, T.L., Myles, K.M., Adelman, Z.N., 2010. Validation of novel promoter sequences derived from two endogenous ubiquitin genes in transgenic *Aedes aegypti*. Insect Mol. Biol. 19, 441–449.

Bateman, J.R., Lee, A.M., Wu, C.T., 2006. Site-specific transformation of *Drosophila* via PhiC31 integrase-mediated cassette exchange. Genetics. 173, 769–777.

Bhatt, S., Weiss, D.J., Cameron, E., Bisanzio, D., Mappin, B., Dalrymple, U., Battle, K.E., Moyes, C.L., Henry, A., Penny, M.A., Smith, T.A., Bennett, A., Yukich, J., Eisele, T.P., Eckhoff, P.A., Wenger, E.A., Brie, O., Griffin, J.T., Fergus, C.A., Lynch, M., Lindgren, F., Cohen, J.M., Murray, C.L.J., Smith, D.L., Hay, S.I., Cibulskis, R.E. & Gething, P.W., 2015. The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. Nature. 526, 207–211.

Brand, A.H., Perrimon, N., 1993. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development (Cambridge, England). 118, 401–15.

Cai, H., Levine, M., 1995. Modulation of enhancer-promoter interactions by insulators in the *Drosophila* embryo. Nature. 376, 533–536.

434 Carballar-Lejarazú, R., Jasinskiene, N., James, A.A, 2013. Exogenous gypsy insulator sequences
435   modulate transgene expression in the malaria vector mosquito, *Anopheles stephensi*. Proc. Natl.
436   Acad. Sci. U. S. A. 110, 7176–7181.

437 Catteruccia, F., Benton, J.P., Crisanti, A., 2005. An *Anopheles* transgenic sexing strain for vector
438   control. Nature Biotechnol. 23, 1414–1417.

439 Franz, A.W.E., Jasinskiene, N., Sanchez-Vargas, I., Isaacs, A.T., Smith, M.R., Khoo, C.C.H.,
440   Heersink, M.S., James, A.A., Olson, K.E., 2011. Comparison of transgene expression in *Aedes*
441   *aegypti* generated by mariner Mos1 transposition and Phic31 site-directed recombination. Insect
442   Mol. Biol. 20, 587–598.

443 Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., Smith, H.O., 2009. Enzymatic
444   assembly of DNA molecules up to several hundred kilobases. Nature Methods. 6, 343–345.

445 Gibson, D.G., Smith, H.O., Hutchison 3rd, C.A., Venter, J.C., Merryman, C, 2010. Chemical synthesis
446   of the mouse mitochondrial genome. Nature methods. 7, 901–903.

447 Giraldo-Calderón, G.I., Emrich, S.J., MacCallum, R.M., Maslen, G., Emrich, S., Collins, F., Dialynas,
448   E., Topalis, P., Ho, N., Gesing, S., Madey, G., Collins, F.H., Lawson, D., Kersey, P., Allen, J.,
449   Christensen, M., Hughes, D., Koscielny, G., Langridge, N., Gallego, E.L., Megy, K., Wilson, D.,
450   Gelbart, B., Emmert, D., Russo, S., Zhou, P., Christophides, G., Brockman, A., Kirmitzoglou, I.,
451   MacCallum, B., Tiirikka, T., Louis, K., Dritsou, V., Mitraka, E., Werner-Washburn, M., Baker, P.,
452   Platero, H., Aguilar, A., Bogol, S., Campbell, D., Carmichael, R., Cieslak, D., Davis, G.,
453   Konopinski, N., Nabrzyski, J., Reinking, C., Sheehan, A., Szakonyi, S., Wieck, R., 2015.
454   VectorBase: An updated Bioinformatics Resource for invertebrate vectors and other organisms
455   related with human diseases. Nucleic Acids Res. 43, 707–713.

456 Grossman, G.L., Rafferty, C.S., Clayton, J.R., Stevens, T.K., Mukabayire, O., Benedict, M.Q., 2001.
457   Germline transformation of the malaria vector, *Anopheles gambiae*, with the piggyBac
458   transposable element. Insect Mol. Biol. 10, 597–604.

459 Groth, A.C., Fish, M., Nusse, R., Calos, M.P., 2004. Construction of Transgenic *Drosophila* by Using
460   the Site-Specific Integrase from Phage PhiC31. Genetics. 166, 1775–1782.

461 Handler, A.M., Harrell, R.A., 1999. Germline transformation of *Drosophila melanogaster* with the
462   piggyBac transposon vector. Insect Mol. Biol. 8, 449–457.

463 Handler, A.M., Harrell, R.A., 2001. Transformation of the Caribbean fruit fly, *Anastrepha suspensa*,
464   with a piggyBac vector marked with polyubiquitin-regulated GFP. Insect Biochem. Mol. Biol. 31,
465   199–205.

466 Handler, A.M., 2002. Use of the piggyBac transposon for germ-line transformation of insects. Insect
467   Biochem. Mol. Biol. 32, 1211-20.

Hammond, A., Galizi, R., Kyrou, K., Simoni, A., Siniscalchi, C., Katsanos, D., Gribble, M., Baker, D., Marois, E., Russell, S., Burt, A., Windbichler, N., Crisanti, A., Nolan, T., 2016. A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector *Anopheles gambiae*. Nature Biotechnol. 34, 78–83.

Heinrich, J.C., Li, X., Henry, R.A., Haack, N., Stringfellow, L., Heath, A.C.G., Scott, M.J., 2002. Germ-line transformation of the Australian sheep blowfly *Lucilia cuprina*. Insect Mol. Biol. 11, 1–10.

Imamura, M., Nakai, J., Inoue, S., Quan, G.X., Kanda, T., Tamura, T., 2003. Targeted gene expression using the GAL4/UAS system in the silkworm *Bombyx mori*. Genetics. 165, 1329–1340.

Kim, W., Koo, H., Richman, A.M., Seeley, D., Vizioli, J., Klocko, A.D., O'Brochta, D.A., 2004. Ectopic expression of a Cecropin transgene in the human malaria vector *Anopheles gambiae* (Diptera: Culicidae): Effects on susceptibility to *Plasmodium*. J. Med. Entomol. 41, 447–455.

Kokoza, V.A., Raikhel, A.S., 2011. Targeted gene expression in the transgenic *Aedes aegypti* using the binary Gal4-UAS system. Insect Biochem. Mol. Biol. 41, 637–644.

Lobo, N.F., Clayton, J.R., Fraser, M.J., Kafatos, F.C., Collins, F.H., 2006. High efficiency germ-line transformation of mosquitoes. Nat. Protoc. 1, 1312–1317.

Lombardo, F., Di Cristina, M., Spanos, L., Louis, C., Coluzzi, M., Arcà, B., 2000. Promoter sequences of the putative *Anopheles gambiae* apyrase confer salivary gland expression in *Drosophila melanogaster*. J. Biol. Chem. 275, 23861–23868.

Lombardo, F., Nolan, T., Lycett, G., Lanfrancotti, A., Stich, N., Catteruccia, F., Louis, C., Coluzzi, M., Arcà, B., 2005. An *Anopheles gambiae* salivary gland promoter analysis in *Drosophila melanogaster* and *Anopheles stephensi*. Insect Mol. Biol. 14, 207–216.

Lombardo, F., Lycett, G.J., Lanfrancotti, A., Coluzzi, M., Arcà, B., 2009. Analysis of apyrase 5' upstream region validates improved *Anopheles gambiae* transformation technique. BMC Res. Notes. 2, 24.

Lorenzen, M.D., Brown, S.J., Denell, R.E., Beeman, R.W., 2002. Transgene expression from the *Tribolium castaneum* Polyubiquitin promoter. Insect Mol. Biol. 11, 399–407.

Lycett, G.J., Amenya, D., Lynd, A., 2012. The *Anopheles gambiae* alpha-tubulin-1b promoter directs neuronal, testes and developing imaginal tissue specific expression and is a sensitive enhancer detector. Insect Mol. Biol. 21, 79–88.

Lynd, A., Lycett, G.J., 2011. Optimization of the Gal4-UAS system in an *Anopheles gambiae* cell line. Insect Mol. Biol. 20, 599–608.

Lynd, A., Lycett, G.J., 2012. Development of the bi-partite Gal4-UAS system in the African malaria mosquito, *Anopheles gambiae*. PLoS ONE. 7, e31552.

502 Markstein, M., Pitsouli, C., Villalta, C., Celniker, S.E., Perrimon, N., 2008. Exploiting position effects
503   and the gypsy retrovirus insulator to engineer precisely expressed transgenes. Nature Genet. 40,
504   476–483.

505 Marois, E., Scali, C., Soichot, J., Kappler, C., Levashina, E.A., Catteruccia, F., 2012. High-throughput
506   sorting of mosquito larvae for laboratory studies and for future vector control interventions. Malar.
507   J. 11, 302.

508 Meredith, J.M., Basu, S., Nimmo, D.D., Larget-Thiery, I., Warr, E.L., Underhill, A., McArthur, C.C.,
509   Carter, V., Hurd, H., Bourgouin, C., Eggleston, P., 2011. Site-specific integration and expression
510   of an anti-malarial gene in transgenic *Anopheles gambiae* significantly reduces *Plasmodium*
511   *infections*. PLoS ONE. 6, 1–9.

512 Meredith, J.M., Underhill, A., McArthur, C.C., Eggleston, P., 2013. Next-Generation Site-Directed
513   Transgenesis in the Malaria Vector Mosquito *Anopheles gambiae*: Self-Docking Strains
514   Expressing Germline-Specific phiC31 Integrase. PLoS ONE. 8, e59264.

515 Moreira, L.A., Iturbe-Ormaetxe, I., Jeffery, J.A., Lu, G., Pyke, A.T., Hedges, L.M., Rocha, B.C., Hall-
516   Mendelin, S., Day, A., Riegler, M., Hugo, L.E., Johnson, K.N., Kay, B.H., McGraw, E.A., van den
517   Hurk, A.F., Ryan, P.A., O'Neill, S.L., 2009. A Wolbachia Symbiont in *Aedes aegypti* Limits
518   Infection with Dengue, Chikungunya, and *Plasmodium*. Cell. 139, 1268–1278.

519 Müller, H.M., Catteruccia, F., Vizioli, J., della Torre, A., Crisanti, A., 1995. Constitutive and blood
520   meal-induced trypsin genes in *Anopheles gambiae*. Exp. Parasitol. 81, 371–85.

521 Nolan, T., Petris, E., Müller, H.M., Cronin, A., Catteruccia, F., Crisanti, A., 2011. Analysis of two novel
522   midgut-specific promoters driving transgene expression in *Anopheles stephensi* mosquitoes.
523   PLoS ONE. 6, e16471.

524 O'Brochta, D.A., Pilitt, K.L., Harrell 2nd, R.A., Aluvihare, C., Alford, R.T., 2012. Gal4-based enhancer-
525   trapping in the malaria mosquito *Anopheles stephensi*. G3 (Bethesda). 2, 1305–1315.

526 Papathanos, P.A., Windbichler, N., Menichelli, M., Burt, A., Crisanti, A., 2009. The vasa regulatory
527   region mediates germline expression and maternal transmission of proteins in the malaria
528   mosquito *Anopheles gambiae*: a versatile tool for genetic control strategies. BMC Mol. Biol. 10,
529   65.

530 Perera, O.P., Harrell, R.A., Handler, A.M., 2002. Germ-line transformation of the South American
531   malaria vector, *Anopheles albimanus*, with a piggyBac/EGFP transposon vector is routine and
532   highly efficient. Insect Mol. Biol. 11, 291–297.

533 Piccin, A., Salameh, A., Benna, C., Sandrelli, F., Mazzotta, G., Zordan, M., Rosato, E., Kyriacou, C.P.,
534   Costa, R., 2001. Efficient and heritable functional knock-out of an adult phenotype in *Drosophila*
535   using a GAL4-driven hairpin RNA incorporating a heterologous spacer. Nucleic Acids Res. 29,
536   E55–E55.

537   Pondeville, E., Puchot, N., Meredith, J.M., Lynd, A., Vernick, K.D., Lycett, G.J., Eggleston, P.,
538   Bourgouin, C., 2014. Efficient ΦC31 integrase–mediated site-specific germline transformation of
539   *Anopheles gambiae*. Nat. Protoc. 9, 1698–1712.

540   Riabinina, O., Task, D., Marr, E., Lin, C.-C., Alford, R., O'Brochta, D.A., Potter, C.J., 2016.
541   Organization of olfactory centres in the malaria mosquito *Anopheles gambiae*. Nat. Commun. 7,
542   13010.

543   Roseman, R.R., Pirrotta, V., Geyer, P.K., 1993. The su(Hw) protein insulates expression of the
544   *Drosophila* melanogaster white gene from chromosomal position-effects. The EMBO journal. 12,
545   435–42.

546   Sarkar, A., Atapattu, A., Belikoff, E.J., Heinrich, J.C., Li, X., Horn, C., Wimmer, E. a, Scott, M.J., 2006.
547   Insulated piggyBac vectors for insect transgenesis. BMC Biotechnol. 6, 27.

548   Schinko, J.B., Weber, M., Viktorinova, I., Kiupakis, A., Averof, M., Klingler, M., Wimmer, E.A., Bucher,
549   G., 2010. Functionality of the GAL4/UAS system in *Tribolium* requires the use of endogenous
550   core promoters. BMC Dev. Biol. 10, 53.

551   Volohonsky, G., Terenzi, O., Soichot, J., Naujoks, D.A., Nolan, T., Windbichler, N., Kapps, D.,
552   Smidler, A.L., Vittu, A., Costa, G., Steinert, S., Levashina, E.A., Blandin, S.A., Marois, E., 2015.
553   Tools for *Anopheles gambiae* Transgenesis. G3 (Bethesda). 5, 1151–1163.

554   Volohonsky, G., Hopp, A.K., Saenger, M., Soichot, J., Scholze, H., Boch, J., Blandin, S.A. & Marois,
555   E., 2017. Transgenic Expression of the Anti-parasitic Factor TEP1 in the Malaria Mosquito
556   *Anopheles gambiae*. PLoS Pathog. 13, 1–26.

557   WHO (2016). World Malaria Report. Geneva, World Health Organization.

558   WHO (2017). World Malaria Report. Geneva, World Health Organization.

559

FIGURE CAPTIONS
561
562
563
564   **Figure 1**. **Constructs used for transgenic line creation and analysis of cell line activity of**

565   **PUBc regulatory regions.**

566   **A)** Schematic of the GAL4/UAS bi-partite system for ectopic gene expression (adapted from Lynd

567   and Lycett, 2011). The driver construct consists of a promoter that induces expression of the

568   transcription activator factor GAL4. The responder construct contains the GAL4 binding site UAS

569   (Upstream Activating Sequence) located upstream of the gene of interest. **B)** Schematic of the

570   PUBc-GAL4 construct inserted in the driver lines. Grey arrows: piggyBac inverted terminal repeats. P

571   orange arrows: φC31 attP sites for RMCE; 3xP3: neuronal promoter. eCFP: enhanced cyan

572   fluorescent protein. Black arrows: Gypsy insulators. PUBc5': PUBc gene 5' region cloned upstream

573   of the GAL4. PUBc3': PUBc gene 3' region cloned downstream of the GAL4. **C)** Schematic of the

574   responder plasmid driving the 14xUAS-regulated expression of luciferase (Lynd and Lycett, 2012)

575   used to co-transfect mosquito cells. UAS(14x): 14 repeats of the Upstream Activating Sequence.

576   Luc: firefly luciferase gene. **D)** Schematic of the plasmid used to measure background activity

577   (blank). The PUB5' region present in this plasmid is a shorter version of the PUBc promoter

578   sequence incorporated in the driver construct fused to eGFP (enhanced green fluorescent protein)

579   gene and SV40 terminator sequences derived from PTubGFP (Lycett et al, 2012) **E)** Schematic of

580   the responder construct present in the UAS-mCD8:mCherry line obtained by φC31-mediated site-

581   specific integration into the E docking line. Orange R and L boxes: attR and attL sites resulting from

582   the recombination of attB and attP site. UAS(10x): 10 UAS repeats. mCD8:mCherry: mCherry

583   fluorescent marker directed to the cell membrane. **F)** Activity of PUBc regulatory regions in *An.*

584   *gambiae* SUA5.1 cells. Promoter activity is shown as luciferase expression (Relative Lights Units –

585   RLU) after co-transfection of the driver plasmid pPUBc-GAL4 with the pUAS-Luc responder plasmid.

586   Top bars represent standard error from mean. Significant differences (****: $P < 0.0001$, ns: $P > 0.05$)

587   were calculated using one-way ANOVA with Tukey's multiple comparisons test on six replicates.

588

589   **Figure 2. Expression profiles driven by the PUBc regulatory regions in A10/mCherry (left) and**

590   **A8/mCherry (right) whole mosquitoes.**

591   **A-B)** L3-4 larvae dorsal view. m: mouthparts; b: brain; ms: muscles; fb: fat body. **C-D)** L3-4 larvae

592   ventral view. nc: nerve cord; id: imaginal discs. **E-F)** Female pupae. an: antenna. **G-H)** Adult females

593   ventral view. p: palps; an: antenna; om: ommatidia; l: leg; t: thorax; ab: abdomen. **I-J)** Adult males

594   ventral view. mCherry signal was detected at 650 nm. Numbers show seconds (s) of exposure.

595
596
597   **Figure 3. Expression profiles driven by the PUBc regulatory regions in dissected and fixed**

598   **female tissues from A10/mCherry and A8/mCherry mosquitoes.**

599   PUBc-driven expression in A10/mCherry (A,E,I,M) and A8/mCherry (C,G,K,O) female adult abdomen

600   (A-D), digestive tract (E-H), salivary glands (I-L), ovaries (M-P). A,C,E,G,I,K,M,O represent mCherry

601   signal at 650 nm; numbers show milliseconds (ms) or seconds (s) of exposure. B,D,F,H,J,L,N,P are

602 the corresponding bright fields. fb: fat body; ms: muscles; nc: nerve cord; tr: trachea; f: follicles; ov:

603 oviduct.

604

605 **Table 1.** Selection strategy to isolate stable transgenic driver lines carrying a single

606 insertion of the PUBc-GAL4 cassette.

607

F_0 pool	F_0 adults Number and sex	Transient eCFP expression	Number of F_1 eCFP + Adults (sex)	% eCFP + F_2 larvae (Number eCFP +/total)
A	13 F	+	34 (17 F, 17 M)	A 2 – 49% (84/173) A 3 – 48% (74/154) A 5 – 62% (5/8) A 6 – 42% (80/190) A 7 – 50% (52/103) A 8 – 49% (67/136) A 10 – 47% (65/139) A 11 – 52% (63/121) A 12 – 53% (102/193) A 15 – 51% (93/181)
B	14 M	+	3 (3 M)	B 1 – 73% (8/11) B 2 – 58% (18/31) B 3 – 44% (31/71) B 7 – 59% (69/116) B 8 – 53% (62/116)
C	18 F	+	5 (2 F, 3 M)	C 1 – 47% (28/59) C 2 – 47% (73/155)
D	20 M	+	3 (2 F, 1 M)	D 2 – 46% (53/115)
E	23 M	–	0	N/A
F	9 F	–	0	N/A

608
609 F_0 pool: designation given to cage of surviving F_0 adults; F_0 adults: the number and sex of F_0 adults
610 crossed with wild type; Transient eCFP expression: F_0 adults either displayed (+) or lacked (-) transient
611 eCFP expression at larval stage; Number of F_1 eCFP + Adults: total number and sex of adults derived
612 from eCFP + larvae from F_0 cage cross; % eCFP + F_2 larvae: % of eCFP + larvae derived from
613 isofemale lines originated from F_1 adults.
614 F: female, M: male. N/A: not applicable.
615 Other isofemales: A1: dead (mosquito died without laying eggs). A4, B9: not mated (eggs did not hatch).
616 A9, A13, A14, B4, B5, B6, B10, D1: sterile (mosquito did not lay eggs).
617
618
619
620

Table 2. Molecular characterisation of transgenic lines carrying single insertions by inverse-PCR.

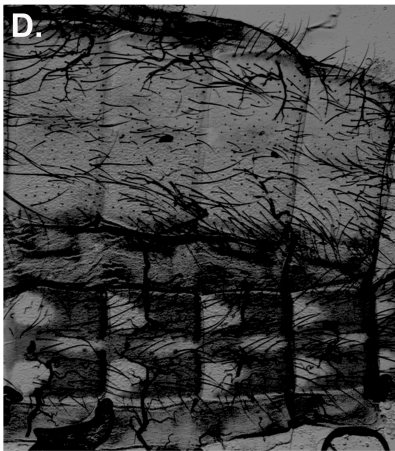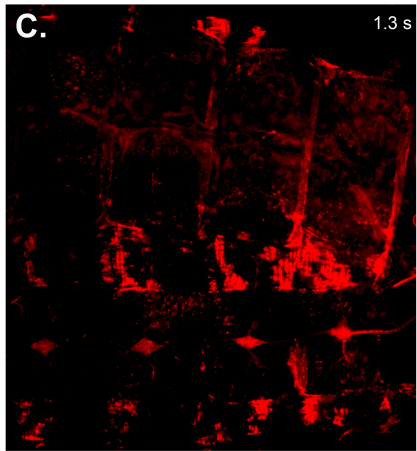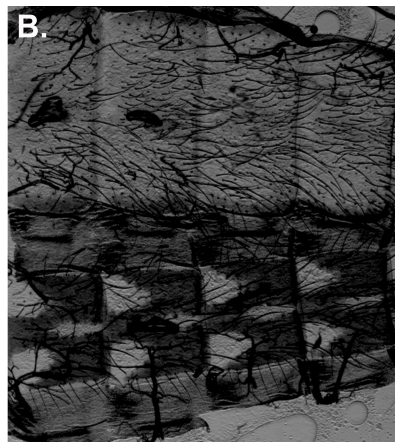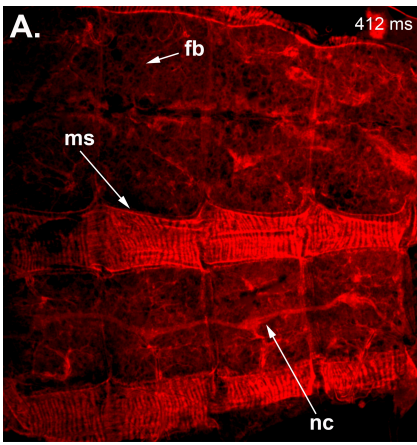| Line	pBAC arm	Enzyme	Sequence	GenBank ID	Query length	Alignment length	Identity %	PEST Chromosome	Position	
A2	Left	TaqαI	TTAAATGGATAGCGGTAGCT	MG934563	795	653-6	95-96	Multiple hits	Undetermined	
	Right	BfuCI	CCGGTGAGTGCGTAGCTTAA	--	79	79	100			
A3	Left	BfuCI	TTAAACTGTTTGCCGGCCGC	--	75	75	96	3 R	6,934,418	Intragenic
	Right	BfuCI	GTCCTCCAGGTAGTCTTTAA	--	33	31	100		6,934,415	AGAP008214
A8	Left	TaqαI	TTAATCTCGGTTCTCGGTAT	MG934564	323	323	100	2 R	32,162,290	Intergenic
	Right	*	ATATTTACGCAGATTCTTAA	--	53	42	90.5		32,162,293	
A10	Left	TaqαI	TTAAAGAACTGATCAATACA	MG934565	404	329**	99.1	2 R	5,816,202	Intergenic
	Right	*	CAAACGCACAGTACCATTAA	MG934566	368	353	98.9		5,816,199	
B3	Left	BfuCI	TTAAAGCTCGTTCATACTCT	MG934567	221	221	98.2	3 R	32,092,431	Intergenic
	Right	BfuCI	GAAGCTGTAAAAAGCTTTAA	--	192	191	95.8		32,092,428	
C1	Left	TaqαI	TTAAAAGAGATCCCGCGAGG	MG934568	652	648	99.2	X	21,936,498	Intergenic
	Right	BfuCI	TAGGTCGATGAACTCCTTAA	--	102	102	100		21,936,501	
D2	Left	TaqαI	TTAAAGTGCTTTCAACGTTA	MG934569	700	700	98.6-99.4	Multiple hits	Undetermined	

Enzyme refers to the enzyme used for genomic DNA cleavage prior to iPCR; sequence refers to the first 20 bp directly adjacent to the TTAA of the insertion site (--: sequence shorter than 200 bp and thus not submitted to GenBank); PEST Chromosome indicates the chromosome of transgene insertion; position indicates the co-ordinates of insertion with reference to the PEST genome sequence AgamP4 reported in VectorBase (Giraldo-Calderón et al 2015).

* the genomic position flanking this arm was obtained by amplification of genomic DNA using primer sequences deduced from the position of the left arm.

** there is an insertion in the PEST sequence that is not present in A10, Pimperena strain or *An. coluzzii*.

A)

Driver   Responder

Promoter — GAL4 ✗ UAS — Gene

B) P | 3xP3 - eCFP | PUBc5' | GAL4 | PUBc3' | P

C) 3xP3 eCFP | UAS(14x) | Luc | UAS(14x) | eYFP

D) PUBc5' | eGFP | SV40

E) 3xP3 - eCFP | R | 3xP3 - eYFP | UAS(10x) | mCD8:mCherry | L

F)

Luciferase activity (RLU)

****

ns

pPUBc-GAL4
pUAS-Luc

pUAS-Luc
(control)

pPUBc-eGFP
(blank)

A. m, b, fb, ms — 5 s
B. 5 s
C. nc, id — 5 s
D. 5 s
E. an — 2.5 s
F. 3.4 s
G. an, p, om, l, t, ab — 7 s
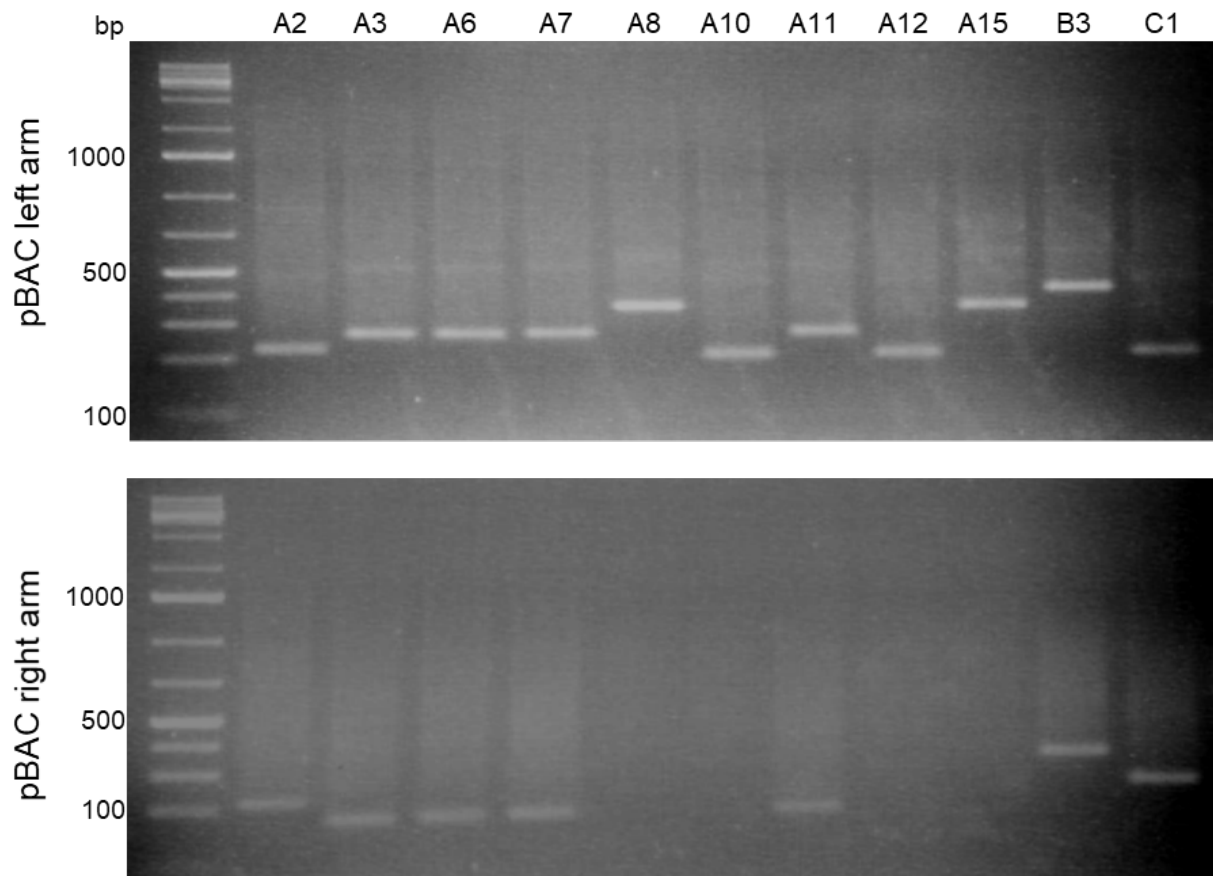**H.** 10.3 s
I. 7 s
J. 10.3 s
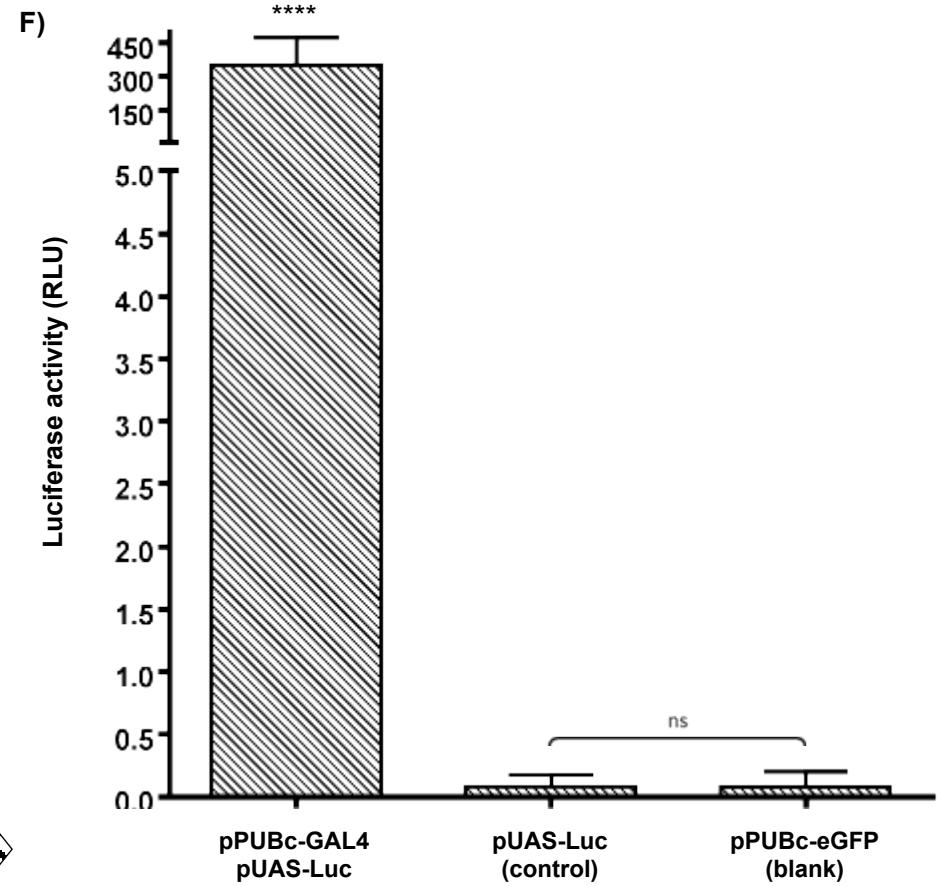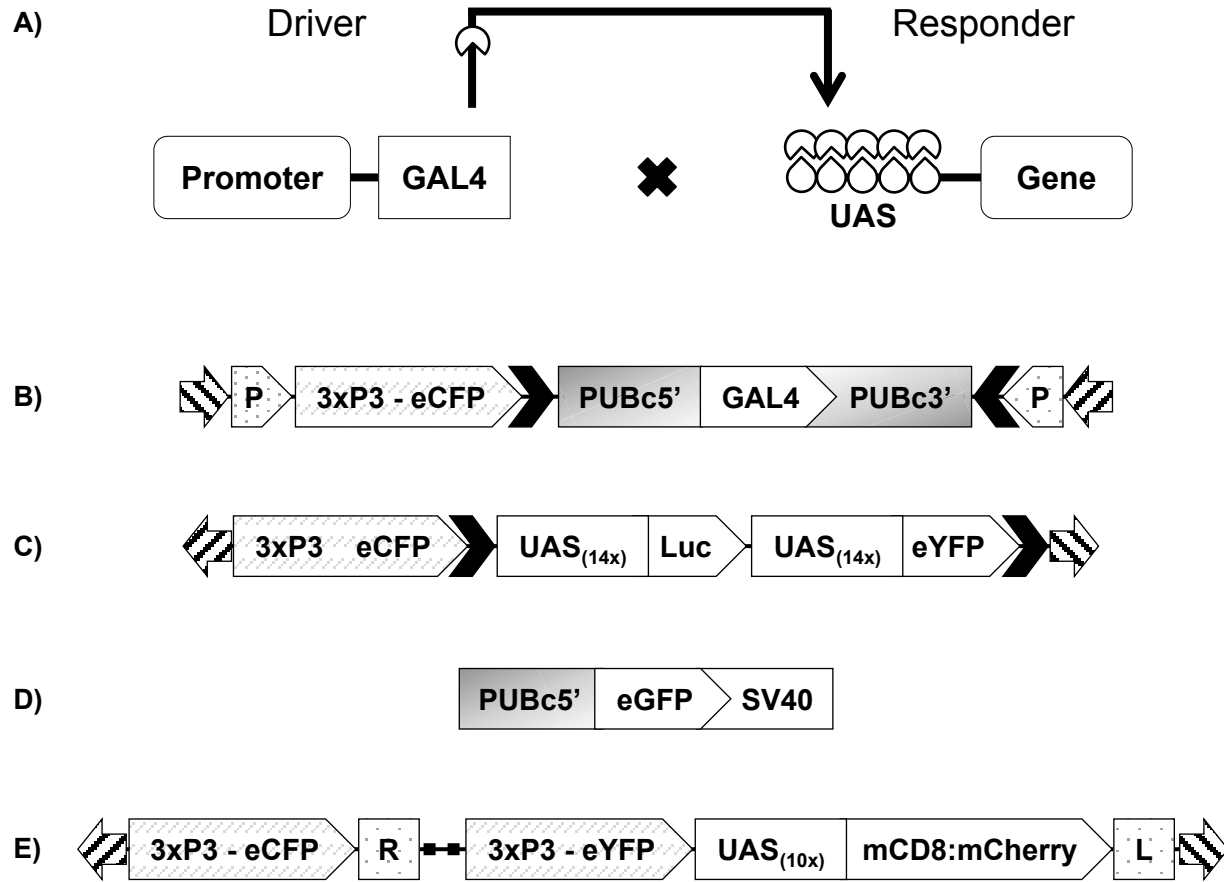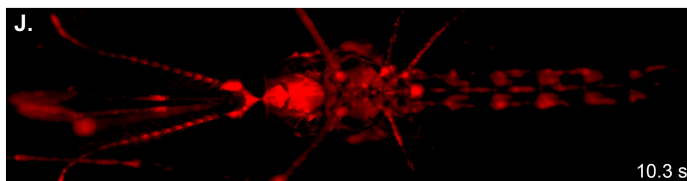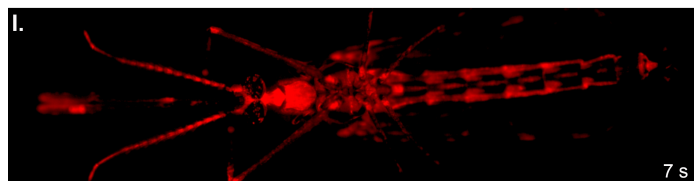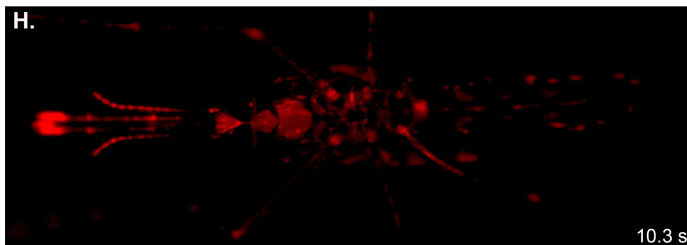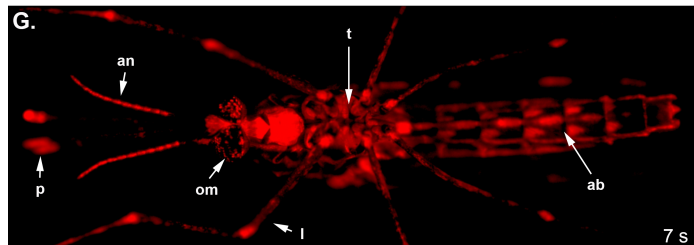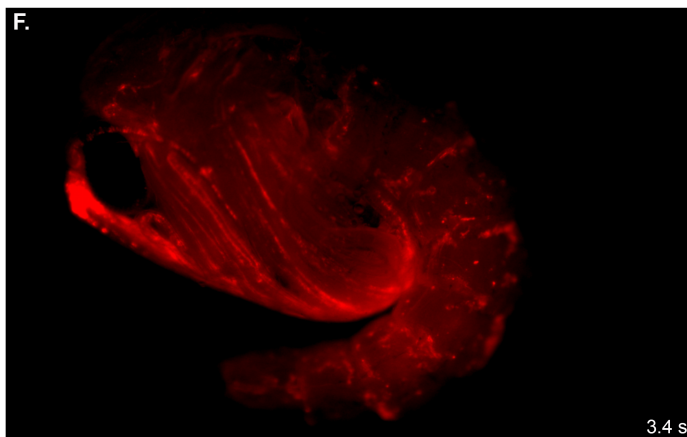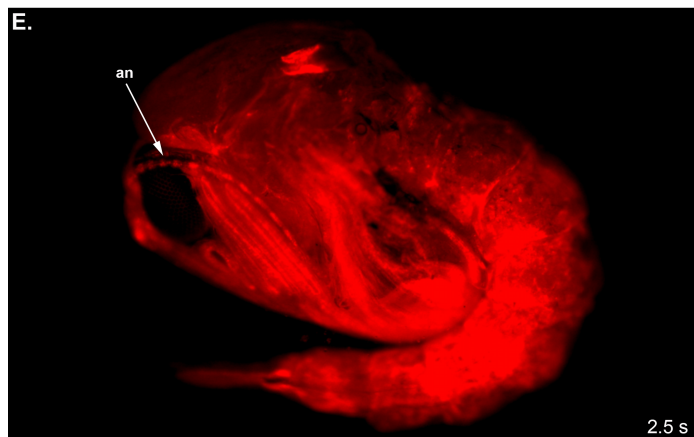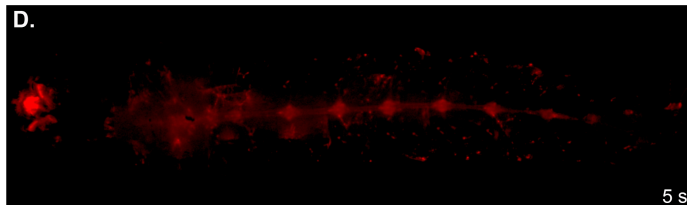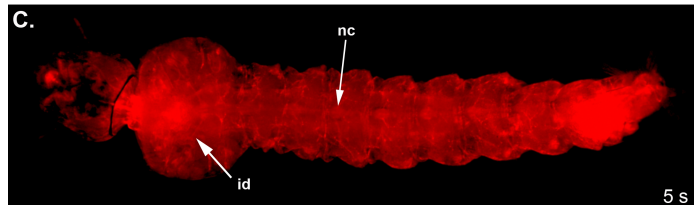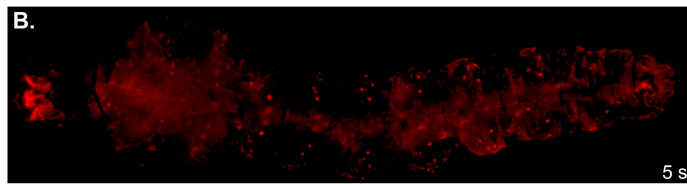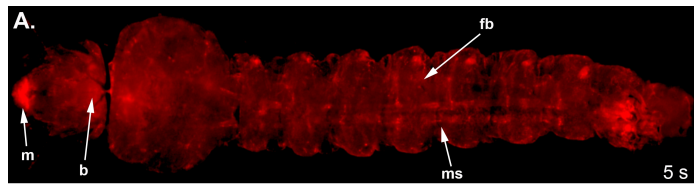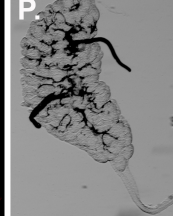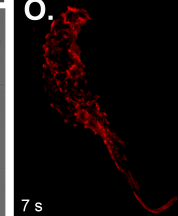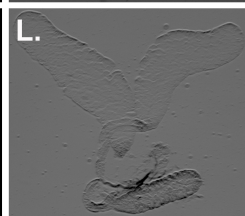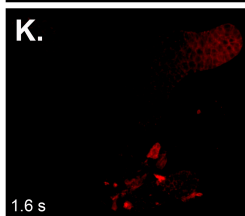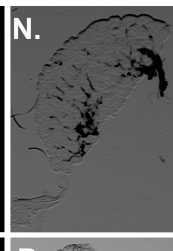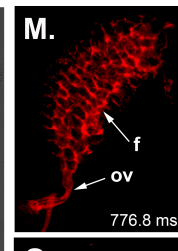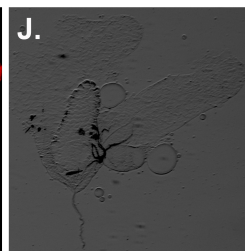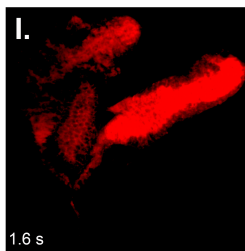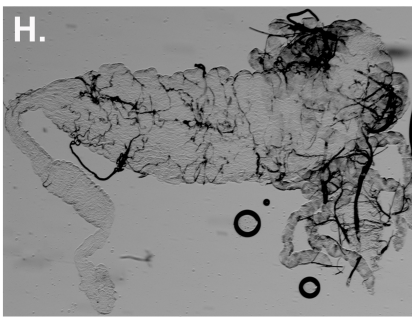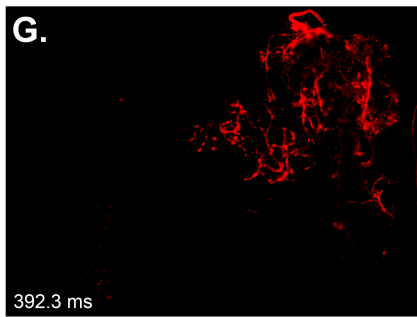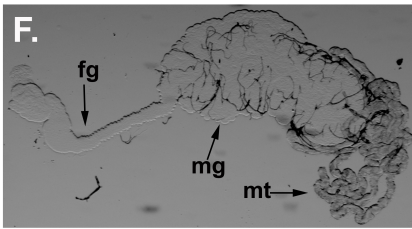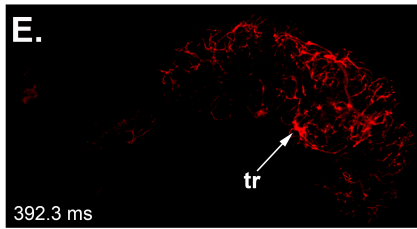
Figure S1. Agarose gel electrophoresis showing results for inverse-PCR on regions flanking the piggyBac arms after BfuCI digestion and self-ligation of genomic DNA isolated from isofemale lines. Among the isofemale lines derived cage A, those sharing the same PCR product size originated from the same F_0 founder individual and thus bear an insertion in the same genomic location. Six different single insertions are shown: 1) A2; 2) A3 = A6 = A7 = A11; 3) A8 = A15; 4) A10 = A12, 5) B3; 6) C1. Ladder is GeneRuler 1 kb Plus (Thermo Scientific).
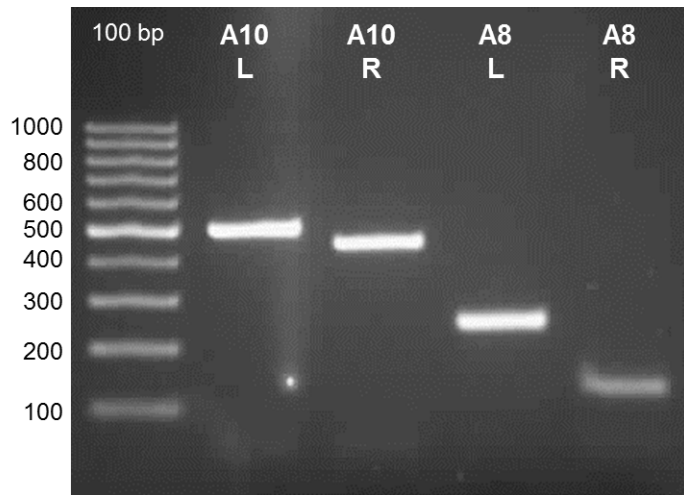
Figure S2. Agarose gel electrophoresis confirming insertion sites in A8 and A10 lines by PCR using a primer designed against regions near the predicted genomic insertion sites and the other annealing to conserved region of the integrated construct. A10: genomic regions external to the left (L) and right (R) piggyBac arms at the insertion site in the A10 line. A8: genomic regions external to the left (L) and right (R) piggyBac arms at the insertion site in the A8 line. Ladder is GeneRuler 100 bp (Thermo Scientific).
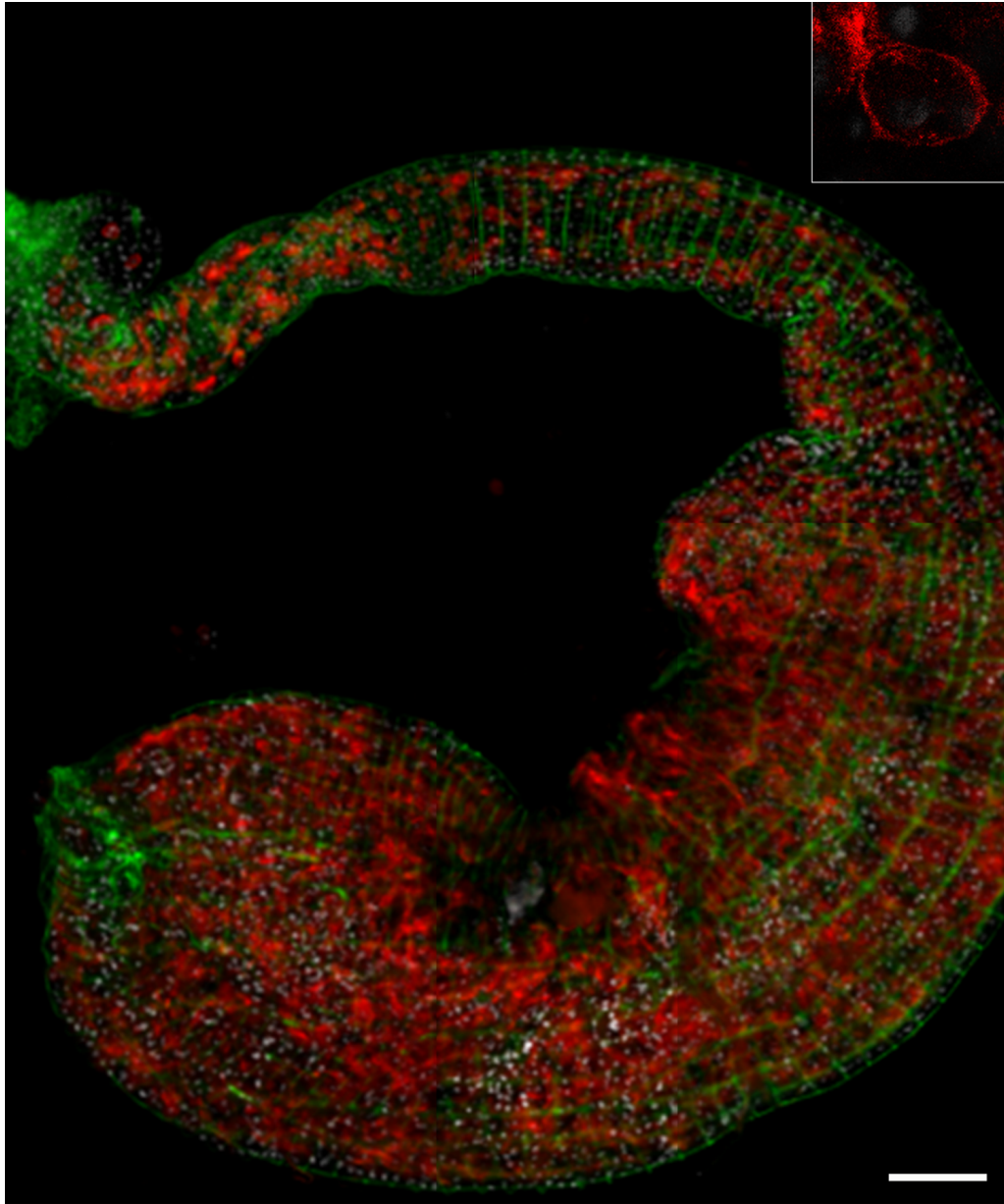
Figure S3. Combined Z stacks from confocal microscopy to include whole midgut from an adult female of the carboxypeptidase-GAL4/mCD8:mCherry line. Fixed midguts were stained with DAPI (white) and phalloidin 488 (green). Red signal shows the endogenous expression of mCD8:mCherry. Scale bar is 250 µm. Inset shows an individual midgut cell expressing mCD8:mCherry at the cell membrane.
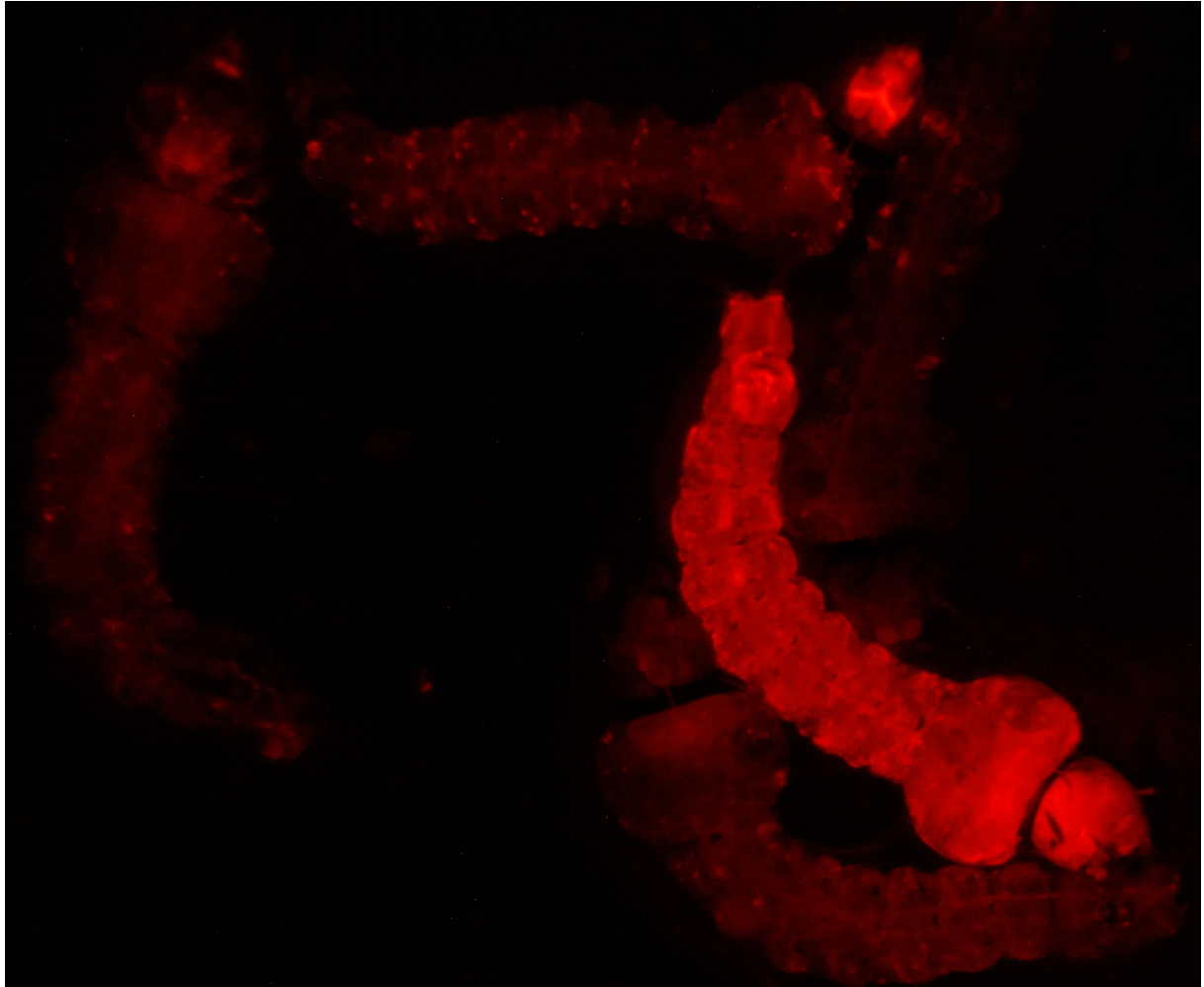
Figure S4. Example of phenotype variegation in the PUBc-GAL4 lines. mCherry phenotypes in L2-3 larvae derived from PUBc-GAL4 F_1 transformant males. These are also representative of the phenotypes observed in the progeny of single laying females.

Table S1. Primers used in this study.

Primer	Fragment	Template	Sequence 5'→3'
pSL-Gypsy1-FW	Gypsy1	attB:UAS:Cyp6m2	GAGGTAATTATAACCCGGGCCCTATATATGACTAGAATTGATCGGCTAAATGGTATGGCA
Gypsy1-PUBc5-RV			GAAATATTTTTTTATTCGCTAGCCAATGTATCTTAACTACTCACGTAATAAGTGTGCG
Gypsy1-PUBc5-FW	PUBc5	Kisumu gDNA	GAGTAGTTAAGATACATTGGCTAGCGAATAAAAAAATATTTCTTAATAATATTCTAAC
PUBc5-GAL4-RV			GAAGACAGTAGCTTCATATTTTATCTGTAAATATAAACGAAAACAAC
PUBc5-GAL4-FW	GAL4	attB:GAL4:DsRed	GTTTTCGTTTATATTTACAGATAAAATATGAAGCTACTGTCTTCTATCGAACAAGCATG
GAL4-pSL-RV			CAGCTGCAGGCGGCCGCCATATGCAAGATCTTTACTCTTTTTTTGGGTTTGGTGGG
pSL-PUBc3-FW	PUBc3	Kisumu gDNA	TAATTATAACCCGGGCCCTATATATGAGATCTTTCGTTGAATAAAGCATATTGAAGCTTC
PUBc3-Gypsy2-RV			AGCCGATCAATTCTAGTCATATGCCGTCGAAATTGTTTTACAATGACAATTTT
PUBc3-Gypsy2-FW	Gypsy2	attB:UAS:Cyp6m2	GTAAAACAATTTCGACGGCATATGACTAGAATTGATCGGCTAAATGGTATGGCA
Gypsy2-attP-RV			GTCAGTCGCGCGAGCGCGCCGCGGCAATGTATCTTAACTACTCACGTAATAAGTGTGCG
Gypsy2-attP-FW	attP	pBAC:attP:eCFP	AGTAGTTAAGATACATTGCCGCGGCGCGCTCGCGCGACTGACGGTCGTAAGCAC
attP-pSL-RV			AGCTGCAGGCGGCCGCCATATGCACGAAGCCCCGGCGGCAACCCTCAGCG
3P3-FW	3xP3-eYFPafm-attB	pBAC[3xP3-eYFPafm]	GTCATCACAGAACACATTTG
attB-RV			CAGGTACCGTCGACGATGTAG
BglII-mCD8-FW	mCD8:cherry	pUAS-mCD8:cherry	TAGCAGCCAGATCTGTCGACGGTATCGATAAGC
Xbal-cherry-RV			TAGCAGCCTCTAGATTACTTGTACAGCTCGTCCATGC
ITRL1F	--	--	ATCAGTGACACTTACCGCATTGACA
ITRL1R	--	--	TGACGAGCTTGTTGGTGAGGATTCT
ITRR1F	--	--	TACGCATGATTATCTTTAACGTA
ITRR1R	--	--	GATGTTTTGTTTTGACGGACCCC