1 **The origin and evolution of *loqs2*: a gene encoding an antiviral dsRNA binding protein in**
2 ***Aedes* mosquitoes**

3 Carlos F. Estevez-Castro¹, Murillo F. Rodrigues², Antinéa Babarit³, Flávia Viana Ferreira¹,
4 Eric Marois³, Rodrigo Cogni⁴, João Trindade Marques^{1,3*}, Roenick Proveti Olmo^{1,3*}

5 ¹ Department of Biochemistry and Immunology, Instituto de Ciências Biológicas, Universidade
6 Federal de Minas Gerais, Belo Horizonte, Brazil

7 ² Institute of Ecology and Evolution, University of Oregon, Eugene, Oregon, United States

8 ³ CNRS UPR9022, Inserm U1257, Université de Strasbourg, Strasbourg, France

9 ⁴ Department of Ecology, Institute of Biosciences, University of São Paulo, São Paulo, Brazil

10 **ORCID:**

11 Carlos F. Estevez-Castro, orcid.org/0000-0002-9872-2006
12 Murillo F. Rodrigues, orcid.org/0000-0001-7508-1384
13 Antinéa Babarit, orcid.org/0000-0002-5896-6321
14 Flávia V. Ferreira, orcid.org/0000-0001-8716-6828
15 Eric Marois, orcid.org/0000-0003-4147-3747
16 Rodrigo Cogni, orcid.org/0000-0001-9907-9297
17 João T. Marques, orcid.org/0000-0002-3457-3320
18 Roenick P. Olmo, orcid.org/0000-0002-3849-8591

19 *** Correspondence:**

20 Roenick Proveti Olmo - rprovetiolmo@unistra.fr

21 João Trindade Marques - jtm@ufmg.br

22 **Keywords:**

23 *loqs2*, RNA interference, *Aedes* mosquitoes, double-stranded RNA binding proteins (dsRBP)

24 Abstract

25 Mosquito borne viruses such as dengue, Zika, yellow fever and Chikungunya cause millions
 26 of infections every year. These viruses are mostly transmitted by two urban-adapted mosquito
 27 species, *Aedes aegypti* and *Aedes albopictus*, that appear to be more permissive to arbovirus
 28 infections compared to closely related species, although mechanistic understanding remains
 29 unknown. *Aedes* mosquitoes may have evolved specialized antiviral mechanisms that
 30 potentially contribute to the low impact of viral infection. Recently, we reported the
 31 identification of an *Aedes* specific double-stranded RNA binding protein (dsRBP), named
 32 Loqs2, that is involved in the control of infection by dengue and Zika viruses in *Ae. aegypti*.
 33 Loqs2 interacts with two important co-factors of the RNA interference (RNAi) pathway,
 34 Loquacious (Loqs) and R2D2, and seems to be a strong regulator of the antiviral defense.
 35 However, the origin and evolution of *loqs2* remains unclear. Here, we describe that *loqs2* likely
 36 originated from two independent duplications of the first dsRNA binding domain (dsRBD) of
 37 *loquacious* that occurred before the radiation of the *Aedes Stegomyia* subgenus. After its origin,
 38 our analyses suggest that *loqs2* evolved by relaxed positive selection towards
 39 neofunctionalization. In fact, *loqs2* is evolving at a faster pace compared to other RNAi
 40 components such as *loquacious*, *r2d2* and *Dicer-2* in *Aedes* mosquitoes. Unlike *loquacious*,
 41 transcriptomic analysis showed that *loqs2* expression is tightly regulated, almost restricted to
 42 reproductive tissues in *Ae. aegypti* and *Ae. albopictus*. Transgenic mosquitoes engineered to
 43 ubiquitously express *loqs2* show massive dysregulation of stress response genes and undergo
 44 developmental arrest at larval stages. Overall, our results uncover the possible origin and
 45 neofunctionalization of a novel antiviral gene, *loqs2*, in *Aedes* mosquitoes that ultimately may
 46 contribute to their effectiveness as vectors for arboviruses.

47 Introduction

48 *Aedes aegypti* and *Aedes albopictus* are the major vectors for arthropod-borne viruses
49 (arboviruses), such as yellow fever, dengue (DENV), Zika (ZIKV), and chikungunya (CHIKV)
50 viruses. DENV alone is estimated to be responsible for approximately 400 million infections
51 and 20,000 deaths per year worldwide [1]. Globalization, urbanization and climate change are
52 contributing to the spread of *Aedes* mosquitoes to previously uncolonized regions impacting
53 virus transmission and emergence of new arboviruses with potential to affect human health [2].
54 Despite the urgent need, there are no treatments or vaccines for most arboviral diseases [3].

55 *Ae. aegypti* and *Ae. albopictus* are not only well adapted to urban settings and anthropophilic
56 feeding but also seem to be naturally more susceptible to arboviruses under laboratory
57 conditions [4–7]. This suggests that these mosquito species may be intrinsically more
58 susceptible to arboviruses, which is a big part of their overall vector competence [7]. It remains
59 unclear whether there are clear genetic markers of vector competence in *Aedes* mosquitoes,
60 especially compared to other mosquito species such as *Culex* and *Anopheles*. One possibility
61 is that the antiviral defense in *Aedes* mosquitoes diverged from closely related species, allowing
62 virus infection without significant harm to the insect host [8–10].

63 RNA interference (RNAi) and, in particular, the small interfering RNA (siRNA) pathway, is a
64 broad antiviral defense mechanism in insects [9,11,12]. Virus-derived double-stranded RNAs
65 (dsRNA) are processed by Dicer-2 (Dcr-2) into siRNAs that are subsequently loaded into the
66 nuclease Argonaute-2 (AGO2) to form the RNA-induced silencing complex (RISC). This
67 complex is then able to target complementary viral RNAs and thus inhibit virus replication.
68 RNAi co-factors such as the dsRNA binding proteins (dsRBPs) Loqs and R2D2 are also
69 essential for the biogenesis and loading of small RNAs to the RISC complex [13] and, in the
70 *Ae. aegypti* cell line Aag2, Loqs and R2D2 also seem to act non-redundantly in the antiviral
71 branch of the siRNA pathway [14]. The arms race between viruses and antiviral RNAi resulted
72 in rapid evolutionary rates of siRNA pathway genes such as *Dcr-2*, *AGO2* and *r2d2*. This
73 adaptive signature has been observed in *Drosophila* and mosquito species [15–17]. In addition,
74 duplications and losses of RNAi genes are observed in multiple insect lineages, resulting in
75 great RNAi divergence and specialization even between related species [18]. Indeed, we have
76 recently identified a novel dsRBP gene in *Aedes* mosquitoes, *loqs2*, that is likely a paralog of
77 *loqs* [19]. Silencing of *loqs2* showed significantly increased levels of DENV RNA during
78 systemic infection and, in addition, ectopic expression of *loqs2* in the midgut resulted in the
79 control of DENV and ZIKV infection in the mosquito [19].

80 Here, we used different bioinformatic approaches to unveil the origin and evolution of *loqs2*.
81 Our results indicate that *loqs2* is a result of independent duplications of the first dsRNA binding
82 domain of *loqs* that occurred before the *Aedes Stegomyia* subgenus radiation. *loqs2* underwent
83 relaxed positive selection after its origin and is evolving at a faster pace than other RNAi
84 components such as *loqs* and *Dicer-2* in *Aedes* mosquitoes. We observed that *loqs2* expression
85 is tightly regulated in *Ae. aegypti* and *Ae. albopictus* and that ectopic ubiquitous expression
86 caused developmental arrest at the larval stages. Taken together, our results corroborate the
87 hypothesis that *loqs2* has evolved through neofunctionalization within some *Aedes* species to
88 regulate different aspects of the antiviral response.

Results

loqs2 is a paralog of *loqs* that originated in the ancestor of the *Aedes Stegomyia* subgenus

To analyze the origin of *loqs2*, we screened the genomes of flies and mosquitoes for proteins with at least one double-stranded RNA-binding domain (dsRBD) (InterPRO accession IPR014720). These were evaluated to determine their phylogenetic relationships (**Figure 1A**, extended version in **Supplementary Figure 1A**). The overall topology of the tree of dsRBPs shows monophyletic origins for each protein ortholog identified, with *Loqs2* proteins located as a single sister clade of *Loqs*. These results confirm that *loqs2* is a paralog of *loquacious* (*loqs*), a component of the RNAi pathway.

Interestingly, we also identified independent cases of gene duplications among the other species and ortholog groups, reinforcing the importance of dsRBPs. The dsRBP divergence and specialization among the species evaluated here also suggest that dsRNA recognition has been tuned by the individual evolutionary history towards specialization, even among closely related species. The clear monophyly found among ortholog groups identified, also shown by the relationship between the *Loqs* and *Loqs2* clade, showed that despite fast diversification there is a functional convergence potentially to maintain key properties. An example is the traceability of dsRBPs from apparently simple organisms to very complex ones, such as the orthologs *TRBPs/PACT* in vertebrates and urochordates, *r2d2/loqs* in arthropods, *RDE-4* in nematodes, and *DRB4* in plants, or also *Dicer* ortholog genes in all aforementioned clades [20]. Of note, it is compelling to expect more duplication events among genes containing dsRBDs that are involved in the antiviral arm of RNAi, as they are expected to be constantly under evolutionary pressure due to their role in immunity [16]. However, we also observed paralog duplications among dsRBD-containing genes that perform housekeeping functions such as the microRNA pathway, which is well recognized to be evolving under strong purifying selection [15,16].

Considering that *loqs2* is only found in *Aedes* mosquitoes, we evaluated the corresponding syntenic regions surrounding *loqs2* in other dipteran species using available genomes at Vectorbase (**Figure 1B**). This analysis showed genetic conservation of synteny structure in the region, in which three genes (*taflb*, *twins*, and *knickkopf*) were present in all dipteran species we analyzed, from *Ae. aegypti* to *Glossina palpalis*. Interestingly, other than the presence of *loqs2* in *Ae. aegypti*, we observed other differences in this genomic region between dipteran species. Two genes appeared between *taflb* and *twins* in *G. palpalis* that are absent in the other dipterans studied. In addition, *Anopheles gambiae*, *Culex quinquefasciatus* and *Ae. aegypti* possess an extra conserved gene inserted between *twins* and *knickkopf*, which is absent in *G. palpalis*. This suggests that this region might be prone to gene insertions and deletions. Our analysis also highlights the large genomic expansion that *Ae. aegypti* has undergone (**Figure 1B**) [21]. The syntenic genomic region from the left and rightmost genes in *Ae. aegypti* is twice as large compared to *C. quinquefasciatus* and five times larger compared to *G. palpalis*.

To further trace the origin of *loqs2* in the *Aedes* genus, we analyzed publicly available high-throughput RNA-seq libraries and whole genome sequencing (WGS) data from various species. Data included four species belonging to different subgenus (i.e., *Georgacraigius*, *Ochlerotatus*, *Dobrotworskyius*, and *Aedimorphus*), and four species from the *Stegomyia*

subgenus. We also analyzed RNA-seq data from the mosquito *Psorophora albipes*, species closely related to the *Aedes* genus (**Figure 1C**). Here we used reference genes from *Ae. aegypti* and *Ae. albopictus* to map libraries of the other mosquitoes, to identify potential orthologous genes. As a control, we were able to detect sequences mapping to the *loqs* gene from *Ae. aegypti* or *Ae. albopictus* in all nine datasets (**Supplementary Figure 1B**). In contrast, we only observed sequences mapping to the *Ae. aegypti* and *Ae. albopictus* *loqs2* genes in libraries derived from species in the *Stegomyia* subgenus. Furthermore, sequence coverages were smaller than that observed for *loqs*, which is expected considering the interspecific diversification of *loqs2*. Altogether, our results suggest that *loqs2* originated by a duplication of the *loqs* gene that occurred before or at the radiation of the *Aedes Stegomyia* subgenus, which is estimated to have occurred between 85.57 to 67.07 million years ago (MYA) (**Figure 1C**) [22]. Such estimation ranges from the late Cretaceous, passing the K-PG mass extinction, up to the beginning of the Paleocene. Thus, the *Aedes* radiation that followed the K-PG mass extinction might have also influenced the origin and evolution of *loqs2* among the *Stegomyia* subgenus. Apart from *Ochlerotatus* and *Dobrotworskyius*, we did not find *loqs2* in the *Aedimorphus* subgenus. This subgenus is phylogenetically closer to the *Stegomyia* than the other two [23]. Thus, we suspect *loqs2* originated near to the last common ancestor of the *Stegomyia* (67.07 MYA). However, as there is no available dating time of the separation between *Aedimorphus* and *Stegomyia* we cannot further narrow down the estimated origin of *loqs2*.

***loqs2* originated from two independent *loqs* dsRBD1 duplication events**

Similar to *loqs* and *r2d2*, the *loqs2* coding sequence is expected to produce a small protein containing two canonical dsRBDs (**Figure 2A**). However, the Loqs2 protein differs from Loqs by the distance between the two dsRBDs and the absence of a third domain, called Staufen C (ST), which is a dsRBD-like predicted to bind to other proteins [24]. Such differences in domain organization prompted us to investigate more carefully the origins of *loqs2* by interrogating its structural features. First, we explored the phylogenetic relationships between individual dsRBDs from *loqs*, *loqs2* and *r2d2*. Although our previous data showed *loqs2* is a paralog of *loqs*, we also included *r2d2* domains to rule out any possible relationship between those proteins. We observed that both the first and second dsRBDs of *loqs2* (D1, D2) grouped together with the first dsRBD of *loqs* but relatively far from the second dsRBD. Local alignment of the dsRBDs and Staufen C (ST) amino acid sequences of Loqs and Loqs2 from *Ae. aegypti* and *Ae. albopictus* (**Figure 2B**) confirmed that both Loqs2 dsRBDs (Loqs2-D1 and Loqs2-D2) were more similar to the first dsRBD of Loqs (Loqs-D1). Importantly, this analysis also showed that they are more similar to Loqs-D1 than to each other (**Figure 2B**) and highlighted the great conservation seen among the evaluated *loqs* orthologs compared to the interspecific diversification of *loqs2* genes (**Supplementary Figure 1B**). We also compared structural properties of Loqs and Loqs2 dsRBDs by using the resolved three-dimensional structures of the dsRBD1 and dsRBD2 from Loqs-PD of *D. melanogaster* (PDB ids 5NPG and 5NPA, respectively) as a reference [25]. We first predicted *in silico* the structures of individual dsRBDs of Loqs and Loqs2 from *Ae. aegypti* and *Ae. albopictus* and used the best model for

comparisons using the DALI server [26]. Both dsRBDs from Loqs2 were more structurally related to the first dsRBD of Loqs (**Figure 2C**). Notably, the second dsRBD of Loqs presented low structural similarity to all the other domains analyzed (Loqs-D1, Loqs2-D1 and Loqs2-D2). Taken together, our findings suggest that *loqs2* originated from two independent tandem duplications of the first dsRBD of *loqs* and not a direct gene duplication. By using *Ae. aegypti* as a model for genome organization in the *Stegomyia* subgenus, we can suggest that such domain duplications from the *loqs* gene located in chromosome 2 were then inserted into chromosome 1, where *loqs2* is now present (**Figure 2D**).

Relaxed positive selection shaped *loqs2* evolution towards neofunctionalization

Gene duplications can have a range of outcomes. Purifying selective constraints are expected to pressure one copy to maintain its original functions, while the other is free to accumulate mutations. In this scenario, the most likely outcome is pseudogenization of one copy. Alternatively, accumulation of mutations can lead to subdivision of ancestral functions between duplicated genes, called subfunctionalization. Occasionally, mutations may lead to neofunctionalization that, if advantageous, will drive fixation of the new gene in the population [27]. We hypothesized that the appearance of *loqs2* might have had little impact on the function of *loqs* (**Figure 3A**), since key functions of Loqs are still conserved even in the presence of Loqs2 [14]. In addition, there is some evidence for the neofunctionalization of *loqs2* [19]. To evaluate the type of selections that shaped the evolution of *loqs* and *loqs2*, we estimated the ratio of nonsynonymous to synonymous substitutions (ω) using a codon-based phylogenetic framework. This allowed us to build a topology by maximum likelihood inference using dipteran species, evaluating *loqs* and *loqs2* as ingroups, and *r2d2* as outgroup (**Figure 3B**). First, we performed a test for relaxation or intensification of selection using the software RELAX [28]. RELAX is based on the selection intensity parameter “ k ”, which establishes a comparison between the test and the reference branches of the phylogenetic tree. $k > 1$ indicates intensified selection while $k < 1$ indicates relaxed selection of test branches compared to the reference. We applied this strategy to *Aedes loqs* or *loqs2* clades as test branches compared to *loqs*, *loqs2* or *r2d2* clades as references. Our results showed that *loqs2* is evolving under relaxed selection compared to either *loqs* ($k = 0.07$) or *r2d2* ($k = 0.09$) (**Figure 3C**). In contrast, *loqs* is evolving under intensified selection compared to *loqs2* ($k = 7.19$) or *r2d2* ($k = 2.82$). These results support the hypothesis that *loqs* and *loqs2* are evolving under different selective regimes.

The unusual origin of *loqs2* may have favored an evolutionary scenario under relaxed selection that continues to the present date, despite *loqs2* origin dated of more than 67 MYA. We aimed here to predict the impact of *loqs2* appearance on selective pressures acting on *loqs*. However, to our understanding, this analysis undergoes conflicting aspects. On one hand, because of prominent differences between both proteins, one could expect that the appearance of *loqs2* might have had little impact on the original selective pressures acting on *loqs*. In this scenario, Loqs should conserve original key functions before and during the evolution of *Stegomyia* subgenus. On the other hand, previous results from our group show that Loqs2 interact with Loqs [19]. This suggests that the appearance of *loqs2* might have influenced *loqs*, affecting

original selective pressures. Altogether, our data favor a hybrid hypothesis, where *Loqs2* had little influence on *Loqs* evolution. However, to provide a stronger conclusion we would need to evaluate other *Aedes* species, whose genomes are not yet available while writing this manuscript. Thus, while *loqs* might have evolved under intense selection to maintain canonical functions [14], *loqs2* is evolving under relaxed selection, which allows diversification and, eventually, innovation (**Figure 3A**).

The type of natural selection acting on a protein can be very useful to understand its functional evolution. For example, a gene evolving towards pseudogenization (loss of function) is expected to evolve under neutral selection, while a gene moving towards neofunctionalization is expected to present sites under positive/negative selection. To investigate the modes of selection acting on both genes, we used two complementary approaches: the branch-site-model implemented in CodeML (CodeML-BSM) [29], and abSREL [30]. Both methods allow the estimation of ω variations among sites and branches, but with different assumptions. Conceptually, the major difference between both methods is that CodeML permits only four specific configurations of branch-specific rate parameters [31], and abSREL infers the optimal number of rate categories for each foreground branch and allows any of the sites of the branch to evolve under any of these categories [30]. As we were mainly interested in the evolution of *loqs* and *loqs2* in the *Aedes* genus, instead of separately testing the terminals for each of the genes (e.g., *Aae-loqs2* or *Aal-loqs2*), we tested branches before the species split, being the closest representatives of the evolutionary history of these proteins at the genus level. Interestingly, both methodologies found significant evidence of positive selection in *loqs* and *loqs2* branches in the *Aedes* genus (**Figure 3D**). Also, both methodologies found similar ω distribution over sites. In the case of the *Aedes loqs*, most sites showed signs of purifying selection (CodeML-BSM site class 0 = 82%, abSREL ω_1 = 92%), and a smaller frequency of sites show signs of positive (CodeML-BSM site class 2a-b= 2%, abSREL ω_2 =8.5%) or neutral selection (CodeML-BSM site class 1 = 16%). In contrast, roughly half of sites of *loqs2* showed evidence of positive selection (CodeML-BSM site class 2a-b= 51%, abSREL ω_2 =46%), while the other half had signs of purifying (CodeML-BSM site class 0 = 41%) or neutral selection (CodeML-BSM site class 1 = 8%, abSREL ω_1 = 54%). These results are consistent with the hypothesis of relaxed selection after the duplication that gave rise to *loqs2*, which was then followed by episodic adaptive evolution (**Figure 3C**).

To better understand the selective forces acting on *loqs* without the presence of *loqs2*, we also analyzed the *loqs* branches corresponding to *Culex* and *Anopheles* (**Figure 3D**). We did not find consistent evidence of positive selection considering both methods. Since the *Culicini* diverged around 130.49 MYA from the *Aedini*, and the *Anophelinae* diverged from both tribes around 182.75 MYA [22], a comparison of the *Aedes loqs* with the *Culex* ortholog is more reliable. Interestingly, this suggests that *loqs* experienced changes in its selective regime following *loqs2* origin and diversification. Possibly, the duplication event that gave rise to *loqs2* may have freed specific sites in *loqs* from previous constraints thus allowing positive directional selection to occur.

We also explored the recent evolutionary history of *loqs* and *loqs2* using the McDonald-Kreitman test (MKT) [32] available in the iMKT web server [33]. MKT infers the presence of

recurrent positive selection by comparing polymorphic sites among a population to divergent sites between the population against an outgroup. Thus, the MKT examines the evolutionary period between the divergence of the outgroup species and the present state of the ingroup population. We used publicly available exome data from *Ae. aegypti* samples from a forest in Senegal [34] compared to *Aedes mascarensis* as an outgroup [35]. Our results showed that *loqs2* has more amino acid substitutions within the Senegal forest population (selected polymorphisms) or compared to *Aedes mascarensis* (selected divergences), than intra- or interspecific neutral substitutions. On the contrary, *loqs* mainly presented neutral substitutions with only a few divergent sites between both species. The rate of nonsynonymous divergence (Ka) to the proportion of neutral fixations (Ks) suggested that *Ae. aegypti loqs2* ($Ka/Ks = 0.63$) is evolving substantially faster than *loqs* ($Ka/Ks = 0.063$) (**Figure 3E** and **Supplementary Figure 2A**). Although we observed an elevated Ka/Ks ratio for *loqs2*, especially when compared with *loqs*, it was not greater than 1, which would provide undeniable evidence of positive selection. In fact, the MKT results did not show significant evidence of positive selection neither for *loqs* nor *loqs2*. This lack of significance is likely due to the limitations of our exome dataset since we used a small mosquito population ($n = 9$) from Senegal as ingroup and a laboratory strain *Ae. aegypti* Liverpool as reference, which certainly does not reflect the diversity of current mosquito populations. Nonetheless, when we compared the Ka/Ks ratio of *loqs2* to the ratio of *dicer2*, a hallmark gene of positive selection in insects [16], including *Ae. aegypti* [17], we found that *loqs2* is also evolving considerably faster. In a similar fashion, fast evolving genes such as *D. melanogaster Dcr-2*, *AGO2* and *r2d2* display similar ω average values comparable to *loqs2* [36]. Of note, we have not compared the evolutionary rates of *Aedes loqs* to its *D. melanogaster* ortholog since it is not required for the antiviral siRNA response in the latter [37,38]. Altogether, our results indicate that *loqs2* is diversifying faster than *loqs* or *dicer-2* in a recent evolutionary scale. Furthermore, analysis of population genetics corroborates that *loqs* and *loqs2* are evolving under different selective regimes. In summary, both long-term and short-term evolutionary histories suggest that *loqs2* has evolved under relaxed positive selection and diversified towards neofunctionalization. Thus, *loqs2* could be fulfilling roles (e.g., antiviral immunity, control of gene expression) that *loqs* had not sufficiently evolved to sustain due to its shared role between the miRNA and the siRNA pathways [14].

Although the identification of conserved regions or motifs between species could indicate functional importance, rapidly evolving genes could escape this trend. We next used FEL (Fixed Effects Likelihood) [39] to explore signals of positive/negative selection at the level of specific sites. This approach estimates the ω rates for each site of a given coding alignment and phylogeny using maximum likelihood methods. We applied this methodology to *loqs* and *loqs2* separately in both *Ae. aegypti* and *Ae. albopictus* (**Figure 3F** and **Supplementary Figure 2B**). We focused on sites that were identified in both species that presumably reflect general trends for each protein. Regarding *loqs*, we observed 35 sites under negative selection that were common between *Ae. aegypti* and *Ae. albopictus*. These results confirm that *loqs* is evolving under strong purifying selection to conserve its functional properties in both species. In contrast, in the case of *loqs2*, we observed only 1 site under negative and 5 sites under positive selection that were common between *Ae. aegypti* and *Ae. albopictus*, 4 of which were located

outside of dsRBDs and were mainly located at the C-terminal region (**Figure 3F** and **Supplementary Figure 2B**).

These findings suggest that the C-terminal region of *loqs2* might have evolved under diversifying selection probably towards new protein-protein interactions and functions, while the dsRBDs mostly remained under pressure to maintain dsRNA binding ability. Similarly, comparison between *Drosophila* species show that positive selection is mostly observed outside conserved domains of *Dcr-2*, *AGO2* and *r2d2* [36]. This intra-protein evolutionary trend could provide novel and yet unknown functions to Loqs2 independent of the constraints of RNAi activity, such as enabling the communication between pathways upon the recognition of specific dsRNAs.

loqs2* has unique pattern of expression that is distinct from *loqs* and *r2d2

Next, we used publicly available RNA-seq data to analyze expression patterns of *loqs*, *loqs2*, and *r2d2* in different tissues of *Ae. aegypti* and *Ae. albopictus* compared to *loqs* and *r2d2* in *An. gambiae*. We observed that the expression of *loqs2* in both *Ae. aegypti* and *Ae. albopictus* was mostly restricted to reproductive tissues (**Figure 4A**). In contrast, *loqs* and *r2d2* showed ubiquitous expression in all three mosquito species (**Figure 4A**). In addition to reproductive tissues, we also noted high expression of *loqs2* in early embryos (0-4h) followed by a sudden decrease in later stages (24-72h) suggesting that *loqs2* mRNA is maternally deposited in the embryo. In addition, we observed no expression of *loqs2* in larval or pupal stages, suggesting tight regulation of its expression during mosquito development. Overall, the unique expression pattern of *loqs2* gives further support for its neofunctionalization after duplication from an ancestral *loqs* gene.

In order to investigate potential functions of *loqs2*, we evaluated if perturbations in *loqs2* expression could impact development in *Ae. aegypti*. For this, we generated two independent transgenic mosquito lines expressing *loqs2* under the control of either the *poly-ubiquitin* (PUB) gene promoter or the baculovirus promoter OpIE2 (**Figure 4B** and **Supplementary Figure 3A**). These promoters drive ubiquitous gene expression at late larval stages of *Aedes* and, in both cases, caused delayed larval development compared to non-transgenic siblings. Most transgenic individuals showed retarded molting with consequent prolonged L3 and L4 larval stages (**Figures 4C,D**). As a comparison, transgenic mosquitoes expressing eGFP or DsRed2 under the control of the same promoters do not show any developmental phenotypes (data not shown). Transcriptomic analysis of transgenic larvae and wild-type siblings at L2 and L4 stages showed significant differences in gene expression between both conditions. In transgenic mosquitoes, gene-set enrichment analysis (GSEA) showed a significant decrease in expression of genes related to metabolism and metabolic pathways, and upregulation of biological processes related to the control of reactive oxygen species (**Figure 4E** and **Supplementary Figure 3B**). Our results highlight the importance of tissue specific control of *loqs2* expression. It is tempting to speculate that the presence of Loqs2 could impact viral tropism since it is absent from tissues well known to host strong arbovirus replication such as the midgut. Loss-of-function experiments will be necessary to validate this hypothesis and better understand tissue-specific roles of Loqs2 in *Aedes* mosquitoes.

Discussion

Our group has recently reported that the ectopic expression of *loqs2* in the midgut of *Ae. aegypti* mosquitoes, where *Loqs2* is naturally absent, enables the control of DENV or ZIKV infection by the mosquito [19]. These results suggest that *Loqs2* could have evolved to be a strong regulator of the siRNA pathway, potentially balancing the response triggered by viruses or exogenous and endogenous dsRNAs. Recognition of viral dsRNA is indeed key to the siRNA pathway, and gene duplications of dsRBPs involved in RNAi are commonly observed in insects (**Supplementary Figure 1A**) [40,41]. In the fly model *D. melanogaster*, *blanks* is a recent paralog exclusive to this species that contains two dsRBDs. *Blanks* is highly expressed in *Drosophila* testes, and knock-out male flies are infertile due to spermiogenesis issues [40,41]. In *D. melanogaster* S2 cells, ablation of *Blanks* leads to defect in the siRNA pathway, suggesting that *Blanks* is a co-factor of RNAi. Strikingly, *Dicer-2* knockout flies are fertile but the role of *Blanks* in fertility seems independent of dsRNA binding, showing the versatility of these proteins [42]. Tracing a parallel with our current work, we observed that *loqs2* is also highly expressed in reproductive tissues of the mosquito. Interestingly, in addition to dsRBPs, duplication of *AGO2* that are only expressed in testis has also been observed in *Drosophila* species [43]. These observations suggest that the siRNA pathway is highly active in the gonads and may affect reproduction in addition to the antiviral defense. Notably, genes affecting host-pathogen interactions and reproduction are both expected to have a strong signal of positive selection.

Of note, *Loqs2* presents the same domain organization as *Loqs-PD*, a *D. melanogaster* *Loqs* isoform that evolved to be a regulator of the siRNA pathway in this species. As mentioned before, *Ae. aegypti* *Loqs* has 3 predicted isoforms (*Loqs-PA*, *-PB* and *-PC*) [14]. Interestingly, *Ae. aegypti* *Loqs-PC* is a predicted isoform that lacks the C-terminal Staufen C domain, similar to *D. melanogaster* *Loqs-PD*. Since *Ae. aegypti* *Loqs-PC* has no detectable expression in tissues of the insect, the role of this isoform could have been somewhat replaced by the appearance of *loqs2*. This could have led to the specialization of *Loqs* by negative selection and the diversification of *Loqs2* towards new tissue specific functions. Identification of isoforms lacking the Staufen C domain in other species belonging to *Aedes*, *Anopheles* and *Culex* genus could provide supporting evidence to this hypothesis. However, except for *Ae. aegypti*, the current genome annotation of these species available in Vectorbase (release 52) only exhibit one transcript correspondent to the long isoform of *loqs* that codes for orthologs of *Ae. aegypti* *Loqs-PA*. The unavailability of different isoforms is possibly a result of poor annotation of RNAi genes in these species, and further identification of isoforms would require more studies.

An outstanding question is whether the molecular functions of *Loqs2* are linked to that of *Loqs*, especially considering that our data indicate its origin from two independent duplications of the first dsRBD of *loqs*. In *D. melanogaster*, both dsRBDs of *Loqs* seem to be equally capable of binding dsRNA [25]. Nevertheless, the first dsRBD of *D. melanogaster* *Loqs-PD* is essential for the cleavage of suboptimal dsRNA substrates by *Dicer-2* *in vitro* [44]. The flexibility and length of the amino acid linker connecting both dsRBDs of *D. melanogaster* *Loqs* seems to allow independent binding of each domain to a dsRNA substrate, with consequent and mobile

reversible interaction [25]. In *Aedes* mosquitoes, Loqs has a linker length between dsRBDs D1 and D2 of 29 amino acids while this distance in Loqs2 is only of 15 amino acids. We speculate that the short linker between the two dsRBDs of Loqs2 result in lower flexibility and restrict binding to the same dsRNA molecule. This may have caused a shift in binding affinities to dsRNA substrates of different origin and conferred evolutionary advantages that helped select Loqs2. This hypothesis is also corroborated by the signs of positive selection observed on the linker of Loqs2 (**Figure 3F**). Taken together, we hypothesize that Loqs2 might have evolved to deal with unique dsRNA substrates possibly in response to the evolutionary arms race with viruses.

In the same perspective, we wondered if Loqs2 could bear similar protein-protein interaction features as others RNAi co-factors such as Loqs and R2D2. In *D. melanogaster*, the interaction of different isoforms of Loqs with other partner proteins occur independently of the two dsRBDs. In *D. melanogaster*, Loqs-PA interacts with Dicer-1 exclusively through its Staufien C domain, while Loqs-PD interacts with Dcr-2 through a 37 aa C-terminal region [37,44,45]. These regions are not present in Loqs2 but we observed strong signals of positive selection in its C-terminal region, which could indicate this may be the main site for interaction with protein partners. Notably, binding of both *D. melanogaster* R2D2 to Dcr-2 is also mediated by its C-terminal region and not dependent on the dsRBDs [44,46]. Nevertheless, we cannot rule out that the interaction between Loqs2 and partner proteins such as R2D2 and Loqs in *Ae. aegypti* are indirect and dependent on binding to dsRNA [19].

This work has begun to shed light into the origin and evolution of *loqs2*, an antiviral gene in *Aedes* mosquitoes. More studies are needed to further characterize *loqs2* functions and how its appearance may have shaped the evolution of vector competence in mosquitoes. With the recent advances in genetic-based strategies, *loqs2* seems like a promising target for genetic interventions since regulation of this *Aedes* specific gene could render mosquitoes resistant to virus infection. Strategies such as population replacement using gene-drive to modify mosquito populations to express *loqs2* in the midgut could be an alternative for targeting the transmission of arboviruses.

412 **Methods**

413 ***loqs* and *loqs2* sequence retrieval from publicly available high-throughput sequencing** 414 **data of different species**

415 Public high-throughput RNA-seq and DNA-seq libraries from various mosquitoes
416 (**Supplementary table 1**) were obtained from the NCBI/SRA database and mapped to the *Ae.*
417 *aegypti* and *Ae. albopictus* reference genomes (Vectorbase release 52) [47] using STAR
418 v2.7.9a [48]. To overcome the expected divergence between the RNA-seq data and the
419 reference genomes, we allowed a maximum alignment mismatch percentage of 50%. Mapping
420 coverage comprising regions of *loqs* and *loqs2* were manually inspected using IGV v2.8.3 [49]
421 to avoid false positives. Data was imported into R v4.1.0 [50] using in-house scripts and plots
422 were generated using the R package ggplot2 v3.3.3.

423 **Transcriptomic analyses of tissue-specific libraries**

424 To quantify the expression profiles of *loqs*, *loqs2* and *r2d2* among different tissues from *Ae.*
425 *aegypti*, *Ae. albopictus* and *An. gambiae* mosquitoes, public RNA-seq libraries were obtained
426 at NCBI/SRA as described above. Subsequently, reads were mapped to the decoyed reference
427 transcriptome of each species (Vectorbase release 52) using Salmon v1.5.2 [51]. Quasi-
428 mapping quantifications were imported into R v4.1.0 [50] and data normalization was
429 performed using the packages EdgeR v3.34.0 and TMM [52,53]. Plots were generated using
430 the R package ggplot2 v3.3.3.

431 **Phylogenetic analyses**

432 For phylogenetic analyses we used a set of mosquitoes and flies to infer the evolutionary history
433 of *loqs2* at three different evolutionary ranges: (i) the dsRBP orthologs (InterPRO accession
434 IPR014720) [54], (ii) the close related evolutionary history of *loqs2*, including *loqs* and *r2d2*
435 and, (iii) the *loqs2*, *loqs* and *r2d2* single domain evolutionary history. Input sequences were
436 aligned using either PRANK₊ (strategies i and iii) [55] or MAFFT-E-INS v7 [56] (strategy ii)
437 and the substitution model and phylogenetic relationships were inferred using IQ-tree2 [57].
438 For all cases, the maximum-likelihood consensus tree was generated with 1000 ultra-fast
439 bootstraps iterations. All trees were rooted at the midpoint using Figtree v1.4.4 [58] for
440 visualization purposes. For the dsRBP evolutionary range, we used VectorBase (release 52) to
441 retrieve amino acid sequences of the dsRBPs (InterPRO accession IPR014720), including
442 Dicer-2. The WAG amino acid exchange matrix under FreeRate heterogeneity model with 5
443 categories (WAG+R5) was determined to be the best fit model based on the Bayesian
444 information criterion (BIC). For the *loqs2* closely related tree we used the coding sequences
445 from the orthologs belonging to the same set of species of strategy i. The KOSI07 empirical
446 codon model with discrete Gamma model with 4 rate categories (KOSI07 + G4) was
447 determined to be the best fit model based on the BIC. For the single domain tree, we used the
448 coding sequences of each domain from *loqs2*, *loqs* and *r2d2* belonging to *Ae. aegypti* and *Ae.*

albopictus. The KOSI07 empirical codon model with amino-acid frequencies given by the protein matrix (KOSI07 + FU) was determined to be the best fit model based on the BIC.

Loqs and Loqs2 dsRBDs tridimensional modeling and alignment, amino acid sequence comparison

Homology models of the dsRBDs of *Ae. aegypti* and *Ae. albopictus* Loqs and Loqs2 were built using a template-based method. The *D. melanogaster* Loqs-PD dsRBD1 structure (PDB id: 5NPG) was used as the structural template to model the dsRBD1 of Loqs and the dsRBDs of Loqs2. For Loqs dsRBD2 models, the *D. melanogaster* Loqs-PD dsRBD2 structure (PDB id: 5NPA) was used as template [25]. Models were made with MODELLER v9.24 [59] using the automodel class. In each case, 100 models were produced, and the model with lowest MODELLER DOPE score was selected. Quality of each selected model was checked by the analysis of its Ramachandran plot built with PROCHECK [60]. All selected models presented at least 88% of their residues located in the most favored regions of the plot and no more than 1 residue located in a disallowed region. Model visualization was done using pymol v2.4.0 [61]. For structural comparisons, the modeled *Ae. aegypti* and *Ae. albopictus* Loqs and Loqs2 dsRBDs were compared to each other using the DALI server [26]. For the amino acid sequence alignment and comparison, the sequences of *Ae. aegypti* and *Ae. albopictus* Loqs and Loqs2 dsRBDs and Staufen C amino acid sequences were pairwise aligned with BLASTp under the BLOSUM62 substitution matrix [62]. Score and coverage of the pairwise alignments were summarized in a correlation matrix plotted using R v4.1.0 and the package ggplot2 v3.3.3.

Evolutionary analyses using a phylogenetic framework

We used the phylogenetic tree including *loqs*, *loqs2* and *r2d2* as evolutionary hypothesis to perform different tests of selection based on the rates of nonsynonymous and synonymous substitutions (ω). A test for relaxation or intensification of selection was performed with RELAX [28] implemented in the datamonkey server [63]. RELAX uses the parameter k to test for relaxed ($k < 1$) or intensified ($k > 1$) selection. RELAX tests for statistical significance (p -value < 0.05) by comparing a null model (k is fixed to 1) to an alternative model (k is a variable parameter) using a likelihood ratio test (LRT). Paired comparisons were made between the *Aedes* orthologs of *loqs* and *loqs2* as foreground branches and *loqs*, *loqs2* and *r2d2* as background branches. All comparisons were made including and excluding the internal branches of the tree. Positive selection at the branch level was evaluated using the branch-site-model implemented in CodeML (CodeML-BSM) [29] and automated in EasyCodeML [64], and abSREL [30] implemented in the datamonkey server. Both methods estimate ω rates at branch- and site-level of the test branches though under different assumptions. CodeML-BSM uses a model that allows positive selection along the foreground branches and sites only permitting four possible configurations of branch-specific rate parameters [31]. abSREL, infers the optimal number of ω classes and models their heterogeneity allowing for all possible configurations of branch-specific rate parameters. Both methods test for selection by using a LRT to compare the adaptive model to a null model that only allows neutral and negative

selection. We used both methods to test for selection acting on the branches of *loqs* or *loqs2* before the interspecific diversification of *Aedes*, *Culex* or *Anopheles* genus. FEL (Fixed Effects Likelihood) [39], implemented in the datamonkey server, was used to evaluate the selective forces acting at the sites-level on each of the *Aedes loqs* and *loqs2* paralogs. This approach estimates the site-specific nonsynonymous (dN) and synonymous (dS) substitution rates using a maximal likelihood approach. Basically, FEL uses the entire dataset to optimize branch lengths and nucleotide substitution parameters, then fits a MG94xREV model to each codon site and infers site-specific dN and dS substitution rates. Finally, it fits a neutral and selection model to every codon and calculates a standard LRT to decide if the site is evolving non-neutrally.

Evolutionary analyses using population genomics

We used public exome data of an *Ae. aegypti* population from a Senegal forest area [34] (SRA project: SRP092518). Low quality reads were removed and sequencing adaptors were trimmed using trimmomatic v0.39 [65], overlapping reads were merged with FLASH v1.2.11 [66] and both paired- and single-end pseudo-reads were mapped to the *Ae. aegypti* reference genome (Vectorbase release 52) using BWA-mem2 v.2.0pre2 [67]. SAMtools v1.13 [68] was used to merge and sort the paired- and single-end alignments into a single BAM file [68]. Read duplicates were marked with Picard v2.21.5 [69] and variant calling was performed with GATK v4.1.4.1 [70] with a heterozygosity prior of 0.003125 as reported by Redmond et al (2020) for *Ae. aegypti*. SNP calling was performed only for *loqs*, *loqs2*, *Dicer1* and *Dicer2* genomic loci including flanks of 2-Mb. Resultant SNP data was hard filtered with VCFtools v0.1.17 and indels were removed [71]. To determine the derived alleles, we performed the same SNP calling pipeline for *Aedes mascarensis* (SRA accession: SRR9959083) whole genome sequence. Finally, we used bcftools and GATK v4.1.4.1 to transform the SNP variants into a fasta sequence and used a custom python script to create a multi-fasta file for *loqs* or *loqs2* coding sequences. These files were used to calculate the nonsynonymous (K_a) to synonymous (K_s) substitutions using the iMKT web server [33].

Mosquito transgenesis and developmental assays

Embryo microinjection was performed with small modifications, as previously described [19]. Briefly, freshly laid *Ae. aegypti* eggs were aligned in parallel against the internal side of a U-shaped nitrocellulose membrane in contact with an overlaying filter paper soaked in demineralized water. Aligned eggs rested in a humid chamber for 30–60 min after alignment, until the eggs turned dark grey. Mixes containing 100 ng/μl of a *piggyBac* transposase helper-plasmid and 400 ng/μl of either plasmid carrying *piggyBac*-flanked cassettes (**Figure 4B**) were diluted in $0.5 \times$ PBS and injected at the embryo posterior pole under a Nikon Eclipse TE2000-S inverted microscope, using a Femtojet injector (Eppendorf) and a TransferMan NK2 micromanipulator. Before allowed to dry, microinjected eggs were kept for 2 days diagonally in a container with 1-cm-deep water. The filter paper had direct contact with water, which kept the eggs moist by capillarity [72]. Surviving eggs were hatched under vacuum and larvae

showing constitutive expression of fluorescent markers were kept for further crossing with wild-type individuals.

Poly-A selection, high-throughput RNA-seq library construction, and transcriptomic analysis

Total RNA from 5 to 10 individual larvae were extracted using TRIzol (Invitrogen) following the recommended protocol. RNA integrity was verified using the 2100 Bioanalyzer system (Agilent). mRNA libraries were constructed using the kits NEBNext Poly(A) mRNA Magnetic Isolation Module and NEBNext Ultra™ II Directional RNA Library Prep Kit for Illumina (New England BioLabs) following the manufacturer protocol. Indexed libraries were pooled and sequenced at the GenomEast sequencing platform at the Institut de Génétique et de Biologie Moléculaire et Cellulaire of Strasbourg, France. Sequenced reads with an average quality score above phred 25 had adaptors removed using Trimmomatic v0.39 and were further mapped to the decoyed transcriptome of *Ae. aegypti* (Vectorbase release 52) using Salmon v1.5.2 [51,65]. Quasi-mapping quantifications were imported into R v4.1.0 (R Core Team 2021) and data normalization was performed using the packages EdgeR v3.28.1 and TMM [52,53]. Ranked lists of gene expression for each comparison (transgenic against wild type larvae) was used as input for Gene Set Enrichment Analysis (GSEA) [73] using the R package fgsea v1.12.0 [74] and in-house developed gene-sets comprising Gene Ontology annotation, KEGG pathways, and genes of interest. Sets with adjusted FDR q -value < 0.1 were considered in our analysis and redundancy was removed if overlapping genes within given gene-sets were higher than 90%.

548 **Acknowledgments**

549 We thank all members of the Marques laboratory and the M3i unit - Insect Models of Innate
550 Immunity for suggestions and discussions.

551 **Funding**

552 This work of the Interdisciplinary Thematic Institute IMCBio, as part of the ITI 2021-2028
553 program of the University of Strasbourg, CNRS and Inserm, was supported by IdEx Unistra
554 (ANR-10-IDEX-0002), by SFRI-STRAT'US project (ANR 20-SFRI-0012), and EUR IMCBio
555 (IMCBio ANR-17-EURE-0023) under the framework of the French Investments for the Future
556 Program as well as from the previous Labex NetRNA (ANR-10-LABX-0036). This work was
557 also funded by grants from Conselho Nacional de Desenvolvimento Científico e Tecnológico
558 (CNPq); Fundação de Amparo a Pesquisa do Estado de Minas Gerais (FAPEMIG); Rede
559 Mineira de Imunobiológicos (grant # REDE-00140-16); Instituto Nacional de Ciência e
560 Tecnologia de Vacinas (INCTV); Institute for Advanced Studies of the University of
561 Strasbourg (USIAS fellowship 2019); and Investissement d'Avenir Programs (ANR-10-
562 LABX-0036 and ANR- 11-EQPX-0022) to JTM. This study was financed in part by the
563 Coordenação de Aperfeiçoamento de Pessoal de Nível Superior — Brasil (CAPES) — Finance
564 Code 001 (JTM).

565 **Author contributions**

566 Conceptualization: CFEC, JTM, RPO
567 Funding acquisition: JTM
568 Methodology: CFEC, MFR, FVF, EM, RC, JTM, RPO
569 Investigation and data curation: CFEC, MFR, AB, FVF, EM, RC, JTM, RPO
570 Formal analysis and visualization: CFEC, MFR, FVF, RC, JTM, RPO
571 Writing: CFEC, MFR, JTM, RPO

572 **Competing interests**

573 Authors declare that they have no competing interests.

574 **Data availability**

575 Transcriptome libraries from this study have been deposited on the Sequence Read Archive
576 (SRA) at NCBI. Other publicly available RNA-seq data sets were obtained from SRA.
577 Accession numbers are provided in **Supplementary Table 1**.

578 References

- 579 1. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global
580 distribution and burden of dengue. *Nature*. 2013;496: 504–507.
581 doi:10.1038/nature12060
- 582 2. Iwamura T, Guzman-Holst A, Murray KA. Accelerating invasion potential of disease
583 vector *Aedes aegypti* under climate change. *Nat Commun*. 2020;11: 2130.
584 doi:10.1038/s41467-020-16010-4
- 585 3. Weaver SC, Charlier C, Vasilakis N, Lecuit M. Zika, Chikungunya, and Other
586 Emerging Vector-Borne Viral Diseases. *Annu Rev Med*. 2018;69: 395–408.
587 doi:10.1146/annurev-med-050715-105122
- 588 4. Azar SR, Weaver SC. Vector Competence: What Has Zika Virus Taught Us? *Viruses*.
589 2019;11: 867. doi:10.3390/v11090867
- 590 5. Ferreira-de-Lima VH, Lima-Camara TN. Natural vertical transmission of dengue virus
591 in *Aedes aegypti* and *Aedes albopictus*: a systematic review. *Parasit Vectors*. 2018;11:
592 77. doi:10.1186/s13071-018-2643-9
- 593 6. Kauffman E, Payne A, Franke MA, Schmid MA, Harris E, Kramer LD. Rearing of
594 *Culex* spp. and *Aedes* spp. Mosquitoes. *Bio-Protoc*. 2017;7: e2542.
595 doi:10.21769/BioProtoc.2542
- 596 7. de Almeida JP, Aguiar ER, Armache JN, Olmo RP, Marques JT. The virome of vector
597 mosquitoes. *Curr Opin Virol*. 2021;49: 7–12. doi:10.1016/j.coviro.2021.04.002
- 598 8. Chamberlain RW, Sudia WD. Mechanism of Transmission of Viruses by Mosquitoes.
599 *Annu Rev Entomol*. 1961;6: 371–390. doi:10.1146/annurev.en.06.010161.002103
- 600 9. Aguiar ERGR, Olmo RP, Marques JT. Virus-derived small RNAs: molecular footprints
601 of host–pathogen interactions. *Wiley Interdiscip Rev RNA*. 2016;7: 824–837.
602 doi:10.1002/wrna.1361
- 603 10. Olmo RP, Martins NE, Aguiar ERGR, Marques JT, Imler J-L. The insect reservoir of
604 biodiversity for viruses and for antiviral mechanisms. *An Acad Bras Ciênc*. 2019;91:
605 e20190122. doi:10.1590/0001-3765201920190122
- 606 11. Rosendo Machado S, van der Most T, Miesen P. Genetic determinants of antiviral
607 immunity in dipteran insects – Compiling the experimental evidence. *Dev Comp*
608 *Immunol*. 2021;119: 104010. doi:10.1016/j.dci.2021.104010
- 609 12. de Faria IJ da S, Olmo RP, Silva EG, Marques JT. dsRNA Sensing During Viral
610 Infection: Lessons from Plants, Worms, Insects, and Mammals. *J Interferon Cytokine*
611 *Res*. 2013;33: 239–253. doi:10.1089/jir.2013.0026
- 612 13. Marques JT, Kim K, Wu P-H, Alleyne TM, Jafari N, Carthew RW. Loqs and R2D2 act
613 sequentially in the siRNA pathway in *Drosophila*. *Nat Struct Mol Biol*. 2010;17: 24–30.
614 doi:10.1038/nsmb.1735
- 615 14. Haac ME, Anderson MAE, Eggleston H, Myles KM, Adelman ZN. The hub protein
616 loquacious connects the microRNA and short interfering RNA pathways in mosquitoes.
617 *Nucleic Acids Res*. 2015;43: 3688–3700. doi:10.1093/nar/gkv152
- 618 15. Obbard DJ, Gordon KHJ, Buck AH, Jiggins FM. The evolution of RNAi as a defence
619 against viruses and transposable elements. *Philos Trans R Soc B Biol Sci*. 2009;364:
620 99–115. doi:10.1098/rstb.2008.0168

- 621 16. Palmer WH, Hadfield JD, Obbard DJ. RNA-Interference Pathways Display High Rates
622 of Adaptive Protein Evolution in Multiple Invertebrates. *Genetics*. 2018;208: 1585–
623 1599. doi:10.1534/genetics.117.300567
- 624 17. Bernhardt SA, Simmons MP, Olson KE, Beaty BJ, Blair CD, Black WC. Rapid
625 Intraspecific Evolution of miRNA and siRNA Genes in the Mosquito *Aedes aegypti*.
626 *PLOS ONE*. 2012;7: e44198. doi:10.1371/journal.pone.0044198
- 627 18. Dowling D, Pauli T, Donath A, Meusemann K, Podsiadlowski L, Petersen M, et al.
628 Phylogenetic Origin and Diversification of RNAi Pathway Genes in Insects. *Genome*
629 *Biol Evol*. 2016;8: 3784–3793. doi:10.1093/gbe/evw281
- 630 19. Olmo RP, Ferreira AGA, Izidoro-Toledo TC, Aguiar ERGR, de Faria IJS, de Souza
631 KPR, et al. Control of dengue virus in the midgut of *Aedes aegypti* by ectopic
632 expression of the dsRNA-binding protein Loqs2. *Nat Microbiol*. 2018;3: 1385–1393.
633 doi:10.1038/s41564-018-0268-6
- 634 20. Murphy D, Dancis B, Brown JR. The evolution of core proteins involved in microRNA
635 biogenesis. *BMC Evol Biol*. 2008;8: 92. doi:10.1186/1471-2148-8-92
- 636 21. Waterhouse RM, Wyder S, Zdobnov EM. The *Aedes aegypti* genome: a comparative
637 perspective. *Insect Mol Biol*. 2008;17: 1–8. doi:10.1111/j.1365-2583.2008.00772.x
- 638 22. da Silva AF, Machado LC, de Paula MB, da Silva Pessoa Vieira CJ, de Moraes Bronzoni
639 RV, de Melo Santos MAV, et al. Culicidae evolutionary history focusing on the
640 Culicinae subfamily based on mitochondrial phylogenomics. *Sci Rep*. 2020;10: 18823.
641 doi:10.1038/s41598-020-74883-3
- 642 23. Wilkerson RC, Linton Y-M, Fonseca DM, Schultz TR, Price DC, Strickman DA.
643 Making Mosquito Taxonomy Useful: A Stable Classification of Tribe Aedini that
644 Balances Utility with Current Knowledge of Evolutionary Relationships. *PLOS ONE*.
645 2015;10: e0133602. doi:10.1371/journal.pone.0133602
- 646 24. Arraes FBM, Martins-de-Sa D, Noriega Vasquez DD, Melo BP, Faheem M, de Macedo
647 LLP, et al. Dissecting protein domain variability in the core RNA interference
648 machinery of five insect orders. *RNA Biol*. 2020;0: 1–29.
649 doi:10.1080/15476286.2020.1861816
- 650 25. Tants J-N, Fesser S, Kern T, Stehle R, Geerlof A, Wunderlich C, et al. Molecular basis
651 for asymmetry sensing of siRNAs by the *Drosophila* Loqs-PD/Dcr-2 complex in RNA
652 interference. *Nucleic Acids Res*. 2017;45: 12536–12550. doi:10.1093/nar/gkx886
- 653 26. Holm L. Using Dali for Protein Structure Comparison. In: Gáspári Z, editor. *Structural*
654 *Bioinformatics: Methods and Protocols*. New York, NY: Springer US; 2020. pp. 29–42.
655 doi:10.1007/978-1-0716-0270-6_3
- 656 27. Prince VE, Pickett FB. Splitting pairs: the diverging fates of duplicated genes. *Nat Rev*
657 *Genet*. 2002;3: 827–837. doi:10.1038/nrg928
- 658 28. Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K. RELAX:
659 Detecting Relaxed Selection in a Phylogenetic Framework. *Mol Biol Evol*. 2015;32:
660 820–832. doi:10.1093/molbev/msu400
- 661 29. Yang Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol Biol Evol*.
662 2007;24: 1586–1591. doi:10.1093/molbev/msm088
- 663 30. Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL.
664 Less Is More: An Adaptive Branch-Site Random Effects Model for Efficient Detection

- of Episodic Diversifying Selection. *Mol Biol Evol.* 2015;32: 1342–1353.
doi:10.1093/molbev/msv022
31. Zhang J, Nielsen R, Yang Z. Evaluation of an Improved Branch-Site Likelihood Method for Detecting Positive Selection at the Molecular Level. *Mol Biol Evol.* 2005;22: 2472–2479. doi:10.1093/molbev/msi237
32. McDonald JH, Kreitman M. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature.* 1991;351: 652–654. doi:10.1038/351652a0
33. Murga-Moreno J, Coronado-Zamora M, Hervás S, Casillas S, Barbadilla A. iMKT: the integrative McDonald and Kreitman test. *Nucleic Acids Res.* 2019;47: W283–W288. doi:10.1093/nar/gkz372
34. Crawford JE, Alves JM, Palmer WJ, Day JP, Sylla M, Ramasamy R, et al. Population genomics reveals that an anthropophilic population of *Aedes aegypti* mosquitoes in West Africa recently gave rise to American and Asian populations of this major disease vector. *BMC Biol.* 2017;15: 16. doi:10.1186/s12915-017-0351-0
35. Redmond SN, Sharma A, Sharakhov I, Tu Z, Sharakhova M, Neafsey DE. Linked-read sequencing identifies abundant microinversions and introgression in the arboviral vector *Aedes aegypti*. *BMC Biol.* 2020;18: 26. doi:10.1186/s12915-020-0757-y
36. Obbard DJ, Jiggins FM, Halligan DL, Little TJ. Natural Selection Drives Extremely Rapid Evolution in Antiviral RNAi Genes. *Curr Biol.* 2006;16: 580–585. doi:10.1016/j.cub.2006.01.065
37. Hartig JV, Förstemann K. Loqs-PD and R2D2 define independent pathways for RISC generation in *Drosophila*. *Nucleic Acids Res.* 2011;39: 3836–3851. doi:10.1093/nar/gkq1324
38. Marques JT, Wang J-P, Wang X, Oliveira KPV de, Gao C, Aguiar ERGR, et al. Functional Specialization of the Small Interfering RNA Pathway in Response to Virus Infection. *PLOS Pathog.* 2013;9: e1003579. doi:10.1371/journal.ppat.1003579
39. Kosakovsky Pond SL, Frost SDW. Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. *Mol Biol Evol.* 2005;22: 1208–1222. doi:10.1093/molbev/msi105
40. Gerbasi VR, Preall JB, Golden DE, Powell DW, Cummins TD, Sontheimer EJ. Blanks, a nuclear siRNA/dsRNA-binding complex component, is required for *Drosophila* spermiogenesis. *Proc Natl Acad Sci.* 2011;108: 3204–3209. doi:10.1073/pnas.1009781108
41. Sanders C, Smith DP. LUMP Is a Putative Double-Stranded RNA Binding Protein Required for Male Fertility in *Drosophila melanogaster*. Dearden PK, editor. *PLoS ONE.* 2011;6: e24151. doi:10.1371/journal.pone.0024151
42. Nitschko V, Kunzelmann S, Fröhlich T, Arnold GJ, Förstemann K. Trafficking of siRNA precursors by the dsRBD protein Blanks in *Drosophila*. *Nucleic Acids Res.* 2020;48: 3906–3921. doi:10.1093/nar/gkaa072
43. Lewis SH, Webster CL, Salmela H, Obbard DJ. Repeated Duplication of Argonaute2 Is Associated with Strong Selection and Testis Specialization in *Drosophila*. *Genetics.* 2016;204: 757–769. doi:10.1534/genetics.116.192336

44. Trettin KD, Sinha NK, Eckert DM, Apple SE, Bass BL. Loquacious-PD facilitates *Drosophila* Dicer-2 cleavage through interactions with the helicase domain and dsRNA. Proc Natl Acad Sci. 2017;114: E7939–E7948. doi:10.1073/pnas.1707063114
45. Heyam A, Coupland CE, Dégut C, Haley RA, Baxter NJ, Jakob L, et al. Conserved asymmetry underpins homodimerization of Dicer-associated double-stranded RNA-binding proteins. Nucleic Acids Res. 2017;45: 12577–12584. doi:10.1093/nar/gkx928
46. Nishida KM, Miyoshi K, Ogino A, Miyoshi T, Siomi H, Siomi MC. Roles of R2D2, a Cytoplasmic D2 Body Component, in the Endogenous siRNA Pathway in *Drosophila*. Mol Cell. 2013;49: 680–691. doi:10.1016/j.molcel.2012.12.024
47. Amos B, Aurrecoechea C, Barba M, Barreto A, Basenko EY, Bazant W, et al. VEuPathDB: the eukaryotic pathogen, vector and host bioinformatics resource center. Nucleic Acids Res. 2021 [cited 15 Nov 2021]. doi:10.1093/nar/gkab929
48. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29: 15–21. doi:10.1093/bioinformatics/bts635
49. Robinson JT. Integrative genomics viewer. C O Rresp O N N Ce. 2011;29: 3.
50. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2021. Available: <https://www.R-project.org/>
51. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017;14: 417–419. doi:10.1038/nmeth.4197
52. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 2010;11: R25. doi:10.1186/gb-2010-11-3-r25
53. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26: 139–140. doi:10.1093/bioinformatics/btp616
54. Blum M, Chang H-Y, Chuguransky S, Grego T, Kandasaamy S, Mitchell A, et al. The InterPro protein families and domains database: 20 years on. Nucleic Acids Res. 2021;49: D344–D354. doi:10.1093/nar/gkaa977
55. Löytynoja A, Goldman N. Uniting Alignments and Trees. Science. 2009;324: 1528–1529. doi:10.1126/science.1175949
56. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform. 2019;20: 1160–1166. doi:10.1093/bib/bbx108
57. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol. 2020;37: 1530–1534. doi:10.1093/molbev/msaa015
58. Andrew Rambaut. FigTree. 2018. Available: <http://tree.bio.ed.ac.uk/software/figtree/>
59. Webb B, Sali A. Comparative Protein Structure Modeling Using MODELLER. Curr Protoc Bioinforma. 2016;54: 5.6.1-5.6.37. doi:10.1002/cpbi.3

60. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr.* 1993;26: 283–291. doi:10.1107/S0021889892009944
61. Schrödinger. The PyMOL Molecular Graphics System. LLC; 2021.
62. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215: 403–410. doi:10.1016/S0022-2836(05)80360-2
63. Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Kosakovsky Pond SL. Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes. *Mol Biol Evol.* 2018;35: 773–777. doi:10.1093/molbev/msx335
64. Gao F, Chen C, Arab DA, Du Z, He Y, Ho SYW. EasyCodeML: A visual tool for analysis of selection using CodeML. *Ecol Evol.* 2019;9: 3891–3898. doi:10.1002/ece3.5015
65. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30: 2114–2120. doi:10.1093/bioinformatics/btu170
66. Magoc T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics.* 2011;27: 2957–2963. doi:10.1093/bioinformatics/btr507
67. Vasimuddin Md, Misra S, Li H, Aluru S. Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS). 2019. pp. 314–324. doi:10.1109/IPDPS.2019.00041
68. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25: 2078–2079. doi:10.1093/bioinformatics/btp352
69. Broad Institute. Picard Toolkit. 2019. Available: <https://broadinstitute.github.io/picard/>
70. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20: 1297–1303. doi:10.1101/gr.107524.110
71. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics.* 2011;27: 2156–2158. doi:10.1093/bioinformatics/btr330
72. Volohonsky G, Terenzi O, Soichot J, Naujoks DA, Nolan T, Windbichler N, et al. Tools for *Anopheles gambiae* Transgenesis. *G3 GenesGenomesGenetics.* 2015;5: 1151–1163. doi:10.1534/g3.115.016808
73. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci.* 2005;102: 15545–15550. doi:10.1073/pnas.0506580102
74. Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A. Fast gene set enrichment analysis. *Bioinformatics;* 2016 Jun. doi:10.1101/060012

788 Figure legends

789 **Figure 1. *loqs2* is a dsRBP paralog of *loqs* that originated in the ancestor of the *Aedes***
 790 ***Stegomyia* subgenus. (A)** Maximum likelihood phylogenetic tree constructed with the amino
 791 acid sequences of the dsRNA-binding proteins (dsRBPs) found in the genomes of *Ae. aegypti*
 792 (*Aae*), *Ae. albopictus* (*Aal*), *Culex quinquefasciatus* (*Cq*), *Anopheles gambiae* (*Ag*), *Anopheles*
 793 *coluzzii* Mali-NIH (*AcM*), *Anopheles coluzzii* Ngousso (*AcN*), *Lutzomyia longipalpis* Jacobina
 794 (*Li*), *Stomoxys calcitrans* USDA (*Sc*) and *Glossina fuscipes* IAEA (*Gf*). Branches
 795 corresponding to the orthologs of *drosha*, *mle*, *DIP1*, *Adar*, *Dicer-2*, *Dicer-1*, *pasha*, *Son*,
 796 *tRNA-dihydrouridine synthase*, *Staufen* and *R2D2* are collapsed to facilitate visualization. Tree
 797 is rooted at the midpoint for visualization purposes. Node values correspond to 1000 ultra-fast
 798 bootstrap iterations. A non-collapsed tree is available at **Supplementary Figure 1A. (B)**
 799 Synteny analysis of *loqs2* flanking genomic regions among *Ae. aegypti*, *C. quinquefasciatus*,
 800 *An. gambiae* and *G. palpalis*. **(C)** Schematic cladogram indicating the probable origin of *loqs2*.
 801 The putative appearance of *loqs2* was inferred from the identification of sequences matching
 802 *loqs2* coding sequence in high-throughput sequencing data publicly available and are illustrated
 803 in **Supplementary Figure 1B**. Phylogenetic relationships and molecular clock were extracted
 804 from Wilkerson et al. [23] and da Silva et al. [22].

805 **Figure 2. *loqs2* originated from two independent *loqs* dsRBD1 duplication events. (A)**
 806 Phylogenetic relationships among the *Ae. aegypti* and *Ae. albopictus* *Loqs2*, *Loqs*, and *R2D2*
 807 double-stranded RNA-binding domains (dsRBDs) *D1* and *D2*, and *Staufen C* (*ST*) domain.
 808 Phylogeny was inferred by maximum likelihood and tree is rooted at the midpoint for
 809 visualization purposes. Node values correspond to the result of 1000 ultra-fast bootstrap
 810 iterations. **(B)** Correlation matrix showing BLASTp alignment score and coverage results from
 811 the comparison of *Ae. aegypti* *Loqs* and *Loqs2* dsRBD amino acid sequences. **(C)** Heatmap
 812 showing the structural comparison between *Loqs* and *Loqs2* dsRBDs *D1* and *D2* models using
 813 DALI server. **(D)** putative model for *loqs2* origin.

814 **Figure 3. Relaxed positive selection shaped *loqs2* evolution towards neofunctionalization.**
 815 **(A)** Proposed model of the evolutionary path followed by *loqs* and *loqs2* after the duplication
 816 events. **(B)** Maximum likelihood phylogenetic tree constructed with the coding sequences of
 817 *loqs*, *loqs2* and *r2d2* orthologs from *Ae. aegypti* (*Aae*), *Ae. albopictus* (*Aal*), *C.*
 818 *quinquefasciatus* (*Cq*), *C. pipiens* (*Cp*), *An. gambiae* (*Ag*), *An. coluzzii* Ngousso (*AcN*), *L.*
 819 *longipalpis* Jacobina (*Li*), *S. calcitrans* USDA (*Sc*) and *G. fuscipes* IAEA (*Gf*). Tree is rooted
 820 at the midpoint for visualization purposes. Node values correspond to the result of 1000 ultra-
 821 fast bootstrap iterations. **(C)** Test for relaxation or intensification of selection performed using
 822 RELAX. **(D)** Test of selection using the branch-site-model implemented in CodeML and
 823 abSREL. **(E)** Divergence metrics and McDonald-Kreitman test (MKT) results from an *Ae.*
 824 *aegypti* population from a forest in Senegal compared to *Ae. mascarensis*. **(F)** Comparative
 825 scheme between *Ae. aegypti* *Loqs* and *Loqs2* proteins. Sites represented in red or blue
 826 correspond to common amino acids under selection among *Ae. aegypti* and *Ae. albopictus*.

Figure 4. Ectopic expression of *loqs2* in larval stages results in larval growth arrest. (A) Heatmap showing the tissue-specific expression of *loqs*, *loqs2* and *r2d2* among *Ae. aegypti*, *Ae. albopictus* and *An. gambiae*. Gene expression between tissues was used for hierarchical clustering. White boxes are indicative of no detected mRNA expression. **(B)** Cassettes for ectopic expression of *loqs2* under the control of the *Ae. aegypti* poly-ubiquitin (*PUB*) promoter or the baculovirus promoter OpIE2. Fluorescent markers under control of the promoters 3xP3 or *PUB* were used to drive eGFP or DsRed2 expression as transgenesis markers as indicated. Images are representative of larvae from each transgenic line. **(C)** Wild-type (non-transgenic) sibling larvae reared at the same water container show normal development until pupal stage while transgenic larvae were halted in L2-L3 stages of development at day 5 post-hatching. **(D)** Stacked bar plots comparing development of wild-type and transgenic mosquitoes ectopically expressing *loqs2* under control of the baculovirus promoter OpIE2. **(E)** Gene set enrichment analysis (GSEA) of larvae ectopically expressing *loqs2* compared to wild-type siblings showed a major transcriptional shutdown on the transcription of metabolic pathways in transgenic individuals.

843 **Supplementary data**

844 **Supplementary figure 1. Phylogenetic tree of the dsRNA binding proteins found in the**
845 **genomes of different mosquitoes and flies, and mapping coverage of RNA-seq and whole-**
846 **genome-seq data from different *Aedini* species mapped to the *Ae. aegypti* and *Ae.***
847 ***albopictus* reference genomes. (A)** Terminals correspond to the VectorBase accession
848 numbers. Tree was rooted at the midpoint for visualization purposes. Node values correspond
849 to the result of 1000 ultra-fast bootstrap iterations. **(B)** *loqs* and *loqs2* exons are colored in
850 purple and blue respectively. Black bars represent the read coverage along gene sequence.
851 Lines represent intronic regions of both gene sequences and reads.

852 **Supplementary figure 2. Rate of nonsynonymous divergence (*Ka*) to the proportion of**
853 **neutral fixations (*Ks*) (ω) among *loqs*, *loqs2*, *Dicer-1* and *Dicer-2* genes, and complete**
854 **comparative scheme between *Ae. aegypti* and *Ae. albopictus* Loqs and Loqs2 proteins. (A)**
855 Analysis was performed comparing a Senegal Forest *Ae. aegypti* population and *Ae.*
856 *mascarensis*. **(B)** Solid lines indicate sites shared between species, as shown in **Figure 4B**,
857 while specific sites are indicated with dotted lines.

858 **Supplementary figure 3. Ectopic expression of *loqs2* in larval stages lead to major**
859 **metabolic shutdown. (A)** Expression levels of *loqs2* in transgenic or wild-type sibling larvae
860 quantified by RT-qPCR. **(B)** GSEA analysis plot showing extended gene sets prior redundancy
861 removal relative to **Figure 4E**.

862 **Supplementary table 1. SRA identifiers of the libraries used in this study. Excel**
863 spreadsheet file.

864

Figure 1

Figure 2

Figure 3

Figure 4

Supplementary figure 1

(A)

(B)

Supplementary figure 2

(A)

(B)

Supplementary figure 3