

Identification of an unannotated early embryonic *single-minded* transcript in the yellow fever mosquito *Aedes aegypti*

William Reid¹, Marc S. Halfon^{1§}

¹Department of Biochemistry, University at Buffalo-State University of New York, Buffalo, New York, United States

[§]To whom correspondence should be addressed: mshalfon@buffalo.edu

Abstract

The yellow fever mosquito, *Aedes aegypti*, is a cosmopolitan species that serves as the vector of multiple disease causing agents including dengue, chikungunya, Zika, and yellow fever viruses. The genome of *Ae. aegypti* has been characterized at the chromosome level, but further manual refinement is required for genes and isoforms with transient expression or low abundance. Here we report on the identification of an early embryonic transcript for the *single-minded* (*sim*) gene in *Ae. aegypti*, and present the putative promoter for the transcript. The identification of an early-driven transcript is consistent with the annotation for *sim* in *Drosophila melanogaster*.

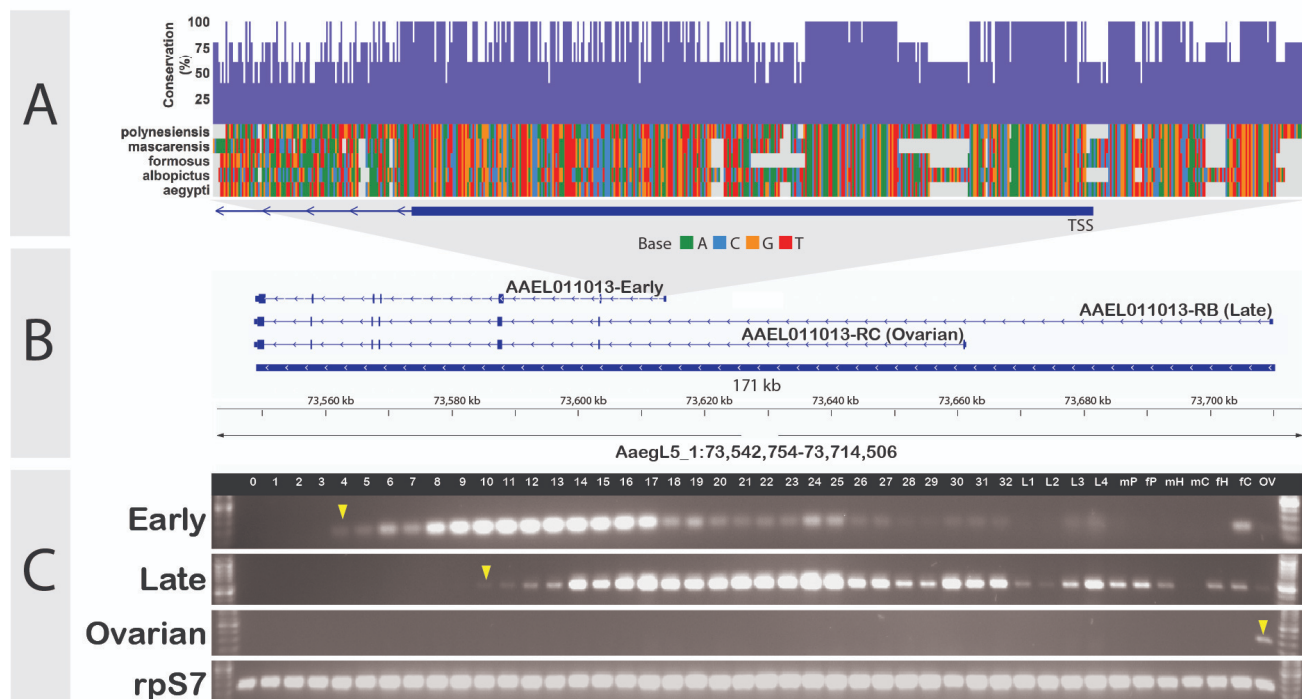


Figure 1. The (revised) *sim* locus of *Ae. aegypti*:

(A) Conservation of the *sim* early transcript promoter region for five species of *Aedes* spp. The upper purple histogram represents the percentage of conservation for all five species; nucleic acid sequences are color-indicated below. (B) Genomic position of the three isoforms of *sim* including the previously un-annotated early promoter. (C) Time-course RT-PCR of the late, ovarian, and early transcripts for the *single-minded* (*sim*) gene in *Aedes aegypti* (AAEL011013). The early transcript is present from 4h AEL and sharply decreases in expression by 18h AEL, while the late (RB) transcript is present from 10h AEL and remains the dominant transcript throughout embryogenesis and the other life stages tested, with the exception of the male carcass, where it is absent, and the ovary. The ovarian (RC) transcript is detected uniquely in the ovaries, where it is the dominant transcript. L1-L4: 1st instar larva through 4th instar larva; mP: male pupae; fP: female pupae; mH: adult male heads; mC: adult male carcass without heads; fH: adult female heads; fC: adult female carcass without heads; OV: 48h post blood meal ovaries.

Description

Accurate characterization of gene structure and transcript isoform expression is an essential prerequisite for studying a gene's function and regulation. During insect embryogenesis, the *single-minded* (*sim*; FlyBaseID:FBgn0004666) gene is a critical and conserved regulator of the central nervous system midline (Nambu et al., 1991; Zinzen et al., 2006; Linne et al., 2012). *Sim*, a bHLH-Pas domain transcription factor, heterodimerizes with the related bHLH-Pas protein Tango to bind to regulatory elements and stimulate the expression of key genes involved in the development of midline neuronal and glial cell lineages (Nambu et al., 1990; Wharton et al., 1994; Sonnenfeld et al., 1997; Watson and Crews, 2012). *sim*'s structure, transcriptional profile, and regulation are all well-characterized in *Drosophila melanogaster*, where *sim* is expressed from the early blastoderm stage and remains expressed throughout embryogenesis (Thomas et al., 1988; Wheeler et al., 2006). Post-embryogenesis, *sim* is involved in axon guidance in the brain, development of the optic lobe and genital discs, and high expression in the follicular cells of the ovaries (Pielage et al., 2002; Freer et al., 2011; Oramas et al., 2023). Three promoters control the expression of *D. melanogaster sim*, giving rise to virtually-identical isoforms: an 'early' transcript, which is the predominant transcript from embryonic stages 5 through 9, a 'late' transcript, which is the predominantly expressed transcript from stage 9 onward throughout embryogenesis and in other developmental stages, and an ovarian isoform, which is expressed in the follicular cells of the developing egg, promoting maturation (Freer et al., 2011; Knapp et al., 2020). The 'late' embryonic promoter also produces a second nearly-identical alternatively-spliced transcript from roughly stage 9 throughout embryogenesis (Freer et al., 2011).

Recently, we have been studying how *sim* is regulated in the distantly-related mosquito *Aedes aegypti* (Schember et al., 2024; Sterling-Lentsch and Halfon, 2024). However, only two transcripts were annotated in the *Ae. aegypti* genome for the *sim* ortholog *AAEL011013*, leading us to wonder if the genome annotation was complete. Although RNA-seq data exist for *Ae. aegypti* embryos (e.g. Biedler et al., 2012; Akbari et al., 2013), an early transcript that is expressed transiently and only in the small population of midline cells could easily be missed in RNA-Seq characterization due to low coverage depth. Here, we report on the identification of such an early *sim* transcript, thus demonstrating equivalency in gene structure and promoter usage between *D. melanogaster* and *Ae. aegypti*.

To test for the presence of a distinct unannotated early *Ae. aegypti sim* transcript, we collected *Ae. aegypti* embryos at 4h after egg laying (AEL) and processed them for 5'-RACE. We identified a novel transcript with a transcription start site (TSS) 10,263 bp upstream of Exon II of *AAEL011013* (Fig. 1B) and confirmed its existence with allele-specific intron-spanning RT-PCR primers (Fig. 1C). This places the early *sim* promoter as the most proximal to the shared Exon II, with the *AAEL011013-RC* (ovarian) and *AAEL011013-RB* (late) TSSs being 58,005 and 106,578 bp upstream, respectively. These results are consistent with the *sim* gene structure of *D. melanogaster*, where the early promoter is the most proximal to the common Exon II of the three defined *sim* promoters. Further testing of the expression of all *sim* transcripts using allele-specific intron-spanning RT-PCR (Fig. 1C) demonstrated that the early *sim* promoter is active as early as 4h AEL and sharply decreases expression by 18h AEL, at which point embryos are approximately equivalent to *Drosophila* stage 10-12. Meanwhile, the late *sim* transcript (*AAEL11013-RB*) is first detectable at 10h AEL (stage 8-10) and continues to be expressed throughout embryogenesis and through all life stages with the exception of the male carcass. The third *sim* isoform, *AAEL11013-RC*, was detectable only in 48h post-bloodmeal ovaries. The onset of transcript expression in *Ae. aegypti* is therefore consistent with that in *D. melanogaster*, where expression from the early promoter tails off by stage 11 and the ovarian promoter drives expression exclusively in the ovaries (Freer et al., 2011). Detection of the early-promoter and *AAEL11013-RB sim* transcripts in female, but not male, carcass may be due to the presence of expression in ovarian tissue, since the assayed carcasses consisted of the entirety of the mosquito minus the head. Three in-frame ATG putative start codons are present within the early *sim* transcript and are conserved among four additional *Aedes* spp: *Ae. albopictus*, *Ae. formosus*, *Ae. mascarensis*, and *Ae. polynesiensis* (Fig. 1A). Alignment and conservation analysis of all five *Aedes* spp. shows two strong blocks of conservation at the beginning of the predicted TSS through the first 100 bp of exon I. A third block of conservation was identified spanning the predicted splice donor site (Fig. 1A), with conservation in the region between the promoter and the splice donor remaining $\geq 60\%$. Conceptual translation of the orthologous exon I in all *Ae.* spp. indicates that either of the two ATG present within exon I could be the start codon, encoding 53 or 61 predicted N-terminal residues respectively, as all species retain in-frame ORFs irrespective of start codon choice. As such, whether the putative exon I of the early *sim* transcript is entirely or only partially 5'UTR cannot be determined from our study. Local alignment of the *D. melanogaster* early-specific N-terminal residues against those predicted for the *Ae. aegypti* early-specific isoform suggests a moderate degree of conservation (28.6% identity; 42.9% similarity), but whether or not these residues are translated or relevant for early Sim function remains to be determined. Our work further refines the structure of the *sim* locus in *Ae. aegypti*, which will facilitate future studies, and underscores the need for careful manual annotation and experiments beyond bulk RNA-seq to ensure reliable description of minor transcripts.

Methods

Mosquito rearing. *Ae. aegypti* (strain Liverpool) were maintained at 27°C and 75% humidity under a 12:12 (L:D) photoperiod. Larval mosquitoes were fed *ad libitum* with a slurry of ground TetraMin fish flake (Tetra, Blacksburg, VA),

8/26/2026 - Open Access

and adults were provided with raisins as a sugar source and defibrinated sheep blood (Hardy Diagnostics, Santa Maria, CA) supplemented with 1 μ M ATP for blood feeding.

Embryo collection and aging. Five day old adult females were provided with a bloodmeal, and four days following, 20-30 hyper gravid females were collected and placed into a 50 mL conical tube containing Whatman #1 filter paper wetted with deionized water. Females were allowed to oviposit for 1h and embryos were subsequently aged on wetted Whatman #1 filter paper at 27°C for the indicated time periods of “hours after egg laying” hours AEL.

RNA extraction, 5'RACE, RT-PCR. Total RNA was collected by homogenizing ~10 mg of embryos in Trizol (ThermoFisher Scientific, Grand Island, NY) according to the manufacturer's instructions. A total of 3 μ g of RNA collected from 4h AEL embryos was used for the 5'-RACE, utilizing the 5'-RLM Generacer kit (ThermoFisher Scientific, Grand Island, NY) and first strand synthesis was conducted using oligo dT and Superscript III reverse transcriptase (ThermoFisher Scientific, Grand Island, NY). PCR products were amplified using *sim*-specific reverse primers and RACE PCR products were sequenced using nanopore sequencing (Plasmidsaurus, Louisville, KY). The cDNA templates for the time course RT-PCR were reverse transcribed in 20 μ L reactions from 500 ng of total RNA using the RevertAid First Strand cDNA synthesis kit (ThermoFisher Scientific, Grand Island, NY) according to the manufacturer's instructions. First strand cDNA was subsequently diluted 5-fold and 1 μ L was used for each of four 15 μ L RT-PCR reactions specific for: 1) intron-spanning *sim* early transcript, 2) intron-spanning *sim* late transcript, 3) intron-spanning *sim* ovarian transcript, and 4) rpS7 control. Five μ L of each PCR product was then run out on a 1.5% agarose gel in 1xTBE and stained with ethidium bromide.

Promoter alignment. The full loci for *sim* for *Ae. aegypti* (AaegL5.0; GCA_002204515.1), *Ae. albopictus* (AalbF5; GCA_035046485.1), *Ae. formosus* (Aaf_Bf05_pri1.0; GCA_052575885.1), *Ae. mascarensis* (Am_MascCH02_pri1.0; GCA_052575835.1), and *Ae. polynesiensis* (Smith_Apoly_02182021; GCA_051529985.1) were downloaded from NCBI (ncbi.nlm.nih.gov) and aligned using MAFFT. The DNA regions matching to the predicted early *sim* transcript promoter from *Ae. aegypti* were then manually extracted and aligned to identify regions of full and partial conservation among the five mosquito species.

Reagents

Strain	Genotype	Available from
Liverpool	<i>Aedes aegypti</i>	BEI reagents
Kit	Source	Description
GeneRacer kit with SuperScript III RT	ThermoFisher Scientific	RNA ligase mediated 5' adapter tagging for RACE reaction
RevertAid First Strand cDNA synthesis kit	ThermoFisher Scientific	Reverse transcriptase for first strand cDNA synthesis from time course samples
HotStar Taq	Qiagen	Taq polymerase for RT-PCR amplification
Sheep blood	Hardy Diagnostics	Blood source for adult female mosquito feeding
Tetramin fish flakes	Tetra	Food source for larval mosquitoes
Primers	5'-3'	
Aegsim_Eearly-F1	ATGACACCGCTGATGGAC TG	

Aegsim_Eearly-II-R	CGTGGCAAGACCCAATAT ATTC	
Aegsim_Ellate-F1	AAGTTTGCCAAAGTCGTC GT	
Aegsim_Ellate-II-R	GCGTGGCAAGACCTTAGA ATTG	
Aegsim_E1ovary-F2	AGGTGGTCAGTTTGTTGC AGA	
Aegsim_E1ovary-II-R	GCGTGGCAAGACTTTTAT AAGAC	
rpS7-F	GGAGAAGAAGTTCTCCGG CAAG	
rpS7-R	TGAAGGTGTCGACCTTGT GTTC	

References

- Akbari OS, Antoshechkin I, Amrhein H, Williams B, Diloreto R, Sandler J, Hay BA. 2013. The developmental transcriptome of the mosquito *Aedes aegypti*, an invasive species and major arbovirus vector. *G3 (Bethesda)*. 3: 1493-509. 16. PubMed ID: [23833213](#)
- Biedler JK, Hu W, Tae H, Tu Z. 2012. Identification of early zygotic genes in the yellow fever mosquito *Aedes aegypti* and discovery of a motif involved in early zygotic genome activation. *PLoS One*. 7: e33933. 17. PubMed ID: [22457801](#)
- Freer SM, Lau DC, Pearson JC, Talsky KB, Crews ST. 2011. Molecular and functional analysis of *Drosophila* single-minded larval central brain expression. *Gene Expr Patterns*. 11: 533-46. 10. PubMed ID: [21945234](#)
- Knapp EM, Li W, Singh V, Sun J. 2020. Nuclear receptor Ftz-f1 promotes follicle maturation and ovulation partly via bHLH/PAS transcription factor Sim. *Elife*. 9 13. PubMed ID: [32338596](#)
- Linne V, Eriksson BJ, Stollewerk A. 2012. Single-minded and the evolution of the ventral midline in arthropods. *Dev Biol*. 364: 66-76. 2. PubMed ID: [22306923](#)
- Nambu JR, Franks RG, Hu S, Crews ST. 1990. The single-minded gene of *Drosophila* is required for the expression of genes important for the development of CNS midline cells. *Cell*. 63: 63-75. 4. PubMed ID: [2242162](#)
- Nambu JR, Lewis JO, Wharton KA, Crews ST. 1991. The *Drosophila* single-minded gene encodes a helix-loop-helix protein that acts as a master regulator of CNS midline development. *Cell*. 67: 1157-67. 1. PubMed ID: [1760843](#)
- Oramas R, Knapp EM, Zeng B, Sun J. 2023. The bHLH-PAS transcriptional complex Sim:Tgo plays active roles in late oogenesis to promote follicle maturation and ovulation. *Development*. 150 11. PubMed ID: [37218521](#)
- Pielage J, Steffes G, Lau DC, Parente BA, Crews ST, Strauss R, Klambt C. 2002. Novel behavioral and developmental defects associated with *Drosophila* single-minded. *Dev Biol*. 249: 283-99. 12. PubMed ID: [12221007](#)
- Schember I, Reid W, Sterling Lentsch G, Halfon MS. 2024. Conserved and novel enhancers in the *Aedes aegypti* single-minded locus recapitulate embryonic ventral midline gene expression. *PLoS Genet*. 20: e1010891. 14. PubMed ID: [38683842](#)
- Sonnenfeld M, Ward M, Nystrom G, Mosher J, Stahl S, Crews S. 1997. The *Drosophila* tango gene encodes a bHLH-PAS protein that is orthologous to mammalian Arnt and controls CNS midline and tracheal development. *Development*. 124: 4571-82. 5. PubMed ID: [9409674](#)
- Sterling Lentsch G, Halfon MS. 2024. Identification of a novel downstream single-minded midline regulatory element in *Drosophila melanogaster*. *MicroPubl Biol*. 2024 15. PubMed ID: [39502425](#)

Thomas JB, Crews ST, Goodman CS. 1988. Molecular genetics of the single-minded locus: a gene involved in the development of the Drosophila nervous system. *Cell*. 52: 133-41. 8. PubMed ID: [3345559](#)

Watson JD, Crews ST. 2012. Formation and specification of a Drosophila dopaminergic precursor cell. *Development*. 139: 3316-25. 6. PubMed ID: [22874915](#)

Wharton KA, Franks RG, Kasai Y, Crews ST. 1994. Control of CNS midline transcription by asymmetric E-box-like elements: similarity to xenobiotic responsive regulation. *Development*. 120: 3563-9. 7. PubMed ID: [7821222](#)

Wheeler SR, Kearney JB, Guardioli AR, Crews ST. 2006. Single-cell mapping of neural and glial gene expression in the developing Drosophila CNS midline cells. *Dev Biol*. 294: 509-24. 9. PubMed ID: [16631157](#)

Zinzen RP, Cande J, Ronshaugen M, Papatsenko D, Levine M. 2006. Evolution of the ventral midline in insect embryos. *Dev Cell*. 11: 895-902. 3. PubMed ID: [17141163](#)

Funding: This work was supported by funds from the University at Buffalo.

Conflicts of Interest: The authors declare that there are no conflicts of interest present.

Author Contributions: William Reid: conceptualization, investigation, writing - original draft. Marc S. Halfon: conceptualization, funding acquisition, writing - review editing.

Reviewed By: Anonymous

History: **Received** June 8, 2026 **Revision Received** August 24, 2026 **Accepted** August 24, 2026 **Published Online** August 26, 2026 **Indexed** September 9, 2026

Copyright: © 2026 by the authors. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Citation: Reid W, Halfon MS. 2026. Identification of an unannotated early embryonic *single-minded* transcript in the yellow fever mosquito *Aedes aegypti*. *microPublication Biology*. [10.17912/micropub.biology.002236](#)