Interactions flavivirus-moustiques : diversité et transmission

Sébastian Lequime

UNIVERSITÉ PIERRE ET MARIE CURIE

ECOLE DOCTORALE 515 « Complexité du Vivant »
Groupe à 5 ans – Interactions Virus-Insectes

Interactions flavivirus – moustiques : diversité et transmission

par

Sébastian LEQUIME

THÈSE DE DOCTORAT
en EVOLUTION VIRALE

Présentée et soutenue publiquement le : 21 Juin 2016

Devant un jury composé de :

M. Samuel ALIZON, Chargé de Recherche	Rapporteur
M. Jacob KOELLA, Professeur	Rapporteur
M. Dominique HIGUET, Professeur	Examinateur
Mme Anna-Bella FAILLOUX, Directeur de Recherche	Examinatrice
M. Serafín GUTIERREZ, Chargé de Recherche	Examinateur
M. Louis LAMBRECHTS, Chargé de Recherche	Directeur de thèse

« *Je sais qu'au point où en est arrivée aujourd'hui la microbiologie, tout nouveau grand pas en avant sera une affaire des plus pénibles et que l'on aura beaucoup de mécomptes et de déceptions.* »

– Alexandre YERSIN, 28 août 1891.

Résumé

Les infections humaines dues aux virus du genre *Flavivirus* constituent depuis longtemps un problème de santé publique majeur à travers le monde, en particulier dans les zones à climat tropical. Ces virus à ARN sont des arbovirus qui infectent alternativement un hôte vertébré et un arthropode « vecteur », dont majoritairement des moustiques de la sous-famille des Culicinae. D'autres flavivirus, en revanche, sont incapables d'infecter les cellules de vertébrés et sont qualifiés de flavivirus spécifiques d'insectes (FSI). L'interaction entre les vecteurs et les flavivirus est centrale dans leur biologie, par l'influence qu'elle a sur leur diversité génétique, leur évolution et leur transmission. Cependant, après plus d'un siècle de recherches scientifiques, certains points de ces aspects fondamentaux restent méconnus, malgré une abondance accrue de données. Les approches basées sur les « mégadonnées » (*big data*) ont été au cœur du travail de cette thèse, qu'elles aient été générées par des technologies modernes ou par compilation de travaux plus anciens.

Dans une première partie, nous avons exploré des génomes de moustiques anophèles disponibles dans les bases de données publiques, à la recherche de traces d'éléments viraux endogènes (EVEs) d'origine flavivirale. Nous avons réussi à identifier *in silico*, puis à confirmer *in vivo*, la présence d'EVEs proches des FSI exogènes, chez les espèces *Anopheles sinensis* et *An. minimus*. Ces résultats suggèrent l'existence de FSI chez les anophèles, habituellement non associés aux , et mettent en lumière la diversité du genre *Flavivirus*, loin d'être restreinte aux seuls arbovirus.

Dans une deuxième partie, nous avons généré et analysé un jeu de données complexe basé sur le séquençage haut-débit d'un flavivirus à l'intérieur de son vecteur. Cette étude nous a permis d'explorer la fine interaction entre le génotype du moustique *Aedes aegypti* et la diversité virale intra-hôte du virus de la dengue-1. En effet,

comme tous les virus à ARN, les flavivirus existent sous la forme d'une population de variants génétiques apparentés, considérée comme critique pour leur *fitness* et leur potentiel adaptatif. Nos résultats ont mis en évidence : (i) un fort effet de la dérive génétique liée à un goulot d'étranglement démographique lors de l'infection initiale du tube digestif, diminuant l'importance relative de la sélection naturelle, et (ii) une modulation de la diversité génétique intra-hôte du virus par le génotype du moustique, indiquant que la diversité génétique du virus, et donc sa *fitness* et son évolution, est inextricablement liée à la variation génétique de l'hôte.

Enfin, nous avons compilé de manière systématique la riche littérature disponible sur la transmission verticale des arbovirus chez le vecteur moustique, c'est-à-dire de la femelle infectée à sa descendance, afin d'identifier des facteurs techniques, environnementaux, taxonomiques et physiologiques sous-jacents. Nos résultats, étayés par une analyse statistique robuste, éclairent d'un jour nouveau ce mode de transmission complémentaire à la transmission horizontale, entre le vecteur et l'hôte vertébré. Ils permettent d'affiner notre compréhension des stratégies employées par les arbovirus pour persister dans leur environnement, tout en fournissant des hypothèses testables sur les processus biologiques impliqués.

Collectivement, nos résultats ont mis en évidence de nouveaux aspects de la complexité des relations entre flavivirus et moustiques. Au-delà, ils soulignent l'intérêt d'étudier les « mégadonnées » générées au cours de plus d'un siècle de recherche, qu'elles soient issues de l'accumulation historique d'études descriptives ou produites par les technologies récentes de séquençage haut-débit. Ces analyses inscrivent résolument l'étude du système flavivirus-moustique dans l'ère du *big data*.

Abstract

Human infections caused by members of the *Flavivirus* genus have long been a major public health burden worldwide, especially in tropical climates. These RNA viruses are arboviruses that infect alternatively vertebrate hosts and arthropod 'vectors', mainly mosquitoes in the Culicinae subfamily. Other flaviviruses, however, are unable to infect vertebrate cells and are accordingly called insect-specific flaviviruses (ISFs). The interaction between vectors and flaviviruses is key to understanding their biology, because it influences their genetic diversity, evolution and transmission. Despite more than a century of scientific research and an ever-increasing amount of data, some fundamental aspects are still poorly understood. Strategies based on 'big data' have been at the heart of the work presented in this dissertation, both by taking advantage of modern technologies and by compiling older literature.

In the first part, we explored publicly available Anopheles mosquito genomes to discover putative endogenous viral elements (EVEs) of flaviviral origin. We identified *in silico* and confirmed *in vivo* one EVE in the genome of *Anopheles minimus* and another in the genome of *Anopheles sinensis*. Both EVEs were related to exogenous ISFs. These results support the existence of ISFs in Anopheles mosquitoes, which are not usually associated with flaviviruses, and highlight the diversity of the *Flavivirus* genus, far from being restricted to arboviruses.

In the second part, we generated and analyzed a complex dataset based on deep sequencing of a flavivirus within its vector. This study allowed us to investigate the fine-tuned interaction between genotypes of the mosquito *Aedes aegypti* and the intra-host diversity of dengue virus 1. Like other RNA viruses, flaviviruses consist of a population of related genetic variants, which is considered critical for their fitness

and adaptive potential. Our results showed : (i) a strong effect of genetic drift due to a population bottleneck during initial infection of the digestive tract, reducing the relative importance of natural selection, and (ii) a modulation of the intra-host viral genetic diversity by the mosquito genotype, indicating that viral genetic diversity, and therefore virus fitness and evolution, is inextricably linked to host genetic variation.

In the last part, we conducted a systematic review of the abundant literature on arbovirus vertical transmission in the mosquito vector, i.e. from an infected female to her offspring, in order to identify technical, environmental, taxonomic and physiological predictors. Our results, based on a robust statistical framework, shed new light on this transmission mode, complementary to horizontal transmission between the vector and the vertebrate host. They contribute to refining our understanding of strategies employed by arboviruses to persist in their environment, while providing testable hypotheses about the underlying biological processes.

Collectively, our results highlight new aspects of the complex relationships between flaviviruses and mosquitoes. Moreover, they underline the new insights brought by the 'big data' compiled during more than one century of research, either extracted from historical, descriptive studies or generated by modern, high-throughput sequencing technologies. Such analyses clearly include the study of flavivirus-mosquito systems in the era of 'big data'.

Remerciements

Mes premiers remerciements vont à Louis LAMBRECHTS. Merci d'avoir accepté de diriger ma thèse. J'ai beaucoup appris à ton contact, que ce soit en biologie évolutive/écologie, en statistiques, ou sur la vie scientifique en général. Merci également de m'avoir fait confiance et laissé évoluer à mon propre rythme. C'est grâce à ton encadrement que je me sens prêt aujourd'hui à voler de mes propres ailes et qui sait, à peut-être envisager sérieusement une carrière de recherche académique.

Je souhaite sincèrement remercier les rapporteurs, Samuel ALIZON et Jacob KOELLA, ainsi que les autres membres du jury, Anna-Bella FAILLOUX, Dominique HIGUET et Serafín GUTIERREZ, d'avoir accepté d'évaluer dans le détail mon travail.

Merci également à tous les membres de l'IVI, passés et présents. Plus que des collègues, ce sont des amis très chers que j'ai trouvés au 4ème étage du Centre François Jacob. Ces trois années sont passées trop vite, tant votre dynamisme, l'entraide et la bonne ambiance que nous avons instauré m'ont permis d'affronter sereinement cette thèse :

- Catherine, je ne sais vraiment pas ce que nous ferions sans ton aide au quotidien ;
- Albin, pour avoir supporté tant bien que mal mes incessantes allées et venues dans ton bureau, et pour avoir piqué en moi ce petit désir de compétition qui m'a fait apprendre à utiliser couramment R, les stats, la bioinfo et m'a fait tant peaufiner mes figures ;
- Vincent, pour nos discussions scientifiques ou non, et pour avoir accepté tant

de fois que je me casse le dos sur ton canapé convertible après de mémorables soirées IVIstes ;

- Isabelle, pour ton aide, ta bonne humeur et ta disponibilité lorsque tu m'appris quelques sagesses et « trucs » de virologiste, mais aussi pour m'avoir révélé que secrétaire particulier pouvait aussi être une option de carrière pour moi ;
- Davy, sans les palpitantes histoires et intrigues de ta vie personnelle, tout à fait dignes d'un transfuge IMIste, les journées auraient souvent été bien plus ennuyeuses ;
- Laura, for your kindness, your crazyness and for sharing your weird Paris experiences ;
- Sarah, pour ta sympathie, ta bonne humeur et ta disponibilité pour discuter de l'après thèse.

J'en profite pour remercier également les étudiants que j'ai encadrés au laboratoire : Clément, Elliott et Marine ; votre efficacité et votre curiosité scientifique ont toujours été des plus stimulantes !

Hors du cercle de l'IVI, j'ai interagi avec de nombreuses personnes à travers le campus de l'Institut Pasteur. La liste serait trop longue à dresser, tant j'ai rencontré de personnes sympathiques et disponibles. Sachez que je vous remercie profondément pour tout.

Mes proches également, méritent des remerciements :

- Mes parents, Patrice et Christine, dont le soutien toujours constant depuis le CP jusqu'à aujourd'hui (pas loin de 22 ans d'études tout de même) m'a permis d'affronter nombre de défis ;
- Mes grands-parents, Liliane, Jean-Marie, Edith et Josef, dont l'intérêt et le soutien dans mes études m'ont toujours été très précieux ;
- Au reste de ma famille et belle-famille, et je réitère ce que j'ai déjà pu écrire : on ne peut vraiment pas en rêver de meilleure !

- À mes amis Chimène et Gaël, parce que maintenant, j'aurais une bonne excuse pour chanter à tue-tête un certain tube de Robert Palmer. À Germain, Nicolas, Julien, Olga, Ivan et Basile pour votre sympathie, votre intérêt dans mes travaux, et surtout les précieux moments de détente.

- À Elsa, sans ta présence, ton aide, ton soutien et ton amour au quotidien, je suis persuadé que ces trois années ne se seraient pas aussi bien déroulées. Je suis plus qu'heureux de savoir que tout ça n'est certainement pas prêt de s'arrêter et que de nombreuses autres aventures nous attendent.

Paris, 31 Mai 2016 S. LEQUIME

Table des matières

Résumé **v**

Abstract **vii**

Remerciements **ix**

Liste des figures **xiv**

I Synthèse **1**

1 Introduction générale 3

2 Diversité des flavivirus 7

2.1 Structure et organisation génomique 7

2.2 Cycle de réplication 9

2.3 Diversité phylogénétique 10

2.4 L'apport des éléments endogènes d'origine flavivirale dans les génomes de moustiques 13

3 Le moustique comme siège de microévolution des flavivirus-arbovirus 16

3.1 Plasticité génétique des arbovirus 16

3.2 Infection arbovirale du moustique 19

3.3 Interactions flavivirus-moustiques 19

4 Transmission inter-hôte des flavivirus et d'autres arbovirus par les moustiques 21

4.1 Transmission des arbovirus 21

4.2 Maintien des arbovirus par transmission verticale chez le moustique 26

5 Conclusions 29

Bibliographie 36

II Annexes 49

A Discovery of flavivirus-derived endogenous viral elements in two *Anopheles* mosquito genomes supports the existence of *Anopheles*-associated insect-specific flaviviruses 51

B Genetic drift, purifying selection and vector genotype shape dengue virus intra-host genetic diversity in mosquitoes 69

C Vertical transmission of arboviruses in mosquitoes: a historical perspective 117

D Determinants of arbovirus vertical transmission in mosquitoes 151

Table des figures

Synthèse

1 Organisation génomique des flavivirus. 7
2 Cycle de réplication des flavivirus. 10
3 Arbre phylogénétique bayésien de la polyprotéine des flavivirus. 11
4 Structure et phylogénie des EVEs dérivés de flavivirus chez *Anopheles minimus* et *Anopheles sinensis*. 15
5 Cycle de la transmission horizontale d'un arbovirus par le moustique vecteur. 20
6 Estimation de la taille du goulot d'étranglement démographique lors de l'infection initiale du mésentéron et niveau de diversité intra-hôte du virus de la dengue-1. 22
7 Réseau de transmission hypothétique du virus du Nil occidental aux États-Unis. 24
8 Évolution des techniques de détection et de la mise en évidence de la transmission verticale des arbovirus chez les moustiques en fonction du temps dans les études expérimentales. 29

Synthèse Partie I

1 Introduction générale

Le terme « arbovirus » (pour *arthropod-borne virus*) décrit un groupe de virus d'animaux dont l'épidémiologie implique une transmission par un arthropode hématophage, appelé « vecteur ». Les moustiques sont de loin les principaux vecteurs d'arbovirus mais d'autres arthropodes hématophages peuvent les transmettre, comme les tiques, les phlébotomes ou les culicoïdes [Gubler, 2001].

Les arbovirus regroupent aujourd'hui plus de 530 virus décrits à travers le monde [Centers for Disease Control and Prevention, 2010] appartenant à plusieurs genres de virus à ARN[1]. Les plus importants en santé humaine sont les *Alphavirus* (famille des Togaviridae, incluant par exemple le virus chikungunya), les *Orthobunyavirus* et *Phlebovirus* (famille des Bunyaviridae, incluant respectivement le virus de l'encéphalite de Californie et le virus de la fièvre de la vallée du Rift) et les *Flavivirus* (famille des Flaviviridae). Ce dernier genre regroupe les principaux arbovirus posant un véritable problème de santé publique au niveau mondial, à l'exception du virus du chikungunya. Ils sont transmis majoritairement par des moustiques de la sous-famille des Culicinae (principalement des genres *Culex* et *Aedes*). Ainsi, on estime que chaque année les virus de la dengue infectent plus de 390 millions de personnes, causant 96 millions de cas symptomatiques [Bhatt *et al.*, 2013]. Le virus de la fièvre jaune cause chaque année 84 000 à 170 000 cas symptomatiques [Barrett et Higgs, 2007] et le virus de l'encéphalite japonaise, quant à lui, affecte annuellement 68 000 personnes [World Health Organization, 2015].

À l'heure où nous écrivons ces lignes, la plus importante épidémie de fièvre jaune depuis presque 30 ans touche l'Angola [Butler, 2016] et une épidémie de fièvre Zika bat son plein en Amérique du Sud, présentant des symptômes jusqu'alors inconnus et inquiétants pour la population résidant en zone d'épidémie. Ce virus a été découvert il y a presque 70 ans en Afrique [Dick *et al.*, 1952]. Longtemps marginale, son importance épidémiologique s'est largement accrue lors de la dernière décennie : épidémie en Micronésie en 2007 [Duffy *et al.*, 2009], dans plusieurs îles du Pacifique entre 2013 et 2015 [Musso *et al.*, 2015], puis le continent américain depuis 2015 [Fauci et Morens, 2016]. Au point qu'aujourd'hui, l'Organisation mondiale de la santé en a fait une « urgence de santé publique de portée internationale » [World Health Organi-

[1] À l'exception des Asfarviridae (un seul virus identifié : African swine fever virus) qui sont des virus à ADN.

zation, 2016]. Cette émergence spectaculaire n'est pas sans rappeler celles d'autres arbovirus comme le virus chikungunya dans l'océan Indien en 2005 [Pialoux *et al.*, 2007], ou, avant, celle d'un autre flavivirus, le virus du Nil occidental (*West Nile virus*) en Amérique du Nord en 1999 [Kilpatrick, 2011].

Ces évènements récents ne doivent pas faire oublier que les arbovirus, et en particulier les flavivirus, font peser un fardeau depuis longtemps sur les populations humaines. Certains auteurs ont vu dans la description par Plutarque de la mort d'Alexandre le Grand en 323 avant J.-C. des indices laissant supposer une fièvre du Nil occidental [Marr et Calisher, 2003] ; ou encore dans une liste de symptômes dressée dans une encyclopédie médicale chinoise de la dynastie Jin (265-420 après J.-C.) la première description de la dengue [Gubler, 1998]. C'est à partir de l'âge des grandes découvertes (XV^{e} – XVII^{e} siècle) puis pendant l'ère coloniale (XVII^{e} – XIX^{e} siècle) que les flavivirus deviennent un problème de santé publique. Les premières épidémies de fièvre jaune et de dengue sont décrites dans les principales cités portuaires du Nouveau Monde [Nogueira, 2009]. Ces pathogènes ont profité de la présence d'une population susceptible d'origine européenne et de l'extension géographique de leur vecteur en zone urbaine, le moustique *Aedes aegypti*, grâce à la traite négrière et au commerce reliant directement les zones tropicales d'Afrique à celles des Amériques [Brown *et al.*, 2014]. La cause et le mode de transmission de ces maladies étaient alors inconnus, faisant planer un certain mystère. Ce n'est qu'à la toute fin du XIX^{e} et au début du XX^{e} siècle que la transmission vectorielle, c'est-à-dire par un vecteur arthropode, est suggérée [Finlay, 1881] puis prouvée pour la fièvre jaune [Reed *et al.*, 1901].

Malgré cette longue histoire commune, il n'y a, à l'heure actuelle, aucun traitement thérapeutique étiologique pour les cas symptomatiques. Des vaccins ont été développés, parfois de longue date contre les virus de la fièvre jaune (en France, Stamaril®), de l'encéphalite japonaise (en France, Ixiaro®) et très récemment contre les virus de la dengue (Dengvaxia®, autorisé au Mexique, aux Philippines, au Brésil et au Salvador) [Sanofi Pasteur, 2016]. L'efficacité de ces vaccins est variable, allant de près de 85% de protection pour le vaccin contre la fièvre jaune [Staples *et al.*, 2010] à 45-66% en fonction de la tranche d'âge pour le vaccin contre les virus de la dengue [Hadinegoro *et al.*, 2015].

En l'absence de traitement efficace et face à une prévention vaccinale limitée à

quelques virus, et qui plus est d'efficacité variable, la lutte anti-vectorielle reste le moyen de prévention le plus efficace pour contrer les infections humaines à flavivirus. Diverses méthodes peuvent être utilisées, seules ou simultanément : gestion de l'environnement afin d'éliminer/réduire les sites larvaires ou éliminer/limiter le contact avec les vecteurs adultes (approche indirecte), utilisation d'agents chimiques larvicides ou adulticides et enfin utilisation d'agents biologiques (approche directe) [World Health Organization, 2009]. Malgré quelques succès, l'efficacité de ces méthodes sur le long terme est régulièrement remise en question [Bowman *et al.*, 2016]. Contrairement à l'approche vaccinale cependant, les méthodes de lutte anti-vectorielle ont l'avantage d'être relativement simples et rapides à mettre en œuvre et ont un impact sur tous les pathogènes pouvant être transmis par les moustiques visés, en particulier *Aedes aegypti* et *Aedes albopictus* dont l'écologie est similaire. Cependant, elles requièrent, pour maximiser leur efficacité, une bonne connaissance de l'écologie, du comportement de l'espèce visée et des interactions entre le pathogène et son vecteur, en plus des contextes culturel, économique et politique local.

Depuis plus d'un siècle maintenant, la recherche scientifique a tenté de comprendre les interactions extrêmement complexes entre les flavivirus arboviraux, leurs hôtes vertébrés, leurs vecteurs arthropodes, le microbiote bactérien, viral et fongique, ainsi que l'environnement dans lequel tous évoluent. La complexité du système implique des approches multidisciplinaires variées allant de la biologie moléculaire à la modélisation mathématique des épidémies, en passant entre autres par l'analyse d'imagerie satellite, la génomique, l'écologie ou l'immunologie. Tirant profit des avancées technologiques et conceptuelles dans les méthodes de détection et d'analyse, la quantité de données générées va croissant. En réponse, ces deux dernières décennies ont vu l'émergence, comme dans d'autres domaines des sciences de la vie, d'une ère des « mégadonnées » (*big data*).

Les études que j'ai menées dans le cadre de ce travail de thèse s'inscrivent clairement dans cette démarche, incontournable aujourd'hui. En utilisant des outils et approches classiquement attribués aux analyses de big data, nous avons tenté d'apporter des éléments de réponse à divers points non éclaircis de l'interaction entre flavivirus et moustiques. Entre l'analyse de données générées par les technologies récentes de séquençage haut-débit ou l'accumulation historique d'études descriptives, les résultats ce de travail ont permis d'apporter des éléments de réponse originaux et

convaincants aussi bien à des questions d'actualité qu'à de « vieilles marottes » qui taraudent la communauté scientifique s'intéressant aux flavivirus et aux arbovirus en général.

La synthèse qui suit ces lignes présente, dans leur contexte bibliographique, les études issues de cette thèse, elle-même divisée en trois parties principales. Dans une première partie, nous nous sommes intéressés à la diversité des flavivirus. Regroupés sous cette dénomination, on retrouve de nombreux virus infectant un large éventail d'hôtes vertébrés et invertébrés, en particulier des moustiques. Alors que les moustiques des genres *Aedes* et *Culex* figurent en bonne place comme hôtes et vecteurs de flavivirus, les moustiques du genre *Anopheles* ne leur sont qu'anecdotiquement associés. En analysant les génomes publiés de plusieurs espèces d'anophèles, nous avons réussi à mettre en évidence l'existence d'éléments viraux endogènes (EVEs) d'origine flavivirale, dont nous avons confirmé l'existence in vivo. Ces EVEs, proches mais divergents des flavivirus spécifiques d'insectes (FSI) exogènes, suggèrent l'existence d'un clade de FSI associé aux anophèles et mettent en lumière la diversité du genre, loin d'être restreinte aux seuls arbovirus.

Dans une deuxième partie, nous avons continué à nous intéresser à la diversité des flavivirus, mais cette fois plus spécifiquement à l'échelle intra-hôte. En générant puis en analysant un jeu de données complexe basé sur le séquençage haut-débit du génome complet du virus de la dengue-1, nous avons exploré la fine interaction entre le génotype du moustique *Aedes aegypti* et la diversité génétique virale. Nos résultats ont confirmé les forts effets de la dérive génétique et de la sélection purifiante déjà suspectés par d'autres études. Au-delà, nous avons pu mettre en évidence une modulation de la diversité génétique intra-hôte du virus par le génotype du moustique. Ce résultat indique clairement que la diversité génétique du virus et donc sa *fitness* et son évolution sont inextricablement liées à la variation génétique de l'hôte.

Finalement, dans une troisième partie, nous nous sommes penchés sur un concept qui date du tout début du XX^{e} siècle : la transmission verticale (TV) des arbovirus chez les moustiques. En effet, contrairement aux deux précédentes questions, la TV est un sujet soulevé de longue date dont l'importance épidémiologique est abondamment discutée dans la littérature. Les conclusions des différentes études, inconstantes voire parfois contradictoires, n'ont pas permis à leur échelle de trancher sur l'importance réelle ou les caractéristiques de la TV. Nous avons donc pris le parti de synthétiser

la littérature au sein de bases de données et d'analyser aussi bien les tendances historiques que les facteurs techniques et biologiques influant sur la TV. Les résultats permettent d'affiner notre compréhension des stratégies employées par les arbovirus pour persister dans leur environnement, tout en fournissant des hypothèses testables sur les processus biologiques sous-jacents.

2 Diversité des flavivirus

2.1 Structure et organisation génomique

Les flavivirus sont des virus enveloppés à ARN simple brin de polarité positive d'environ 11 kb. Les virions matures sont sphériques, d'un diamètre compris entre 40 et 70 nm [Lindenbach *et al.*, 2013]. Le génome ne présente qu'un seul cadre ouvert de lecture codant pour une polyprotéine. Cette dernière est clivée pendant ou après la traduction par des protéases virales (protéine non-structurale 3) ou cellulaires en 3 protéines structurales et 5 (ou 7 en comptant les subdivisions) protéines non-structurales (Figure 1).

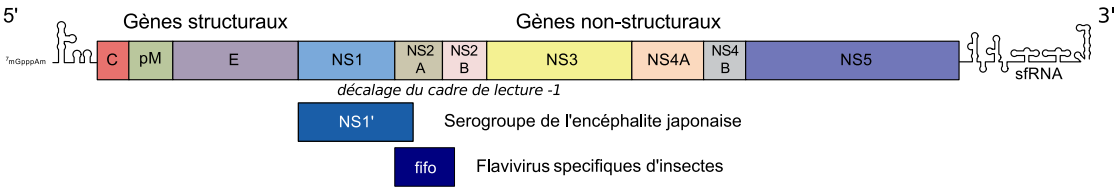

FIGURE 1 – Organisation génomique des flavivirus. C : protéine de capside ; E : glycoprotéine d'enveloppe ; M : glycoprotéine de membrane ; NS1 : glycoprotéine non-structurale 1 ; NS2A : protéine non-structurale 2A ; NS2B : protéine non-structurale 2B ; NS3 : protéine non-structurale 3 (protéase/hélicase) ; NS4A : protéine non-structurale 4A ; NS4B : protéine non-structurale 4B ; NS5 : protéine non-structurale 5 (*RdRP, RNA-dependent-RNA polymerase*) ; sfRNA : ARN subgénomique flaviviral (*subgenomic flavivirus RNA*) (figure d'après [Lindenbach *et al.*, 2013]).

La protéine de capside (C) protège et empaquète l'ARN viral dans une capside à symétrie icosaédrique. Le précurseur de la glycoprotéine de membrane (prM) protège la protéine d'enveloppe d'une fusion prématurée pendant le transit du virus dans les voies de sécrétions de la cellule. C'est l'exposition au pH acide du trans-Golgi qui, en dévoilant un site de clivage, permet la maturation du virion lorsque prM est clivé en M et pr, ce dernier étant libéré dans l'espace extracellulaire. La glycoprotéine d'enveloppe

(E) est une protéine de fusion de classe II, impliquée dans la reconnaissance du ou des récepteur(s) et la fusion membranaire [Lindenbach *et al.*, 2013].

La glycoprotéine non-structurale 1 (NS1) est impliquée dans la réplication de l'ARN et la production des particules infectieuses, en interaction avec NS4A. Un décalage du cadre de lecture -1 au cours de la traduction permet la production d'une protéine NS1 allongée, NS1', chez les virus du sérogroupe de l'encéphalite japonaise. NS1' a été impliquée dans l'invasion du tissu nerveux par les virus de ce groupe [Melian *et al.*, 2010]. La protéine non-structurale 2 est clivée en NS2A et NS2B par la protéase virale NS3. La protéine NS2A a été impliquée dans l'inhibition de l'interféron chez les vertébrés mais aussi dans l'assemblage du virion. NS2B, quant à elle, est un cofacteur de la protéine NS3 avec qui elle forme un complexe stable, permettant de l'ancrer à la membrane cellulaire [Lindenbach *et al.*, 2013]. Là aussi, un décalage du cadre de lecture -1 au cours de la traduction permet la production d'une protéine NS2A/B alternative appelée *fifo* (*fairly interesting flavivirus ORF*) chez les FSI, dont le rôle est actuellement inconnu [Firth *et al.*, 2010]. La protéine non-structurale 3 (NS3) est une protéine multifonction : à l'extrémité N-terminale, une protéase à sérine permet le clivage de NS2A/B, NS2B/NS3, NS3/NS4A et NS4B/NS5 ; à l'extrémité C-terminale, une ARN hélicase-nucléoside triphosphatase (NTPase) permet de dérouler le génome, étape essentielle pour la réplication. Elle possède également une activité ARN triphosphatase (RTPase) permettant de déphosphoryler l'extrémité 5' du génome avant l'addition d'une coiffe. La protéine non-structurale 4 est clivée en NS4A et NS4B par une protéase cellulaire. NS4A est impliquée dans la réplication de l'ARN grâce à son interaction avec NS1. NS4B semble également impliquée dans la réplication et a été associée à un blocage de l'interféron de type I chez certains virus. La protéine non-structurale 5 (NS5) est elle aussi une protéine multifonction : à l'extrémité N-terminale la formation de la coiffe de l'ARN (*RNA capping*) ; en à l'extrémité C-terminale l'activité ARN-polymérase-ARN-dépendante (*RNA-dependant-RNA-polymerase* ; RdRP) permettant la copie de l'ARN génomique [Lindenbach *et al.*, 2013].

Le cadre ouvert de lecture des flavivirus est flanqué de deux séquences non codantes. À l'extrémité 5', la séquence d'environ 100 nucléotides est impliquée dans la traduction et la réplication du génome et possède des structures secondaires communes entre flavivirus [Gritsun et Gould, 2007]. À l'extrémité 3', la séquence est de taille variable, entre 400 et 700 nucléotides. Sa séquence et son organisation diffèrent

entre FSI, flavivirus transmis par les tiques ou les moustiques ou flavivirus spécifiques de vertébrés [Villordo *et al.*, 2015]. Une boucle finale 3'SL est cependant systématiquement retrouvée et impliquée dans la traduction et l'interaction avec NS2A, NS3 et NS5 [Lindenbach *et al.*, 2013]. En amont, la présence de pseudo-nœuds confère une résistance à l'action d'une exoribonucléase cellulaire (XRN-1) impliquée dans la dégradation des ARN messagers, ce qui induit la production de plusieurs ARN subgénomiques appelés sfRNA (*subgenomic flavivirus RNA*) [Chapman *et al.*, 2014]. Le rôle des sfRNA dans l'échappement aux systèmes immunitaires des hôtes vertébrés [Manokaran *et al.*, 2015] et invertébrés [Moon *et al.*, 2015] est actuellement activement exploré [Roby *et al.*, 2014, Kieft *et al.*, 2015, Clarke *et al.*, 2015].

2.2 Cycle de réplication

Le cycle de réplication des flavivirus présente les mêmes caractéristiques, qu'il prenne place dans une cellule de vertébré ou d'arthropode. Plusieurs protéines candidates au rôle de récepteurs ou corécepteurs, variant selon le type cellulaire ou l'hôte, ont été identifiées [Mukhopadhyay *et al.*, 2005]. Après sa fixation, le virus est internalisé par endocytose. Le pH acide de la voie endosomale entraîne des changements dans la structure du virion qui déclenchent la fusion de l'enveloppe virale à la membrane de l'endosome, permettant la libération de la capside. L'ARN génomique, une fois libéré suite à la décapsidation, remplit trois rôles : (i) ARN messager pour la traduction de la polyprotéine virale, (ii) matrice pour la réplication et (iii) matériel génomique encapsidé pour la formation de nouvelles particules virales (Figure 2) [Mukhopadhyay *et al.*, 2005, Lindenbach *et al.*, 2013].

La réplication du génome viral a lieu dans des structures vésiculaires liées aux membranes du réticulum endoplasmique, reliées au cytoplasme par des pores de taille réduite [Welsch *et al.*, 2009], où se retrouvent entre autres les protéines non-structurales du virus, en particulier NS5, la RdRP. La réplication débute par la synthèse du brin complémentaire ARN de polarité négative (ARN-) de longueur génomique sans coiffe, servant de matrice à la production de nouveaux ARN génomiques (ARN+). Un large excès (environ 10 fois plus) de ces derniers est observé vis-à-vis des brins d'ARN- dans les cellules infectées [Lindenbach *et al.*, 2013] ; plusieurs brins d'ARN+ étant synthétisés simultanément sur le même brin matriciel ARN- [Selisko *et al.*, 2014].

Finalement, l'assemblage du virus a lieu à la surface du réticulum endoplasmique,

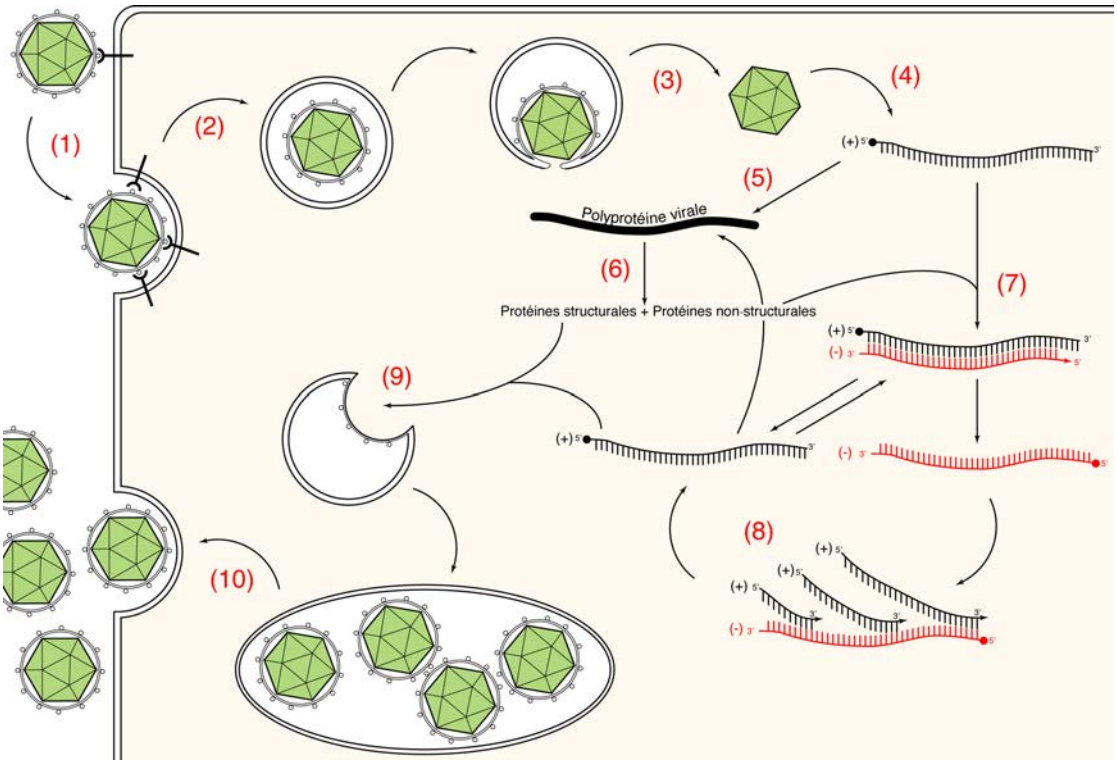

FIGURE 2 – Cycle de réplication des flavivirus. (1) Fixation du virion à son récepteur. (2) Endocytose puis (3) fusion et libération de la capside dans le cytoplasme. (4) Décapsidation. (5) Traduction du précurseur polyprotéique qui est (6) clivé en protéines structurales et non structurales. (7) Synthèse du brin complémentaire de polarité (-) servant de matrice (8) à la synthèse de nouveaux brins (+). (9) Assemblage du virion puis (10) libération. (figure d'après [Lindenbach *et al.*, 2013]).

lors d'un bourgeonnement dans le lumen. Les particules immatures, non infectieuses, transitent alors par les voies de sécrétions de l'appareil de Golgi, dont le pH graduellement acide permet la maturation de la particule par clivage de la glycoprotéine d'enveloppe. Les virions matures sont alors libérés par exocytose dans le compartiment extracellulaire [Mukhopadhyay *et al.*, 2005, Lindenbach *et al.*, 2013].

2.3 Diversité phylogénétique

Les flavivirus font partie de la famille des Flaviviridae dont ils partagent les caractères généraux, que ce soit au niveau de leur morphologie, de la taille du génome (entre 9,5 et 12,5 kb) et de l'organisation génomique, ou de leur stratégie de réplication. Quatre genres viraux ont été décrits et sont actuellement reconnus : les *Flavivirus*, les *Hepacivirus* (ex : virus de l'hépatite C), les *Pestivirus* et les *Pegivirus* [Lindenbach *et al.*,

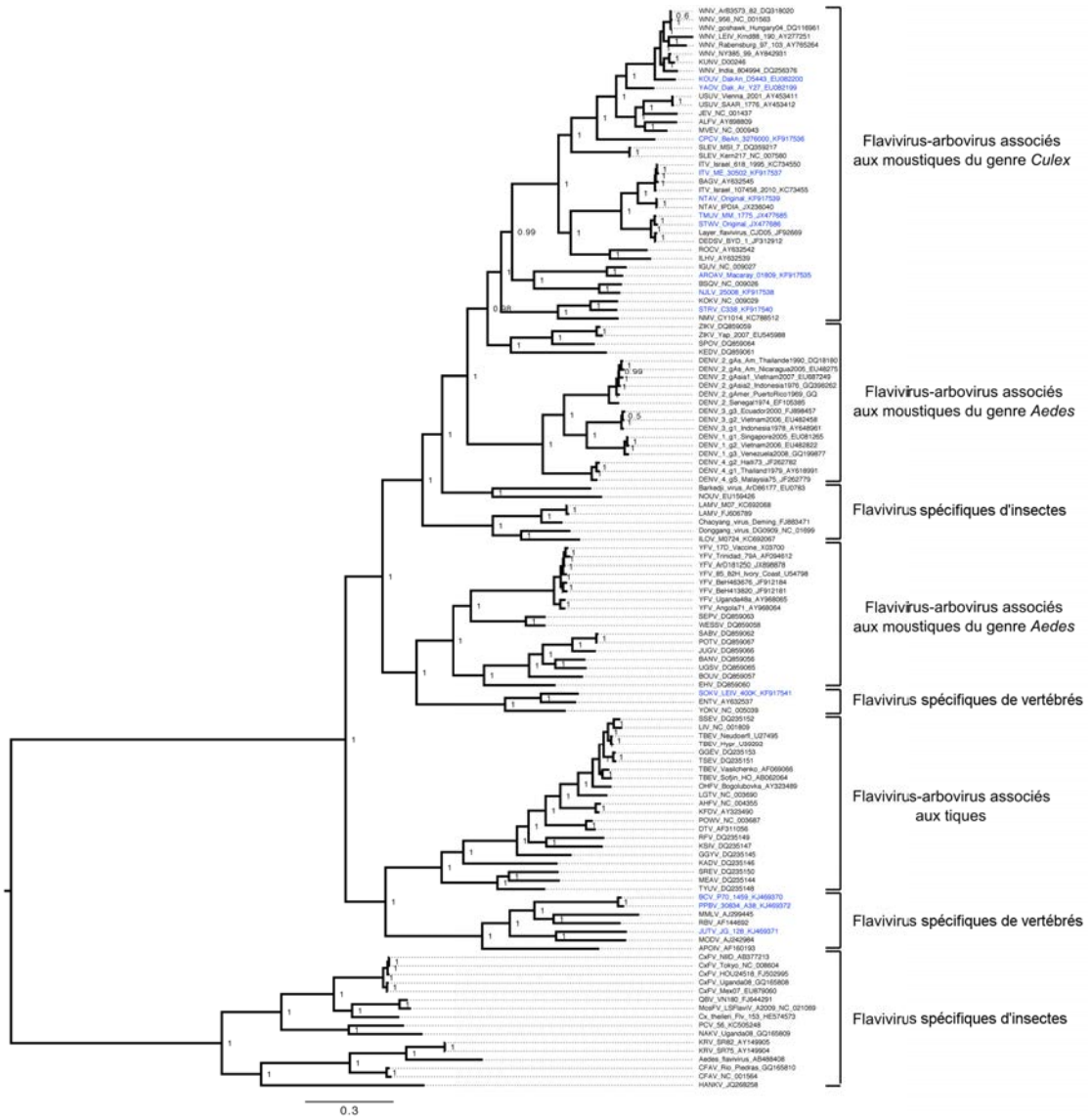

FIGURE 3 – Arbre phylogénétique bayésien de la polyprotéine des flavivirus (figure adaptée de [Moureau *et al.*, 2015]).

2013]. De nombreux nouveaux virus classés parmi les Flaviviridae ont été découverts récemment grâce aux technologies de séquençage haut-débit appliquées à des organismes non-modèles. Si certains étoffent les rameaux déjà connus de la phylogénie de cette famille (comme pour les flavivirus), d'autres en ajoutent de nouveaux qui contestent les caractères d'identification des Flaviviridae ou qui apportent de surprenantes considérations sur l'évolution des virus en général. Ainsi, l'éventail de taille des génomes s'est vu considérablement allongé, avec la découverte de virus de 16 à 24 kb dans des hôtes aussi variés que des plantes, des nématodes, des arthropodes ou

des vertébrés [Kobayashi *et al.*, 2013, Bekal *et al.*, 2014, Shi *et al.*, 2015, Teixeira *et al.*, 2016]. Plus surprenante encore a été la découverte de virus segmentés chez des tiques dont deux segments sont similaires aux gènes codant pour NS3 (protéase/hélicase) et NS5 (RdRP) des Flaviviridae, les deux autres codant pour des protéines structurelles jusqu'alors inconnues [Qin *et al.*, 2014, Shi *et al.*, 2015].

Le genre *Flavivirus* lui-même n'est pas épargné par la découverte de nombreux nouveaux virus, en particulier dans certains clades, comme les FSI [Calzolari *et al.*, 2015, Blitvich et Firth, 2015]. Il inclut à l'heure actuelle plus de 50 virus taxonomiquement reconnus [Moureau *et al.*, 2015] mais on estime qu'il contiendrait plus de 2 000 taxons [Pybus *et al.*, 2002]. Les analyses phylogénétiques distinguent à l'heure actuelle 4 grands groupes de flavivirus, en lien avec leur spectre d'hôte : les flavivirus-arbovirus associés aux moustiques (FAM), les flavivirus-arbovirus associés aux tiques (FAT), les flavivirus sans vecteur connus ou spécifiques de vertébrés (FSV) et les flavivirus spécifiques d'insectes (FSI) (Figure 3) [Moureau *et al.*, 2015].

Les FAM regroupent la grande majorité des flavivirus les plus étudiés ; ils comptent en effet parmi eux les arbovirus ayant le plus fort impact en santé publique. Deux grands clades sont décrits, liés au genre de moustiques les transmettant : ceux transmis par des *Culex* (dont les virus du Nil occidental, de l'encéphalite japonaise ou de l'encéphalite de Saint Louis) et ceux transmis par des *Aedes* (incluant les virus Zika, de la dengue et de la fièvre jaune). Plusieurs virus en leur sein, décrits récemment, sont considérés à l'heure actuelle comme spécifiques d'insectes. Ces virus ont été isolés chez des moustiques, principalement des *Aedes* ou genres proches, et ont été incapables d'infecter des cellules de vertébrés [Moureau *et al.*, 2015].

Les FAT forment un clade homogène dont la principale caractéristique est de regrouper des arbovirus transmis par des tiques. Là aussi, deux clades sont classiquement décrits, liés cette fois à leur hôte vertébré : les virus infectant les mammifères (dont les virus de l'encéphalite à tique et de l'encéphalomyélite de Powassan) et les virus infectant les oiseaux marins (incluant les virus Tyuleniy ou Meaban) [Grard *et al.*, 2007].

Les FSV sont des virus spécifiques de vertébrés incapables d'infecter des cellules d'arthropodes. Ce groupe est mal connu et peu décrit, et regroupe plusieurs virus isolés chez des rongeurs (virus Modoc) ou des chauves-souris (virus de la leucoencéphalite des myotis du Montana) [Moureau *et al.*, 2015]. Un autre clade, proche des

FAM, a été associé aux FSV. Ce dernier inclut des virus de chauve-souris (virus de chauve-souris d'Entebbe, ou virus Yokose par exemple) qui contrairement au groupe précédent peuvent se répliquer dans des cellules de moustique [Varelas-Wesley et Calisher, 1982].

Les FSI au sens strict quant à eux forment un clade lui aussi plutôt homogène et divergent au sein du genre *Flavivirus*. Ces virus sont incapables d'infecter des cellules vertébrées. Le premier FSI ayant été identifié, le *cell fusing agent virus* (CFAV), a été découvert dans les années 1970, grâce à l'effet cytopathogène qu'un surnageant de culture de cellules d'*Aedes aegypti* a eu sur une lignée cellulaire originaire d'*Aedes albopictus* [Stollar et Thomas, 1975]. Longtemps resté orphelin, il a été rejoint en 2003 par le Kamiti River virus (KRV) [Sang *et al.*, 2003], puis par de nombreux autres virus [Calzolari *et al.*, 2015, Blitvich et Firth, 2015]. On distingue à l'heure actuelle deux grands clades de FSI : les virus associés aux moustiques du genre *Aedes* et ceux associés aux moustiques du genre *Culex*. Un troisième clade, associé aux moustiques de la tribu des Mansoniini est parfois suggéré suite à la découverte de FSI dans des moustiques des genres *Mansonia* et *Coquillettidia* [Moureau *et al.*, 2015].

L'origine évolutive des flavivirus reste encore incertaine. Ils ont d'abord été suspectés d'être ancestralement des virus de vertébrés, à cause de leur lien phylogénétique avec les autres Flaviviridae, tous, à l'époque, trouvés chez des vertébrés et l'existence des FSV [Gould *et al.*, 2003]. La découverte de nombreux FSI, d'autres Flaviviridae chez des arthropodes [Conway, 2015, Teixeira *et al.*, 2016, Shi *et al.*, 2015] ainsi que de potentiels flavivirus chez les drosophiles [Webster *et al.*, 2015] ou des chironomes [Cook *et al.*, 2013], insectes non piqueurs et donc moins ou pas exposés aux virus de vertébrés, semblent au contraire suggérer qu'ils sont originellement des virus d'arthropodes [Gubler, 2014].

2.4 L'apport des éléments endogènes d'origine flavivirale dans les génomes de moustiques

Les éléments viraux endogènes (EVE) sont des insertions permanentes de tout ou partie du matériel génétique d'un virus dans le génome de son hôte. L'insertion (transitoire) fait partie du cycle viral des rétrovirus et par conséquent de nombreux EVEs identifiés sont d'origine rétrovirale. Plusieurs EVEs non-rétroviraux ont été identifiés dans les génomes de nombreuses espèces d'hôtes, dont des vertébrés et des arthro-

podes [Feschotte et Gilbert, 2012]. Le mode d'endogénisation n'est pour l'instant pas élucidé, même si une interaction avec des rétro-éléments du génome de l'hôte semble être privilégiée [Holmes, 2011]. La conservation d'une transcription en ARN de certains EVEs suggère que leur présence pourrait procurer un avantage sélectif à l'hôte. Ainsi, des EVEs exprimés pourraient être à l'origine d'une protection ou d'une tolérance face à des virus exogènes génétiquement proches [Flegel, 2009, Holmes, 2011, Bell-Sakyi et Attoui, 2013, Fujino *et al.*, 2014]. Des EVEs d'origine flavivirale, parfois toujours exprimés, ont ainsi été identifiés dans les génomes des moustiques *Aedes aegypti* [Crochu *et al.*, 2004, Katzourakis et Gifford, 2010] et *Aedes albopictus* [Roiz *et al.*, 2009, Chen *et al.*, 2015].

Quel que soit le mécanisme d'intégration ou le rôle que peuvent jouer les EVEs dans leur hôte, ce sont des témoins précieux dans l'étude de l'histoire évolutive des virus et leur possible spectre d'hôte. En effet, les EVEs sont très probablement le produit d'une interaction ancienne et intime entre le virus et son hôte. Ils nécessitent au moins deux évènements considérés comme rares : (i) une rétrotranscription et (ii) une intégration dans le génome des lignées germinales [Holmes, 2011, Aiewsakun et Katzourakis, 2015], ce qui explique la faible abondance d'EVEs non-rétroviraux dans les génomes eucaryotes.

Comme nous l'avons précédemment évoqué, parmi les plus de 3 500 espèces de moustiques (famille des Culicidae) [White, 2008], ceux de la sous-famille des Culicinae (principalement les genres *Aedes* et *Culex*) sont les principaux vecteurs d'arbovirus d'importance en santé publique. L'autre sous-famille des Anophelinae (principalement le genre *Anopheles*) est quant à elle généralement associée à la transmission des parasites du paludisme mais très rarement suspectée d'être impliquée dans la transmission de flavivirus, et d'arbovirus en général à l'exception du virus O'nyong'nyong (*Alphavirus*) [Brault *et al.*, 2004]. Ces derniers ont cependant été détectés chez des anophèles sur le terrain, que ce soient des arbovirus pathogènes pour l'Homme [Bernard *et al.*, 2001, Kulasekera *et al.*, 2001, Feng *et al.*, 2012, Liu *et al.*, 2013, Kemenesi *et al.*, 2014, Calzolari *et al.*, 2010, Calzolari *et al.*, 2013] ou des FSI précédemment décrits, infectant habituellement des moustiques du genre *Culex* [Aranda *et al.*, 2009, Zuo *et al.*, 2014, Liang *et al.*, 2015]. Ces études ne permettent cependant pas de savoir si ces anophèles sont bien des hôtes de ces virus, c'est-à-dire avec une réplication active, ou s'ils n'ont été que transitoirement contaminés par un repas sanguin sur

hôte virémique, sans établissement d'une infection, ou sujets à une contamination expérimentale. La détection d'EVE d'origine flavivirale permettrait ainsi d'apporter de nouvelles preuves de possibles infections flavivirales naturelles chez les anophèles.

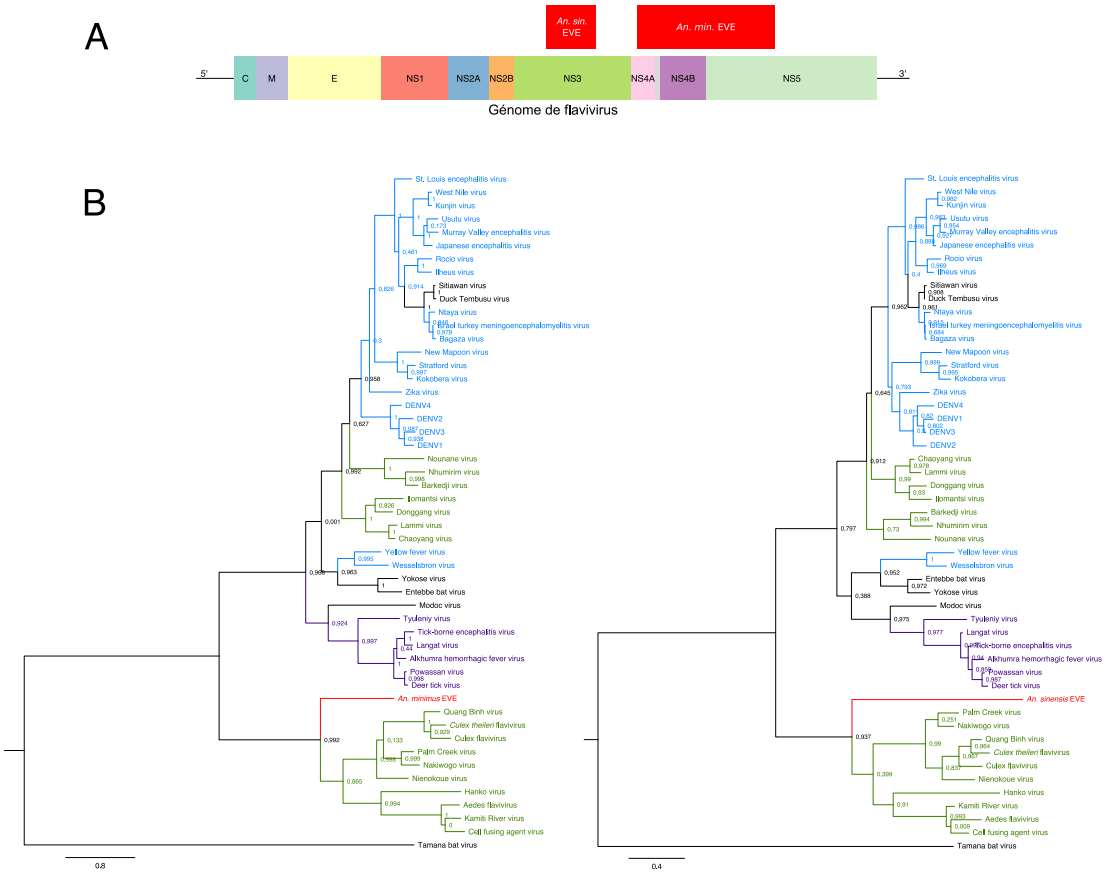

FIGURE 4 – Structure (A) et phylogénie par maximum de vraisemblance (B) des EVEs dérivés de flavivirus chez *Anopheles minimus* et *Anopheles sinensis*. Les couleurs dans les phylogénies représentent la spécificité d'hôte : en vert les FSI, en violet les FAT, en noir les FVS, en bleu les FAM et en rouge, les EVEs. L'échelle indique le nombre de substitutions par site acide-aminé et les valeurs de nœud le SH-like branch support (figure adaptée de l'Annexe A).

Nous avons donc recherché, grâce à des approches bioinformatiques, la présence de séquences proches des flavivirus dans les génomes de 21 espèces d'anophèles (Annexe A). Cette étude nous a permis de détecter la présence d'EVEs d'origine flavivirale dans les génomes de deux espèces asiatiques, *Anopheles minimus* et *Anopheles sinensis*. L'EVE détecté chez *An. minimus* correspond aux protéines non-structurales NS4A, NS4B et NS5 et a une longueur de 1 881 nucléotides (Figure 4-A). Celui détecté chez *An. sinensis* correspond à la protéine non-structurale NS3 et a une longueur de

792 nucléotides (Figure 4-A). Nous avons ensuite confirmé leur existence *in vivo* par PCR sur ADN génomique de moustique et nous avons été capables de montrer leur transcription par RT-PCR. Cette dernière a été confirmée par les analyses de jeux de données de séquençage massif d'ARN messager précédemment publiés. L'analyse phylogénétique des EVEs que nous avons découverts montre la proximité des deux éléments avec les FSI (Figure 4-B), ce qui pourrait suggérer l'existence d'un clade de FSI associés aux anophèles, comparable aux FSI associés aux moustiques du genre *Culex* ou aux FSI associés au genre *Aedes*. La transmission verticale des FSI favorise en effet la co-divergence avec leur hôte [Jackson et Charleston, 2004] et la présence d'un clade associé aux anophèles à la racine des FSI serait cohérente avec l'histoire évolutive des moustiques ; les anophèles ayant divergé antérieurement à la différenciation entre *Culex* et *Aedes* [Reidenbach *et al.*, 2009]. Ces résultats soulignent à nouveau la diversité du genre, loin d'être restreinte aux seuls arbovirus.

3 Le moustique comme siège de microévolution des flavivirus-arbovirus

3.1 Plasticité génétique des arbovirus

Nous avons vu dans la partie précédente que les flavivirus, bien que partageant des caractères communs en terme de structure ou de cycle de réplication, constituent un genre viral très diversifié, infectant un important éventail d'hôtes vertébrés et arthropodes, parfois même alternant entre les deux dans le cas des arbovirus. On associe classiquement cette capacité de se répliquer dans deux hôtes très différents à la flexibilité génomique et au fort taux de mutation des virus à ARN [Weaver, 2006].

En effet, comme pour d'autres virus à ARN, la RdRP virale est caractérisée par une fidélité médiocre due à l'absence d'activité exonucléase permettant la correction des erreurs de réplication, entraînant un taux de mutation d'environ 10^{-6} à 10^{-4} substitution par nucléotide et par cycle de réplication [Sanjuán *et al.*, 2010]. Une étude en culture cellulaire de la souche vaccinale du virus de la fièvre jaune a estimé un taux d'erreur entre $1{,}9\times10^{-7}$ et $2{,}3\times10^{-7}$ par nucléotide copié [Pugachev *et al.*, 2004]. Pour le virus de la dengue-2, une étude *in vitro* a estimé un taux d'erreur de la RdRP entre $2{,}94\times10^{-5}$ et $7{,}41\times10^{-6}$ par nucléotide copié [Jin *et al.*, 2011]. Une récente publication a

par ailleurs montré que ce taux de mutation pouvait être variable pour un même isolat viral en fonction de l'environnement cellulaire, lui-même influencé par l'hôte [Combe et Sanjuán, 2014]. Quelle que soit de la valeur exacte du taux d'erreur de la RdRP, ce dernier, couplé à une réplication rapide et une population de grande taille, a donné naissance au concept de « quasi-espèce » virale. Ce terme définit la population de virus génétiquement apparentés (« nuage de mutants ») et distribués autour d'une séquence dite « consensus », moyenne pour chaque position du nucléotide le plus fréquent dans la population [Lauring *et al.*, 2013].

Cette structure en quasi-espèce s'est montrée critique pour plusieurs aspects fondamentaux de la *fitness* générale des virus à ARN ou leur pathogénèse [Domingo *et al.*, 2012]. Pour les virus, la *fitness* est souvent définie comme étant la capacité de produire de nouvelles particules infectieuses dans un environnement donné [Domingo et Holland, 1997]. Cette définition, centrée sur le succès réplicatif (*replicative fitness*), n'intègre pas totalement la définition usuelle du terme telle qu'elle est comprise en biologie évolutive : la *fitness* est la quantité de matériel génétique passée à la génération suivante [Wargo et Kurath, 2012]. Ainsi, à cause de leur parasitisme strict, la *fitness* des virus repose également sur le succès de leur transmission (*transmission fitness*). La structure en quasi-espèce des virus à ARN s'est ainsi montrée importante pour l'adaptation à de nouveaux environnements cellulaires, pressions de sélection ou hôtes, caractéristiques de la transmission [Domingo *et al.*, 2012]. Par ailleurs, réplication et transmission s'intègrent, au niveau supérieur, dans la capacité d'un virus (quelque soit le niveau considéré : sérotype, génotype, clade, isolat ou variant) a être plus prévalent sur le terrain qu'un autre virus. On parle alors de *fitness* épidémiologique [Wargo et Kurath, 2012].

Contrairement à la majorité des virus à ARN qui n'infectent qu'une ou plusieurs espèces d'hôtes biologiquement proches, les flavivirus-arbovirus doivent être capables de se répliquer dans deux types d'hôtes très différents. L'adaptation à l'un pouvant nuire à la *fitness* dans l'autre hôte, les arbovirus se doivent de maintenir un « compromis » (*trade-off*) [Ciota et Kramer, 2010] qui expliquerait leur évolution plus lente comparée aux virus à ARN à transmission directe [Jenkins *et al.*, 2002]. Un nuage de mutants assez large, en explorant l'« espace génétique » permet ainsi de maintenir au sein d'une même population des variants génétiques plus ou moins adaptés aux différents types d'hôtes. Le maintien d'une diversité génétique est donc crucial pour

la transmission des arbovirus sur le long terme. Les forces évolutives qui modulent la diversité génétique virale au sein des différents hôtes ne sont pas bien connues. On s'attend à ce que de manière générale la sélection naturelle et la dérive génétique s'opposent à la production de mutations *de novo* et tendent à réduire la diversité génétique virale. Ces forces évolutives peuvent par ailleurs très bien s'appliquer de manière opposée ou du moins différemment entre les types d'hôte. Le *trade-off* peut alors être conçu comme une adaptation alternée entre vertébré et arthropode, en « zig-zag », maintenant l'arbovirus, au long terme, sur une voie centrale. On comprend dès lors l'importance de l'étude des forces évolutives s'appliquant sur les populations d'arbovirus, aussi bien chez l'hôte vertébré que chez le vecteur arthropode, afin de mieux appréhender leur biologie, leur potentiel adaptatif et épidémique, ainsi que leur évolution en général [Coffey *et al.*, 2013].

La dynamique du nuage de mutants, c'est à dire l'évolution intra-hôte, a été particulièrement étudiée dans le cas du virus du Nil occidental et plus récemment de la dengue chez les moustiques des genres *Culex* et *Aedes*, respectivement. Ces études ont montré que la diversité génétique du virus du Nil occidental était bien associée à un gain de *fitness* chez le moustique *in vitro* [Ciota *et al.*, 2007, Ciota *et al.*, 2012] ou *in vivo* [Fitzpatrick *et al.*, 2010] et que le nuage de mutants était sous influence de la sélection purifiante [Jerzak *et al.*, 2005, Jerzak *et al.*, 2007] malgré la persistance de mutations délétères [Grubaugh *et al.*, 2016]. Elles ont également mis en évidence que la diversité génétique du virus était globalement maintenue lors de son passage chez le moustique [Brackney *et al.*, 2011] ou légèrement diminuée suite à la présence de goulots d'étranglement démographiques [Ciota *et al.*, 2012]. L'existence d'un goulot d'étranglement changeant la composition du nuage de mutants a également été détectée dans le cas du virus de la dengue-2 [Sim *et al.*, 2015]. Les goulots d'étranglement démographiques (*population bottleneck* en anglais) sont d'importantes réductions transitoires de la taille de la population, pouvant être dues à (i) un évènement tirant aléatoirement quelques variants du nuage ou (ii) un important épisode de sélection naturelle. Dans les deux cas, la stochasticité introduite, par principe dans le premier cas ou sur d'autres marqueurs génétiques que celui sous sélection dans le deuxième, peut conduire à une réduction de la *fitness* par accumulation de mutations délétères et par la diminution de la variabilité génétique de la population [Duarte *et al.*, 1992, Gutiérrez *et al.*, 2012]. Les barrières anatomiques rencontrées par le virus lors

de l'infection du moustique sont suspectées d'être à l'origine d'importants goulots d'étranglement démographiques de nature aléatoire [Forrester *et al.*, 2014].

3.2 Infection arbovirale du moustique

La femelle moustique s'infecte lors d'un repas sanguin sur un hôte vertébré virémique. L'arbovirus entre alors dans le système digestif du moustique avec le sang (Figure 5) et gagne le mésentéron, ou intestin moyen (*midgut* en anglais). Afin d'établir une infection dans les cellules épithéliales du mésentéron, l'arbovirus doit avoir été ingéré en quantité suffisante. Il entreprend alors une phase intensive de réplication. Les virions néoformés sont libérés dans la cavité générale de l'insecte, ou hémocœle, et se disséminent dans divers organes internes où il peuvent éventuellement se répliquer : système nerveux central, cordon nerveux ventral, corps gras, tractus génital et glandes salivaires [Salazar *et al.*, 2007]. L'infection des organes secondaires permet à l'arbovirus d'être ensuite transmis, soit horizontalement *via* les glandes salivaires lors d'un nouveau repas sanguin, soit verticalement, *via* le tractus génital (voir section 3). Le moustique reste infecté toute sa vie. L'infection ne semble avoir qu'un impact très légèrement délétère sur la survie du moustique vecteur [Lambrechts et Scott, 2009] et parfois la fécondité [Styer *et al.*, 2007, Ciota *et al.*, 2013].

Tous les moustiques ne sont pas égaux face à l'infection arbovirale. Leur aptitude intrinsèque (génétique) à s'infecter, permettre la réplication puis transmettre un arbovirus à un nouvel hôte est désignée sous le nom de « compétence vectorielle » [Hardy *et al.*, 1983]. Les barrières anatomiques rencontrées lors de l'infection font, entre autres, partie des facteurs influençant la compétence vectorielle. Quatre grandes barrières ont été décrites : la barrière d'infection du mésentéron (*midgut infection barrier* ou MIB), la barrière d'échappement du mésentéron (*midgut escape barrier* ou MEB), la barrière d'infection des glandes salivaires (*salivary gland infection barrier* ou SGIB), et la barrière d'échappement des glandes salivaires (*salivary gland escape barrier* ou SGEB) [Franz *et al.*, 2015]. Le succès de l'infection dépend donc de la capacité de l'arbovirus à surmonter ces barrières.

3.3 Interactions flavivirus-moustiques

La compétence vectorielle, étant déterminée génétiquement, varie entre les espèces de moustiques mais également entre populations au sein de la même espèce.

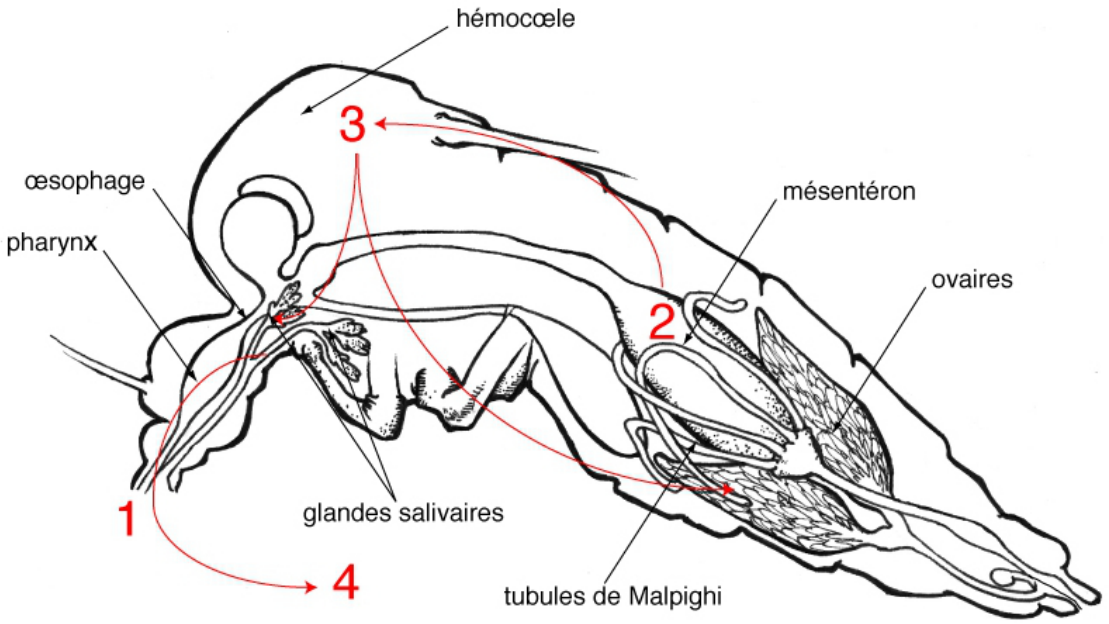

FIGURE 5 – Cycle de la transmission horizontale d'un arbovirus par le moustique vecteur. (1) Repas sanguin infectieux. (2) Infection du mésentéron. (3) Dissémination du virus dans l'hémocœle d'où il infecte d'autres organes dont les glandes salivaires. (4) Inoculation du virus via la salive infectée lors d'un nouveau repas sanguin (figure d'après les données de [Salazar *et al.*, 2007]).

Ainsi, différentes populations du moustique *Aedes aegypti*, génétiquement distinctes, diffèrent dans leur compétence vectorielle pour les mêmes isolats de virus de la dengue (DENV) [Anderson et Rico-Hesse, 2006, Armstrong et Rico-Hesse, 2003, Bennett *et al.*, 2002, Gubler *et al.*, 1979, Rosen *et al.*, 1985, Tardieux *et al.*, 1990, Vazeille-Falcoz *et al.*, 1999]. Plus qu'une simple variation du côté du moustique, plusieurs études ont montré que la compétence vectorielle est en fait le produit d'une interaction spécifique entre des génotypes d'*Ae. aegypti* et des variants génétiques de DENV [Lambrechts *et al.*, 2009, Lambrechts, 2011, Lambrechts *et al.*, 2013, Fansiri *et al.*, 2013, Dickson *et al.*, 2014].

L'impact de la diversité génétique du vecteur sur celle des flavivirus au cours de l'infection commence seulement à être exploré. Une récente étude a montré que différentes espèces de moustiques du genre *Culex* présentaient des niveaux de diversité génétique intra-hôte du virus du Nil Occidental différents au cours de l'infection [Grubaugh *et al.*, 2016]. Aucun travail cependant ne s'est intéressé à l'impact de la diversité génétique au sein d'une même espèce sur les niveaux de diversité virale.

Afin d'apporter des éléments de réponse à cette question, nous avons utilisé une technologie de séquençage à haut-débit pour suivre la diversité intra-hôte du virus de la dengue-1 lors de l'infection de plusieurs fonds génétiques de moustique *Ae. aegypti* sous la forme de lignées isofemelles (Annexe B). Ces lignées isofemelles, initiées par un mâle et une femelle uniques, maintenues par croisements consanguins pour plusieurs générations consécutives ont une variation génétique réduite ; les individus de la lignées sont plus proches entre eux que d'autres individus extérieurs à celle-ci.

Nos résultats ont confirmé les forts effets de la sélection purifiante et de la dérive génétique, déjà évoqués par les études menées sur le virus du Nil occidental [Jerzak *et al.*, 2005, Jerzak *et al.*, 2007, Ciota *et al.*, 2012, Grubaugh *et al.*, 2016]. Nous avons en effet détecté un goulot d'étranglement démographique au moment de l'infection initiale du mésentéron, avec une taille effective de population estimée à quelques dizaines de génomes viraux (Figure 6-A). Au-delà, nous avons également pu mettre en évidence une modulation de la diversité génétique intra-hôte du virus par le génotype du moustique. L'une des lignées isofemelles montre en effet un niveau de diversité génétique accru dans le mésentéron comparé aux deux autres (Figure 6-B et C), sans que cela soit lié à un différentiel dans le niveau de la sélection purifiante ou de la taille du goulot d'étranglement à l'infection. Notre étude n'ayant pas pu mesurer convenablement la diversité génétique dans les glandes salivaires des différentes lignées isofemelles, nous ne pouvons pas à l'heure actuelle lier directement cette observation à la compétence vectorielle. Ces résultats indiquent cependant clairement que la diversité génétique du virus, donc sa *fitness* et son évolution, sont inextricablement liées à la variation génétique de l'hôte, au moins dans le mésentéron.

4 Transmission inter-hôte des flavivirus et d'autres arbovirus par les moustiques

4.1 Transmission des arbovirus

Comme nous l'avons évoqué dans l'introduction, de nombreux genres viraux sont regroupés sous le qualificatif d'arbovirus. Ces genres, malgré une biologie différente, partagent la caractéristique d'être transmis entre hôtes vertébrés par des arthropodes

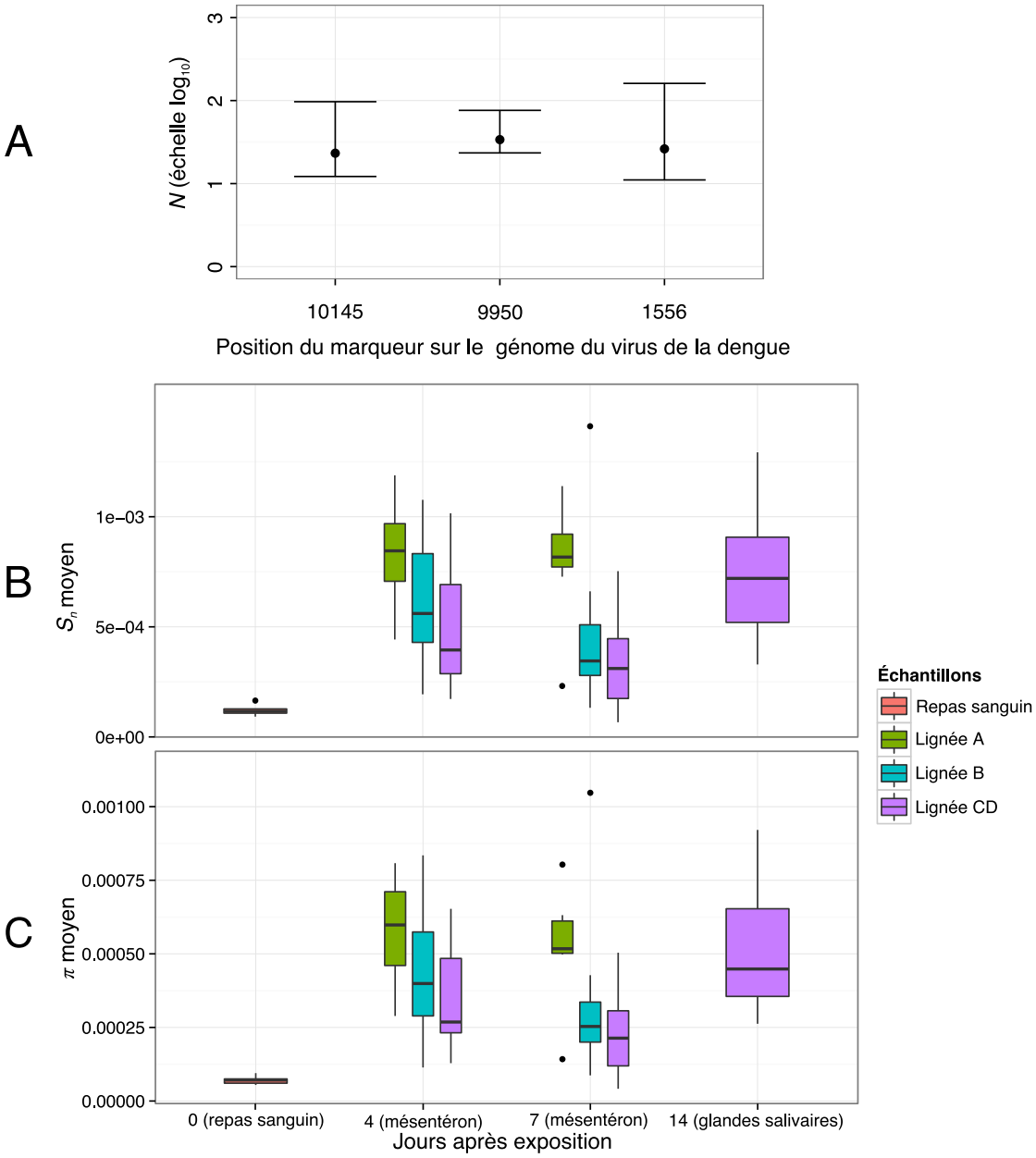

FIGURE 6 – Estimation de la taille du goulot d'étranglement démographique lors de l'infection initiale du mésentéron et niveau de diversité intra-hôte du virus de la dengue-1. (A) Estimation du nombre de génomes établissant l'infection (N) pour 3 marqueurs nucléotidiques identifiés par leur position sur le génome du virus de la dengue (1556, 9950, 10145) à partir d'échantillons collectés 4 jours après infection. Les marqueurs sont des variants nucléotidiques dont on admet leur neutralité ou quasi-neutralité. Les barres horizontales indiquent les intervalles de confiance de l'estimation de N par bootstrapping. (B) Diversité virale intra-hôte estimée par l'entropie de Shannon (S_n) par site et par échantillon. (C) Diversité virale intra-hôte estimée par la diversité nucléotidique (π) par site et par échantillon (figure adaptée de l'Annexe B).

hématophages, souvent des moustiques.

Les arbovirus sont généralement zoonotiques et sont par conséquent maintenus dans la nature par des cycles impliquant divers hôtes vertébrés et arthropodes. Plus qu'une simple alternance entre un seul hôte vertébré et un seul arthropode vecteur, la transmission arbovirale s'inscrit le plus souvent dans un réseau de transmissions complexe, impliquant de nombreux hôtes et vecteurs (Figure 7) [Diaz *et al.*, 2012]. Certains hôtes ou vecteurs de ce réseau peuvent par ailleurs être impliqués dans la maintenance des arbovirus lors des périodes inter-épidémiques. Certains auteurs ont par exemple suggéré le cas de l'écureuil fauve pouvant participer à la maintenance du virus du Nil occidental dans un contexte péri-urbain en Amérique du Nord [Root *et al.*, 2006, Root *et al.*, 2007]. Les êtres humains ne sont pas forcément centraux dans ces réseaux de transmission et ne sont même parfois que des hôtes accidentels, soit en cas d'une transmission tangentielle à l'Homme (virus du Nil occidental), ou lorsque des sujets susceptibles entrent dans un foyer de transmission enzoonotique selvatique (virus de la fièvre jaune ou de l'encéphalite équine vénézuélienne) [Weaver et Barrett, 2004]. Pour les virus de la dengue, deux cycles distincts mais reliés dans l'histoire évolutive coexistent : un cycle selvatique, ancestral, enzoonotique impliquant des primates non-humains et des moustiques zoophiles et un cycle urbain endémique ou épidémique impliquant l'Homme comme seul hôte vertébré et des moustiques anthropophiles comme vecteurs (comme *Aedes aegypti* ou *Ae. albopictus*) [Holmes et Twiddy, 2003, Vasilakis *et al.*, 2011].

Le rôle épidémiologique de chaque espèce vectrice potentielle au sein de ces réseaux ou cycles de transmission peut être décrite par la notion de capacité vectorielle, englobant de nombreux facteurs génétiques, environnementaux ou comportementaux. En effet, malgré leur souplesse génétique et leur diversité que nous avons évoquées dans les chapitres précédents en prenant l'exemple des flavivirus, les arbovirus ne sont généralement transmis que par quelques espèces vectrices principales. La capacité vectorielle peut être nulle si l'arthropode est réfractaire à l'infection ou incapable de transmettre le virus, que ce soit dû à une composante génétique (du virus ou du vecteur, c'est-à-dire la compétence vectorielle), environnementale (la température, par exemple) ou comportementale (un arthropode hématophage ne piquant pas l'hôte vertébré du virus). La capacité vectorielle peut également varier significativement au sein d'une espèce vectrice. La capacité vectorielle (**C**) est classiquement

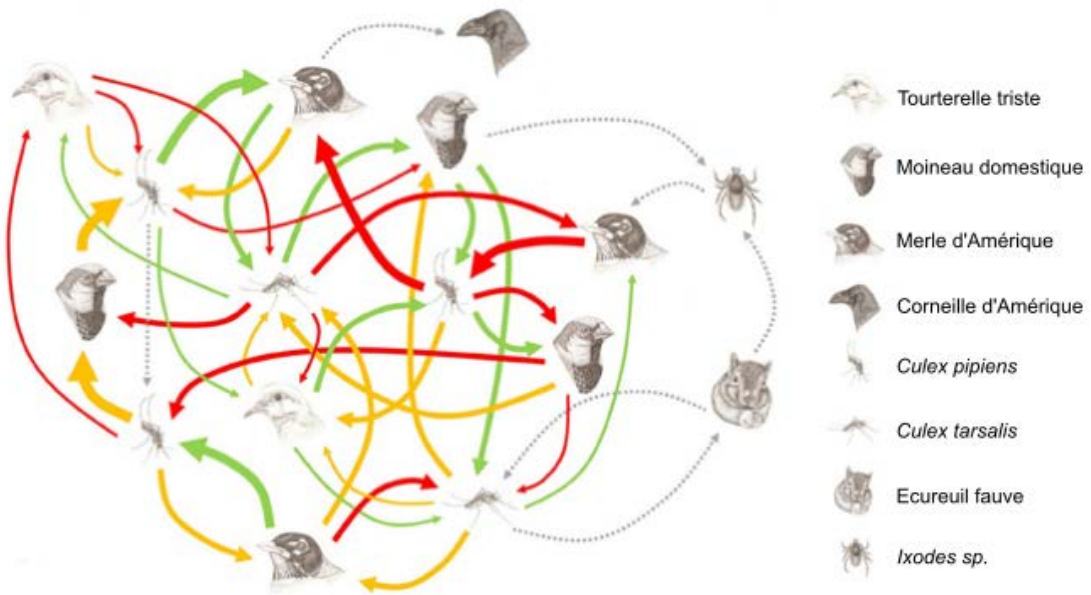

FIGURE 7 – Réseau de transmission hypothétique du virus du Nil occidental aux États-Unis. Les flèches représentent le lien de transmission entre vecteurs et hôtes impliqués dans le réseau et leur épaisseur l'importance de cette liaison (en termes de préférence trophique du vecteur, de la densité de population, de la compétence vectorielle, etc.). Leur couleur représente la saison de la transmission figurée (vert : printemps ; rouge : été ; orange : automne). Les lignes pointillées sont des voies de transmission alternatives (transmission verticale chez le vecteur par exemple) ou des hôtes ou vecteurs secondaires (tiques ou vertébrés) (figure adaptée de [Diaz *et al.*, 2012].

modélisée mathématiquement par l'équation de Ross-MacDonald, initialement développée pour le paludisme mais également applicable en arbovirologie [Kramer et Ebel, 2003] :

$$C = \frac{m \times a^2 \times p^n \times b}{-\ln(p)}$$

où **m** représente la densité du vecteur relative à celle de l'hôte ; **a** la probabilité que le vecteur pique l'hôte en un jour, élevée au carré car deux piqûres sont nécessaires (une pour l'infection du vecteur, l'autre pour la transmission au vertébré) ; **p** la probabilité de survie quotidienne du vecteur ; **n** la durée de la période d'incubation extrinsèque, c'est-à-dire le temps (en jours) séparant l'infection du vecteur du moment où ce dernier est capable de transmettre le virus à un nouvel hôte vertébré ; $\frac{1}{-\ln(p)}$ la survie du vecteur en jours au-delà de la durée de la période d'incubation

extrinsèque ; et finalement **b**, la compétence vectorielle. Les facteurs élevés au carré (**a**) ou en exponentiel (**p** et **n**) sont ceux qui influencent le plus **C** [Kramer et Ebel, 2003, Ebel et Kramer, 2009].

Cette équation ne peut estimer l'importance d'un vecteur d'arbovirus que dans le cadre d'une transmission « classique », horizontale, entre vecteur et hôte vertébré. D'autres modes de transmission ont été décrits, qu'ils soient considérés comme « mécaniques » ou « biologiques », c'est-à-dire impliquant l'infection active du vecteur par le virus, comme la transmission horizontale. Au contraire, la transmission mécanique n'implique pas l'infection du vecteur. Ce sont alors les pièces buccales souillées lors d'un repas sanguin sur hôte virémique qui infectent un nouvel hôte, dans le court laps de temps avant que le virus ne soit inactivé [Higgs et Beaty, 2005]. Ce mode de transmission, commun pour les virus de plantes transmis par des insectes comme les pucerons [Blanc *et al.*, 2014], est considéré comme plus rare pour les virus animaux, même s'il a été décrit pour le virus de l'encéphalite équine de l'Ouest, le virus Ross River ou le virus de la fièvre de la vallée du Rift [Turell, 1988]. L'efficacité de la transmission mécanique augmente avec la densité de vertébrés infectés mêlés à une population de vertébrés naïfs et la fréquence de piqûre du ou des arthropodes hématophages. Ce mode de transmission, en effet, sans infection du vecteur n'implique donc pas de notion de compétence vectorielle (paramètre **b** dans l'équation de Ross-MacDonald), et en théorie un grand nombre d'arthropodes hématophages peuvent être impliqués. Certains auteurs ont suggéré que la transmission mécanique s'est trouvée impliquée dans certaines grandes épidémies de fièvre de la vallée du rift en Afrique ou d'encéphalite équine vénézuélienne en Colombie [Kuno et Chang, 2005].

La transmission verticale (TV) est l'une des alternatives biologiques dans laquelle la femelle arthropode infectée transmet le virus à sa descendance, sans passage par un hôte vertébré. La TV a été décrite dans la littérature pour de nombreux virus de différentes familles et genres. Nous avons vu précédemment que la dissémination du virus dans son vecteur lui permet d'infecter les glandes salivaires, ce qui rend possible une transmission à un nouvel hôte vertébré *via* la salive. D'autres organes peuvent également être infectés, comme le tractus génital de la femelle ouvrant dans ce cas la possibilité d'une TV. Deux mécanismes différents de TV ont été décrits : une voie trans-ovarienne et une voie trans-œuf.

La TV par voie trans-ovarienne implique l'infection par le virus des follicules ovariens de la femelle et la traversée de l'épithélium folliculaire pour gagner l'accès aux oocytes en développement [Leake, 1984, Turell, 1988]. La voie trans-œuf se caractérise quant à elle par l'infection de l'œuf *via* le micropyle lors de sa fécondation dans l'oviducte. Cette TV trans-/oeuf peut être d'origine maternelle, si le tractus génital de la femelle est infecté, ou paternelle, *via* le fluide séminal, avec ou sans réplication dans le tractus génital de la femelle [Higgs, 2004]. Dans ce dernier cas, il est à noter que le mâle, ne prenant pas de repas sanguin et n'étant donc pas exposé aux arbovirus par voie horizontale « classique », ne peut avoir été infecté que verticalement ou par contact sexuel, un autre mode de transmission horizontale biologique qualifié de « vénérien » [Higgs, 2004].

4.2 Maintien des arbovirus par transmission verticale chez le moustique

On comprend dès lors l'importance potentielle de la TV pour le maintien des arbovirus, d'autant plus qu'une femelle infectée verticalement est potentiellement infectieuse pour un nouvel hôte vertébré en l'absence d'un repas sanguin sur hôte virémique ou de période d'incubation extrinsèque pour transmettre le virus [Higgs et Beaty, 2005]. En effet, les conditions nécessaires à la transmission horizontale ne sont pas toujours réunies, comme cela peut être le cas lors d'une saison sèche en zone tropicale ou froide en zone tempérée, ou encore, pendant une campagne de lutte antivectorielle réduisant ou diminuant la densité de moustiques adultes. De plus, les infections arbovirales chez les vertébrés sont généralement de courte durée, suivies d'une immunité protectrice. Au décours d'une épidémie, la densité de vertébrés susceptibles décroit donc et peut entraîner la fin d'une épidémie grâce à une immunité dite de groupe. Au sujet des mécanismes par lesquels ces virus peuvent se maintenir lors de conditions défavorables en zone d'endémie ou suite à une épidémie, trois hypothèses sont classiquement proposées et considérées comme les plus probables [Reeves, 2004] :

- persistance du virus dans des femelles arthropodes adultes en diapause (forme de vie ralentie, génétiquement déterminée, avec diminution des activités métaboliques et arrêt du développement) infectées durant l'automne. En survivant pendant l'hiver, elles peuvent ainsi ré-initier un cycle de transmission horizon-

tale au printemps ;

- réintroduction annuelle du virus par la migration des vertébrés ou des vecteurs depuis une zone endémique où le cycle de transmission horizontale est continu ou les saisons inversées ;

- persistance du virus dans les formes immatures du moustique (œuf ou larves) grâce à la TV. Les descendants infectés gardent l'infection pendant tous les stades de développement et ré-initient le cycle de transmission horizontale une fois les conditions environnementales favorables réunies.

Cependant, les taux de TV observés sont généralement trop bas pour pouvoir, selon des modèles mathématiques, expliquer le maintien au long terme des arbovirus. Dans le cas des virus de la dengue par exemple, un taux de transmission verticale aux alentours de 1% de descendants est observé expérimentalement pour une femelle infectée. Selon un modèle mathématique, ce taux devrait être 5 à 30 fois plus élevé pour que la TV ait un réel impact sur la persistance au long terme du virus [Adams et Boots, 2010]. En revanche, certains auteurs ont proposé l'existence d'une infection « stabilisée ». Celle-ci serait favorisée dans le cadre d'une transmission transovarienne, en particulier observée chez les Bunyaviridae, et l'existence d'une sous-population dont les lignées germinales seraient infectées de manière permanente [Turell *et al.*, 1982], même si la prévalence totale dans la population est faible [Tesh, 1984].

Si l'existence en tant que telle de la TV, quel que soit le mécanisme impliqué, n'est pas sujette à débat, l'importance qu'elle peut revêtir d'un point de vue épidémiologique est toujours très discutée. De même, les facteurs biologiques ou environnementaux favorisant ou défavorisant ce mode de transmission sont peu caractérisés. De nombreuses recherches ont été menées depuis la première évocation de la TV à la fin du XIX^{e} siècle jusqu'à nos jours. Ces études aux résultats inconstants, voire parfois contradictoires, ne permettent pas de résoudre ces questions.

C'est dans ce contexte que nous avons mené notre propre revue systématique sur la TV des arbovirus chez les moustiques ; la dernière synthèse descriptive de la littérature sur la question ayant été menée il y a 28 ans [Turell, 1988]. Plus qu'une synthèse descriptive, nous avons extrait des différentes publications des informations méthodologiques, expérimentales et environnementales afin de s'appuyer sur une analyse statistique robuste pour apporter un regard neuf, sans parti pris, et à jour sur

la question. L'important jeu de données que nous avons compilé au cours de ce travail nous a permis d'explorer différents aspects liés à la TV. Les résultats de cette étude ont été publiés dans deux articles, disponibles en annexes C et D.

Nous nous sommes intéressés dans un premier temps à la dimension historique de la recherche sur la TV (Annexe C). Nous avons ainsi constaté que des évènements épidémiologiques majeurs, comme l'épidémie de dengue à fièvre hémorragique à Cuba en 1981 ou l'émergence du virus du Nil occidental en Amérique du Nord en 1999, avaient fortement stimulé la recherche sur la TV. Notre analyse révèle aussi une association entre l'évolution des méthodes de détection de virus et la détection de la transmission verticale (Figure 8), que nous avons pu attribuer à l'évidente augmentation de la sensibilité mais également à l'augmentation de la taille des échantillons, plus simples à traiter avec les méthodes de détection récentes. Ces résultats soulignent ainsi que l'interprétation d'études ayant trait à la TV doit être faite avec précaution, car elles sont dépendantes du contexte et de la méthodologie.

Ces résultats ont été pris en compte dans la deuxième série d'analyses que nous avons menées, portant plus spécifiquement sur l'identification de déterminants ou facteurs favorisant la TV (Annexe C). Ainsi, nos analyses statistiques ont systématiquement pris en compte la méthode de détection comme co-variable. Nous avons pu mettre en évidence l'influence de nombreux facteurs déjà fortement suspectés, ou non, d'avoir un impact sur la TV, tels que des facteurs taxonomiques, que ce soit du côté du virus (famille virale, sérotype pour les virus de la dengue) ou du moustique (genre de moustique ou espèce dans le cas de la dengue). Des facteurs biologiques, comme le cycle gonotrophique (développement des œufs après un repas sanguin) ont également été impliqués. En effet, l'oogenèse démarrant rapidement après un repas sanguin, la probabilité que la TV ait lieu lors du premier cycle gonotrophique est plus faible que lors du second (suivant un deuxième repas sanguin), le virus n'ayant pas eu le temps de disséminer et infecter le tractus génital [Leake, 1984, Turell, 1988]. L'ensemble des résultats de ces deux études permet ainsi d'affiner notre compréhension des stratégies employées par les arbovirus pour se maintenir dans leur environnement lors des périodes défavorables à la transmission horizontale. Par ailleurs, ils mettent en lumière des facteurs, dont le rôle doit être expérimentalement élucidé, influant sur la TV et devant être pris en compte lors de l'estimation de l'importance épidémiologique qu'elle revêt.

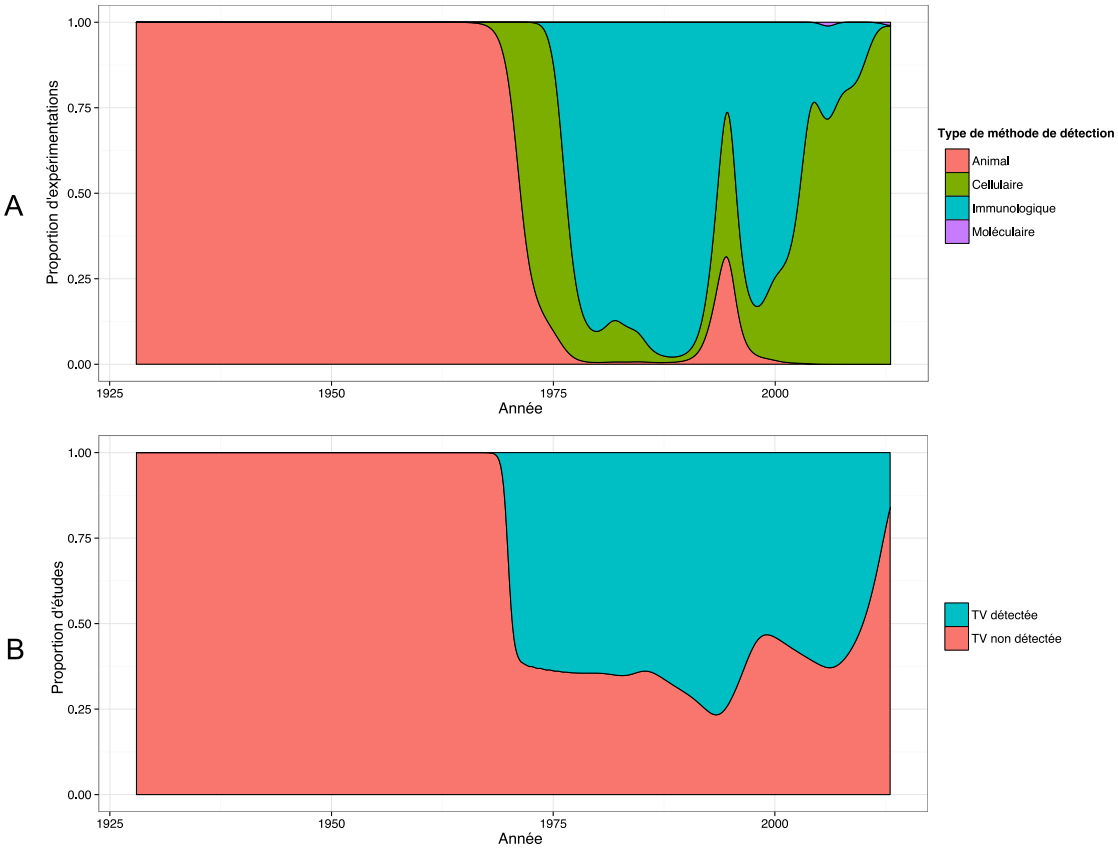

FIGURE 8 – Évolution des techniques de détection (A) et de la mise en évidence de la transmission verticale des arbovirus chez les moustiques en fonction du temps dans les études expérimentales. (A) Évolution des techniques de détection. Les aires colorées représentent la proportion relative de chaque catégorie de méthode de détection en fonction du temps. La proportion annuelle est pondérée par le nombre total de moustiques testés pour la transmission verticale. (B) Évolution de la détection de la transmission verticale. La courbe représentent la probabilité relative de détection de transmission verticale en fonction du temps (figure adaptée de l'Annexe C).

5 Conclusions

Au fil des pages de ce manuscrit s'est dessinée une recherche tournée vers le monde des « mégadonnées » ou *big data*, que ce soit au travers de la littérature ou des études que j'ai moi-même menées (voir Annexes A-D). Les technologies de séquençage haut-débit, par exemple, ont révolutionné non seulement l'étude des interactions flavivirus-moustique mais plus largement l'ensemble de la recherche en biologie. *Transcriptomics*, *(meta)-genomics*, *epigenomics*, autant de mots en « *omics* »

qui fleurissent sur les publications récentes. Et la tendance n'est probablement pas prête à s'arrêter : entre 2010 et 2015, la capacité de séquençage annuelle au niveau mondial a doublé tous les 7 mois, atteignant mi-2015 près de 35 peta-bases[2] [Stephens *et al.*, 2015]. Une projection pour les dix années à venir fait de la génomique le plus gros consommateur de capacité de stockage (estimé entre 2 et 40 exa-bytes (EB) par an, contre 1 EB/an pour l'astronomie ou 1 à 2 EB/an pour la plateforme de partage de vidéos YouTube) [Stephens *et al.*, 2015]. Loin de ne toucher que la biologie, évidemment, c'est l'ensemble des données produites et accumulées par l'humanité qui croît : on estime ainsi qu'entre l'invention de l'écriture et 2006, 180 EB de données ont été générées ; un total qui croit à 1 600 EB entre 2006 et 2011, et dont on estime qu'il quadruple tous les 3 ans [Floridi, 2012].

L'abondance des données est une caractéristique fondamentale du monde du *big data* mais elle n'est pas la seule : elle en côtoie d'autres comme la vitesse de calcul nécessaire à l'analyse, la variabilité ainsi que la véracité des sources de ces données. Ainsi, loin de se limiter aux seuls aspects techniques ayant lien à la gestion du stockage ou du traitement de ces données, l'ère du *big data* oblige les scientifiques à gérer une information complexe de sources hétérogènes [Li et Chen, 2014]. Plus encore, elle les invite à repenser la méthode scientifique à la lumière de l'abondance des données, autrefois si difficiles et coûteuses à obtenir et aujourd'hui de plus en plus simples et bon marché.

On peut distinguer deux approches principales de l'usage du *big data*. L'une est intégrée dans la démarche scientifique classique : l'approche hypothético-déductive. À partir d'une théorie générale, une hypothèse est formulée puis testée par une méthode expérimentale. En fonction du résultat, l'hypothèse peut se retrouver confirmée, ce qui soutient un peu plus la théorie, ou infirmée, ce qui peut amener à affiner la théorie ou la remettre en question. Dans les deux premières études présentées dans ce manuscrit de thèse, l'utilisation de la bioinformatique et du séquençage haut-débit permettait de répondre à une question précise (Y a-t-il trace d'une ancienne infection à flavivirus chez des anophèles ? La diversité virale intra-hôte du virus de la dengue-1 est-elle influencée par le génotype du vecteur ?). Dans cette optique donc, le *big data* n'est qu'une manière de caractériser l'abondance des données à exploiter et les défis techniques qui y sont liés, sans la dimension polémique qui y est parfois associée. La

[2] Le suffixe peta- représente 10^{15} et le suffixe exa- 10^{18}.

démarche reste hypothético-déductive mais se trouve « dopée » par l'abondance de données et à la puissance de calcul [Frické, 2014].

La dimension polémique du *big data* est généralement associée à une approche purement inductive que l'on retrouve dans l'exploration des données ou *data mining*. Certains auteurs, généralement extérieurs au monde des sciences expérimentales, ont clamé qu'avec une importante quantité de données, l'approche hypothético-déductive tombait dans l'archaïsme. Ainsi, avec des jeux de données assez grands, couplés au travail d'algorithmes statistiques, la détection de corrélations entre variable explicative et variable expliquée, de « motifs », est suffisante pour faire avancer la science, ou du moins pour de nombreuses applications pratiques de cette découverte ; le *big data* représenterait la « fin de la théorie » [Anderson, 2008, Graham, 2012].

L'approche inductive part du principe que le *big data*, en capturant toutes ou la grande majorité des données disponibles, permet de s'affranchir d'hypothèses ou de modèles pour les analyser, permettant ainsi d'exclure tout biais venant de l'analyste [Kitchin, 2014]. Une partie des études sur la TV que j'ai menées dans le cadre de cette thèse s'approche de cette conception polémique du *big data* ; paradoxalement, car contrairement aux études précédentes analysant plusieurs centaines de GB de données, l'ensemble des données de cette étude ne représente qu'un peu plus de 200 kB. Lors de ces travaux, j'ai compilé les variables de très nombreuses études, de sources variées et j'ai cherché à analyser les motifs et tendances des variables associées à la TV des arbovirus.

Si les prémices posées par les approches inductives basées sur le *big data* semblent par certains aspects séduisantes, elles n'en sont pas moins fausses : les données sont générées suites à un raisonnement scientifique et ne sont donc pas exemptes de théorie sous-jacente [Frické, 2014, Mazzocchi, 2015]. Ainsi, dans le cas de la TV, les données compilées par les auteurs des différentes études sont celles dont l'effet pouvait être suspecté, du moins ayant une importance dans la biologie des moustiques ou des virus. Le « biais » théorique est donc présent avant même l'analyse des données [Kitchin, 2014]. De plus, l'analyse, elle aussi, est souvent biaisée par l'analyste : dans le cas de la TV par exemple, lors de la construction des modèles statistiques, n'ont été incluses que les variables que je suspectais pouvoir avoir un réel effet. D'autres, pourtant compilées directement ou indirectement dans le jeu de données (par exemple le sexe du premier auteur), ont été écartées car considérées comme non pertinentes dans le

contexte de l'étude. Ainsi, l'idée même d'une « expérimentation » sans idée préconçue rêvée par les défenseurs du *big data* inductiviste serait sans fondement, comme le souligne Henri Poincaré dans *La science et l'hypothèse* :

> « On dit souvent qu'il faut expérimenter sans idée préconçue. Cela n'est pas possible ; non seulement ce serait rendre toute expérience stérile, mais on le voudrait qu'on ne le pourrait pas. Chacun porte en soi sa conception du monde dont il ne peut se défaire si aisément. Il faut bien, par exemple, que nous nous servions du langage, et notre langage n'est pétri que d'idées préconçues et ne peut l'être d'autre chose. Seulement ce sont des idées préconçues inconscientes, mille fois plus dangereuses que les autres. »

Ainsi cette approche purement inductive s'est rapidement vue opposée une fin de non-recevoir. Par ailleurs, les corrélations détectées entre les données, fer de lance des résultats de *data mining*, n'impliquent pas causalité et la compréhension du mécanisme unissant variable explicative et variable expliquée est au cœur de l'avancée scientifique [Pigliucci, 2009]. Si cette compréhension n'est pas nécessaire pour certaines applications d'une découverte scientifique, elle n'en reste pas moins fondamentale pour appréhender l'ensemble des applications[3] et implications possibles [Mazzocchi, 2015]. La théorie n'est donc pas morte ni archaïque mais l'ère du *big data* lui offre de nouvelles opportunités [Mazzocchi, 2015].

Plutôt qu'une approche purement inductive, le *big data* a donné naissance à une approche qualifiée par certains auteurs comme « conduite par les données » (*data-driven*) [Leonelli, 2012]. Quasi-similaire à l'approche hypothético-déductive, elle y insuffle de l'induction, permettant ainsi de générer des hypothèses non pas de la théorie mais des données elles-mêmes (on parle alors d'abduction) [Kelling *et al.*, 2009]. L'induction n'est donc pas ici une finalité mais est contextualisée dans un domaine théorique : ainsi, une corrélation statistique entre variable expliquée et une variable explicative peut être rapidement écartée si considérée comme absurde, alors qu'une autre peut être mise en exergue [Kitchin, 2014]. Même les approches les plus descriptives, parfois critiquées pour leur côté purement naturaliste plus proche du catalogage que de l'entreprise scientifique, visent à répondre à une question (par

[3] Ceci si l'on considère la recherche scientifique comme une entreprise finaliste ; dans le cas contraire d'une recherche inscrite dans la quête d'une meilleure compréhension du monde, l'argument des enthousiastes du *big data* précédemment évoqué est sans objet.

exemple, quelle est la diversité du vivant dans les océans ? Quels sont les virus infectant des insectes dans leur milieu naturel ?). Elles fourniront par la suite aux chercheurs matière à penser et à construire de nouvelles théories et hypothèses [Pigliucci, 2009].

Cette approche n'est évidemment pas sans risques de biais, en particulier statistique, augmentant le risque de détection de faux positifs dans la masse de données. Ces risques sont parfois perçus comme bien trop grands pour défendre pleinement l'approche *data-driven*, au point qu'un épistémologiste suggère que la « science a besoin de plus de théorie et moins de données » [Frické, 2014]. D'autres, que personnellement je rejoins, voient dans l'approche *data-driven* un nouveau paradigme de la méthode scientifique [Kitchin, 2014]. La traditionnelle méthode hypothético-déductive était particulièrement bien adaptée aux conditions précédant l'ère du *big data* : peu de données et peu de puissance de calcul. L'approche *data-driven*, quant à elle, permet d'extraire de nouveaux concepts et idées, tout en permettant d'accéder à des modèles plus complets et interdisciplinaires [Kelling *et al.*, 2009, Kitchin, 2014]. Pour citer les mots de l'épistémologiste Fulvio Mazzocchi [Mazzocchi, 2015] :

« Framing the issue of Big Data in terms of oppositions, that is, deduction versus induction, hypothesis-driven versus data-driven or human versus machine, misses the point that both strategies are necessary and can complement each other. As others have argued, the inductive and deductive phases should be seen as an iterative cycle of knowledge acquisition. »

« Formuler le problème du *big data* en terme d'oppositions, c'est-à-dire déduction contre induction, conduite par l'hypothèse ou conduite par les données, humain contre machine, passe à côté du fait que ces deux stratégies sont nécessaires et peuvent se complémenter. Comme d'autres l'ont soutenu, les phases inductive et déductive doivent être vues comme un cycle itératif d'acquisition de la connaissance. » (ma traduction)

Les études menées sur la TV des arbovirus chez les moustiques présentées dans cette thèse s'inscrivent ainsi résolument dans ce cadre. L'analyse a été nourrie de certaines hypothèses que nous avions, comme par exemple certaines variables dont l'effet était suspecté, inscrites dans une théorie générale de la TV. Mais elle a permis de faire émerger d'autres concepts et de nouvelles hypothèses plus inattendues, comme par exemple des différences entre sérotypes des virus de la dengue. L'approche véri-

tablement originale de ce travail est finalement plus l'utilisation de données parfois anciennes que la méthode d'analyse.

L'arbovirologie au sens large a une longue histoire : plus d'un siècle sépare les travaux de Walter Reed des nôtres. Mais contrairement à d'autres sciences, et souvent à nos idées préconçues, certaines des observations et opinions sont toujours intrigantes et loin d'être dépassées. S'il est indéniable que la technologie a évolué, que nos démonstrations peuvent être plus étayées, plus solides, il y a, je pense, énormément de données, d'hypothèses et de théories à tirer des travaux anciens. À titre d'exemple, certaines de nos conclusions concernant la TV ne sont pas fondamentalement différentes de celles de Marchoux et Simond [Marchoux et Simond, 1906] 110 ans plus tôt, même si, évidemment, elles reposent sur bien plus de données et sont plus solides que le sentiment profond des deux scientifiques français. Au-delà, notre analyse, grâce aux outils proches de la philosophie du big data, nous a permis de tirer des données contenues dans ces publications plus ou moins anciennes de nouvelles hypothèses, sans que les auteurs originaux ne les aient, ne serait-ce qu'en partie, conceptualisées lors de l'acquisition de ces données.

D'autres études ont également pris le parti d'analyser ou de ré-analyser ces données anciennes, que ce soit au détour d'une synthèse descriptive de la littérature [Kuno et Chang, 2005, Kuno, 2009], d'une nouvelle analyse de la littérature publiée [Nishiura et Halstead, 2007] ou même non publiées à l'époque [Snow *et al.*, 2014], apportant elles aussi des connaissances et hypothèses nouvelles. L'un des principaux freins à l'utilisation des travaux anciens reste évidemment la difficulté de trouver une copie des publications, dormant sur les étagères des bibliothèques universitaires et n'étant, bien souvent, pas référencées dans les bases de données bibliographiques. On peut espérer que les campagnes de numérisation menées par exemple par Google puissent un jour y remédier, de même que les logiciels de reconnaissance de caractères de plus en plus sophistiqués puissent à terme optimiser la recherche d'information dans ces articles.

Nous vivons une ère tout à fait intéressante pour la recherche scientifique. Si l'on peut se lamenter à juste titre du peu de considération politique, du moins dans les pays occidentaux, pour la recherche, des financements en baisse, de la précarité croissante des emplois de chercheur, en particulier chez les jeunes diplômés, on ne peut en revanche qu'être inspiré par les perspectives futures en termes de connaissances et

d'applications. Les systèmes de plus en plus complexes comme, entre bien d'autres, les interactions flavivirus-moustiques, livrent peu à peu leurs secrets. Le *big data*, en synergie avec d'autres approches, a et aura son rôle à jouer. La route est encore longue, de plus en plus escarpée et nécessite des outils technologiques et conceptuels de plus en plus avancés. Mais n'est-ce pas quelque part enthousiasmant ?

Je me demande ce qu'en dirait Alexandre Yersin.

Bibliographie

[Adams et Boots, 2010] ADAMS, B. et BOOTS, M. (2010). How important is vertical transmission in mosquitoes for the persistence of dengue ? Insights from a mathematical model. *Epidemics*, 2(1):1–10.

[Aiewsakun et Katzourakis, 2015] AIEWSAKUN, P. et KATZOURAKIS, A. (2015). Endogenous viruses : Connecting recent and ancient viral evolution. *Virology*, 479-480:26–37.

[Anderson, 2008] ANDERSON, C. (2008). The End of Theory : The Data Deluge Makes the Scientific Method Obsolete. *Wired magazine*, pages 1–10.

[Anderson et Rico-Hesse, 2006] ANDERSON, J. R. et RICO-HESSE, R. (2006). Aedes aegypti vectorial capacity is determined by the infecting genotype of dengue virus. *American Journal of Tropical Medicine and Hygiene*, 75(5):886–892.

[Aranda *et al.*, 2009] ARANDA, C., SÁNCHEZ-SECO, M. P., CÁCERES, F., ESCOSA, R., GÁLVEZ, J. C., MASIÀ, M., MARQUÉS, E., RUÍZ, S., ALBA, A., BUSQUETS, N., VÁZQUEZ, A., CASTELLÀ, J. et TENORIO, A. (2009). Detection and monitoring of mosquito flaviviruses in Spain between 2001 and 2005. *Vector borne and zoonotic diseases (Larchmont, N.Y.)*, 9(2):171–178.

[Armstrong et Rico-Hesse, 2003] ARMSTRONG, P. M. et RICO-HESSE, R. (2003). Efficiency of dengue serotype 2 virus strains to infect and disseminate in Aedes aegpti. *American Journal of Tropical Medicine and Hygiene*, 68(5):539–544.

[Barrett et Higgs, 2007] BARRETT, A. D. T. et HIGGS, S. (2007). Yellow fever : a disease that has yet to be conquered. *Annual review of entomology*, 52(1):209–229.

[Bekal *et al.*, 2014] BEKAL, S., DOMIER, L. L., GONFA, B., MCCOPPIN, N. K., LAMBERT, K. N. et BHALERAO, K. (2014). A novel flavivirus in the soybean cyst nematode. *The Journal of general virology*, 95(Pt 6):1272–1280.

[Bell-Sakyi et Attoui, 2013] BELL-SAKYI, L. et ATTOUI, H. (2013). Endogenous tick viruses and modulation of tick-borne pathogen growth. *Frontiers in cellular and infection microbiology*, 3:25.

[Bennett *et al.*, 2002] BENNETT, K. E., OLSON, K. E., MUNOZ, M. d. L., FERNANDEZ-SALAS, I., FARFAN-ALE, J. A., HIGGS, S., BLACK, W. C. et BEATY, B. J. (2002). Variation in vector competence for dengue 2 virus among 24 collections of Aedes aegypti from Mexico and the United States. *American Journal of Tropical Medicine and Hygiene*, 67(1):85–92.

[Bernard *et al.*, 2001] BERNARD, K. A., MAFFEI, J. G., JONES, S. A., KAUFFMAN, E. B., EBEL, G., DUPUIS, A. P., NGO, K. A., NICHOLAS, D. C., YOUNG, D. M., SHI, P. Y., KULASEKERA, V. L., EIDSON, M., WHITE, D. J., STONE, W. B., KRAMER, L. D. et NY STATE WEST NILE VIRUS SURVEILLANCE TEAM (2001). West Nile virus infection in birds and mosquitoes, New York State, 2000. *Emerging infectious diseases*, 7(4):679–685.

[Bhatt *et al.*, 2013] BHATT, S., GETHING, P. W., BRADY, O. J., MESSINA, J. P., FARLOW, A. W., MOYES, C. L., DRAKE, J. M., BROWNSTEIN, J. S., HOEN, A. G., SANKOH, O., MYERS, M. F., GEORGE, D. B., JAENISCH, T., WINT, G. R. W., SIMMONS, C. P., SCOTT, T. W., FARRAR, J. J. et HAY, S. I. (2013). The global distribution and burden of dengue. *Nature*, 496(7446):504–507.

[Blanc *et al.*, 2014] BLANC, S., DRUCKER, M. et UZEST, M. (2014). Localizing viruses in their insect vectors. *Annual Review of Phytopathology*, 52:403–425.

[Blitvich et Firth, 2015] BLITVICH, B. J. et FIRTH, A. E. (2015). Insect-specific flaviviruses : a systematic review of their discovery, host range, mode of transmission, superinfection exclusion potential and genomic organization. *Viruses*, 7(4):1927–1959.

[Bowman *et al.*, 2016] BOWMAN, L. R., DONEGAN, S. et MCCALL, P. J. (2016). Is Dengue Vector Control Deficient in Effectiveness or Evidence ? : Systematic Review and Meta-analysis. *PLoS neglected tropical diseases*, 10(3):e0004551.

[Brackney *et al.*, 2011] BRACKNEY, D. E., PESKO, K. N., BROWN, I. K., DEARDORFF, E. R., KAWATACHI, J. et EBEL, G. D. (2011). West Nile Virus Genetic Diversity is Maintained during Transmission by Culex pipiens quinquefasciatus Mosquitoes. *Plos One*, 6(9).

[Brault *et al.*, 2004] BRAULT, A. C., FOY, B. D., MYLES, K. M., KELLY, C. L. H., HIGGS, S., WEAVER, S. C., OLSON, K. E., MILLER, B. R. et POWERS, A. M. (2004). Infection patterns of o'nyong nyong virus in the malaria-transmitting mosquito, Anopheles gambiae. *Insect molecular biology*, 13(6):625–635.

[Brown *et al.*, 2014] BROWN, J. E., EVANS, B. R., ZHENG, W., OBAS, V., BARRERA-MARTINEZ, L., EGIZI, A., ZHAO, H., CACCONE, A. et POWELL, J. R. (2014). Human impacts have shaped historical and recent evolution in Aedes aegypti, the dengue and yellow fever mosquito. *Evolution*, 68(2):514–525.

[Butler, 2016] BUTLER, D. (2016). **Fears rise over yellow fever's next move**. *Nature*, 532:155–156.

[Calzolari *et al.*, 2013] CALZOLARI, M., BONILAURI, P., BELLINI, R., ALBIERI, A., DEFILIPPO, F., TAMBA, M., TASSINARI, M., GELATI, A., CORDIOLI, P., ANGELINI, P. et DOTTORI, M. (2013). Usutu virus persistence and West Nile virus inactivity in the Emilia-Romagna region (Italy) in 2011. *Plos One*, 8(5):e63978.

[Calzolari *et al.*, 2010] CALZOLARI, M., BONILAURI, P., BELLINI, R., CAIMI, M., DEFILIPPO, F., MAIOLI, G., ALBIERI, A., MEDICI, A., VERONESI, R., PILANI, R., GELATI, A., ANGELINI, P., PARCO, V., FABBI, M., BARBIERI, I., LELLI, D., LAVAZZA, A., CORDIOLI, P. et DOTTORI, M. (2010). Arboviral survey of mosquitoes in two northern Italian regions in 2007 and 2008. *Vector borne and zoonotic diseases (Larchmont, N.Y.)*, 10(9):875–884.

[Calzolari *et al.*, 2015] CALZOLARI, M., ZÉ-ZÉ, L., VÁZQUEZ, A., SÁNCHEZ SECO, M. P., AMARO, F. et DOTTORI, M. (2015). Insect-specific flaviviruses, a worldwide widespread group of viruses only detected in insects. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*.

[Centers for Disease Control and Prevention, 2010] CENTERS FOR DISEASE CONTROL AND PREVENTION (2010). Arbovirus Catalog.

[Chapman *et al.*, 2014] CHAPMAN, E. G., MOON, S. L., WILUSZ, J. et KIEFT, J. S. (2014). RNA structures that resist degradation by Xrn1 produce a pathogenic Dengue virus RNA. *eLife*, 3:e01892.

[Chen *et al.*, 2015] CHEN, X.-G., JIANG, X., GU, J., XU, M., WU, Y., DENG, Y., ZHANG, C., BONIZZONI, M., DERMAUW, W., VONTAS, J., ARMBRUSTER, P., HUANG, X., YANG, Y., ZHANG, H., HE, W., PENG, H., LIU, Y., WU, K., CHEN, J., LIRAKIS, M., TOPALIS, P., VAN LEEUWEN, T., HALL, A. B., JIANG, X., THORPE, C., MUELLER, R. L., SUN, C., WATERHOUSE, R. M., YAN, G., TU, Z. J., FANG, X.

et JAMES, A. A. (2015). Genome sequence of the Asian Tiger mosquito, Aedes albopictus, reveals insights into its biology, genetics, and evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 112(44):E5907–15.

[Ciota *et al.*, 2013] CIOTA, A. T., EHRBAR, D. J., MATACCHIERO, A. C., VAN SLYKE, G. A. et KRAMER, L. D. (2013). The evolution of virulence of West Nile virus in a mosquito vector : implications for arbovirus adaptation and evolution. *BMC evolutionary biology*, 13(1):71.

[Ciota *et al.*, 2012] CIOTA, A. T., EHRBAR, D. J., VAN SLYKE, G. A., PAYNE, A. F., WILLSEY, G. G., VISCIO, R. E. et KRAMER, L. D. (2012). Quantification of intrahost bottlenecks of West Nile virus in Culex pipiens mosquitoes using an artificial mutant swarm. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 12(3):557–564.

[Ciota et Kramer, 2010] CIOTA, A. T. et KRAMER, L. D. (2010). Insights into arbovirus evolution and adaptation from experimental studies. *Viruses*, 2(12):2594–2617.

[Ciota *et al.*, 2007] CIOTA, A. T., NGO, K. A., LOVELACE, A. O., PAYNE, A. F., ZHOU, Y., SHI, P.-Y. et KRAMER, L. D. (2007). Role of the mutant spectrum in adaptation and replication of West Nile virus. *The Journal of general virology*, 88(Pt 3):865–874.

[Clarke *et al.*, 2015] CLARKE, B. D., ROBY, J. A., SLONCHAK, A. et KHROMYKH, A. A. (2015). Functional non-coding RNAs derived from the flavivirus 3' untranslated region. *Virus research*, 206:53–61.

[Coffey *et al.*, 2013] COFFEY, L. L., FORRESTER, N., TSETSARKIN, K., VASILAKIS, N. et WEAVER, S. C. (2013). Factors shaping the adaptive landscape for arboviruses : implications for the emergence of disease. *Future microbiology*, 8(2):155–176.

[Combe et Sanjuán, 2014] COMBE, M. et SANJUÁN, R. (2014). Variation in RNA virus mutation rates across host cells. *PLoS Pathogens*, 10(1):e1003855.

[Conway, 2015] CONWAY, M. J. (2015). Identification of a Flavivirus Sequence in a Marine Arthropod. *Plos One*, 10(12):e0146037.

[Cook *et al.*, 2013] COOK, S., CHUNG, B. Y. W., BASS, D., MOUREAU, G., TANG, S., MCALISTER, E., CULVERWELL, C. L., GLÜCKSMAN, E., WANG, H., BROWN, T. D. K., GOULD, E. A., HARBACH, R. E., DE LAMBALLERIE, X. et FIRTH, A. E. (2013). Novel virus discovery and genome reconstruction from field RNA samples reveals highly divergent viruses in dipteran hosts. *Plos One*, 8(11):e80720.

[Crochu *et al.*, 2004] CROCHU, S., COOK, S., ATTOUI, H., CHARREL, R. N., DE CHESSE, R., BELHOUCHET, M., LEMASSON, J.-J., de MICCO, P. et DE LAMBALLERIE, X. (2004). Sequences of flavivirus-related RNA viruses persist in DNA form integrated in the genome of Aedes spp. mosquitoes. *The Journal of general virology*, 85(Pt 7):1971–1980.

[Diaz *et al.*, 2012] DIAZ, L. A., FLORES, F. S., QUAGLIA, A. et CONTIGIANI, M. S. (2012). Intertwined arbovirus transmission activity : reassessing the transmission cycle paradigm. *Frontiers in physiology*, 3:493.

[Dick *et al.*, 1952] DICK, G. W. A., KITCHEN, S. F. et HADDOW, A. J. (1952). Zika virus. I. Isolations and serological specificity. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 46(5):509–520.

[Dickson *et al.*, 2014] DICKSON, L. B., SANCHEZ-VARGAS, I., SYLLA, M., FLEMING, K. et BLACK, W. C. (2014). Vector competence in West African Aedes aegypti Is Flavivirus species and genotype dependent. *PLoS neglected tropical diseases*, 8(10):e3153.

[Domingo et Holland, 1997] DOMINGO, E. et HOLLAND, J. J. (1997). RNA virus mutations and fitness for survival. *Annual review of microbiology*, 51(1):151–178.

[Domingo *et al.*, 2012] DOMINGO, E., SHELDON, J. et PERALES, C. (2012). Viral quasispecies evolution. *Microbiology and molecular biology reviews : MMBR*, 76(2):159–216.

[Duarte *et al.*, 1992] DUARTE, E., CLARKE, D., MOYA, A., DOMINGO, E. et HOLLAND, J. (1992). Rapid fitness losses in mammalian RNA virus clones due to Muller's ratchet. *Proceedings of the National Academy of Sciences of the United States of America*, 89(13):6015–6019.

[Duffy *et al.*, 2009] DUFFY, M. R., CHEN, T.-H., HANCOCK, W. T., POWERS, A. M., KOOL, J. L., LANCIOTTI, R. S., PRETRICK, M., MARFEL, M., HOLZBAUER, S., DUBRAY, C., GUILLAUMOT, L., GRIGGS, A., BEL, M., LAMBERT, A. J., LAVEN, J., KOSOY, O., PANELLA, A., BIGGERSTAFF, B. J., FISCHER, M. et HAYES, E. B. (2009). Zika virus outbreak on Yap Island, Federated States of Micronesia. *The New England journal of medicine*, 360(24):2536–2543.

[Ebel et Kramer, 2009] EBEL, G. D. et KRAMER, L. D. (2009). West Nile virus : molecular epidemiology and diversity. *In* DIAMOND, M. S., éditeur : *West Nile Encephalitis Virus Infection*, pages 25–43. New York.

[Fansiri *et al.*, 2013] FANSIRI, T., FONTAINE, A., DIANCOURT, L., CARO, V., THAISOMBOONSUK, B., RICHARDSON, J. H., JARMAN, R. G., PONLAWAT, A. et LAMBRECHTS, L. (2013). Genetic Mapping of Specific Interactions between Aedes aegypti Mosquitoes and Dengue Viruses. *PLoS Genet*, 9(8):e1003621.

[Fauci et Morens, 2016] FAUCI, A. S. et MORENS, D. M. (2016). Zika Virus in the Americas–Yet Another Arbovirus Threat. *The New England journal of medicine*, 374(7):601–604.

[Feng *et al.*, 2012] FENG, Y., FU, S., ZHANG, H., LI, M., ZHOU, T., WANG, J., ZHANG, Y., WANG, H., TANG, Q. et LIANG, G. (2012). Distribution of mosquitoes and mosquito-borne viruses along the China-Myanmar border in Yunnan Province. *Japanese journal of infectious diseases*, 65(3):215–221.

[Feschotte et Gilbert, 2012] FESCHOTTE, C. et GILBERT, C. (2012). Endogenous viruses : insights into viral evolution and impact on host biology. *Nature reviews. Genetics*, 13(4):283–296.

[Finlay, 1881] FINLAY, C. J. (1881). The Mosquito Hypothetically Considered as an Agent in the Transmission of Yellow Fever Poison. *The Yale journal of biology and medicine*, 9(6):589–604.

[Firth *et al.*, 2010] FIRTH, A. E., BLITVICH, B. J., WILLS, N. M., MILLER, C. L. et ATKINS, J. F. (2010). Evidence for ribosomal frameshifting and a novel overlapping gene in the genomes of insect-specific flaviviruses. *Virology*, 399(1): 153–166.

[Fitzpatrick *et al.*, 2010] FITZPATRICK, K. A., DEARDORFF, E. R., PESKO, K., BRACKNEY, D. E., ZHANG, B., BEDRICK, E., SHI, P.-Y. et EBEL, G. D. (2010). Population variation of West Nile virus confers a host-specific fitness benefit in mosquitoes. *Virology*, 404(1):89–95.

[Flegel, 2009] FLEGEL, T. W. (2009). Hypothesis for heritable, anti-viral immunity in crustaceans and insects. *Biology direct*, 4:32.

[Floridi, 2012] FLORIDI, L. (2012). Big Data and Their Epistemological Challenge. *Philosophy & Technology*, 25(4):435–437.

[Forrester *et al.*, 2014] FORRESTER, N. L., COFFEY, L. L. et WEAVER, S. C. (2014). Arboviral bottlenecks and challenges to maintaining diversity and fitness during mosquito transmission. *Viruses*, 6(10):3991–4004.

[Franz *et al.*, 2015] FRANZ, A. W. E., KANTOR, A. M., PASSARELLI, A. L. et CLEM, R. J. (2015). Tissue Barriers to Arbovirus Infection in Mosquitoes. *Viruses*, 7(7):3741–3767.

[Frické, 2014] FRICKÉ, M. (2014). Big data and its epistemology. *Journal of the Association for Information Science and Technology*, 66(4):651–661.

[Fujino *et al.*, 2014] FUJINO, K., HORIE, M., HONDA, T., MERRIMAN, D. K. et TOMONAGA, K. (2014). Inhibition of Borna disease virus replication by an endogenous bornavirus-like element in the ground squirrel genome. *Proceedings of the National Academy of Sciences of the United States of America*, 111(36):13175–13180.

[Gould *et al.*, 2003] GOULD, E. A., DE LAMBALLERIE, X., ZANOTTO, P. M. et HOLMES, E. C. (2003). Origins, evolution, and vector/host coadaptations within the genus Flavivirus. *Advances in virus research*, 59:277–314.

[Graham, 2012] GRAHAM, M. (2012). Big data and the end of theory ?

[Grard *et al.*, 2007] GRARD, G., MOUREAU, G., CHARREL, R. N., LEMASSON, J.-J., GONZALEZ, J.-P., GALLIAN, P., GRITSUN, T. S., HOLMES, E. C., GOULD, E. A. et DE LAMBALLERIE, X. (2007). Genetic characterization of tick-borne flaviviruses : new insights into evolution, pathogenetic determinants and taxonomy. *Virology*, 361(1):80–92.

[Gritsun et Gould, 2007] GRITSUN, T. S. et GOULD, E. A. (2007). Origin and evolution of flavivirus 5'UTRs and panhandles : trans-terminal duplications ? *Virology*, 366(1):8–15.

[Grubaugh *et al.*, 2016] GRUBAUGH, N. D., WEGER-LUCARELLI, J., MURRIETA, R. A., FAUVER, J. R., GARCIA-LUNA, S. M., PRASAD, A. N., BLACK, W. C. et EBEL, G. D. (2016). Genetic Drift during Systemic Arbovirus Infection of Mosquito Vectors Leads to Decreased Relative Fitness during Host Switching. *Cell host & microbe*.

[Gubler, 1998] GUBLER, D. J. (1998). Dengue and dengue hemorrhagic fever. *Clinical microbiology reviews*, 11(3):480–496.

[Gubler, 2001] GUBLER, D. J. (2001). Human arbovirus infections worldwide. *Annals of the New York Academy of Sciences*, 951:13–24.

[Gubler, 2014] GUBLER, D. J. (2014). Dengue viruses : their evolution, history and emergence as a global public health problem. *In* GUBLER, D. J., OOI, E. E., VASUDEVAN, S. et FARRAR, J., éditeurs : *Dengue and dengue hemorrhagic fever*, pages 1–29. CABI, Wallingford.

[Gubler *et al.*, 1979] GUBLER, D. J., NALIM, S., TAN, R., SAIPAN, H. et SULIANTI SAROSO, J. (1979). Variation in Susceptibility to Oral Infection with Dengue Viruses Among Geographic Strains of Aedes-Aegypti. *American Journal of Tropical Medicine and Hygiene*, 28(6):1045–1052.

[Gutiérrez *et al.*, 2012] GUTIÉRREZ, S., MICHALAKIS, Y. et BLANC, S. (2012). Virus population bottlenecks during within-host progression and

host-to-host transmission. *Current opinion in virology*, 2(5):546–555.

[Hadinegoro *et al.*, 2015] HADINEGORO, S. R., ARREDONDO-GARCÍA, J. L., CAPEDING, M. R., DESEDA, C., CHOTPITAYASUNONDH, T., DIETZE, R., MUHAMMAD ISMAIL, H. I. H., REYNALES, H., LIMKITTIKUL, K., RIVERA-MEDINA, D. M., TRAN, H. N., BOUCKENOOGHE, A., CHANSINGHAKUL, D., CORTÉS, M., FANOUILLERE, K., FORRAT, R., FRAGO, C., GAILHARDOU, S., JACKSON, N., NORIEGA, F., PLENNEVAUX, E., WARTEL, T. A., ZAMBRANO, B., SAVILLE, M. et CYD-TDV DENGUE VACCINE WORKING GROUP (2015). Efficacy and Long-Term Safety of a Dengue Vaccine in Regions of Endemic Disease. *The New England journal of medicine*, 373(13):1195–1206.

[Hardy *et al.*, 1983] HARDY, J. L., HOUK, E. J., KRAMER, L. D. et REEVES, W. C. (1983). Intrinsic factors affecting vector competence of mosquitoes for arboviruses. *Annual review of entomology*, 28(1):229–262.

[Higgs, 2004] HIGGS, S. (2004). How do mosquito vectors live with their viruses ? *In* GILLESPIE, S. H., SMITH, G. L. et OSBOURN, A., éditeurs : *Microbe-vector interactions in vector-borne diseases*, pages 103–137. New York.

[Higgs et Beaty, 2005] HIGGS, S. et BEATY, B. J. (2005). Natural Cycles of Vector-Borne Pathogens. *In* MARQUARDT, W. C., éditeur : *Biology of Disease Vectors*, pages 167–185.

[Holmes, 2011] HOLMES, E. C. (2011). The evolution of endogenous viral elements. *Cell host & microbe*, 10(4):368–377.

[Holmes et Twiddy, 2003] HOLMES, E. C. et TWIDDY, S. S. (2003). The origin, emergence and evolutionary genetics of dengue virus. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 3(1):19–28.

[Jackson et Charleston, 2004] JACKSON, A. P. et CHARLESTON, M. A. (2004). A cophylogenetic perspective of RNA-virus evolution. *Molecular biology and evolution*, 21(1):45–57.

[Jenkins *et al.*, 2002] JENKINS, G. M., RAMBAUT, A., PYBUS, O. G. et HOLMES, E. C. (2002). Rates of molecular evolution in RNA viruses : A quantitative phylogenetic analysis. *Journal of Molecular Evolution*, 54(2):156–165.

[Jerzak *et al.*, 2005] JERZAK, G., BERNARD, K. A., KRAMER, L. D. et EBEL, G. D. (2005). Genetic variation in West Nile virus from naturally infected mosquitoes and birds suggests quasispecies structure and strong purifying selection. *The Journal of general virology*, 86(Pt 8):2175–2183.

[Jerzak *et al.*, 2007] JERZAK, G. V. S., BERNARD, K., KRAMER, L. D., SHI, P.-Y. et EBEL, G. D. (2007). The West Nile virus mutant spectrum is host-dependant and a determinant of mortality in mice. *Virology*, 360(2):469–476.

[Jin *et al.*, 2011] JIN, Z., DEVAL, J., JOHNSON, K. A. et SWINNEY, D. C. (2011). Characterization of the elongation complex of dengue virus RNA polymerase : assembly, kinetics of nucleotide incorporation, and fidelity. *The Journal of biological chemistry*, 286(3):2067–2077.

[Katzourakis et Gifford, 2010] KATZOURAKIS, A. et GIFFORD, R. J. (2010). Endogenous viral elements in animal genomes. *PLoS Genet*, 6(11): e1001191.

[Kelling *et al.*, 2009] KELLING, S., HOCHACHKA, W. M., FINK, D., RIEDEWALD, M., CARUANA, R., BALLARD, G. et HOOKER, G. (2009). Data-intensive Science : A New Paradigm for Biodiversity Studies. *BioScience*, 59(7):613–620.

[Kemenesi *et al.*, 2014] KEMENESI, G., KRTINIĆ, B., MILANKOV, V., KUTAS, A., DALLOS, B., OLDAL, M., SOMOGYI, N., NEMETH, V., BANYAI, K. et JAKAB, F. (2014). West Nile virus surveillance in mosquitoes, April to October 2013, Vojvodina province, Serbia : implications for the 2014 season. *Euro surveillance : bulletin Européen sur les maladies transmissibles = European communicable disease bulletin*, 19(16):20779.

[Kieft *et al.*, 2015] KIEFT, J. S., RABE, J. L. et CHAPMAN, E. G. (2015). New hypotheses derived from the structure of a flaviviral Xrn1-resistant RNA : Conservation, folding, and host adaptation. *RNA biology*, 12(11):1169–1177.

[Kilpatrick, 2011] KILPATRICK, A. M. (2011). Globalization, land use, and the invasion of West Nile virus. *Science*, 334(6054):323–327.

[Kitchin, 2014] KITCHIN, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big data & society*, 1(1):2053951714528481.

[Kobayashi *et al.*, 2013] KOBAYASHI, K., ATSUMI, G., IWADATE, Y., TOMITA, R., CHIBA, K.-i., AKASAKA, S., NISHIHARA, M., TAKAHASHI, H., YAMAOKA, N., NISHIGUCHI, M. et SEKINE, K.-T. (2013). Gentian Kobu-sho-associated virus : a tentative, novel double-stranded RNA virus that is relevant to gentian Kobu-sho syndrome. *Journal of general plant pathology*.

[Kramer et Ebel, 2003] KRAMER, L. D. et EBEL, G. D. (2003). Dynamics of flavivirus infection in mosquitoes. *Advances in virus research*, 60:187–232.

[Kulasekera *et al.*, 2001] KULASEKERA, V. L., KRAMER, L., NASCI, R. S., MOSTASHARI, F., CHERRY, B., TROCK, S. C., GLASER, C. et MILLER, J. R. (2001). West Nile virus infection in mosquitoes, birds, horses, and humans, Staten Island, New York, 2000. *Emerging infectious diseases*, 7(4):722–725.

[Kuno, 2009] KUNO, G. (2009). Emergence of the Severe Syndrome and Mortality Associated with Dengue and Dengue-Like Illness : Historical Records (1890 to 1950) and Their Compatibility with Current Hypotheses on the Shift of Disease Manifestation. *Clinical microbiology reviews*, 22(2):186–201.

[Kuno et Chang, 2005] KUNO, G. et CHANG, G.-J. J. (2005). Biological transmission of arboviruses : reexamination of and new insights into components, mechanisms, and unique traits as well as their evolutionary trends. *Clinical microbiology reviews*, 18(4):608–637.

[Lambrechts, 2011] LAMBRECHTS, L. (2011). Quantitative genetics of Aedes aegypti vector competence for dengue viruses : towards a new paradigm ? *Trends in parasitology*, 27(3):111–114.

[Lambrechts *et al.*, 2009] LAMBRECHTS, L., CHEVILLON, C., ALBRIGHT, R. G., THAISOMBOONSUK, B., RICHARDSON, J. H., JARMAN, R. G. et SCOTT, T. W. (2009). Genetic specificity and potential for local adaptation between dengue viruses and mosquito vectors. *BMC evolutionary biology*, 9(1).

[Lambrechts *et al.*, 2013] LAMBRECHTS, L., QUILLERY, E., NOEL, V., RICHARDSON, J. H., JARMAN, R. G., SCOTT, T. W. et CHEVILLON, C. (2013). Specificity of resistance to dengue virus isolates is associated with genotypes of the mosquito antiviral gene Dicer-2. *Proceedings of the Royal Society B-Biological Sciences*, 280(1751):–20122437.

[Lambrechts et Scott, 2009] LAMBRECHTS, L. et SCOTT, T. W. (2009). Mode of transmission and the evolution of arbovirus virulence in mosquito vectors. *Proceedings of the Royal Society B-Biological Sciences*, 276(1660):1369–1378.

[Lauring *et al.*, 2013] LAURING, A. S., FRYDMAN, J. et ANDINO, R. (2013). The role of mutational robustness in RNA virus evolution. *Nature reviews. Microbiology*, 11(5):327–336.

[Leake, 1984] LEAKE, C. J. (1984). Transovarial Transmission of Arboviruses by Mosquitoes. *In Vectors in virus biology*, pages 63–91. Academic Press, New-York & London.

[Leonelli, 2012] LEONELLI, S. (2012). Introduction : Making sense of data-driven research in the biological and biomedical sciences. *Studies in history and philosophy of biological and biomedical sciences*, 43(1):1–3.

[Li et Chen, 2014] LI, Y. et CHEN, L. (2014). Big biological data : challenges and opportunities. *Genomics, proteomics & bioinformatics*, 12(5): 187–189.

[Liang *et al.*, 2015] LIANG, W., HE, X., LIU, G., ZHANG, S., FU, S., WANG, M., CHEN, W., HE, Y., TAO, X., JIANG, H., LIN, X., GAO, X., HU, W., LIU, Y., FENG, L., CAO, Y., YANG, G., JING, C., LIANG, G. et WANG, H. (2015). Distribution and phylogenetic analysis of Culex flavivirus in mosquitoes in China. *Archives of virology*, 160(9):2259–2268.

[Lindenbach *et al.*, 2013] LINDENBACH, B. D., MURRAY, C. L., THIEL, H.-J. et RICE, C. M. (2013). *Flaviviridae*. *In* KNIPE, D. M. et HOWLEY, P. M., éditeurs : *Fields Virology*, pages 712–746. Lippincott Williams & Wilkins, Philadelphia.

[Liu *et al.*, 2013] LIU, H., LU, H.-J., LIU, Z.-J., JING, J., REN, J.-Q., LIU, Y.-Y., LU, F. et JIN, N.-Y. (2013). Japanese encephalitis virus in mosquitoes and swine in Yunnan province, China 2009-2010. *Vector borne and zoonotic diseases (Larchmont, N.Y.)*, 13(1):41–49.

[Manokaran *et al.*, 2015] MANOKARAN, G., FINOL, E., WANG, C., GUNARATNE, J., BAHL, J., ONG, E. Z., TAN, H. C., SESSIONS, O. M., WARD, A. M., GUBLER, D. J., HARRIS, E., GARCIA-BLANCO, M. A. et OOI, E. E. (2015). Dengue subgenomic RNA binds TRIM25 to inhibit interferon expression for epidemiological fitness. *Science*, 350(6257):217–221.

[Marchoux et Simond, 1906] MARCHOUX, É. et SIMOND, P.-L. (1906). Etudes sur la fièvre jaune - Deuxième mémoire. *Annales de l'Institut Pasteur*, 20(1):1–25.

[Marr et Calisher, 2003] MARR, J. S. et CALISHER, C. H. (2003). Alexander the Great and West Nile virus encephalitis. *Emerging infectious diseases*, 9(12):1599–1603.

[Mazzocchi, 2015] MAZZOCCHI, F. (2015). Could Big Data be the end of theory in science ? A few remarks on the epistemology of data-driven science. *EMBO reports*, 16(10):1250–1255.

[Melian *et al.*, 2010] MELIAN, E. B., HINZMAN, E., NAGASAKI, T., FIRTH, A. E., WILLS, N. M., NOUWENS, A. S., BLITVICH, B. J., LEUNG, J., FUNK, A., ATKINS, J. F., HALL, R. et KHROMYKH, A. A. (2010). NS1' of flaviviruses in the Japanese encephalitis virus serogroup is a product of ribosomal frameshifting and plays a role in viral neuroinvasiveness. *Journal of Virology*, 84(3):1641–1647.

[Moon *et al.*, 2015] MOON, S. L., DODD, B. J. T., BRACKNEY, D. E., WILUSZ, C. J., EBEL, G. D. et WILUSZ, J. (2015). Flavivirus sfRNA suppresses antiviral RNA interference in cultured cells and mosquitoes and directly interacts with the RNAi machinery. *Virology*, 485:322–329.

[Moureau *et al.*, 2015] MOUREAU, G., COOK, S., LEMEY, P., NOUGAIREDE, A., FORRESTER, N. L., KHASNATINOV, M., CHARREL, R. N., FIRTH, A. E., GOULD, E. A. et DE LAMBALLERIE, X. (2015).

New insights into flavivirus evolution, taxonomy and biogeographic history, extended by analysis of canonical and alternative coding sequences. *Plos One*, 10(2):e0117849.

[Mukhopadhyay *et al.*, 2005] MUKHOPADHYAY, S., KUHN, R. J. et ROSSMANN, M. G. (2005). A structural perspective of the flavivirus life cycle. *Nature reviews. Microbiology*, 3(1):13–22.

[Musso *et al.*, 2015] MUSSO, D., CAO-LORMEAU, V.-M. et GUBLER, D. J. (2015). Zika virus : following the path of dengue and chikungunya ? *Lancet (London, England)*, 386(9990):243–244.

[Nishiura et Halstead, 2007] NISHIURA, H. et HALSTEAD, S. B. (2007). Natural History of Dengue Virus (DENV)–1 and DENV-4 Infections : Reanalysis of Classic Studies. *The Journal of Infectious Diseases*, 195(7):1007–1013.

[Nogueira, 2009] NOGUEIRA, P. (2009). The Early History of Yellow Fever. *In Yellow fever, a symposium in commemoration of Carlos Juan Finlay,*, pages 1–15.

[Pialoux *et al.*, 2007] PIALOUX, G., GAÜZÈRE, B.-A., JAURÉGUIBERRY, S. et STROBEL, M. (2007). Chikungunya, an epidemic arbovirosis. *The Lancet. Infectious diseases*, 7(5):319–327.

[Pigliucci, 2009] PIGLIUCCI, M. (2009). The end of theory in science ? *EMBO reports*, 10(6):534–534.

[Pugachev *et al.*, 2004] PUGACHEV, K. V., GUIRAKHOO, F., OCRAN, S. W., MITCHELL, F., PARSONS, M., PENAL, C., GIRAKHOO, S., POUGATCHEVA, S. O., ARROYO, J., TRENT, D. W. et MONATH, T. P. (2004). High fidelity of yellow fever virus RNA polymerase. *Journal of Virology*, 78(2):1032–1038.

[Pybus *et al.*, 2002] PYBUS, O. G., RAMBAUT, A., HOLMES, E. C. et HARVEY, P. H. (2002). New inferences from tree shape : numbers of missing taxa and population growth rates. *Systematic biology*, 51(6):881–888.

[Qin *et al.*, 2014] QIN, X.-C., SHI, M., TIAN, J.-H., LIN, X.-D., GAO, D.-Y., HE, J.-R., WANG, J.-B., LI, C.-X., KANG, Y.-J., YU, B., ZHOU, D.-J., XU, J., PLYUSNIN, A., HOLMES, E. C. et ZHANG, Y.-Z. (2014). A tick-borne segmented RNA virus contains genome segments derived from unsegmented viral ancestors. *Proceedings of the National Academy of Sciences of the United States of America*, 111(18):6744–6749.

[Reed *et al.*, 1901] REED, W., CARROLL, J. et AGRAMONTE, A. (1901). Experimental yellow fever. *American Medicine*, 2:15–23.

[Reeves, 2004] REEVES, W. C. (2004). Overwintering of arboviruses. *In Epidemiology and Control of Mosquito-borne Arboviruses in California*, pages 357–382. Calif Mosq Vector Control Assoc, Sacramento, California.

[Reidenbach *et al.*, 2009] REIDENBACH, K. R., COOK, S., BERTONE, M. A., HARBACH, R. E., WIEGMANN, B. M. et BESANSKY, N. J. (2009). Phylogenetic analysis and temporal diversification of mosquitoes (Diptera : Culicidae) based on nuclear genes and morphology. *BMC evolutionary biology*, 9(1):298.

[Roby *et al.*, 2014] ROBY, J. A., PIJLMAN, G. P., WILUSZ, J. et KHROMYKH, A. A. (2014). Noncoding subgenomic flavivirus RNA : multiple functions in West Nile virus pathogenesis and modulation of host responses. *Viruses*, 6(2):404–427.

[Roiz *et al.*, 2009] ROIZ, D., VÁZQUEZ, A., SECO, M. P. S., TENORIO, A. et RIZZOLI, A. (2009). Detection of novel insect flavivirus sequences integrated in Aedes albopictus (Diptera : Culicidae) in Northern Italy. *Virology journal*, 6:93.

[Root *et al.*, 2006] ROOT, J. J., OESTERLE, P. T., NEMETH, N. M., KLENK, K., GOULD, D. H., MCLEAN,

R. G., CLARK, L. et HALL, J. S. (2006). Experimental infection of fox squirrels (Sciurus niger) with West Nile virus. *American Journal of Tropical Medicine and Hygiene*, 75(4):697–701.

[Root *et al.*, 2007] ROOT, J. J., OESTERLE, P. T., SULLIVAN, H. J., HALL, J. S., MARLENEE, N. L., MCLEAN, R. G., MONTENIERI, J. A. et CLARK, L. (2007). Fox squirrel (Sciurus niger) associations with West Nile virus. *American Journal of Tropical Medicine and Hygiene*, 76(4):782–784.

[Rosen *et al.*, 1985] ROSEN, L., ROSEBOOM, L. E., GUBLER, D. J., LIEN, J. C. et CHANIOTIS, B. N. (1985). Comparative Susceptibility of Mosquito Species and Strains to Oral and Parenteral Infection with Dengue and Japanese Encephalitis Viruses. *American Journal of Tropical Medicine and Hygiene*, 34(3):603–615.

[Salazar *et al.*, 2007] SALAZAR, M. I., RICHARDSON, J. H., SANCHEZ-VARGAS, I., OLSON, K. E. et BEATY, B. J. (2007). Dengue virus type 2 : replication and tropisms in orally infected Aedes aegypti mosquitoes. *Bmc Microbiology*, 7(1).

[Sang *et al.*, 2003] SANG, R. C., GICHOGO, A., GACHOYA, J., DUNSTER, M. D., OFULA, V., HUNT, A. R., CRABTREE, M. B., MILLER, B. R. et DUNSTER, L. M. (2003). Isolation of a new flavivirus related to cell fusing agent virus (CFAV) from field-collected flood-water Aedes mosquitoes sampled from a dambo in central Kenya. *Archives of virology*, 148(6):1085–1093.

[Sanjuán *et al.*, 2010] SANJUÁN, R., NEBOT, M. R., CHIRICO, N., MANSKY, L. M. et BELSHAW, R. (2010). Viral mutation rates. *Journal of Virology*, 84(19):9733–9748.

[Sanofi Pasteur, 2016] SANOFI PASTEUR (2016). Premières vaccinations contre la dengue : un moment historique dans la prévention des maladies infectieuses. Sanofi Pasteur.

[Selisko *et al.*, 2014] SELISKO, B., WANG, C., HARRIS, E. et CANARD, B. (2014). Regulation of Flavivirus RNA synthesis and replication. *Current opinion in virology*, 9:74–83.

[Shi *et al.*, 2015] SHI, M., LIN, X.-D., VASILAKIS, N., TIAN, J.-H., LI, C.-X., CHEN, L.-J., EASTWOOD, G., DIAO, X.-N., CHEN, M.-H., CHEN, X., QIN, X.-C., WIDEN, S. G., WOOD, T. G., TESH, R. B., XU, J., HOLMES, E. C. et ZHANG, Y.-Z. (2015). Divergent Viruses Discovered in Arthropods and Vertebrates Revise the Evolutionary History of the Flaviviridae and Related Viruses. *Journal of Virology*, 90(2):659–669.

[Sim *et al.*, 2015] SIM, S., AW, P. P. K., WILM, A., TEOH, G., HUE, K. D. T., NGUYEN, N. M., NAGARAJAN, N., SIMMONS, C. P. et HIBBERD, M. L. (2015). Tracking Dengue Virus Intra-host Genetic Diversity during Human-to-Mosquito Transmission. *PLoS neglected tropical diseases*, 9(9): e0004052.

[Snow *et al.*, 2014] SNOW, G. E., HAALAND, B., OOI, E. E. et GUBLER, D. J. (2014). Review article : Research on dengue during World War II revisited. *American Journal of Tropical Medicine and Hygiene*, 91(6):1203–1217.

[Staples *et al.*, 2010] STAPLES, J. E., GERSHMAN, M., FISCHER, M. et CENTERS FOR DISEASE CONTROL AND PREVENTION (CDC) (2010). Yellow fever vaccine : recommendations of the Advisory Committee on Immunization Practices (ACIP).

[Stephens *et al.*, 2015] STEPHENS, Z. D., LEE, S. Y., FAGHRI, F., CAMPBELL, R. H., ZHAI, C., EFRON, M. J., IYER, R., SCHATZ, M. C., SINHA, S. et ROBINSON, G. E. (2015). Big Data : Astronomical or Genomical ? *PLoS biology*, 13(7):e1002195.

[Stollar et Thomas, 1975] STOLLAR, V. et THOMAS, V. L. (1975). An agent in the Aedes aegypti cell

line (Peleg) which causes fusion of Aedes albopictus cells. *Virology*, 64(2):367–377.

[Styer *et al.*, 2007] STYER, L. M., MEOLA, M. A. et KRAMER, L. D. (2007). West Nile virus infection decreases fecundity of Culex tarsalis females. *Journal of medical entomology*, 44(6): 1074–1085.

[Tardieux *et al.*, 1990] TARDIEUX, I., POUPEL, O., LAPCHIN, L. et RODHAIN, F. (1990). Variation Among Strains of Aedes-Aegypti in Susceptibility to Oral Infection with Dengue Virus Type-2. *American Journal of Tropical Medicine and Hygiene*, 43(3):308–313.

[Teixeira *et al.*, 2016] TEIXEIRA, M., SELA, N., NG, J., CASTEEL, C. L., PENG, H.-C., BEKAL, S., GIRKE, T., GHANIM, M. et KALOSHIAN, I. (2016). A novel virus from Macrosiphum euphorbiae with similarities to members of the family Flaviviridae. *The Journal of general virology*.

[Tesh, 1984] TESH, R. B. (1984). Transovarial Transmission of Arboviruses in Their Invertebrate Vectors. *In Current topics in vector research*, pages 57–76. New-York, NY : Praeger.

[Turell, 1988] TURELL, M. J. (1988). Horizontal and vertical transmission of viruses by insect and tick vectors. *In The Arboviruses*, pages 127–152. CRC PressI Llc, Boca Raton, Florida.

[Turell *et al.*, 1982] TURELL, M. J., HARDY, J. L. et REEVES, W. C. (1982). Stabilized Infection of California Encephalitis Virus in Aedes dorsalis, and Its Implications for Viral Maintenance in Nature. *American Journal of Tropical Medicine and Hygiene*, 31(6):1252–1259.

[Varelas-Wesley et Calisher, 1982] VARELAS-WESLEY, I. et CALISHER, C. H. (1982). Antigenic relationships of flaviviruses with undetermined arthropod-borne status. *American Journal of Tropical Medicine and Hygiene*, 31(6):1273–1284.

[Vasilakis *et al.*, 2011] VASILAKIS, N., CARDOSA, J., HANLEY, K. A., HOLMES, E. C. et WEAVER, S. C. (2011). Fever from the forest : prospects for the continued emergence of sylvatic dengue virus and its impact on public health. *Nature reviews. Microbiology*, 9(7):532–541.

[Vazeille-Falcoz *et al.*, 1999] VAZEILLE-FALCOZ, M., MOUSSON, L., RODHAIN, F., CHUNGUE, E. et FAILLOUX, A.-B. (1999). Variation in oral susceptibility to dengue type 2 virus of populations of Aedes aegypti from the islands of Tahiti and Moorea, French Polynesia. *American Journal of Tropical Medicine and Hygiene*, 60(2):292–299.

[Villordo *et al.*, 2015] VILLORDO, S. M., FILOMATORI, C. V., SANCHEZ-VARGAS, I., BLAIR, C. D. et GAMARNIK, A. V. (2015). Dengue virus RNA structure specialization facilitates host adaptation. *PLoS Pathogens*, 11(1):e1004604.

[Wargo et Kurath, 2012] WARGO, A. R. et KURATH, G. (2012). Viral fitness : definitions, measurement, and current insights. *Current opinion in virology*, 2(5):538–545.

[Weaver, 2006] WEAVER, S. C. (2006). Evolutionary influences in arboviral disease. *Quasispecies : Concept and Implications for Virology*, 299:285–314.

[Weaver et Barrett, 2004] WEAVER, S. C. et BARRETT, A. D. T. (2004). Transmission cycles, host range, evolution and emergence of arboviral disease. *Nature reviews. Microbiology*, 2(10):789–801.

[Webster *et al.*, 2015] WEBSTER, C. L., WALDRON, F. M., ROBERTSON, S., CROWSON, D., FERRARI, G., QUINTANA, J. F., BROUQUI, J.-M., BAYNE, E. H., LONGDON, B., BUCK, A. H., LAZZARO, B. P., AKORLI, J., HADDRILL, P. R. et OBBARD, D. J.

(2015). The Discovery, Distribution, and Evolution of Viruses Associated with Drosophila melanogaster. *PLoS biology*, 13(7):e1002210.

[Welsch *et al.*, 2009] WELSCH, S., MILLER, S., ROMERO-BREY, I., MERZ, A., BLECK, C. K. E., WALTHER, P., FULLER, S. D., ANTONY, C., KRIJNSE-LOCKER, J. et BARTENSCHLAGER, R. (2009). Composition and three-dimensional architecture of the dengue virus replication and assembly sites. *Cell host & microbe*, 5(4):365–375.

[White, 2008] WHITE, G. B. (2008). Mosquitoes. *In* COOK, G. C. et ZUMLA, A., éditeurs : *Manson's Tropical Diseases*, pages 139–180.

[World Health Organization, 2009] WORLD HEALTH ORGANIZATION (2009). Dengue : guidelines for diagnosis, treatment, prevention and control. New edition.

[World Health Organization, 2015] WORLD HEALTH ORGANIZATION (2015). Japanese encephalitis - Fact Sheet.

[World Health Organization, 2016] WORLD HEALTH ORGANIZATION (2016). Déclaration sur la 1ère réunion du Comité d'urgence du Règlement sanitaire international concernant le virus Zika.

[Zuo *et al.*, 2014] ZUO, S., ZHAO, Q., GUO, X., ZHOU, H., CAO, W. et ZHANG, J. (2014). Detection of Quang Binh virus from mosquitoes in China. *Virus research*, 180:31–38.

Annexes Partie II

A Discovery of flavivirus-derived endogenous viral elements in two *Anopheles* mosquito genomes supports the existence of *Anopheles*-associated insect-specific flaviviruses

Sebastian Lequime & Louis Lambrechts. Discovery of flavivirus-derived endogenous viral elements in two *Anopheles* mosquito genomes supports the existence of *Anopheles*-associated insect-specific flaviviruses. *In prep*, 2016

Abstract

The *Flavivirus* genus encompasses several arboviruses of public health significance such as dengue, yellow fever, and Zika viruses. It also includes insect-specific flaviviruses (ISFs) that are only capable of infecting insect hosts. The vast majority of mosquito-infecting flaviviruses have been associated with mosquito species of the *Aedes* and *Culex* genera in the Culicinae subfamily, which also includes most arbovirus vectors. Mosquitoes of the Anophelinae subfamily are not considered significant arbovirus vectors, however flaviviruses have occasionally been detected in field-caught *Anopheles* specimens. Whether such observations reflect occasional spillover or laboratory contamination or whether *Anopheles* mosquitoes are natural hosts of flaviviruses is unknown. Here, we provide *in silico* and *in vivo* evidence of transcriptionally active, flavivirus-derived endogenous viral elements (EVEs) in the

genome of *Anopheles minimus* and *Anopheles sinensis*. Such non-retroviral endogenization of RNA viruses is consistent with a shared evolutionary history between flaviviruses and *Anopheles* mosquitoes. Phylogenetic analyses of these two newly described EVEs support the existence of a distinct *Anopheles*-associated clade of ISFs.

1 Introduction

Flaviviruses are positive single stranded RNA viruses that infect a broad range of hosts including vertebrates (e.g., birds, primates) and arthropods (e.g., ticks, mosquitoes). In addition to major arboviruses of public health significance such as dengue, Zika, West Nile and yellow fever viruses, the *Flavivirus* genus also includes vertebrate-specific (not known vector; NKV) and insect-specific (ISF; insect-specific flavivirus) members [Moureau et al., 2015]. The majority of mosquito-infecting flaviviruses have been associated with members of the Culicinae subfamily, mainly from the *Culex* and *Aedes* genera. *Anopheles* mosquitoes in the Anophelinae subfamily are well known for their role in the transmission of human malaria parasites, but they are not considered important vectors of arboviruses in general, and of flaviviruses in particular. Nevertheless, several studies have detected flaviviruses in field-caught *Anopheles* mosquitoes from different parts of the world. In North America, West Nile virus (WNV) was detected in *Anopheles punctipennis* [Bernard et al., 2001, Kulasekera et al., 2001]. In Asia, Japanese encephalitis virus was detected in *Anopheles sinensis* [Feng et al., 2012, Liu et al., 2013]. In Europe, *Anopheles maculipennis* was found positive for WNV [Filipe, 1972, Kemenesi et al., 2014], Usutu virus [Calzolari et al., 2013] and Batai virus [Calzolari et al., 2010]. Interestingly, some ISFs were also detected in *An. sinensis* [Zuo et al., 2014, Liang et al., 2015] and *Anopheles atroparvus* [Aranda et al., 2009]. It is unknown whether these detections reflect from occasional spillover or laboratory contamination, or whether *Anopheles* mosquitoes are in fact natural hosts of flaviviruses.

Endogenous viral elements (EVEs) are chromosomal integrations of partial or full viral genetic material into the genome of a host species. Not only retroviruses, whose replication cycle includes incorporation of a DNA form of the RNA viral genome into the host cell genome, but virtually all types of eukaryotic viruses can become endogenous [Feschotte and Gilbert, 2012].Non-retroviral EVEs have been documented in the genome of a wide variety of host species, including vertebrates and arthro-

pods [Feschotte and Gilbert, 2012]. Unlike detection of exogenous viruses, subject to possible laboratory or environmental contamination, EVEs are likely to reflect a long-lasting evolutionary relationship between an RNA virus and its natural host. This is because endogenization, for a single-stranded RNA virus, requires (1) reverse transcription, (2) integration of the virus-derived DNA into the genome of germinal host cells and (3) fixation of the integrated sequence in the host population [Holmes, 2011, Aiewsakun and Katzourakis, 2015]. The low probability of this combination of events makes endogenization exceedingly unlikely to occur unless the viral infection is common in the host population over long evolutionary times. For example, flavivirus-derived EVEs have been reported in the genome of *Aedes aegypti* [Crochu et al., 2004, Katzourakis and Gifford, 2010] and *Aedes albopictus* [Roiz et al., 2009, Chen et al., 2015]. These EVEs are phylogenetically related to the clade of *Aedes*-associated ISFs [Crochu et al., 2004, Katzourakis and Gifford, 2010, Roiz et al., 2009], which is consistent with the ancient evolutionary relationship between *Aedes* mosquitoes and ISFs.

Here, we report the discovery of two flavivirus-derived EVEs in the genomes of *An. minimus* and *An. sinensis* mosquitoes. We screened 21 publicly available *Anopheles* genomes [Holt et al., 2002, Neafsey et al., 2015] for flaviviral sequences, and validated *in silico* hits both at the DNA and RNA levels *in vivo*. The two newly described flavivirus-derived EVEs are phylogenetically related to ISFs, and support the existence of a previously unsuspected *Anopheles*-associated clade of ISFs.

2 Material and Methods

2.1 *In silico* survey

Genome screen

Twenty-one *Anopheles* genomes (full list and accession numbers are provided in Table 1) were screened by tBLASTn (BLAST+ 2.2.28) [Camacho et al., 2009] for the presence of flavivirus-derived EVEs using a collection of 50 full flavivirus polyproteins queries (full list and accession numbers are provided in Table S1) representing the currently known diversity of the *Flavivirus* genus. The sequences of hits whose E-value was > E^{-4} were extracted using an in-house bash shell script. In order to reconstruct

putative flavivirus-derived EVEs, BLAST hits were clustered using CD-HIT v.4.6.1 [Li and Godzik, 2006] and aligned using MAFFT v7.017 [Katoh et al., 2002]. Putative EVEs were extracted and used as query for a reciprocal tBLASTn (BLAST+ 2.2.30) against the National Center for Biotechnology Information (NCBI) nucleotide database (E-value > Eup-4). Genetic features of identified EVE were analyzed using the NCBI Conserved Domain Database [Marchler-Bauer et al., 2015].

Species	**GenBank WGS Project**	**Assembly**	**GenBank Assembly ID**
Anopheles albimanus	APCK01	AalbS1	GCA_000349125.1
Anopheles arabiensis	APCN01	AaraD1	GCA_000349185.1
Anopheles atroparvus	AXCP01	AatrE1	GCA_000473505.1
Anopheles christyi	APCM01	AchrA1	GCA_000349165.1
Anopheles coluzzii	ABKP02	AcolM1	GCA_000150765.1
Anopheles culicifacies A	AXCM01	AculA1	GCA_000473375.1
Anopheles darlingi	ADMH02	AdarC3	GCA_000211455.3
Anopheles dirus A	APCL01	AdirW1	GCA_000349145.1
Anopheles epiroticus	APCJ01	AepiE1	GCA_000349105.1
Anopheles farauti	AXCN01	AfarF1	GCA_000473445.1
Anopheles funestus	APCI01	AfunF1	GCA_000349085.1
Anopheles gambiae	AAAB01	AgamP4	GCA_000005575.2
Anopheles maculatus B	AXCL01	AmacM1	GCA_000473185.1
Anopheles melas	ACXO01	AmelC1	GCA_000473525.1
Anopheles merus	AXCQ01	AmerM1	GCA_000473845.1
Anopheles minimus A	APHL01	AminM1	GCA_000349025.1
Anopheles nili	ATLZ01	Anili1	GCA_000439205.1
Anopheles quadriannulatus A	APCH01	AquaS1	GCA_000349065.1
Anopheles sinensis	AXCK01	AsinS1	GCA_000472065.1
Anopheles stephensi: Indian	ALPR002	AsteI2	GCA_000300775.2
Anopheles stephensi: SDA-500	APCG01	AsteS1	GCA_000349045.1

Table 1 – *Anopheles* genomes screened in this study.

Phylogenetic analyses

Translated EVE sequences were aligned to the corresponding sections of several flaviviral polyproteins (Table S1) with MAFFT v7.017 and phylogenetically uninformative positions were trimmed using TrimAI v.1.3 [Capella-Gutiérrez et al., 2009] accessed through the webserver Phylemon 2 [Sánchez et al., 2011]. The trimmed alignments were used to construct phylogenetic trees with PhyML Best AIC Tree [Sánchez et al., 2011]. Best substitution models were Blosum62+I+G+F for *An. minimus* and Blosum62+I+G for *An. sinensis*.

Transcriptome screen

Published RNA-sequencing (RNA-seq) data were retrieved from NCBI Sequence Read Archive [Leinonen et al., 2011] and explored for the presence of previously identified EVE sequences. Only one *An. minimus* transcriptome sequence read archive was found under accession number SRX265162. Six *An. sinensis* transcriptome sequence read archives were found under accession numbers SRX448985, SRX449003, SRX449006 and SRX277584 for experiments using Illumina sequencing technology, and SRX218691 and SRX218721 for experiments using Roche 454 sequencing technology. RNA-seq reads were mapped to the EVE nucleotide sequence using Bowtie2 v2.1.0 [Langmead and Salzberg, 2012]. The alignment file was converted, sorted and indexed with Samtools v0.1.19 [Li et al., 2009]. Coverage was assessed using bedtools v2.17.0 [Quinlan and Hall, 2010].

2.2 *In vivo* validation

Mosquitoes

Anopheles minimus and*An. sinensis* mosquitoes were obtained through BEI Resources (www.beiresources.org), NIAID, NIH (*An. minimus* MINIMUS1, MRA-729; *An. sinensis* SINENSIS, MRA-1154). *Anopheles minimus* and *An. sinensis* mosquitoes were from the 132nd and 65th generations, respectively. Eggs were hatched in filtered tap water, reared in 24×34×9 cm plastic trays and fed with fish food (TetraMin, Tetra, Melle, Germany). Adults were maintained in 30×30×30 cm screened cages under controlled insectary conditions (28±1°C, 75±5% relative humidity, 12:12 hour light-dark cycle). They were provided with cotton soaked in a 10% (m/v) sucrose solution *ad libitum*. *Anopheles stephensi* nucleic acids, used as a reaction control, were provided by the Genetics and Genomics of Insect Vectors unit, Institut Pasteur, Paris.

EVE genomic integration

Mosquitoes were homogenized in pools of 10 separated by sex in 300 μL of Dulbecco's phosphate-buffered saline (DPBS) during two rounds of 30 sec at 5,000 rpm in a mixer mill (Precellys 24, Bertin Technologies, Montigny le Bretonneux, France). DNA was extracted using All Prep DNA/RNA Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. EVE presence in genomic DNA was assessed by 35 cycles

of PCR using Taq Polymerase (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA) (Table S2). PCR primers were designed to generate an amplicon spanning part of the EVE sequence and the flanking host sequence. Identity of the EVE sequence was confirmed by Sanger sequencing of the PCR product.

EVE transcription level

Mosquitoes were homogenized in pools of 5 separated by sex or development stage in 300 μL of DPBS during two rounds of 30 sec at 5,000 rpm in a mixer mill (Precellys 24). RNA was extracted from mosquito homogenates separated by sex using TRIzol Reagent (Life Technologies, Thermo Fisher Scientific, Waltham, MA, USA) following the manufacturer's instructions. Samples were treated by Turbo DNA-free kit (Life Technologies) and reverse transcribed using random hexamers (M-MLV, Invitrogen). Complementary DNA was amplified with 35 cycles of PCR for *An. minimus* and 40 cycles of PCR for *An. sinensis*, respectively, using DreamTaq polymerase (Thermo Fisher Scientific) and primers located within the EVE sequence (Table S2). To verify that RNA samples were free of DNA contamination, two sets of primers spanning exon 3 and 4 of the RPS7 gene of both *Anopheles* species (under VectorBase annotation number AMIN008193 and ASIC017918 for *An. minimus* and *An. sinensis*, respectively) were designed (Table S2). Because the corresponding DNA sequence includes intron 3, DNA contamination is expected to result in a larger PCR product. The length of intron 3 is 252 base pairs (bp) for *An. minimus* and 295 bp for *An. sinensis*.

3 Results

The *in silico* screen of 21 *Anopheles* genomes identified two flavivirus-derived EVEs, one in the genome of *An. minimus* and one in the genome of *An. sinensis* (Table 2). The *An. minimus* EVE is 1,881 bp long (627 amino acid residues) with Nienokoue virus as the closest BLAST hit (44% amino acid identity). The integrated sequence spans non-structural protein 4A (NS4A), NS4B and NS5 (Figure 1). Conserved domain search identified the NS5-methyltransferase domain involved in RNA capping and part of the RNA-directed RNA-polymerase domain. The *An. sinensis* EVE is 792 bp long (264 amino acid residues) and the closest BLAST hit is *Culex* flavivirus (45% amino acid identity). The integrated sequence corresponds to the middle part of NS3 (Figure 1). Conserved domain search identified the presence of a P-loop containing

the nucleoside triphosphate hydrolase domain found in the NS3 protein of exogenous flaviviruses. Phylogenetically, both EVEs are sister to the ISF clade (Figure 2).

Element	GenBank accession no.	Protein homology	Supercontig no.	Coordinates			
				Supercontig		Viral genome*	
				Start	End	Start	End
An. minimus EVE	Pending	NS4-NS5	supercont1.455	107	1,987	6,567	8,432
An. sinensis EVE	Pending	NS3	cont1.25450	822	1,613	5,214	5,822

Table 2 – Newly described flavivirus-derived EVEs in *An. minimus* and *An. sinensis* genomes. * Viral genomes coordinates are based on the closest tBLASTn hits: Nienokoue virus (GenBank accession no. JQ957875) for the *An. minimus* EVE and *Culex* flavivirus (GenBank accession no. JQ308188) for the *An. sinensis* EVE.

The presence of both EVEs was verified *in vivo* by PCR on genomic DNA (Figure 3), followed by sequencing to confirm identity. EVEs were found in both male and female genomic DNA. The *An. minimus* EVE was transcriptionally expressed for all combinations of sex and development stages tested (Figure 4-A). Evidence for transcriptional activity of the *An. minimus* EVE was confirmed in published RNA-seq data (Figure S1-A). The *An. sinensis* EVE was also transcriptionally expressed, although less abundantly, for all combinations of sex and development stages tested, especially in L_4 larvae (Figure 4-B). Low expression observed for the *An. sinensis* EVE is consistent with barely detectable transcriptional activity in published RNA-seq data (Figure S1-B).

4 Discussion

ISFs have attracted substantial interest in recent years after some of them were shown to enhance or suppress the replication of medically important flaviviruses in co-infected mosquito cells [Blitvich and Firth, 2015]. Over a dozen ISFs have been identified to date, mainly in *Aedes* and *Culex* genera of the Culicinae subfamily [Blitvich and Firth, 2015]. ISFs were also reported in *Anopheles* mosquitoes of the Anophelinae subfamily [Aranda et al., 2009, Zuo et al., 2014, Liang et al., 2015]. However, these *Anopheles*-associated ISFs are thought to infect a broad range of hosts including several mosquito species, mainly in the *Culex* genus, and are phylogenetically related to *Culex*-associated ISFs. Therefore, it is unclear whether *Anopheles* mosquitoes are true natural hosts of flaviviruses. Detection of ISFs in field-caught mosquitoes could result from incidental infection, or from a laboratory artifact. In this study, we discovered

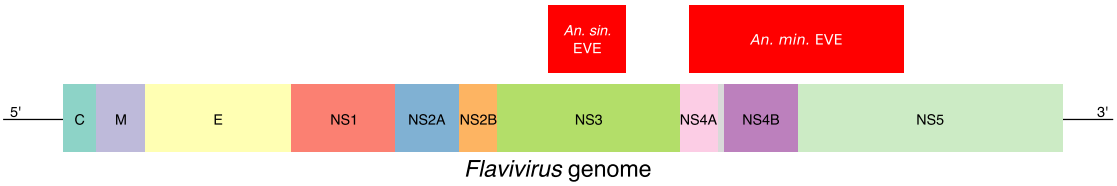

Figure 1 – Location of *An. minimus* and *An. sinensis* flavivirus-derived EVEs in a generic *Flavivirus* genome. Positioning is based on the sequence of Nienokoue virus (GenBank accession no. JQ957875) for the *An. minimus* EVE and of *Culex* flavivirus (GenBank accession no. JQ308188) for the *An. sinensis* EVE, as they were the closest tBLASTn hits. C=capsid protein, E=envelope glycoprotein, M=membrane glycoprotein, NS1=non-structural glycoprotein 1; NS2A=non-structural protein 2A; NS2B= non-structural protein 2B; NS3=non-structural protein 3 (protease/helicase); NS4A=non-structural protein 4A; NS4B=non-structural glycoprotein 4B; NS5=non-structural protein 5 (RNA-dependent-RNA polymerase).

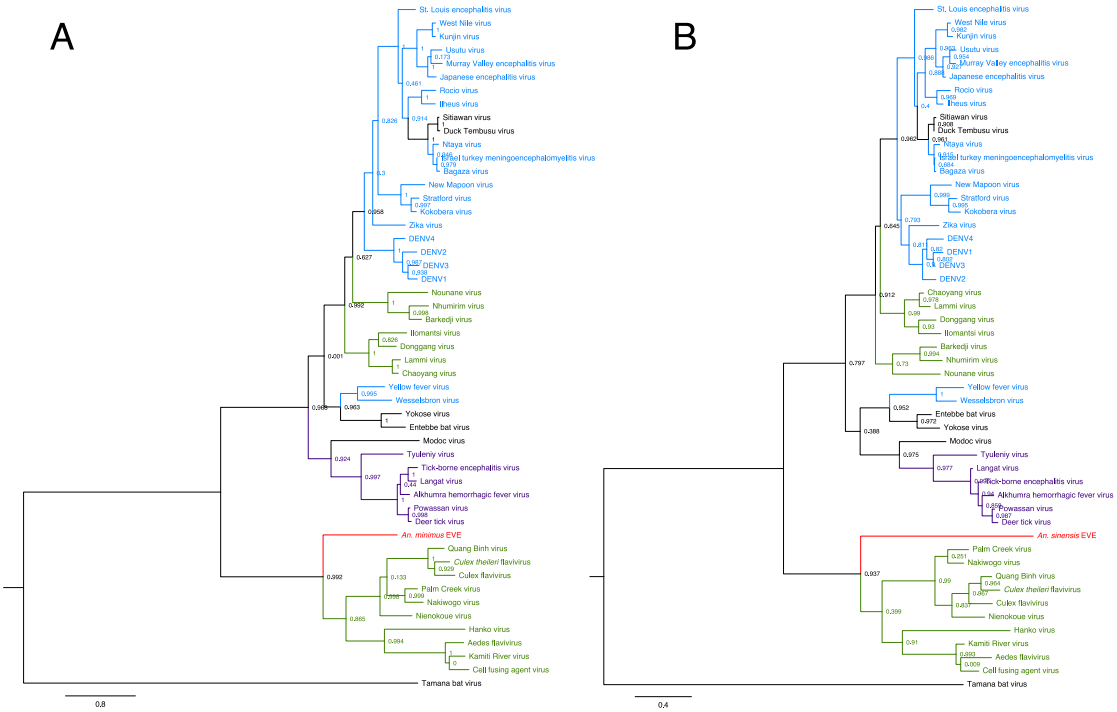

Figure 2 – Phylogenetic relationships of (A) *An. minimus* and (B) *An. sinensis* flavivirus-derived EVEs with exogenous flaviviruses. Maximum likelihood trees were constructed based on the translated EVE sequences. Clades are color-coded according to known host specificity: green, ISFs; purple, tick-borne arboviruses; black, 'not known vector' (vertebrate specific); blue, mosquito-borne arboviruses; red: EVEs. Scale bar indicates the number of substitutions per amino-acid site. Node values represent Shimodaira-Hasegawa (SH)-like branch support.

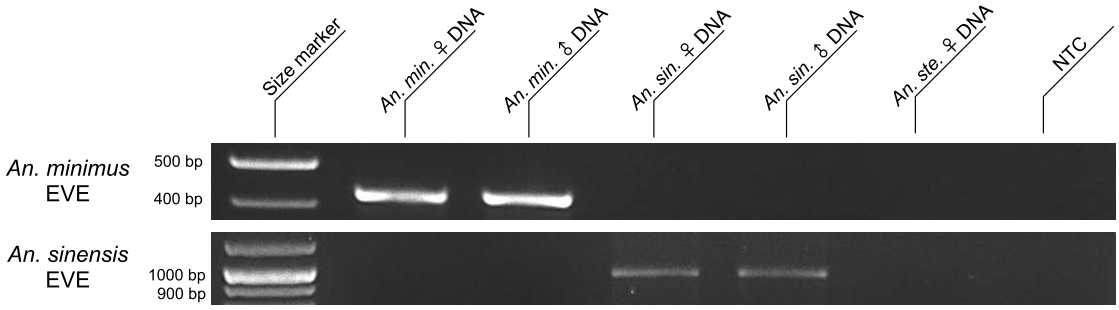

Figure 3 – Detection of *An. minimus* (top) and *An. sinensis* (bottom) flavivirus-derived EVEs in genomic DNA. Lane 1: size marker; lane 2: amplified genomic DNA from a pool of 10 *An. minimus* adult females; lane 3: amplified genomic DNA from a pool of 10 *An. minimus* adult males; lane 4: amplified genomic DNA from a pool of 10 *An. sinensis* adult females; lane 5: amplified genomic DNA from a pool of 10 *An. sinensis* adult males; lane 6: amplified DNA from a pool of 10 *An. stephensi* females; lane 7: no template control (NTC).

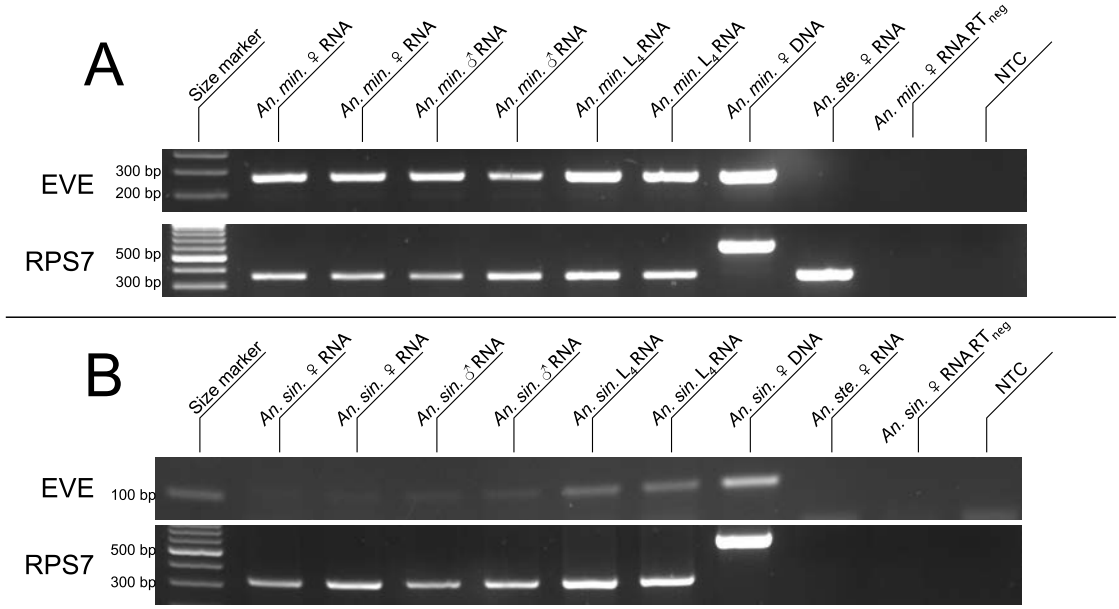

Figure 4 – Detection of (A) *An. minimus* and (B) *An. sinensis* flavivirus-derived EVEs RNA. Lane 1: size marker; lanes 2 and 3: amplified cDNA from pools of 5 adult females; lanes 4 and 5: amplified cDNA from pools of 5 adult males; lanes 6 and 7: amplified cDNA from pools of 5 L4 larvae; lane 8: amplified DNA from a pool of 10 females; lane 9: amplified cDNA from a pool of 5 *An. stephensi* females; lane 10: DNA contamination control (no reverse transcription) using the same pool of 5 adult females as lane 2; lane 11: no template control (NTC). First row: EVE; second row: RPS7 (control gene). The RPS7 target DNA sequence includes an intron, so that DNA contamination is expected to result in a larger PCR product.

flavivirus-derived EVEs in the genomes of two *Anopheles* species. Phylogenetic analyses indicated that both EVEs are related to ISFs but belong to a clade that is distinct from *Aedes*-associated and *Culex*-associated ISFs.

Presence of flavivirus-derived EVEs in *Anopheles* genomes supports the hypothesis that *Anopheles* mosquitoes are natural hosts of flaviviruses. Endogenization of non-retroviral RNA viruses is unlikely to occur in the absence of recurrent host-virus interactions over a long evolutionary time scale. Endogenization requires reverse transcription, germ line integration and fixation in the host population, three steps whose combined frequency is exceedingly rare [Holmes, 2011, Aiewsakun and Katzourakis, 2015]. Thus, our discovery of flavivirus-derived EVEs in *Anopheles* genomes is consistent with a long=lasting host-virus interaction between flaviviruses and mosquitoes of the Anophelinae subfamily.

ISFs are thought to be mainly maintained through vertical transmission from an infected female to its offspring [Blitvich and Firth, 2015]. Vertical transmission is likely to favor co-divergence of pathogens and hosts [Jackson and Charleston, 2004], as illustrated by the existence of *Aedes*-associated and *Culex*-associated clades of ISFs [Moureau et al., 2015]. Although extrapolation is limited by the scarcity of data on ISF host range and diversity, phylogenetic position of *Anopheles*-associated ISFs as sister to all other ISFs is consistent with the co-divergence hypothesis. During the evolutionary history of mosquitoes, the Anophelinae diverged from the Culicinae prior to the separation of *Culex* and *Aedes* genera [Reidenbach et al., 2009]. Further investigations will be necessary to determine whether an *Anopheles*-associated clade of exogenous ISFs exists, or existed.

Finally, our observation that both EVEs exhibited transcriptional activity may reflect a selective advantage for the host [Holmes, 2011]. Transcriptionally active EVEs have been suggested to confer protection or tolerance against related exogenous viruses [Flegel, 2009, Holmes, 2011, Bell-Sakyi and Attoui, 2013, Fujino et al., 2014]. Despite the lack of empirical evidence so far, flavivirus-derived EVEs could contribute to antiviral immunity and arbovirus vector competence in mosquitoes.

Acknowledgements

This work was supported by the French Government's Investissement d'Avenir program Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases grant ANR-10-LABX-62-IBEID, and the City of Paris Emergence(s) program in Biomedical Research. S.L. was supported by a doctoral fellowship from University Pierre and Marie Curie. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

The authors thank Inge Holm and Guillaume Carrissimo for providing *Anopheles stephensi* nucleic acids, Albin Fontaine for help with the *in silico* screen, Davy Jiolle and Elliott Miot for technical assistance, Clément Gilbert for advice, and members of the Lambrechts lab for insightful comments and discussions.

References

[Aiewsakun and Katzourakis, 2015] Aiewsakun, P. and Katzourakis, A. (2015). Endogenous viruses: Connecting recent and ancient viral evolution. *Virology*, 479-480:26–37.

[Aranda et al., 2009] Aranda, C., Sánchez-Seco, M. P., Cáceres, F., Escosa, R., Gálvez, J. C., Masià, M., Marqués, E., Ruíz, S., Alba, A., Busquets, N., Vázquez, A., Castellà, J., and Tenorio, A. (2009). Detection and monitoring of mosquito flaviviruses in Spain between 2001 and 2005. *Vector borne and zoonotic diseases (Larchmont, N.Y.)*, 9(2):171–178.

[Bell-Sakyi and Attoui, 2013] Bell-Sakyi, L. and Attoui, H. (2013). Endogenous tick viruses and modulation of tick-borne pathogen growth. *Frontiers in cellular and infection microbiology*, 3:25.

[Bernard et al., 2001] Bernard, K. A., Maffei, J. G., Jones, S. A., Kauffman, E. B., Ebel, G., Dupuis, A. P., Ngo, K. A., Nicholas, D. C., Young, D. M., Shi, P. Y., Kulasekera, V. L., Eidson, M., White, D. J., Stone, W. B., Kramer, L. D., and NY State West Nile Virus Surveillance Team (2001). West Nile virus infection in birds and mosquitoes, New York State, 2000. *Emerging infectious diseases*, 7(4):679–685.

[Blitvich and Firth, 2015] Blitvich, B. J. and Firth, A. E. (2015). Insect-specific flaviviruses: a systematic review of their discovery, host range, mode of transmission, superinfection exclusion potential and genomic organization. *Viruses*, 7(4):1927–1959.

[Calzolari et al., 2013] Calzolari, M., Bonilauri, P., Bellini, R., Albieri, A., Defilippo, F., Tamba, M., Tassinari, M., Gelati, A., Cordioli, P., Angelini, P., and Dottori, M. (2013). Usutu virus persistence and West Nile virus inactivity in the Emilia-Romagna region (Italy) in 2011. *Plos One*, 8(5):e63978.

[Calzolari et al., 2010] Calzolari, M., Bonilauri, P., Bellini, R., Caimi, M., Defilippo, F., Maioli, G., Albieri, A., Medici, A., Veronesi, R., Pilani, R., Gelati, A., Angelini, P., Parco, V., Fabbi, M., Barbieri, I., Lelli, D., Lavazza, A., Cordioli, P., and Dottori, M. (2010). Arboviral survey of mosquitoes in two northern Italian regions in 2007 and 2008. *Vector borne and zoonotic diseases (Larchmont, N.Y.)*, 10(9):875–884.

[Camacho et al., 2009] Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L. (2009). BLAST+: architecture and applications. *Bmc Bioinformatics*, 10(1):421.

[Capella-Gutiérrez et al., 2009] Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25(15):1972–1973.

[Chen et al., 2015] Chen, X.-G., Jiang, X., Gu, J., Xu, M., Wu, Y., Deng, Y., Zhang, C., Bonizzoni, M., Dermauw, W., Vontas, J., Armbruster, P., Huang, X., Yang, Y., Zhang, H., He, W., Peng, H., Liu, Y., Wu, K., Chen, J., Lirakis, M., Topalis, P., Van Leeuwen, T., Hall, A. B., Jiang, X., Thorpe, C., Mueller, R. L., Sun, C., Waterhouse, R. M., Yan, G., Tu, Z. J., Fang, X., and James, A. A. (2015). Genome sequence of the Asian Tiger mosquito, Aedes albopictus, reveals insights into its biology, genetics, and evolution. *Proceedings of*

the National Academy of Sciences of the United States of America, 112(44):E5907–15.

[Crochu et al., 2004] Crochu, S., Cook, S., Attoui, H., Charrel, R. N., De Chesse, R., Belhouchet, M., Lemasson, J.-J., de Micco, P., and De Lamballerie, X. (2004). Sequences of flavivirus-related RNA viruses persist in DNA form integrated in the genome of Aedes spp. mosquitoes. *The Journal of general virology*, 85(Pt 7):1971–1980.

[Feng et al., 2012] Feng, Y., Fu, S., Zhang, H., Li, M., Zhou, T., Wang, J., Zhang, Y., Wang, H., Tang, Q., and Liang, G. (2012). Distribution of mosquitoes and mosquito-borne viruses along the China-Myanmar border in Yunnan Province. *Japanese journal of infectious diseases*, 65(3):215–221.

[Feschotte and Gilbert, 2012] Feschotte, C. and Gilbert, C. (2012). Endogenous viruses: insights into viral evolution and impact on host biology. *Nature reviews. Genetics*, 13(4):283–296.

[Filipe, 1972] Filipe, A. R. (1972). Isolation in Portugal of West Nile virus from Anopheles maculipennis mosquitoes. *Acta virologica*, 16(4):361.

[Flegel, 2009] Flegel, T. W. (2009). Hypothesis for heritable, anti-viral immunity in crustaceans and insects. *Biology direct*, 4:32.

[Fujino et al., 2014] Fujino, K., Horie, M., Honda, T., Merriman, D. K., and Tomonaga, K. (2014). Inhibition of Borna disease virus replication by an endogenous bornavirus-like element in the ground squirrel genome. *Proceedings of the National Academy of Sciences of the United States of America*, 111(36):13175–13180.

[Holmes, 2011] Holmes, E. C. (2011). The evolution of endogenous viral elements. *Cell host & microbe*, 10(4):368–377.

[Holt et al., 2002] Holt, R. A., Subramanian, G. M., Halpern, A., Sutton, G. G., Charlab, R., Nusskern, D. R., Wincker, P., Clark, A. G., Ribeiro, J. M. C., Wides, R., Salzberg, S. L., Loftus, B., Yandell, M., Majoros, W. H., Rusch, D. B., Lai, Z., Kraft, C. L., Abril, J. F., Anthouard, V., Arensburger, P., Atkinson, P. W., Baden, H., de Berardinis, V., Baldwin, D., Benes, V., Biedler, J., Blass, C., Bolanos, R., Boscus, D., Barnstead, M., Cai, S., Center, A., Chaturverdi, K., Christophides, G. K., Chrystal, M. A., Clamp, M., Cravchik, A., Curwen, V., Dana, A., Delcher, A., Dew, I., Evans, C. A., Flanigan, M., Grundschober-Freimoser, A., Friedli, L., Gu, Z., Guan, P., Guigo, R., Hillenmeyer, M. E., Hladun, S. L., Hogan, J. R., Hong, Y. S., Hoover, J., Jaillon, O., Ke, Z., Kodira, C., Kokoza, E., Koutsos, A., Letunic, I., Levitsky, A., Liang, Y., Lin, J.-J., Lobo, N. F., Lopez, J. R., Malek, J. A., McIntosh, T. C., Meister, S., Miller, J., Mobarry, C., Mongin, E., Murphy, S. D., O'Brochta, D. A., Pfannkoch, C., Qi, R., Regier, M. A., Remington, K., Shao, H., Sharakhova, M. V., Sitter, C. D., Shetty, J., Smith, T. J., Strong, R., Sun, J., Thomasova, D., Ton, L. Q., Topalis, P., Tu, Z., Unger, M. F., Walenz, B., Wang, A., Wang, J., Wang, M., Wang, X., Woodford, K. J., Wortman, J. R., Wu, M., Yao, A., Zdobnov, E. M., Zhang, H., Zhao, Q., Zhao, S., Zhu, S. C., Zhimulev, I., Coluzzi, M., della Torre, A., Roth, C. W., Louis, C., Kalush, F., Mural, R. J., Myers, E. W., Adams, M. D., Smith, H. O., Broder, S., Gardner, M. J., Fraser, C. M., Birney, E., Bork, P., Brey, P. T., Venter, J. C., Weissenbach, J., Kafatos, F. C., Collins, F. H., and Hoffman, S. L. (2002). The genome sequence of the malaria mosquito Anopheles gambiae. *Science*, 298(5591):129–149.

[Jackson and Charleston, 2004] Jackson, A. P. and Charleston, M. A. (2004). A cophylogenetic perspective of RNA-virus evolution. *Molecular bi-*

ology and evolution, 21(1):45–57.

[Katoh et al., 2002] Katoh, K., Misawa, K., Kuma, K.-i., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30(14):3059–3066.

[Katzourakis and Gifford, 2010] Katzourakis, A. and Gifford, R. J. (2010). Endogenous viral elements in animal genomes. *PLoS Genet*, 6(11):e1001191.

[Kemenesi et al., 2014] Kemenesi, G., Krtinić, B., Milankov, V., Kutas, A., Dallos, B., Oldal, M., Somogyi, N., Nemeth, V., Banyai, K., and Jakab, F. (2014). West Nile virus surveillance in mosquitoes, April to October 2013, Vojvodina province, Serbia: implications for the 2014 season. *Euro surveillance : bulletin Européen sur les maladies transmissibles = European communicable disease bulletin*, 19(16):20779.

[Kulasekera et al., 2001] Kulasekera, V. L., Kramer, L., Nasci, R. S., Mostashari, F., Cherry, B., Trock, S. C., Glaser, C., and Miller, J. R. (2001). West Nile virus infection in mosquitoes, birds, horses, and humans, Staten Island, New York, 2000. *Emerging infectious diseases*, 7(4):722–725.

[Langmead and Salzberg, 2012] Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4):357–U54.

[Leinonen et al., 2011] Leinonen, R., Sugawara, H., Shumway, M., and International Nucleotide Sequence Database Collaboration (2011). The sequence read archive. *Nucleic Acids Research*, 39(Database issue):D19–21.

[Li et al., 2009] Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):2078–2079.

[Li and Godzik, 2006] Li, W. and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22(13):1658–1659.

[Liang et al., 2015] Liang, W., He, X., Liu, G., Zhang, S., Fu, S., Wang, M., Chen, W., He, Y., Tao, X., Jiang, H., Lin, X., Gao, X., Hu, W., Liu, Y., Feng, L., Cao, Y., Yang, G., Jing, C., Liang, G., and Wang, H. (2015). Distribution and phylogenetic analysis of Culex flavivirus in mosquitoes in China. *Archives of virology*, 160(9):2259–2268.

[Liu et al., 2013] Liu, H., Lu, H.-J., Liu, Z.-J., Jing, J., Ren, J.-Q., Liu, Y.-Y., Lu, F., and Jin, N.-Y. (2013). Japanese encephalitis virus in mosquitoes and swine in Yunnan province, China 2009-2010. *Vector borne and zoonotic diseases (Larchmont, N.Y.)*, 13(1):41–49.

[Marchler-Bauer et al., 2015] Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., Geer, R. C., He, J., Gwadz, M., Hurwitz, D. I., Lanczycki, C. J., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Wang, Z., Yamashita, R. A., Zhang, D., Zheng, C., and Bryant, S. H. (2015). CDD: NCBI's conserved domain database. *Nucleic Acids Research*, 43(Database issue):D222–6.

[Moureau et al., 2015] Moureau, G., Cook, S., Lemey, P., Nougairede, A., Forrester, N. L., Khasnatinov, M., Charrel, R. N., Firth, A. E., Gould, E. A., and De Lamballerie, X. (2015). New insights into flavivirus evolution, taxonomy and biogeographic history, extended by analysis of canonical and alternative coding sequences. *Plos One*, 10(2):e0117849.

[Neafsey et al., 2015] Neafsey, D. E., Waterhouse, R. M., Abai, M. R., Aganezov, S. S., Alekseyev,

M. A., Allen, J. E., Amon, J., Arcà, B., Arensburger, P., Artemov, G., Assour, L. A., Basseri, H., Berlin, A., Birren, B. W., Blandin, S. A., Brockman, A. I., Burkot, T. R., Burt, A., Chan, C. S., Chauve, C., Chiu, J. C., Christensen, M., Costantini, C., Davidson, V. L. M., Deligianni, E., Dottorini, T., Dritsou, V., Gabriel, S. B., Guelbeogo, W. M., Hall, A. B., Han, M. V., Hlaing, T., Hughes, D. S. T., Jenkins, A. M., Jiang, X., Jungreis, I., Kakani, E. G., Kamali, M., Kemppainen, P., Kennedy, R. C., Kirmitzoglou, I. K., Koekemoer, L. L., Laban, N., Langridge, N., Lawniczak, M. K. N., Lirakis, M., Lobo, N. F., Lowy, E., MacCallum, R. M., Mao, C., Maslen, G., Mbogo, C., McCarthy, J., Michel, K., Mitchell, S. N., Moore, W., Murphy, K. A., Naumenko, A. N., Nolan, T., Novoa, E. M., O'Loughlin, S., Oringanje, C., Oshaghi, M. A., Pakpour, N., Papathanos, P. A., Peery, A. N., Povelones, M., Prakash, A., Price, D. P., Rajaraman, A., Reimer, L. J., Rinker, D. C., Rokas, A., Russell, T. L., Sagnon, N., Sharakhova, M. V., Shea, T., Simão, F. A., Simard, F., Slotman, M. A., Somboon, P., Stegniy, V., Struchiner, C. J., Thomas, G. W. C., Tojo, M., Topalis, P., Tubio, J. M. C., Unger, M. F., Vontas, J., Walton, C., Wilding, C. S., Willis, J. H., Wu, Y.-C., Yan, G., Zdobnov, E. M., Zhou, X., Catteruccia, F., Christophides, G. K., Collins, F. H., Cornman, R. S., Crisanti, A., Donnelly, M. J., Emrich, S. J., Fontaine, M. C., Gelbart, W., Hahn, M. W., Hansen, I. A., Howell, P. I., Kafatos, F. C., Kellis, M., Lawson, D., Louis, C., Luckhart, S., Muskavitch, M. A. T., Ribeiro, J. M., Riehle, M. A., Sharakhov, I. V., Tu, Z., Zwiebel, L. J., and Besansky, N. J. (2015). Mosquito genomics. Highly evolvable malaria vectors: the genomes of 16 Anopheles mosquitoes. *Science*, 347(6217):1258522–1258522.

[Quinlan and Hall, 2010] Quinlan, A. R. and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842.

[Reidenbach et al., 2009] Reidenbach, K. R., Cook, S., Bertone, M. A., Harbach, R. E., Wiegmann, B. M., and Besansky, N. J. (2009). Phylogenetic analysis and temporal diversification of mosquitoes (Diptera: Culicidae) based on nuclear genes and morphology. *BMC evolutionary biology*, 9(1):298.

[Roiz et al., 2009] Roiz, D., Vázquez, A., Seco, M. P. S., Tenorio, A., and Rizzoli, A. (2009). Detection of novel insect flavivirus sequences integrated in Aedes albopictus (Diptera: Culicidae) in Northern Italy. *Virology journal*, 6:93.

[Sánchez et al., 2011] Sánchez, R., Serra, F., Tárraga, J., Medina, I., Carbonell, J., Pulido, L., de María, A., Capella-Gutiérrez, S., Huerta-Cepas, J., Gabaldón, T., Dopazo, J., and Dopazo, H. (2011). Phylemon 2.0: a suite of web-tools for molecular evolution, phylogenetics, phylogenomics and hypotheses testing. *Nucleic Acids Research*, 39(Web Server issue):W470–4.

[Zuo et al., 2014] Zuo, S., Zhao, Q., Guo, X., Zhou, H., Cao, W., and Zhang, J. (2014). Detection of Quang Binh virus from mosquitoes in China. *Virus research*, 180:31–38.

Supporting information

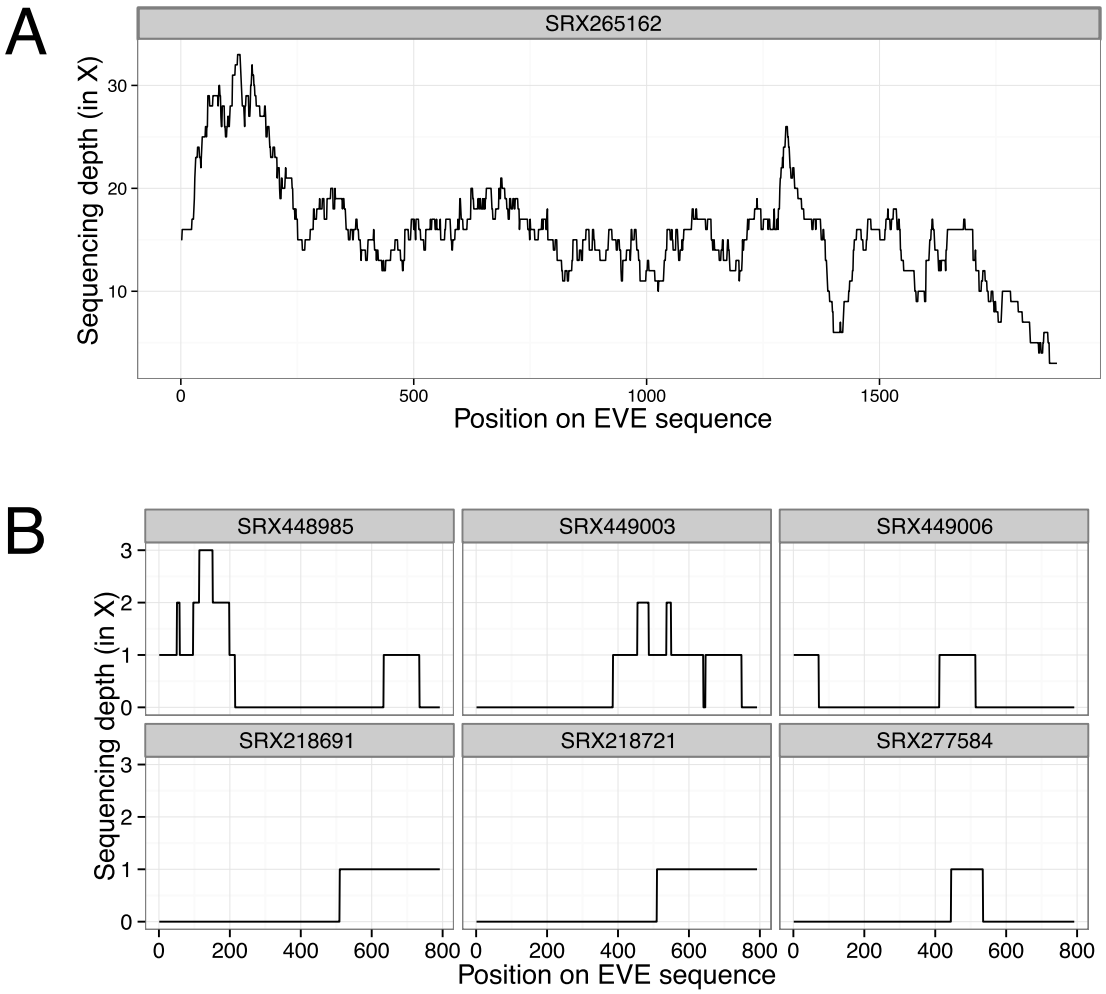

Figure S1 – Sequencing read coverage of (A) *An. minimus* and (B) *An. sinensis* EVEs in published RNA-seq experiments. For each experiment, the Sequence Read Archive (NCBI) accession number is indicated above the graph.

Accession no.	Virus name
AAV34152	Bussuquara virus
AAV34158	Rocio virus
ABC87765	St. Louis encephalitis virus
ACN73462	Nounane virus
ACV04605	Nakiwogo virus
AEK75355	Japanese encephalitis virus
AEY84723	Hanko virus
AAL32169	Deer tick virus
AFD50747	Chaoyang virus
AGG76014	Palm Creek virus
AGI03959	New Mapoon virus
AGJ84083	Ilheus virus
AGN32859	Louping ill virus
AGS41451	Barkedji virus
AGV15509	Israel turkey meningoencephalomyelitis virus
AGX15374	Dengue virus 2
AHF50490	Dengue virus 1
AHF70999	Ilomantsi virus
AHL17465	Dengue virus 3
AHN50410	Dengue virus 4
AHW82954	Nhumirim virus
AIJ19432	Stratford virus
AIM49244	Aedes flavivirus
AIM49245	Cell fusing agent virus
AIN44012	Duck Tembusu virus
AIN76232	Usutu virus
AIQ82561	West Nile virus
BAA00176	Kunjin virus
BAM74427	Culex flavivirus
CCC55433	Culex theileri flavivirus RP-2011
NP_041726	Yellow fever virus
NP_043135	Tick-borne encephalitis virus
NP_051124	Murray Valley encephalitis virus
NP_619758	Modoc virus
NP_620099	Powassan virus
NP_620108	Langat virus
NP_658908	Tamana bat virus
NP_722551	Alkhumra hemorrhagic fever virus
NP_872627	Yokose virus
NP_891560	Kamiti River virus
YP_950477	Entebbe bat virus
YP_001040007	Kokobera virus
YP_002790881	Zika virus
YP_002790883	Bagaza virus
YP_002922020	Wesselsbron virus
YP_005352889	Donggang virus
YP_006846328	Ntaya virus
YP_009001464	Tyuleniy virus
YP_009041466	Nienokoue virus
YP_009056848	Lammi virus

(a) Screening

Accession no.	Virus name
AIM49244	Aedes flavivirus
NP_722551	Alkhumra hemorrhagic fever virus
YP_002790883	Bagaza virus
AGS41451	Barkedji virus
AIM49245	Cell fusing agent virus
AFD50747	Chaoyang virus
BAM74427	Culex flavivirus
CCC55433	Culex theileri flavivirus RP-2011
AAL32169	Deer tick virus
AHF50490	DENV1
AGX15374	DENV2
AHL17465	DENV3
AHN50410	DENV4
YP_005352889	Donggang virus
AIN44012	Duck Tembusu virus
YP_950477	Entebbe bat virus
AEY84723	Hanko virus
AGJ84083	Ilheus virus
AHF70999	Ilomantsi virus
AGV15509	Israel turkey meningoencephalomyelitis virus
AEK75355	Japanese encephalitis virus
NP_891560	Kamiti River virus
YP_001040007	Kokobera virus
BAA00176	Kunjin virus
NP_051124	Murray Valley encephalitis virus
ACV04605	Nakiwogo virus
AGI03959	New Mapoon virus
AHW82954	Nhumirim virus
YP_009041466	Nienokoue virus
ACN73462	Nounane virus
YP_006846328	Ntaya virus
AGG76014	Palm Creek virus
NP_620099	Powassan virus
ACQ55297	Quang Binh virus
AAV34158	Rocio virus
AFP95929	Sitiawan virus
ABC87765	St. Louis encephalitis virus
AIJ19432	Stratford virus
NP_658908	Tamana bat virus
NP_043135	Tick-borne encephalitis virus
YP_009001464	Tyuleniy virus
AIN76232	Usutu virus
YP_002922020	Wesselsbron virus
AIQ82561	West Nile virus
NP_041726	Yellow fever virus
NP_872627	Yokose virus
YP_002790881	Zika virus

(b) Phylogeny

Table S1 – Name and accession number of flavivirus polyprotein sequences used as query for the screen in *Anopheles* genomes and for phylogenetic analysis of EVE sequences.

Name	**Sequence**	**Annealing temp.**	**Amplicon size**
An. min. EVE-in (R)	GTATGCGTGATTTTAAGATC	56°C	423 bp
An. min. EVE-out (F)	TGTATACTTCAGCTGTTG		
An. sin. EVE-in (R)	TGATCTAGACAAGAACTATG	56°C	945 bp
An. sin. EVE-out (F)	GATTCTCTGTTATTCTGTTG		

(a) Primers and reaction conditions used to amplify EVE targets from DNA samples.

Name	**Sequence**	**Annealing temp.**	**Amplicon size**
An. min. EVE-in (R)	GTATGCGTGATTTTAAGATC	55°C	272 bp
An. min. EVE-in (F)	CTCAACTAGAAGTTACATC		
An. min. S7 (R)	CCTCTTATTTTACAGGTAG	55°C	351 bp (RNA) 603 bp (DNA)
An. min. S7 (F)	GAATTGGAGAAGAAGTTC		
An. sin. EVE-in (R)	TGATCTAGACAAGAACTATG	55°C	111 bp
An. sin. EVE-in (F)	CACGATAGAATAGAATCTG		
An. sin. S7 (R)	CTCAAAGATTTACAGGTAG	55°C	321 bp (RNA) 616 bp (DNA)
An. sin. S7 (F)	GTAGTGTTCATTGGTGAG		

(b) Primers and reaction conditions used to amplify EVE and RPS7 targets from cDNA samples.

Table S2 – Primers and reaction conditions used to amplify EVE and RPS7 targets from DNA or cDNA samples.

B Genetic drift, purifying selection and vector genotype shape dengue virus intra-host genetic diversity in mosquitoes

> Sebastian Lequime, Albin Fontaine, Meriadeg Ar Gouilh, Isabelle Moltini-Conclois & Louis Lambrechts. Genetic drift, purifying selection and vector genotype shape dengue virus intra-host genetic diversity in mosquitoes. *PLoS Genetics, in press* 2016

Abstract

Due to their error-prone replication, RNA viruses typically exist as a diverse population of closely related genomes, which is considered critical for their fitness and adaptive potential. Intra-host demographic fluctuations that stochastically reduce the effective size of viral populations are a challenge to maintaining genetic diversity during systemic host infection. Arthropod-borne viruses (arboviruses) traverse several anatomical barriers during infection of their arthropod vectors that are believed to impose population bottlenecks. These anatomical barriers have been associated with both maintenance of arboviral genetic diversity and alteration of the variant repertoire. Whether these patterns result from stochastic sampling (genetic drift) rather than natural selection, and/or from the influence of vector genetic heterogeneity has not been elucidated. Here, we used deep sequencing of full-length viral genomes to monitor the intra-host evolution of a wild-type dengue virus isolate during infection of several mosquito genetic backgrounds. We estimated a bottleneck size ranging from

5 to 42 founding viral genomes at initial midgut infection, irrespective of mosquito genotype, resulting in stochastic reshuffling of the variant repertoire. The observed level of genetic diversity increased following initial midgut infection but significantly differed between mosquito genetic backgrounds despite a similar initial bottleneck size. Natural selection was predominantly negative (purifying) during viral population expansion. Taken together, our results indicate that dengue virus intra-host genetic diversity in the mosquito vector is shaped by genetic drift and purifying selection, and point to a novel role for vector genetic factors in the genetic breadth of virus populations during infection. Identifying the evolutionary forces acting on arboviral populations within their arthropod vector provides novel insights into arbovirus evolution.

1 Introduction

Due to the low fidelity of their RNA-dependent RNA polymerase, rapid replication kinetics and large population size, RNA viruses consist of a heterogeneous intra-host population of related mutants, sometimes referred to as a quasispecies [Lauring and Andino, 2010]. This mutant swarm as a whole defines the properties of the viral population, and is considered to be critical for the fitness and adaptive potential of RNA viruses [Lauring and Andino, 2010]. For example, high fidelity poliovirus mutants are attenuated in mice *in vivo*, demonstrating the functional importance of intra-host genetic diversity for pathogenesis [Vignuzzi et al., 2006].

Arthropod-borne viruses (arboviruses) are maintained by transmission between vertebrate hosts and blood-feeding arthropods such as mosquitoes that serve as vectors. Although arboviruses span a wide range of viral taxa in the Togaviridae, Flaviviridae, Bunyaviridae, Rhabdoviridae and Orthomyxoviridae families, the vast majority are RNA viruses, with the single known exception of a DNA arbovirus (African swine fever virus). The genetic plasticity of an RNA genome may confer arboviruses the remarkable ability to alternate between two fundamentally different hosts, and to quickly adapt to novel hosts [Coffey et al., 2013]. Like other RNA viruses, high levels of intra-host genetic diversity are critical for arboviral fitness, as demonstrated in both host types for chikungunya virus [Coffey et al., 2011, Rozen-Gagnon et al., 2014] and West Nile virus [Ciota et al., 2007, Ciota et al., 2012b, Jerzak et al., 2007].

Arboviruses usually rely on horizontal transmission between vertebrate hosts and

arthropod vectors, although vertical transmission from an infected female arthropod to her offspring may occur [Lequime and Lambrechts, 2014, Lequime et al., 2016]. After being ingested in a blood meal taken from a viremic vertebrate, arboviruses initially establish infection in the midgut epithelial cells of the arthropod vector. Transmission to another vertebrate host occurs after an extrinsic incubation period during which the arthropod develops a systemic infection that results in the release of viral particles in saliva. During infection of the arthropod vector, arboviruses are confronted with several anatomical barriers that are believed to impose severe population bottlenecks on viral populations [Forrester et al., 2014]. Bottlenecks are dramatic reductions in population size, resulting in stochastic sampling of a small number of viral genomes from the mutant swarm. Population bottlenecks can significantly reduce the fitness of RNA viruses through accumulation of deleterious mutations that cannot be efficiently removed by purifying selection [Duarte et al., 1992]. Initial infection and traversal of midgut cells, followed by virus dissemination and invasion of the salivary glands are expected to result in strong population drops that represents a challenge to maintaining arboviral genetic diversity during systemic vector infection [Forrester et al., 2012].

Despite such population bottlenecks, arboviruses typically maintain high levels of genetic diversity during transmission by their arthropod vectors [Forrester et al., 2014]. For example, analyses of West Nile virus populations in the midgut, hemolymph and saliva of *Culex* mosquitoes failed to document reductions in genetic diversity [Brackney et al., 2011]. However, the authors of this study did not determine whether a large effective population size was maintained, or if viral genetic diversity was quickly replenished by mutation and demographic expansion following population bottlenecks. In a recent study of dengue virus (DENV), genetic diversity was maintained during human to mosquito transmission but the variant repertoire changed substantially between venous blood and different organs of *Aedes* mosquitoes that became infected by feeding on the person [Sim et al., 2015]. Over 90% of DENV genetic variants were lost upon transition from venous blood to mosquito abdomen, as well as from abdomen to salivary glands, which led the authors to estimate that about a hundred viral genomes initially established a productive midgut infection [Sim et al., 2015]. However, this number could have been underestimated because the calculation assumed that the observed change in variant frequency was due to chance alone (i.e., it did not account

for the effect of natural selection). Genetic drift and purifying selection, for example, can result in a similar loss of genetic diversity.

The relative strength of natural selection and genetic drift is informed by the effective population size (N_e), defined as the size of an idealized population that would drift at the same rate as the observed population [Wright, 1931]. Ne indicates whether the evolution of a population is better described as a deterministic (selection) or stochastic (drift) process. When N_e is large, competition between variants occurs with little interference of random processes. When N_e is small, stochastic sampling of variants counteracts selection and hinders adaptation. For example, genetic drift plays a limited role during systemic infection of the plant host by Cauliflower mosaic virus, as viral populations maintain an effective size of several hundreds of viral genomes [Monsion et al., 2008]. Understanding the relative role of genetic drift and natural selection is critical to evaluate the risk of arboviral emergence through adaptive processes [Coffey et al., 2013]. For example, limited epidemic potential of an Asian lineage of chikungunya virus was associated with fixation of a deleterious deletion likely due to a founder effect [Chen et al., 2013].

In the present study, we investigated the intra-host evolution of DENV in the main mosquito vector *Aedes aegypti* by deep sequencing the full genome of viral populations at different time points of infection. Importantly, we accounted for the potential role of mosquito genetic variation on DENV intra-host genetic diversity. DENV intra-host genetic diversity has attracted considerable interest since the confirmation of its quasispecies nature [Wang et al., 2002]. Until now, however, most of this research has focused on viral genetic diversity in humans [Chao et al., 2005, Descloux et al., 2009, Parameswaran et al., 2012, Thai et al., 2012]. A few studies examined DENV intra-host genetic diversity in the mosquito vector [Lin et al., 2004, Sessions et al., 2015, Sim et al., 2015], but these studies did not account for vector genetic heterogeneity. There is substantial evidence for genetic variation in *Ae. aegypti* vector competence for DENV [Anderson and Rico-Hesse, 2006, Armstrong and Rico-Hesse, 2003, Bennett et al., 2002, Gubler et al., 1979, Rosen et al., 1985, Tardieux et al., 1990, Vazeille-Falcoz et al., 1999], as well as specific interactions between *Ae. aegypti* genotypes and DENV genetic variants [Lambrechts et al., 2009, Lambrechts, 2011, Lambrechts et al., 2013, Fansiri et al., 2013, Dickson et al., 2014].

Our objectives were three-fold: (i) measure the bottleneck size during initial

midgut infection of *Ae. aegypti* mosquitoes by DENV; (ii) monitor DENV intra-host genetic diversity during population expansion ans systemic infection; and (iii) determine the influence of the vector genotype on bottleneck size and intra-host DENV genetic diversity.

2 Material and methods

2.1 Ethics statement

The Institut Pasteur animal facility has received accreditation from the French Ministry of Agriculture to perform experiments on live animals in compliance with the French and European regulations on care and protection of laboratory animals. This study was approved by the Institutional Animal Care and Use Committee at Institut Pasteur.

2.2 Virus and mosquitoes

This study used a wild-type DENV-1 isolate (KDH0026A) that was originally recovered from the serum of a clinically ill dengue patient attending Kamphaeng Phet Provincial Hospital, Thailand as previously described [Fansiri et al., 2013]. Informed consent of the patient was not necessary because the virus was isolated in laboratory cell culture for diagnostic purposes (unrelated to this study) and, therefore, was no longer a human sample. The isolate was passaged three times in *Aedes albopictus* C6/36 cells prior to its use in this study. The full-length consensus genome sequence of the isolate is available from GenBank under accession number HG316481.

Aedes aegypti females used in this study belonged to the 16th generation of four isofemales lines (referred to as A, B, C, and D thereafter) derived from wild *Ae. aegypti* specimens collected in Kamphaeng Phet Province, Thailand. The lines were initiated by single mating pairs of field-caught males and females as previously described [Fansiri et al., 2013]. One male and one female from different collection sites (subdistricts) of the Muang district, Kamphaeng Phet Province, were randomly paired. The mothers of lines A and B, and the father of line C were collected in Thep Na Korn. The fathers of lines A, B and D were collected in Mae Na Ree. The mothers of lines C and D were collected in Nhong Ping Kai. They were maintained in the laboratory by mass sib-mating and collective oviposition at each subsequent generation. Quantification of genetic variation within and between the four isofemale lines was conducted as

part of this study (see below).

To initiate the experiment, eggs were hatched in filtered tap water. Larvae were reared in 24×34×9 cm plastic trays at a density of about 200 larvae per tray. Adults were maintained in 30×30×30 cm screened cages under controlled insectary conditions (28±1°C, 75±5% relative humidity, 12:12 hour light-dark cycle). They were provided with cotton soaked in a 10% (m/v) sucrose solution ad libitum and allowed to mate for 6-7 days before the experimental infection.

2.3 Restriction-site Associated DNA (RAD) sequencing of mosquitoes

Genetic characterization of the *Ae. aegypti* isofemale lines used single nucleotide polymorphism (SNP) markers identified and genotyped by Restriction-site Associated DNA (RAD) sequencing [Miller et al., 2007]. Ten females from the 16th generation of each isofemale line (i.e., from the same generation that was used in the experimental infection) and 10 females from the 1st generation of an outbred population collected in 2013 in Thep Na Korn, Kamphaeng Phet Province, Thailand (i.e., the region where the isofemale lines originated) were genotyped using RAD sequencing.

Mosquito genomic DNA was purified using the procedure developed by Pat Roman's laboratory at the University of Toronto [Black and DuTeau, 1997]. DNA concentration was measured with Qubit fluorometer and Quant-iT dsDNA Assay kit (Life Technologies, Paisley, UK). A modified version of the original double-digest Restriction-site Associated DNA (ddRAD) sequencing protocol [Peterson et al., 2012] was used as previously described [Rašić et al., 2014] with minor additional modifications. Briefly, 350 ng of genomic DNA from each mosquito was digested in a 50-μl reaction containing 50 units each of *NlaIII* and *MluCI* restriction enzymes (New England Biolabs, Herts, UK), 1X CutSmart Buffer and water for 3 hours at 37°C, without a heat-kill step. Digestion products were cleaned with 1.5X volume of Ampure XP paramagnetic beads (Beckman Coulter, Brea, CA, USA) and ligated to the modified Illumina P1 and P2 adapters with overhangs complementary to *NlaIII* and *MluCI* cutting sites, respectively. Each mosquito was uniquely labeled with a combination of P1 and P2 barcodes of variable lengths to increase library diversity at 5' and 3' ends (File S1). Ligation reactions were set up in a 45-μl volume with 2 μl of 4 μM P1 and 12 μM P2 adapters, 1,000 units of T4 ligase and 1X T4 buffer (New England Biolabs)

and were incubated at 16°C overnight. Ligations were heat-inactivated at 65°C for 10 minutes and cooled down to room temperature (20-25°C) in a thermocycler at a rate of 1.5°C per 2 minutes. Adapter-ligated DNA fragments from all individuals were pooled and cleaned with 1.5X bead solution. Fragments from 350 to 440 base pairs (bp) were selected using a Pippin-Prep 2% gel cassette (Sage Sciences, Beverly, MA, USA). Finally, 1 μl of the size-selected DNA was used as a template in a 10-μl PCR reaction with 5 μl of Phusion High Fidelity 2X Master mix (New England Biolabs) and 1 μl of 50 μM P1 and P2 primers (File S1). To reduce bias due to PCR duplicates, 8 PCR reactions were run in parallel, pooled, and cleaned with a 0.8X bead solution to make the final library. At this step, final libraries were quantified by quantitative PCR using the QPCR NGS Library Quantification Kit (Agilent Technologies, Palo Alto, CA, USA).

Libraries containing multiplexed DNA fragments from 50 mosquitoes were sequenced on an Illumina NextSeq platform using a NextSeq 500 High Output 300 cycles v2 kit (Illumina, San Diego, CA, USA) to obtain 150-bp paired-end reads. An optimized final library concentration of 1.1 pM, spiked with 15% PhiX, was loaded onto the flow cell.

2.4 RAD markers for mosquito genotyping

A previously developed bash script [Rašić et al., 2014] was used to process raw sequencing reads with minor modifications. Briefly, the DDemux program was used for demultiplexing fastq files according to the P1 and P2 barcodes combinations. Sequence quality scores were automatically converted into Sanger format. Sequences were filtered with FASTX-Toolkit. The first 5 bp (i.e., the restriction enzyme cutting site) and last 11 bp of P1 and P2 reads were trimmed. All reads with Phred scores <25 were discarded. P1 and P2 reads were then matched and unpaired reads were sorted as orphans.

Paired reads were aligned to the reference *Ae. aegypti* genome (AaegL3, February 2016) [Nene et al., 2007] using Bowtie version 0.12.7 [Langmead et al., 2009]. Parameters for the ungapped alignment included a maximum of three mismatches allowed in the seed, suppression of alignments if more than one reportable alignment existed, and a "try-hard" option to find valid alignments. Orphans were combined with all unaligned paired reads and single-end alignment was attempted. All aligned Bowtie output files were merged per individual and were imported into the Stacks pipeline. A

catalog of RAD loci used for SNP discovery was created using the ref_ map.pl pipeline in Stacks version 1.37 [Catchen et al., 2011, Catchen et al., 2013]. First, sequences aligned to the same genomic location were stacked together and merged to form loci using Pstacks. Only loci with a sequencing depth ≥3 reads per individual were retained. Cstacks was used to create a catalog of consensus loci, merging alleles together and Sstacks was used to match all identified loci. The Stacks pipeline identified a total of 899,892 RAD loci. The "populations" module was used to export markers with a sequencing depth ≥10X that were present in ≥98% of samples. The phylogenetic analysis was performed with the resulting subset of 2,321 SNPs belonging to 1,319 RAD loci (0.15%).

2.5 Phylogenetic analysis of mosquitoes

Phylogenetic trees were constructed using a Bayesian Markov Chain Monte Carlo (MCMC) algorithm, implemented in the BEAST 1.8.3 package [Drummond and Rambaut, 2007]. Inferences were produced under the coalescent model (constant size), and under the GTR+G (global time reversible with gamma distribution and no invariable sites) and the HKY+G (Hasegawa-Kishino-Yano) substitution models. Heterozygote positions were considered in calculations by enabling the use of IUPAC code and associated degeneracy within the substitution model. The length of MCMC was set at 3×10^7 states to obtain Effective Sampling Size (ESS) values >200.

2.6 Experimental mosquito infection

Six- to seven-day-old *Ae. aegypti* females were deprived of water and sucrose for 24h prior to the infectious blood meal. The virus stock was diluted in cell culture medium (Leibovitz's L-15 medium + 10% heat-inactivated fetal calf serum + non-essential amino-acids + 0.1% penicillin/streptomycin + 1% sodium bicarbonate) to reach an expected infectious titer of 3×10^6 focus-forming units (FFU) per mL. One volume of virus suspension was mixed with two volumes of freshly drawn rabbit erythrocytes washed in distilled phosphate buffer saline (DPBS). After gentle mixing, 2.5 mL of the infectious blood meal was placed in each of several Hemotek membrane feeders (Hemotek Ltd, Blackburn, UK) maintained at 37°C and covered with a piece of desalted porcine intestine as a membrane. Sixty μL of 0.5 M ATP were added to each feeder as a phagostimulant. Each isofemale line was allowed to feed during two rounds of 15

min on different feeders to ensure randomization of a potential feeder effect. Actual virus titer in the blood meal was measured by standard focus-forming assay in C6/36 cells [Lambrechts et al., 2009]. After feeding, mosquitoes were cold anesthetized on ice and fully engorged females were transferred to 1-pint cardboard cups. They were incubated under controlled conditions (28±1°C, 75±5% relative humidity, 12:12 hour light-dark cycle) in a climatic chamber.

At 4, 7 and 14 days post exposure (dpe), the midgut of 8-12 individuals from each isofemale line (i.e., biological replicates) were dissected. Midguts were homogenized individually in 140 μL of DPBS + 560 μL of QIAamp Viral RNA Mini Kit lysis buffer (Qiagen, Hilden, Germany) during two rounds of 30 sec at 5,000 rpm in a mixer mill (Precellys 24, Bertin Technologies, Montigny le Bretonneux, France). At 7 and 14 dpe, the legs of midgut-dissected mosquitoes were removed and homogenized as described above. At 14 dpe, the salivary glands of the midgut- and leg-less individuals were harvested and processed as above.

2.7 Virus deep sequencing

Total RNA was extracted from mosquito homogenates using QIAamp Viral RNA Mini Kit (Qiagen) and reverse transcribed using Transcriptor High Fidelity cDNA Synthesis Kit (Roche Applied Science, Penzberg, Germany) and a specific reverse primer located at the 3' end of the viral genome (Table S1). Presence and amount of viral cDNA was assessed by quantitative PCR using the LightCycler DNA Master SyberGreen I kit (Roche Applied Science) and custom primer pairs (Table S1). Absolute quantification used a standard curve generated with serial dilutions of PCR amplicons of known concentration. Selected samples were amplified by 40 cycles of PCR in 10 overlapping amplicons with Q5 High Fidelity DNA polymerase (New England Biolabs) and custom primer pairs (File S2).

PCR products were purified with Agencourt AMPure XP magnetic beads (Beckman Coulter) and their concentration was measured by Quant-iT PicoGreen dsDNA fluorometric quantification (Invitrogen). Equal amounts of each amplicon were pooled by sample and brought to a final concentration of 0.2 ng/μL. Multiplexed libraries were prepared using Nextera XT DNA Library Preparation Kit (Illumina) and single-end sequenced on an Illumina NextSeq 500 platform using a high-output 75 cycles v1 kit (Illumina). Sequencing reads were demultiplexed using bcl2fastq v2.15.0 (Illumina).

Raw sequences were deposited in the NCBI Short Read Archive (accession numbers pending).

After demultiplexing, reads were trimmed to remove Illumina adaptor sequences using Trimmomatic v0.33 [Bolger et al., 2014] and amplification primers if matching sequences were found on either the 5' or 3' end of the reads using Cutadapt v.1.8.3 [Martin, 2011]. Reads shorter than 32 bases were discarded and remaining reads were then mapped to the reference DENV genome sequence using Bowtie2 v2.1.0 [Langmead and Salzberg, 2012]. The alignment file was converted, sorted and indexed using Samtools v0.1.19 [Li et al., 2009]. Coverage and sequencing depth were assessed using bedtools v2.17.0 [Quinlan and Hall, 2010]. Single nucleotide variants (SNVs) and their proportion among all reads were called using LoFreq* v2.1.1 [Wilm et al., 2012] and their effect at the amino-acid level assessed by SNPdat v.1.0.5 [Doran and Creevey, 2013].

2.8 Viral genetic diversity analyses

Two sets of SNV markers were used for analyses of genetic diversity and natural selection. The 'full' marker set excluded all nucleotide positions in a given sample that had (i) a sequencing depth <500X or (ii) where potential sequencing or library preparation artifacts [Costello et al., 2013] were detected. The 'conservative' marker set excluded all nucleotide positions in all samples that had (i) a sequencing depth <500X or (ii) where potential sequencing or library preparation artifacts were detected in a least one sample. The conservative marker set minimized the potential bias owing to the unique mutational profile of each nucleotide position. However, because some of the overlapping fragments covering the viral genome could not be successfully amplified in a few samples (S1 Fig), the conservative marker set failed to cover large fractions of the viral genome (S2 Fig A). The full marker set, conversely, minimized the potential bias owing to distinct evolutionary properties of the different regions of the viral genome.

Genetic complexity of the viral population was estimated using normalized Shannon entropy (Sn) for each nucleotide site [Gregori et al., 2014]:

$$S_n = \frac{-p(\ln(p) + (1-p) \times \ln(1-p)}{\ln(4)}$$

where p is the SNV minor allele frequency at the considered position, and $\ln(4)$ corresponds to maximum complexity (i.e., four possible nucleotides at each position). For individual SNVs, S_n values range from 0 to 1. For diallelic SNVs, S_n values range from 0 (no diversity) to 0.5 (maximum complexity, when the two alternative nucleotides are present at equal frequency). For each sample S_n was averaged over all nucleotide sites included in either the full or the conservative set of SNV markers (i.e., total genome length minus number of excluded positions).

Genetic diversity of the viral population was also estimated using nucleotide diversity at each nucleotide site [Cornman et al., 2013]:

$$\pi = \frac{D}{D-1} \times (1 - (p^2 + (p-1)^2)$$

where D is the sequencing depth at the considered position and p is the SNV minor allele frequency. Like for S_n, π values for a diallelic SNV range from 0 (no polymorphism) to 0.5 (when the two alternative nucleotides are present at equal frequency). For each sample, π was averaged over all nucleotide sites included in either the full or the conservative set of markers. This index of genetic diversity is less sensitive to low-frequency variants than S_n, due to the lack of log-transformation of the frequencies.

2.9 Natural selection assessment

The magnitude and direction of natural selection were assessed using the d_N/d_S ratio, which is the ratio between the number of non-synonymous substitutions per non-synonymous site (d_N) over the number of synonymous substitutions per synonymous site (d_S) of a coding sequence, assuming synonymous substitutions are selectively neutral:

$$d_S = \frac{-3 \times \ln\left(\frac{4 \times \frac{S_d}{S_s}}{3}\right)}{4} \qquad d_N = \frac{-3 \times \ln\left(\frac{4 \times \frac{N_d}{N_s}}{3}\right)}{4}$$

where S_d is the number of synonymous substitutions in the sequence, S_s is the number of synonymous sites, N_d is the number of non-synonymous substitutions

in the sequence and N_s is the number of non-synonymous sites [Nei and Gojobori, 1986]. A d_N/d_S ratio >1 means that there is an excess of normalized number of non-synonymous substitutions relative to the normalized number of synonymous substitutions and is interpreted as evidence for positive selection (i.e., driving change). A d_N/d_S ratio <1 means that there is an excess of normalized number of synonymous substitutions relative to the normalized number of non-synonymous substitutions and is interpreted as evidence for negative selection (i.e., acting against change). A d_N/d_S ratio equal to 1 is interpreted as evidence for the absence of natural selection (i.e, neutral evolution).

The d_N/d_S ratio was computed using the Nei-Gojobori method [Nei and Gojobori, 1986] with suggested modifications for high-throughput sequencing data [Grubaugh et al., 2015]. Briefly, N_d and S_d were calculated for each sample as the sum of SNV frequencies. Mean N_d and S_d were computed for each isofemale line at each time point and used for d_N and d_S calculation, respectively. Numbers of synonymous and non-synonymous sites from the initial population consensus sequence were estimated using MEGA v.7.0.16 [Kumar et al., 2016] by computing the number of 0-, 2-, 3- and 4-fold degenerate sites following the Nei-Gojobori method [Nei and Gojobori, 1986]. The full SNV marker set had a variable number of synonymous and non-synonymous sites depending of the number of nucleotide sites retained or excluded for each sample. The conservative marker set had 328.67 synonymous and 1,481.33 non-synonymous sites for all samples.

2.10 Statistical testing

Statistical analyses were performed in the statistical environment R, version 3.2.0 (http://www.r-project.org/) using the packages car [Fox and Weisberg, 2011], plyr [Wickham, 2011], ggplot2 [Wickham, 2009], stringr [Wickham, 2015], reshape2 [Wickham, 2007], gridExtra [Auguie, 2015], fitdistrplus [Delignette-Muller et al., 2015] and boot [Canty and Ripley, 2015]. In all the analysis, the individual mosquito sample was considered a biological unit of replication.

Infection prevalence and cDNA copy numbers were compared among isofemale lines at each time point by pairwise Pearson χ^2 tests and pairwise Wilcoxon signed-rank tests, respectively, followed by a Holm correction of p-values for multiple testing.

The proportion of SNVs per position, mean S_n and mean π estimates were com-

pared between the input and later time points using pairwise Wilcoxon signed-rank tests and a Holm p-value adjustment. The proportion of SNVs per position, S_n, π and d_N/d_S estimates in midgut samples were analyzed between 4 and 7 dpe as a function of the combined effects of time point and mosquito genotype using a linear model with an identity link function and a normal error distribution. Model validity was verified with quantile-quantile (Q-Q) plots of residuals and by computing Cook's distance to assess influence of observations. Statistically significant effects (p<0.05) of time point, mosquito genotype and their interactions were determined using type-II analysis of variance. Statistically insignificant interactions were removed from the model, subsequently repeating model validation. Statistical testing of pairwise differences between isofemale lines used the linear regression coefficients. Estimated regression coefficients were extracted and their 95% confidence intervals and p-values were calculated based on their standard errors compared to a reference level. Isofemale line A was arbitrarily chosen as the reference level.

2.11 Bottleneck size estimation

Following a published method [Monsion et al., 2008], bottleneck size at initial midgut infection was estimated by analyzing the change in frequency distribution of neutral markers between blood meal (initial) and midgut (final) samples. Under the assumption of neutrality (i.e., absence of natural selection), the idealized number of founding genomes (N) initiating the midgut infection can be computed as:

$$N = \frac{p \times (1-p)}{Var(p') - Var(p)}$$

where p is the marker allele frequency in the initial population and p' is the marker allele frequency in the final population [Monsion et al., 2008]. This method considers that changes in the genetic variance between sequential samples result exclusively from genetic drift and therefore requires neutral or quasi-neutral markers.

SNVs that were presumably neutral were selected based on the following set of criteria: (i) synonymous change at the third codon position, (ii) no significant change in mean frequency between sampling time points, (iii) SNV detected in ≥80% of the five viral input replicates (viral stock and blood meal samples), and (iv) mean frequency >0.02 in the input population. Confidence intervals of N estimates were

computed using a bootstrapping procedure as described in [Monsion et al., 2008]. Briefly, for each bootstrap all individuals were sampled with replacement to calculate N. This was repeated 1,000 times to generate a distribution of N values and derive 95% confidence intervals corresponding to the 2.5 and 97.5 percentiles of the distribution.

2.12 Bottleneck simulation

The effect of the initial midgut infection bottleneck on viral diversity indices was simulated in R based on 100 sampling events from an initial viral population containing 100 independent SNVs (Code S1). SNV minor allele frequency was randomly drawn from an exponential distribution (λ=100). Initial viral population size (equivalent to the infectious dose ingested in the blood meal) was drawn from a normal distribution (mean=2,000; standard deviation=200). Bottleneck size was drawn from a normal distribution (mean=28; standard deviation=5). Mean S_n and mean π were computed for all samples in the presence or the absence of a detection threshold arbitrarily set at an SNV minor allele frequency of 0.01.

3 Results

3.1 Mosquito genetic variation

A genome-wide set of 2,321 SNPs generated by RAD sequencing was used to genetically characterize the four *Ae. aegypti* isofemale lines (A, B, C and D). These markers had a sequencing depth ≥10X per sample and were missing in <2% of samples. An outbred *Ae. aegypti* population from the same geographic location where the lines were created was also genotyped to provide a phylogenetic background. Phylogenetic relationships among individuals from the four isofemale lines and the outbred population were determined with a Bayesian method (Figure 1). As expected, the outbred mosquito population was paraphyletic, reflecting its genetic diversity. Mosquitoes from isofemale lines A and B clustered independently with strong statistical support, confirming their distinct genetic identity. Unexpectedly, mosquitoes from isofemale line C grouped with mosquitoes from isofemale line D within the same clade. This could be the result of relatedness between the parents randomly chosen to initiate the lines, as the mothers of lines C and D came from the same collection site and may have been siblings. Two different substitution models for the phylogenetic reconstruction

gave similar clustering patterns. Similar results were also obtained when testing a variable number of markers (allowing from 0% to 30% of missing genotypes) with the same method. Because isofemale lines C and D were not unambiguously assigned to different monophyletic groups, they could not be considered as distinct genotypes and were thus combined in all subsequent analyses. They are jointly referred to as line CD hereafter.

3.2 Infection time course and sample selection

Mosquitoes from the three different genotypes (A, B, and CD) were exposed to DENV through an artificial blood meal at a final titer of 1.52×10^6 focus-forming units (FFU)/mL. Assuming a blood meal size of approximately 2 μL, the infectious dose ingested by each mosquito was about 3,000 infectious viral particles. The proportion of mosquitoes that acquired a midgut infection ranged from 75 to 100% and did not differ significantly between time points or isofemale lines (Figure 2-A). The proportion of mosquitoes with a DENV infection that disseminated to their legs increased from 10-40% at 7 days post exposure (dpe) to 60-100% at 14 dpe, but the rate of virus dissemination to the legs did not differ significantly between isofemale lines (Figure 2-A). However, the proportion of mosquitoes with a disseminated infection in the salivary glands was significantly higher for line CD (87.5%) than for line A (37.5%) and line B (41.7%) at 14 dpe (line A vs. line CD, p=0.037; line B vs. line CD, p=0.037). Among infected mosquitoes, viral load ranged from 8.9×10^2 to 2.8×10^6 DENV genome copies per sample with no significant difference between isofemale lines at any of the time points, with the exception of lines B and CD that had significantly different viral loads (p=0.037) in their salivary glands at 14 dpe (Figure 2-B).

Deep sequencing of viral genomes was performed for a subset of 78 infected samples at selected time points (Figure 2-B) that were processed individually and treated as biological replicates. Some samples were excluded because their low concentration of viral RNA resulted in unsuccessful RT-PCR amplification. A total of 4, 7 and 13 infected midguts at 4 dpe and 7, 11 and 21 infected midguts at 7 dpe were analyzed for lines A, B, and CD, respectively. Ten infected salivary glands at 14 dpe were analyzed in line CD. In addition, DENV genomes were deep sequenced in the initial viral stock and in four replicates of the artificial infectious blood meal. On average, 3,615,466 sequencing reads per sample aligned to the reference DENV genome. Mean DENV

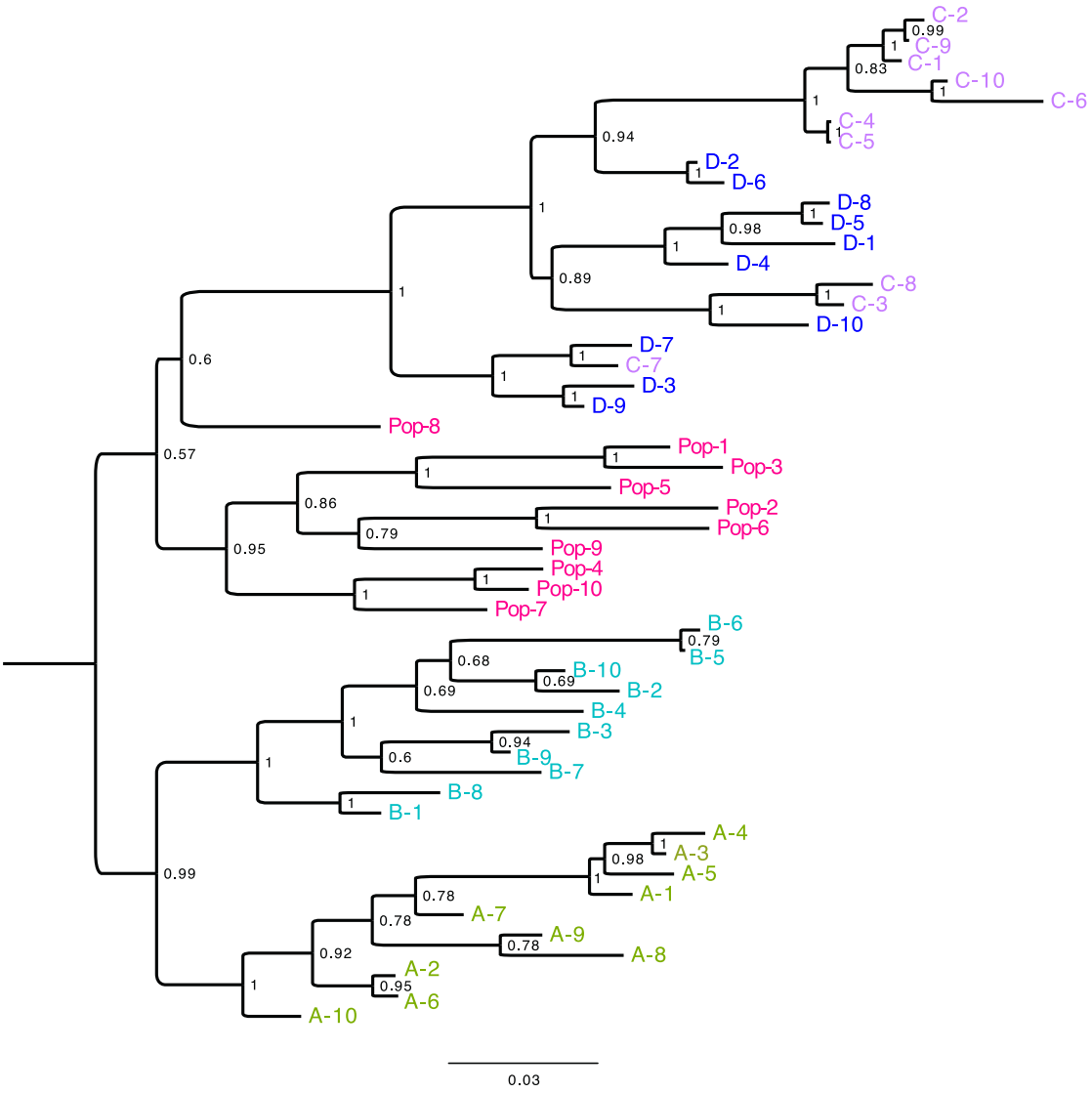

Figure 1 – Phylogenetic relationships between *Aedes aegypti* isofemale lines. Bayesian phylogenetic tree representing the genetic diversity across individuals from the four isofemale lines (A, B, C and D) and from a field-derived outbred population (Pop) from the same geographic location where the lines were created. The phylogenetic analysis was based on a (GTR+G) substitution model of 2,321 SNPs. Putative populations are depicted in different colors. The scale bar indicates the number of substitutions and posterior probabilities are displayed at relevant nodes.

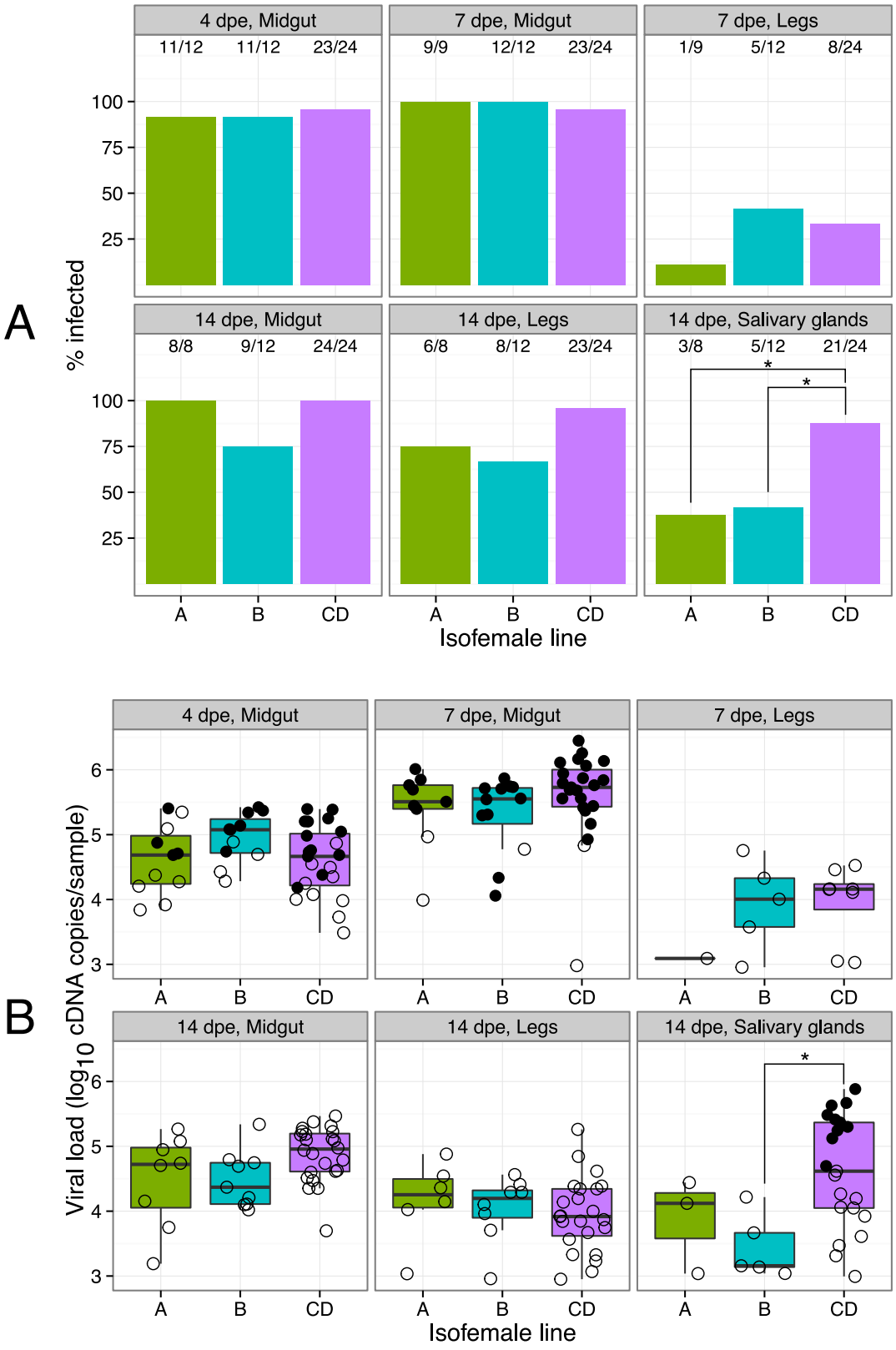

Figure 2 – Time-course of DENV prevalence and viral load. (A) Bar graphs show the percentage of DENV-infected samples stratified by time point, tissue and isofemale line. Relative numbers of positive samples are indicated above the bars. (B) Box plots show the number of DENV genome copies per infected sample stratified by time point, tissue and isofemale line. Solid dots represent individual samples selected for deep sequencing and open dots represent samples not sequenced. dpe=days post exposure. * p<0.05.

genome coverage with a sequencing depth >500X was 10,594 nucleotides per sample, which represents 98.8% of the 10,718 nucleotides of the total genome length. Mean sequencing depth was >24,212X per nucleotide position per sample (Figure S1).

3.3 Patterns of viral genetic diversity

The full set of SNV markers retained for population genetic analyses included an average of 5,843 nucleotide sites across the DENV genome whereas a more conservative set (see Materials and Methods) was restricted to 1,810 nucleotides (Figure S2-B). The SNVs of the full marker set were randomly distributed across the genome without obvious mutation hot or cold spot (Figure 3). A new variant reached consensus level (frequency>0.5) in one midgut sample at 4 dpe and one midgut sample at 7 dpe, but the SNV was different in each case. In salivary glands collected at 14 dpe, new variants reached consensus level at 11 positions, none of which was shared among individuals within or between isofemale lines (Figure 3). In the more restricted conservative set of markers, no variant reached consensus level at any time point (Figure S2-B).

To determine the effect of initial midgut infection on DENV genetic diversity, a first series of analyses compared viral genetic diversity observed in the input samples with genetic diversity observed at any of the later time points. In the full marker set, initial infection of the midgut was associated with an increase in viral genetic diversity relative to the input (Figure 4) both when measured with normalized Shannon entropy S_n (0 dpe vs. 4 dpe, p=0.0001; 0 dpe vs. 7 dpe, p=0.002; 0 dpe vs. 14 dpe, p=0.003) (Figure 4-A) and when measured with nucleotide diversity π (0 dpe vs. 4 dpe, p=0.0001; 0 dpe vs. 7 dpe, p=0.0004; 0 dpe vs. 14 dpe, p=0.003) (Figure 4-B). Viral diversity was also significantly higher in the salivary glands at 14 dpe than in the midgut at 7 dpe (S_n: p=0.012; π: p=0.012). The proportion of variable sites detected also increased following initial midgut infection (Figure S3) although differences were only statistically significant between 0 dpe and 4 dpe (p=0.0029) and between 0 dpe and 14 dpe (p=0.0067). Similarly, in the conservative set of markers, mosquito infection was associated with a relative increase in viral genetic diversity following initial midgut infection, albeit more modestly due to the smaller number of markers, both when measured with normalized Shannon entropy S_n (0 dpe vs. 4 dpe, p=0.046; 0 dpe vs. 14 dpe, p=0.008) (Figure S4-B) and when measured with nucleotide diversity π (0 dpe vs. 14 dpe, p=0.008) (Figure S4-C). The proportion of variable sites detected, however,

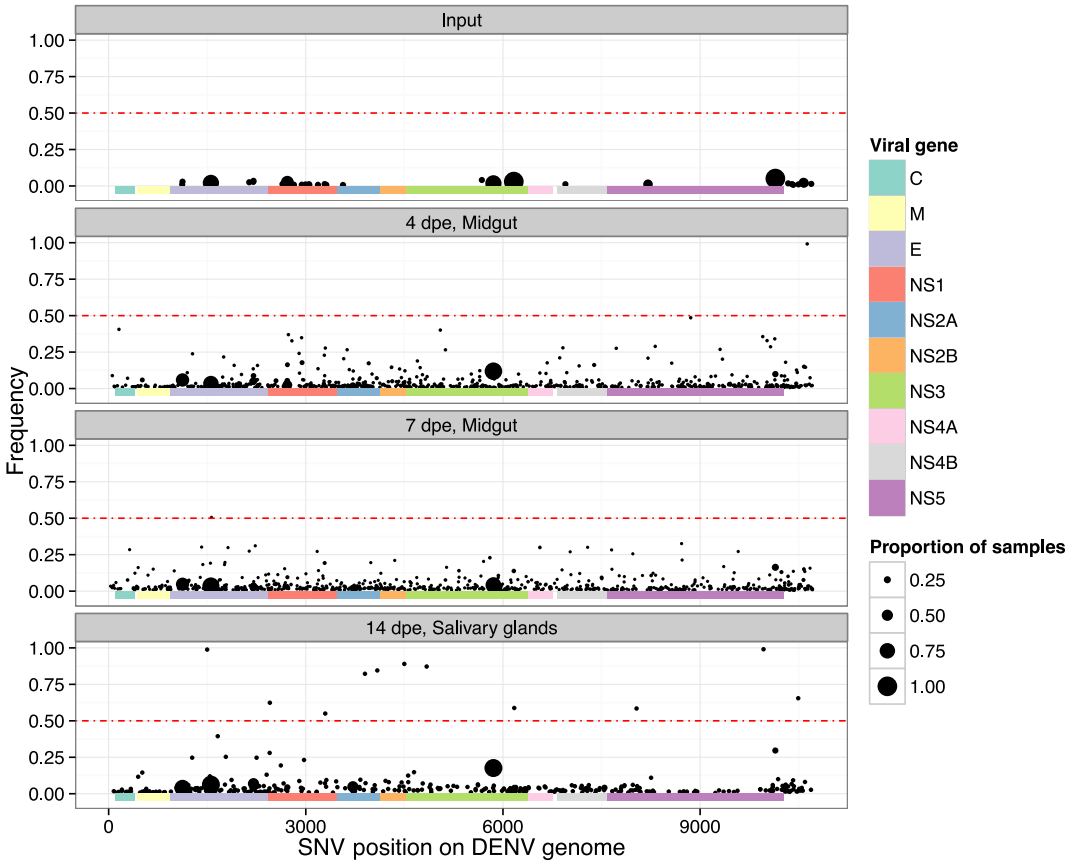

Figure 3 – Distribution of SNV positions and their mean detected frequencies in the full marker set. Each dot represents the minor allele frequency of a single SNV along the DENV reference genome indicated on the x-axis, averaged over all samples in which the SNV was detected. Dot size corresponds to the proportion of samples from the same time point in which the SNV was detected. The horizontal red dashed line represents a frequency of 0.5 over which a new variant becomes the consensus sequence. SNV distributions are stratified by time point. C=capsid protein, E=envelope glycoprotein, M=membrane glycoprotein, NS1=non-structural glycoprotein 1; NS2A=non-structural protein 2A; NS2B= non-structural protein 2B; NS3=non-structural protein 3 (protease/helicase); NS4A=non-structural protein 4A; NS4B=non-structural protein 4B; NS5=non-structural protein 5 (RNA-dependent-RNA polymerase).

did not differ statistically between time points (Figure S4-A).

To evaluate the dynamics of DENV genetic diversity during viral population expansion in the midgut, a second series of analyses compared viral genetic diversity between 4 and 7 dpe, accounting for potential differences between mosquito genotypes. In the full set of markers, both the time point (proportion of variable sites: p=0.03; S_n: p=0.04; π: p=0.04) and the isofemale line (proportion of variable sites: p=0.0035; S_n: p=0.0002; π: p=0.0005) significantly influenced viral genetic diversity. Overall, DENV genetic diversity slightly decreased between 4 dpe and 7 dpe. Isofemale line A displayed significantly higher viral genetic diversity than lines B and CD, for all three indices: proportion of variable sites (p=0.012 and p=0.0008, respectively), S_n (p=0.006 and p=0.0005 respectively) and π (p=0.017 and p=0.0001, respectively). Similar results were obtained with the conservative set of markers. Both the time point (proportion of variable sites: p=0.01; S_n: p=0.013; π: p=0.015) and the isofemale line (proportion of variable sites: p=0.0011; S_n: p=0.0006; π: p=0.0029) significantly influenced viral genetic diversity. Overall, DENV genetic diversity slightly decreased between 4 dpe and 7 dpe. Isofemale line A displayed significantly higher viral genetic diversity than lines B and CD, for all three indices: proportion of variable sites (p=0.017 and p=0.0002, respectively), S_n (p=0.0047 and p=0.0001, respectively) and π (p=0.012 and p=0.0007, respectively).

3.4 Natural selection

Based on the full set of SNV markers, d_N/d_S ratios were predominantly negative indicating strong purifying selection (Figure S4-C). There was no statistically significant difference in d_N/d_S ratios between time points or mosquito isofemale lines. Computing d_N/d_S ratios was not possible with the conservative set of markers because the smaller number of SNVs resulted in a large proportion of samples with d_S=0. Analysis of d_N/d_S ratios calculated per isofemale line, however, provided results consistent with predominantly purifying selection using the conservative set of markers. Average d_N/d_S ratios were remarkably similar among lines and time points around 0.2218 (Table S1).

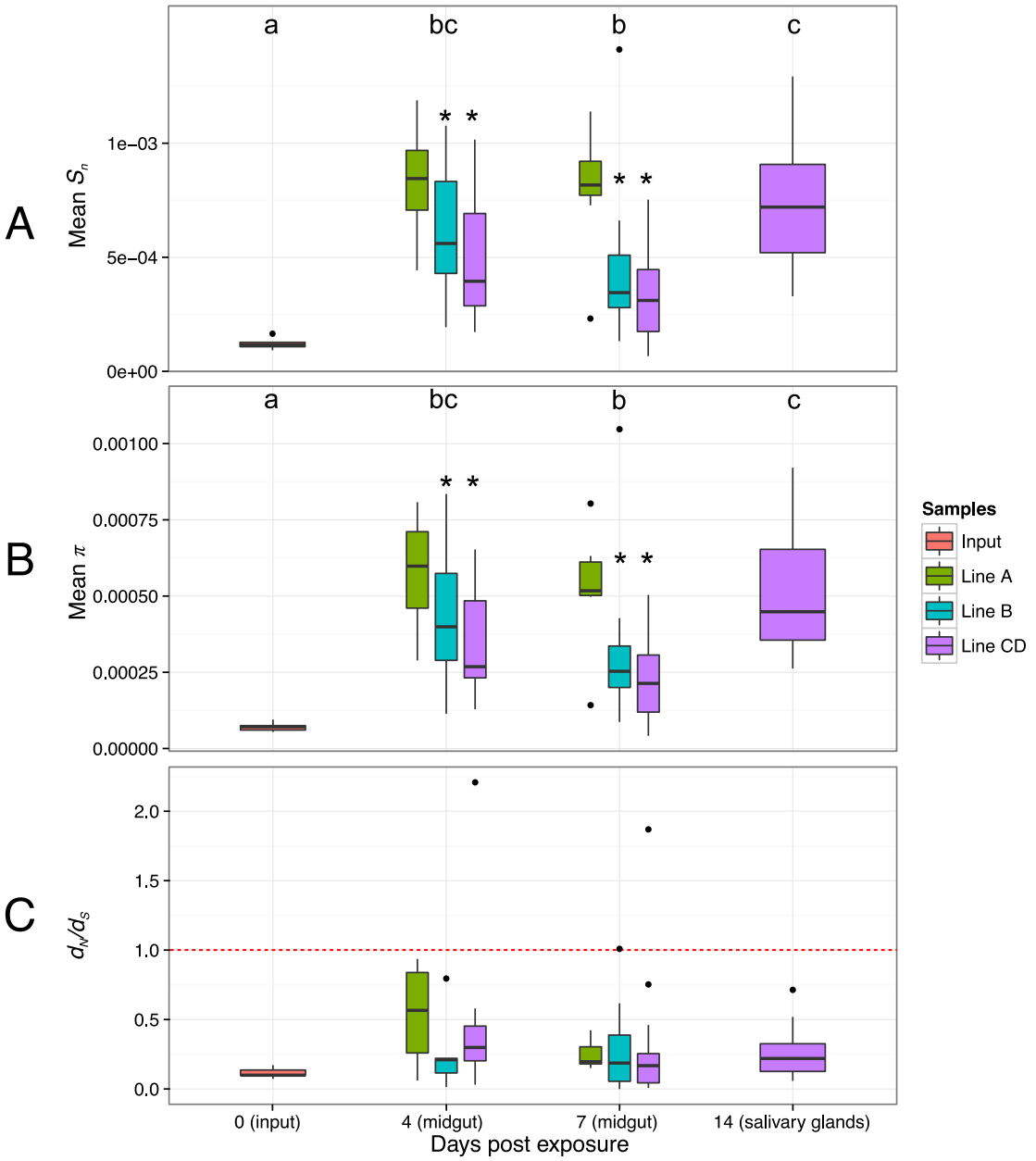

Figure 4 – Observed levels of DENV intra-host genetic diversity and natural selection assessment. (A) Averaged Shannon entropy (S_n) per site over all positions per sample. (B) Averaged nucleotide diversity (π) over all positions per sample. (C) d_N/d_S ratios over all coding positions per sample. The horizontal, dashed red line represents a d_N/d_S ratio of 1, which is interpreted as evidence for neutral evolution (i.e., absence of natural selection). A d_N/d_S ratio >1 is interpreted as evidence for positive selection; a d_N/d_S ratio <1 it is interpreted as evidence for negative (purifying) selection. Letters above the graph indicate statistically significant pairwise differences between time points. For midgut samples, stars above the bars indicate statistically significant pairwise differences between isofemale lines, with line A as the reference level.

3.5 Bottleneck size estimates

Three SNVs that complied with criteria of quasi-neutral evolution were selected to estimate the idealized number of founding viral genomes (N) initiating the midgut infection based on changes in the variance of their frequency between input and midgut samples (Table 1). Based on the three markers, initial midgut infection was founded by 23-34 genomes when estimated at 4 dpe (Figure 5-A) and 5-42 genomes when estimated at 7 dpe (Figure 5-B). Collectively, 95% confidence intervals ranged from 2 to 161 founding genomes. N estimates and their confidence intervals were consistent across time points, especially for marker at position 1556. For this marker, 4 dpe and 7 dpe data were pooled to compute isofemale line-specific N estimates. There were no statistically significant differences among lines in the estimated bottleneck size (Figure 5-C), ranging from 83 (95% confidence interval: 52–396) founding genomes for line A, to 23 (9–220) for line B and 33 (16–108) for line CD.

SNV position	Mutation	Position in codon	Amino acid	Viral gene	Initial frequency		Final frequency (4 dpe)		Final frequency (7 dpe)	
					Mean	sd	Mean	sd	Mean	sd
1556	A → C	3rd	L	E	0.021	0.004	0.026	0.028	0.031	0.023
9950	C → A	3rd	T	NS5	0.117	0.022	0.108	0.059	0.095	0.054
10145	C → T	3rd	T	NS5	0.051	0.011	0.021	0.047	0.043	0.097

Table 1 – SNV markers used for bottleneck size estimation. dpe=days post exposure; sd=standard deviation.

3.6 Bottleneck simulations

Simulations were performed to model the effect of population bottlenecks on DENV intra-host genetic diversity. The simulation randomly assigned SNV minor allele frequency, initial viral population size and bottleneck size to explore whether a minimum threshold for SNV detection would alter the observed genetic diversity following a population bottleneck compared to the true genetic diversity. When 100 SNVs were present in the input viral population and no minimum detection threshold was set, mean S_n and π estimated in 100 replicate samples decreased following the bottleneck (Figure 6-A). However, when only SNVs with a minor allele frequency >1% were detected, mean S_n and π estimates increased after the bottleneck (Figure 6-B).

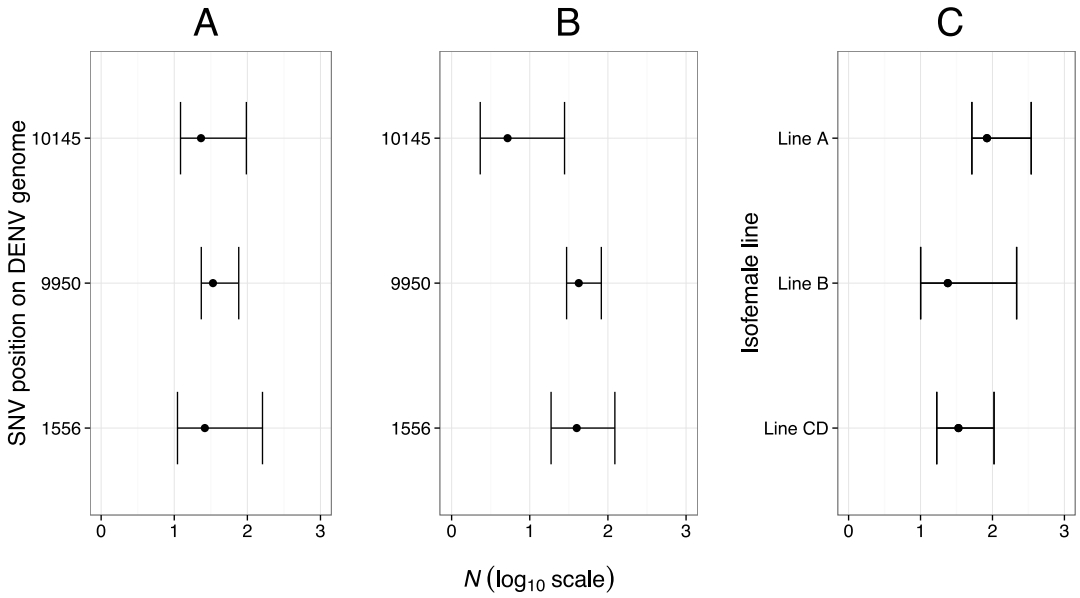

Figure 5 – Estimates of bottleneck size at initial midgut infection. The estimated number of founding genomes (N) that contribute to initial midgut infection is shown for three markers identified by their position on the DENV genome (1556, 9950, 10145). The markers are SNVs that are assumed to evolve neutrally or quasi-neutrally. Horizontal bars indicate confidence intervals of N estimates computed by bootstrapping. (A) N estimates based on samples collected at 4 dpe. (B) N estimates based on samples collected at 7 dpe. (C) N estimates for each isofemale line obtained for marker 1556 using combined 4 dpe and 7 dpe samples.

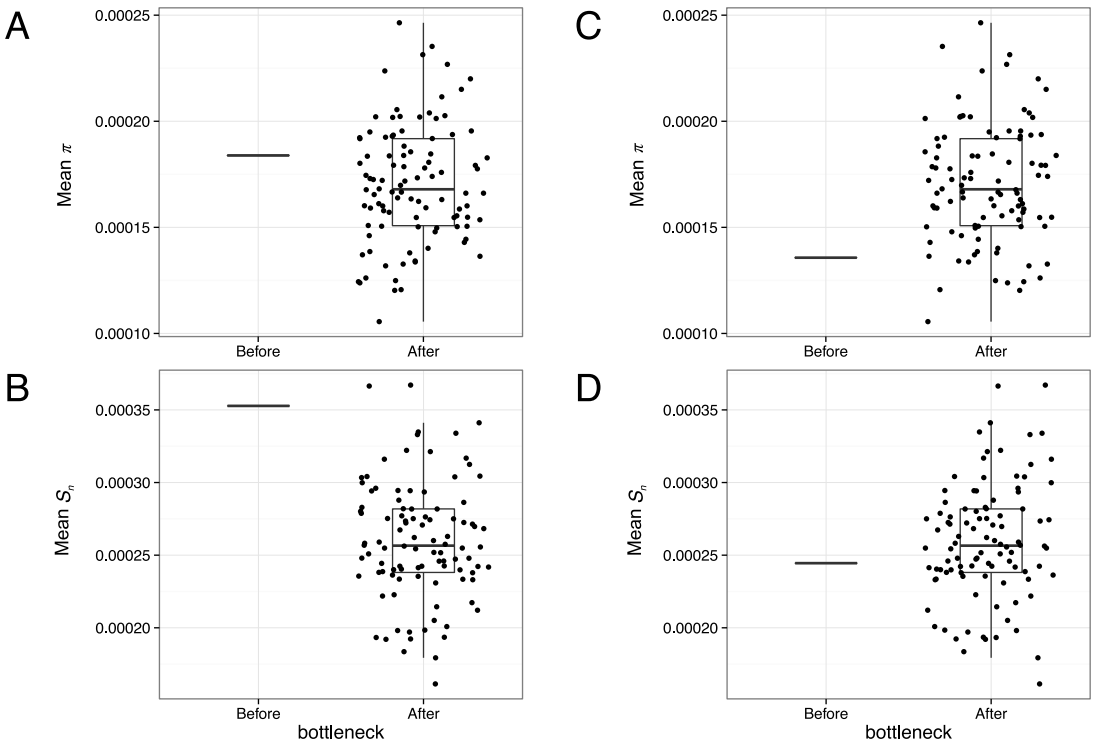

Figure 6 – Simulated effects of a population bottleneck and SNV detection threshold on observed levels of genetic diversity. The simulation considered 100 SNVs present in the input population, which were sampled 100 times (infection events) with randomly assigned SNV frequency, initial viral population size and bottleneck size. In the population sampled following the bottleneck, mean π (A, C) and S_n (B, D) are shown when no frequency detection threshold was set (A, B), and when a 1% frequency detection threshold was set (C, D).

4 Discussion

We investigated the evolutionary forces acting on DENV populations within their arthropod vector. Specifically, we evaluated the relative effects of natural selection and genetic drift on DENV intra-host evolution in the midgut of *Ae. aegypti*. In addition, we assessed the influence of vector genetic heterogeneity on intra-host viral genetic diversity. Our results show that DENV intra-host genetic diversity in *Ae. aegypti* is shaped by the combined effects of genetic drift, purifying selection and vector genotype. Reshuffling of the variant repertoire during initial infection of the midgut was associated with a bottleneck size ranging from 5 to 42 founding viral genomes, irrespective of the mosquito genotype. DENV genetic diversity increased significantly following initial infection, but was restricted by strong purifying selection during DENV population expansion in the midgut. Observed levels of DENV genetic diversity in the midgut differed significantly between mosquito isofemale lines despite a similar bottleneck size at initial infection.

Arboviruses typically maintain high levels of genetic diversity during transmission by their arthropod vectors despite anatomical barriers that often result in severe population drops [Forrester et al., 2014]. Such population bottlenecks have been documented for several arboviruses in their vectors using artificial mutant swarms [Ciota et al., 2012a], marked viral clones [Forrester et al., 2012], viral replicons [Gutiérrez et al., 2015] or stochastic simulations based on observed changes in variant frequencies [Sim et al., 2015]. Although the overall level of arboviral genetic diversity is usually maintained during vector infection [Brackney et al., 2011], the viral variant repertoire can be significantly altered [Ciota et al., 2012a, Stapleford et al., 2014, Sim et al., 2015]. Presumably, viral genetic diversity is quickly replenished by mutation and demographic expansion following population bottlenecks [Forrester et al., 2014]. However, whether changes in the viral variant repertoire are due to stochastic sampling (i.e., genetic drift), purifying selection (i.e., removal of variants with lower fitness), or vector genetic heterogeneity combined with specific interactions between vector and virus genotypes [Lambrechts et al., 2009, Lambrechts, 2011, Lambrechts et al., 2013, Fansiri et al., 2013, Dickson et al., 2014] has remained largely unresolved. Our analysis used neutral or quasi-neutral genetic markers to estimate the effective DENV population size during initial infection of the *Ae. aegypti* midgut. This approach rules out natural

selection and only measures the effect of genetic drift due to random sampling. It is worth noting, however, that true neutral mutation may not exist because even synonymous mutations can have a fitness effect, especially in RNA viruses [Cuevas et al., 2012]. Deviation from our assumption of neutrality or quasi-neutrality of the chosen markers may have overestimated the bottleneck size. Indeed, both positive selection and negative selection are expected to decrease the variance of marker frequency and therefore result in a larger estimate of N_e with our method. Therefore, our conclusion that DENV populations undergo a strong population bottleneck during initial midgut infection should be robust to any undetected departure from neutrality. Moreover, we chose markers whose average frequency was similar before and after the bottleneck, supporting the assumption that they were not under directional selection. Our estimation that initial midgut infection is founded by only a few tens of DENV genomes is consistent with previous estimations for DENV based on stochastic simulations using empirical data [Sim et al., 2015]. We went one step further by demonstrating that genetic drift, rather than natural selection, is the main evolutionary force underlying this population bottleneck.

Although our estimated bottleneck size is larger than for other RNA viruses during host-to-host transmission [Gutiérrez et al., 2012], it is expected to substantially reduce the genetic breadth of the viral quasispecies [Escarmis et al., 2006]. This finding has important implications for DENV evolution in general. A small effective population size means that natural selection will be relatively inefficient during human-to-mosquito transmission. It will prevent adaptive evolution especially if beneficial SNVs are present at low frequencies in the mutant swarm transmitted from the human host [Novella et al., 1999]. On the other hand, the population bottleneck associated with initial midgut infection may not be small enough to prevent the long-term maintenance of defective viral genomes through complementation by co-infection of host cells with functional viruses. Such a phenomenon was previously documented in the case of a stop-codon mutation that became widespread in DENV populations sampled in Myanmar in 2001 [Aaskov et al., 2006]. The frequency of the stop-codon mutation was likely high enough to overcome the effect of population bottlenecks during multiple host-to-host transmission events.

During DENV population expansion following initial midgut infection, natural selection was predominantly negative (i.e., acting against change). Accordingly, the

consensus sequence remained unchanged in most of the midgut samples. Only in the salivary glands did several SNVs reach consensus level (frequency >0.5), but with no evidence of evolutionary convergence. As was observed for West Nile virus [Ciota et al., 2012a], DENV intra-host genetic diversity in midguts slightly decreased between 4 and 7 dpe. Importantly, we found that overall levels of DENV intra-host genetic diversity differed significantly between distinct mosquito genetic backgrounds. Both the initial bottleneck size and the census size of the viral population did not significantly vary among mosquito genotypes, and thus are unlikely to explain this difference. The mechanistic basis of this difference remains to be determined, but we speculate that viral populations may undergo different selective constraints in different mosquito genotypes. Mosquito genotypes could vary in the intensity of purifying selection (i.e., variation in the efficiency of mechanisms that remove deleterious *de novo* mutations), but this was not supported by our data. Likewise, the overall lack of positive selection that we observed indicates that it is unlikely to be the underlying mechanism. Alternatively, mosquito genotypes may differ in the level of balancing selection (i.e., mechanisms that act to promote genetic diversity such as negative frequency-dependent selection). The antiviral RNA interference (RNAi) pathway of mosquitoes was suggested to play a role in viral genetic 'diversification' [Brackney et al., 2009, Brackney et al., 2015], by promoting escape to complementary base-pairing required for RNAi-mediated cleavage. Variation in host factors could also result in differences in viral intra-host genetic diversity through subtle changes in viral RNA-dependent RNA polymerase fidelity [Combe and Sanjuán, 2014]. Mutation rates of RNA viruses are not only determined by virus-encoded factors, by also by host-dependent processes. Replicase fidelity of a plant RNA virus was found to differ according to the host type [Pita et al., 2007]. Replication fidelity in retroviruses can be affected by intracellular dNTP imbalance [Bebenek et al., 1992, Julias and Pathak, 1998]. Viral mutation rate can also be influenced by the expression of host genes, such as cellular deaminases that promote hypermutation in RNA viruses [Hajjar and Linial, 1995, Cattaneo et al., 1988, O'Hara et al., 1984].

Interestingly, the isofemale line that displayed the lowest level of DENV genetic diversity in the midgut (i.e., line CD) was also associated with the highest prevalence and highest viral load in salivary glands. Because we did not examine viral populations in saliva samples, whether this translates in differences of virus transmission potential

is unknown. It is tempting to speculate that the vector competence phenotype relates to the level of viral genetic diversity. Unfortunately, we could not compare DENV intra-host diversity in salivary glands between mosquito isofemale lines because DENV amplification was unsuccessful in two out of three lines due to low template concentration. A recent study found differences in the intra-host genetic diversity of West Nile virus among different species of *Culex* mosquitoes [Grubaugh et al., 2016]. Here, we provided evidence that such differences exist at the intra-specific level. The potential relationship between viral intra-host genetic diversity and vector competence variation among mosquito genotypes deserves further investigation. It will be intereseing to determine in future experiments whether the effect of the vector genotype varies according to the mosquito generation, the virus strain, and/or the specific combinations of mosquito lines and virus strains.

Finally, we introduced a non-exclusinve, alternative scenario to the 'diversification' hypothesis that may contribute to explain why the level levels of arboviral genetic diversity increases despite a population bottleneck. Our proposed scenario is based on the counter-intuitive idea that a strong initial population bottleneck may actually result in a higher observed level of genetic diversity if low-frequency SNVs go undetected for methodological reasons. We used a model based on stochastic simulations to illustrate the effect of a minimum detection threshold of low-frequency SNVs on observed genetic diversity. When all SNVs present were detected regardless of their frequency (i.e., no detection threshold), mean viral genetic diversity indices decreased following a simulated population bottleneck. Conversely, mean genetic diversity indices increased when only SNVs present at a frequency >1% were successfully detected. In our empirical data, it was not possible to ascertain whether SNVs newly detected after the initial population bottleneck resulted from *de novo* mutations or were already present prior to the bottleneck at frequencies lower than the detection threshold. However, our model indicated that a change in the SNV frequency spectrum following the population bottleneck combined with a minimum detection threshold is a potential explanation to the observed increased genetic diversity following the initial bottleneck.

Taken together, our results show that DENV intra-host genetic diversity in the mosquito vector is shaped by stochastic events during initial midgut infection due to a sharp reduction in population size, followed by predominantly purifying selection

during population expansion and diversification in the midgut. Differential diversification between mosquito isofemale lines indicates a genetic foundation, but the lack of convergent SNVs does not support the existence of mosquito genotype-specific directional selection. We conclude that the evolution of DENV intra-host genetic diversity in mosquitoes is not only driven by genetic drift and purifying selection, but is also modulated by vector genetic factors. Characterizing the evolutionary forces that govern arboviral genetic diversity contributes to understanding their unique biology and adaptive potential.

Acknowledgements

The authors thank Marco Vignuzzi, Kenneth Stapleford, Hervé Blanc, Etienne Patin, Maud Fagny, Guillaume Laval, and all members of the Lambrechts lab for insightful discussions. We thank Laura Dickson for critical reading of an earlier version of the manuscript, Catherine Lallemand for assistance with mosquito rearing, and Gordana Rašić and Igor Filipović for help with mosquito genotyping. We are grateful to Rosmari Rodríguez-Roche for providing some of the PCR primer sequences, and to Alongkot Ponlawat and Thanyalak Fansiri for the initial field collection of mosquitoes. We also thank two anonymous reviewers for their constructive comments, which helped improve an earlier version of the manuscript.

References

[Aaskov et al., 2006] Aaskov, J., Buzacott, K., Thu, H. M., Lowry, K., and Holmes, E. C. (2006). Long-term transmission of defective RNA viruses in humans and Aedes mosquitoes. *Science*, 311(5758):236–238.

[Anderson and Rico-Hesse, 2006] Anderson, J. R. and Rico-Hesse, R. (2006). Aedes aegypti vectorial capacity is determined by the infecting genotype of dengue virus. *American Journal of Tropical Medicine and Hygiene*, 75(5):886–892.

[Armstrong and Rico-Hesse, 2003] Armstrong, P. M. and Rico-Hesse, R. (2003). Efficiency of dengue serotype 2 virus strains to infect and disseminate in Aedes aegpti. *American Journal of Tropical Medicine and Hygiene*, 68(5):539–544.

[Auguie, 2015] Auguie, B. (2015). Miscellaneous Functions for "Grid" Graphics.

[Bebenek et al., 1992] Bebenek, K., Roberts, J. D., and Kunkel, T. A. (1992). The effects of dNTP pool imbalances on frameshift fidelity during DNA replication. *The Journal of biological chemistry*, 267(6):3589–3596.

[Bennett et al., 2002] Bennett, K. E., Olson, K. E., Munoz, M. d. L., Fernandez-Salas, I., Farfan-Ale, J. A., Higgs, S., Black, W. C., and Beaty, B. J. (2002). Variation in vector competence for dengue 2 virus among 24 collections of Aedes aegypti from Mexico and the United States. *American Journal of Tropical Medicine and Hygiene*, 67(1):85–92.

[Black and DuTeau, 1997] Black, W. C. and DuTeau, N. M. (1997). RAPD-PCR and SSCP analysis for insect population genetic studies. In Crampton, J. M., Beard, C. B., and Louis, C., editors, *The Molecular Biology of Insect Disease Vectors*, pages 361–373. Chapman & Hall, London.

[Bolger et al., 2014] Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15):2114–2120.

[Brackney et al., 2009] Brackney, D. E., Beane, J. E., and Ebel, G. D. (2009). RNAi targeting of West Nile virus in mosquito midguts promotes virus diversification. *PLoS Pathogens*, 5(7):e1000502.

[Brackney et al., 2011] Brackney, D. E., Pesko, K. N., Brown, I. K., Deardorff, E. R., Kawatachi, J., and Ebel, G. D. (2011). West Nile Virus Genetic Diversity is Maintained during Transmission by Culex pipiens quinquefasciatus Mosquitoes. *Plos One*, 6(9).

[Brackney et al., 2015] Brackney, D. E., Schirtzinger, E. E., Harrison, T. D., Ebel, G. D., and Hanley, K. A. (2015). Modulation of flavivirus population diversity by RNA interference. *Journal of Virology*, 89(7):4035–4039.

[Canty and Ripley, 2015] Canty, A. and Ripley, B. (2015). boot: Bootstrap R (S-Plus) Functions. .

[Catchen et al., 2013] Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., and Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular ecology*, 22(11):3124–3140.

[Catchen et al., 2011] Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., and Postlethwait, J. H. (2011). Stacks: building and genotyping Loci de novo from short-read sequences. *G3 (Bethesda, Md.)*, 1(3):171–182.

[Cattaneo et al., 1988] Cattaneo, R., Schmid, A., Eschle, D., Baczko, K., ter Meulen, V., and Billeter, M. A. (1988). Biased hypermutation and other genetic changes in defective measles viruses in human brain infections. *Cell*, 55(2):255–265.

[Chao et al., 2005] Chao, D.-Y., King, C.-C., Wang, W.-K., Chen, W.-J., Wu, H.-L., and Chang, G.-J. J. (2005). Strategically examining the full-genome of dengue virus type 3 in clinical isolates reveals its mutation spectra. *Virology journal*, 2:72.

[Chen et al., 2013] Chen, R., Wang, E., Tsetsarkin, K. A., and Weaver, S. C. (2013). Chikungunya virus 3' untranslated region: adaptation to mosquitoes and a population bottleneck as major evolutionary forces. *PLoS Pathogens*, 9(8):e1003591.

[Ciota et al., 2012a] Ciota, A. T., Ehrbar, D. J., Van Slyke, G. A., Payne, A. F., Willsey, G. G., Viscio, R. E., and Kramer, L. D. (2012a). Quantification of intrahost bottlenecks of West Nile virus in Culex pipiens mosquitoes using an artificial mutant swarm. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 12(3):557–564.

[Ciota et al., 2012b] Ciota, A. T., Ehrbar, D. J., Van Slyke, G. A., Willsey, G. G., and Kramer, L. D. (2012b). Cooperative interactions in the West Nile virus mutant swarm. *BMC evolutionary biology*, 12(1):58.

[Ciota et al., 2007] Ciota, A. T., Ngo, K. A., Lovelace, A. O., Payne, A. F., Zhou, Y., Shi, P.-Y., and Kramer, L. D. (2007). Role of the mutant spectrum in adaptation and replication of West Nile virus. *The Journal of general virology*, 88(Pt 3):865–874.

[Coffey et al., 2011] Coffey, L. L., Beeharry, Y., Bordería, A. V., Blanc, H., and Vignuzzi, M. (2011). Arbovirus high fidelity variant loses fitness in mosquitoes and mice. *Proceedings of the National Academy of Sciences of the United States of America*, 108(38):16038–16043.

[Coffey et al., 2013] Coffey, L. L., Forrester, N., Tsetsarkin, K., Vasilakis, N., and Weaver, S. C. (2013). Factors shaping the adaptive landscape for arboviruses: implications for the emergence of disease. *Future microbiology*, 8(2):155–176.

[Combe and Sanjuán, 2014] Combe, M. and Sanjuán, R. (2014). Variation in RNA virus mutation rates across host cells. *PLoS Pathogens*, 10(1):e1003855.

[Cornman et al., 2013] Cornman, R. S., Boncristiani, H., Dainat, B., Chen, Y., vanEngelsdorp, D., Weaver, D., and Evans, J. D. (2013). Population-genomic variation within RNA viruses of the Western honey bee, Apis mellifera, inferred from deep sequencing. *BMC genomics*, 14(1):154.

[Costello et al., 2013] Costello, M., Pugh, T. J., Fennell, T. J., Stewart, C., Lichtenstein, L., Meldrim, J. C., Fostel, J. L., Friedrich, D. C., Perrin, D., Dionne, D., Kim, S., Gabriel, S. B., Lander, E. S., Fisher, S., and Getz, G. (2013). Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic Acids Research*, 41(6):–e67.

[Cuevas et al., 2012] Cuevas, J. M., Domingo-Calap, P., and Sanjuán, R. (2012). The fitness effects of synonymous mutations in DNA and RNA viruses. *Molecular biology and evolution*, 29(1):17–20.

[Delignette-Muller et al., 2015] Delignette-Muller, M.-L., Dutang, C., Pouillot, R., and Denis, J.-B. (2015). Help to Fit of a Parametric

Distribution to Non-Censored or Censored Data.

[Descloux et al., 2009] Descloux, E., Cao-Lormeau, V.-M., Roche, C., and De Lamballerie, X. (2009). Dengue 1 diversity and microevolution, French Polynesia 2001-2006: connection with epidemiology and clinics. *PLoS neglected tropical diseases*, 3(8):e493.

[Dickson et al., 2014] Dickson, L. B., Sanchez-Vargas, I., Sylla, M., Fleming, K., and Black, W. C. (2014). Vector competence in West African Aedes aegypti Is Flavivirus species and genotype dependent. *PLoS neglected tropical diseases*, 8(10):e3153.

[Doran and Creevey, 2013] Doran, A. G. and Creevey, C. J. (2013). Snpdat: Easy and rapid annotation of results from de novo snp discovery projects for model and non-model organisms. *Bmc Bioinformatics*, 14(1).

[Drummond and Rambaut, 2007] Drummond, A. J. and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology*, 7(1):214.

[Duarte et al., 1992] Duarte, E., Clarke, D., Moya, A., Domingo, E., and Holland, J. (1992). Rapid fitness losses in mammalian RNA virus clones due to Muller's ratchet. *Proceedings of the National Academy of Sciences of the United States of America*, 89(13):6015–6019.

[Escarmis et al., 2006] Escarmis, C., Lazaro, E., and Manrubia, S. C. (2006). Population bottlenecks in quasispecies dynamics. *Quasispecies: Concept and Implications for Virology*, 299:141–170.

[Fansiri et al., 2013] Fansiri, T., Fontaine, A., Diancourt, L., Caro, V., Thaisomboonsuk, B., Richardson, J. H., Jarman, R. G., Ponlawat, A., and Lambrechts, L. (2013). Genetic Mapping of Specific Interactions between Aedes aegypti Mosquitoes and Dengue Viruses. *PLoS Genet*, 9(8):e1003621.

[Forrester et al., 2014] Forrester, N. L., Coffey, L. L., and Weaver, S. C. (2014). Arboviral bottlenecks and challenges to maintaining diversity and fitness during mosquito transmission. *Viruses*, 6(10):3991–4004.

[Forrester et al., 2012] Forrester, N. L., Guerbois, M., Seymour, R. L., Spratt, H., and Weaver, S. C. (2012). Vector-borne transmission imposes a severe bottleneck on an RNA virus population. *PLoS Pathogens*, 8(9):e1002897.

[Fox and Weisberg, 2011] Fox, J. and Weisberg, S. (2011). *An R Companion to Applied Regression*. SAGE.

[Gregori et al., 2014] Gregori, J., Salicru, M., Domingo, E., Sanchez, A., Esteban, J. I., Rodriguez-Frias, F., and Quer, J. (2014). Inference with viral quasispecies diversity indices: clonal and NGS approaches. *Bioinformatics*, 30(8):1104–1111.

[Grubaugh et al., 2015] Grubaugh, N. D., Smith, D. R., Brackney, D. E., Bosco-Lauth, A. M., Fauver, J. R., Campbell, C. L., Felix, T. A., Romo, H., Duggal, N. K., Dietrich, E. A., Eike, T., Beane, J. E., Bowen, R. A., Black, W. C., Brault, A. C., and Ebel, G. D. (2015). Experimental evolution of an RNA virus in wild birds: evidence for host-dependent impacts on population structure and competitive fitness. *PLoS Pathogens*, 11(5):e1004874.

[Grubaugh et al., 2016] Grubaugh, N. D., Weger-Lucarelli, J., Murrieta, R. A., Fauver, J. R., Garcia-Luna, S. M., Prasad, A. N., Black, W. C., and Ebel, G. D. (2016). Genetic Drift during Systemic Arbovirus Infection of Mosquito Vectors Leads to Decreased Relative Fitness during Host Switching. *Cell host & microbe*.

[Gubler et al., 1979] Gubler, D. J., Nalim, S., Tan, R., Saipan, H., and Sulianti Saroso, J. (1979). Variation in Susceptibility to Oral Infection with Dengue Viruses Among Geographic Strains of Aedes-Aegypti. *American Journal of Tropical Medicine and Hygiene*, 28(6):1045–1052.

[Gutiérrez et al., 2012] Gutiérrez, S., Michalakis, Y., and Blanc, S. (2012). Virus population bottlenecks during within-host progression and host-to-host transmission. *Current opinion in virology*, 2(5):546–555.

[Gutiérrez et al., 2015] Gutiérrez, S., Thébaud, G., Smith, D. R., Kenney, J. L., and Weaver, S. C. (2015). Demographics of natural oral infection of mosquitos by Venezuelan equine encephalitis virus. *Journal of Virology*, 89(7):4020–4022.

[Hajjar and Linial, 1995] Hajjar, A. M. and Linial, M. L. (1995). Modification of retroviral RNA by double-stranded RNA adenosine deaminase. *Journal of Virology*, 69(9):5878–5882.

[Jerzak et al., 2007] Jerzak, G. V. S., Bernard, K., Kramer, L. D., Shi, P.-Y., and Ebel, G. D. (2007). The West Nile virus mutant spectrum is host-dependant and a determinant of mortality in mice. *Virology*, 360(2):469–476.

[Julias and Pathak, 1998] Julias, J. G. and Pathak, V. K. (1998). Deoxyribonucleoside triphosphate pool imbalances in vivo are associated with an increased retroviral mutation rate. *Journal of Virology*, 72(10):7941–7949.

[Kumar et al., 2016] Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular biology and evolution*, page msw054.

[Lambrechts, 2011] Lambrechts, L. (2011). Quantitative genetics of Aedes aegypti vector competence for dengue viruses: towards a new paradigm? *Trends in parasitology*, 27(3):111–114.

[Lambrechts et al., 2009] Lambrechts, L., Chevillon, C., Albright, R. G., Thaisomboonsuk, B., Richardson, J. H., Jarman, R. G., and Scott, T. W. (2009). Genetic specificity and potential for local adaptation between dengue viruses and mosquito vectors. *BMC evolutionary biology*, 9(1).

[Lambrechts et al., 2013] Lambrechts, L., Quillery, E., Noel, V., Richardson, J. H., Jarman, R. G., Scott, T. W., and Chevillon, C. (2013). Specificity of resistance to dengue virus isolates is associated with genotypes of the mosquito antiviral gene Dicer-2. *Proceedings of the Royal Society B-Biological Sciences*, 280(1751):–20122437.

[Langmead and Salzberg, 2012] Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4):357–U54.

[Langmead et al., 2009] Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*, 10(3):R25.

[Lauring and Andino, 2010] Lauring, A. S. and Andino, R. (2010). Quasispecies Theory and the Behavior of RNA Viruses. *PLoS Pathogens*, 6(7).

[Lequime and Lambrechts, 2014] Lequime, S. and Lambrechts, L. (2014). Vertical transmission of arboviruses in mosquitoes: a historical perspective. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 28:681–690.

[Lequime et al., 2016] Lequime, S., Paul, R. E., and Lambrechts, L. (2016). Determinants of arbovirus vertical transmission in mosquitoes. *PLoS Pathogens*.

[Li et al., 2009] Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):2078–2079.

[Lin et al., 2004] Lin, S.-R., Hsieh, S.-C., Yueh, Y.-Y., Lin, T.-H., Chao, D.-Y., Chen, W.-J., King, C.-C., and Wang, W.-K. (2004). Study of sequence variation of dengue type 3 virus in naturally infected mosquitoes and human hosts: implications for transmission and evolution. *Journal of Virology*, 78(22):12717–12721.

[Martin, 2011] Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1):pp. 10–12.

[Miller et al., 2007] Miller, M. R., Dunham, J. P., Amores, A., Cresko, W. A., and Johnson, E. A. (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome research*, 17(2):240–248.

[Monsion et al., 2008] Monsion, B., Froissart, R., Michalakis, Y., and Blanc, S. (2008). Large bottleneck size in Cauliflower Mosaic Virus populations during host plant colonization. *PLoS Pathogens*, 4(10):e1000174.

[Nei and Gojobori, 1986] Nei, M. and Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular biology and evolution*, 3(5):418–426.

[Nene et al., 2007] Nene, V., Wortman, J. R., Lawson, D., Haas, B., Kodira, C., Tu, Z. J., Loftus, B., Xi, Z., Megy, K., Grabherr, M., Ren, Q., Zdobnov, E. M., Lobo, N. F., Campbell, K. S., Brown, S. E., Bonaldo, M. F., Zhu, J., Sinkins, S. P., Hogenkamp, D. G., Amedeo, P., Arensburger, P., Atkinson, P. W., Bidwell, S., Biedler, J., Birney, E., Bruggner, R. V., Costas, J., Coy, M. R., Crabtree, J., Crawford, M., Debruyn, B., Decaprio, D., Eiglmeier, K., Eisenstadt, E., El-Dorry, H., Gelbart, W. M., Gomes, S. L., Hammond, M., Hannick, L. I., Hogan, J. R., Holmes, M. H., Jaffe, D., Johnston, J. S., Kennedy, R. C., Koo, H., Kravitz, S., Kriventseva, E. V., Kulp, D., Labutti, K., Lee, E., Li, S., Lovin, D. D., Mao, C., Mauceli, E., Menck, C. F. M., Miller, J. R., Montgomery, P., Mori, A., Nascimento, A. L., Naveira, H. F., Nusbaum, C., O'leary, S., Orvis, J., Pertea, M., Quesneville, H., Reidenbach, K. R., Rogers, Y.-H., Roth, C. W., Schneider, J. R., Schatz, M., Shumway, M., Stanke, M., Stinson, E. O., Tubio, J. M. C., Vanzee, J. P., Verjovski-Almeida, S., Werner, D., White, O., Wyder, S., Zeng, Q., Zhao, Q., Zhao, Y., Hill, C. A., Raikhel, A. S., Soares, M. B., Knudson, D. L., Lee, N. H., Galagan, J., Salzberg, S. L., Paulsen, I. T., Dimopoulos, G., Collins, F. H., Birren, B., Fraser-Liggett, C. M., and Severson, D. W. (2007). Genome sequence of Aedes aegypti, a major arbovirus vector. *Science*, 316(5832):1718–1723.

[Novella et al., 1999] Novella, I. S., Quer, J., Domingo, E., and Holland, J. J. (1999). Exponential fitness gains of RNA virus populations are limited by bottleneck effects. *Journal of Virology*, 73(2):1668–1671.

[O'Hara et al., 1984] O'Hara, P. J., Nichol, S. T., Horodyski, F. M., and Holland, J. J. (1984). Vesicular stomatitis virus defective interfering particles can contain extensive genomic sequence

rearrangements and base substitutions. *Cell*, 36(4):915–924.

[Parameswaran et al., 2012] Parameswaran, P., Charlebois, P., Tellez, Y., Nunez, A., Ryan, E. M., Malboeuf, C. M., Levin, J. Z., Lennon, N. J., Balmaseda, A., Harris, E., and Henn, M. R. (2012). Genome-wide patterns of intrahuman dengue virus diversity reveal associations with viral phylogenetic clade and interhost diversity. *Journal of Virology*, 86(16):8546–8558.

[Peterson et al., 2012] Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., and Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *Plos One*, 7(5):e37135.

[Pita et al., 2007] Pita, J. S., de Miranda, J. R., Schneider, W. L., and Roossinck, M. J. (2007). Environment determines fidelity for an RNA virus replicase. *Journal of Virology*, 81(17):9072–9077.

[Quinlan and Hall, 2010] Quinlan, A. R. and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842.

[Rašić et al., 2014] Rašić, G., Filipović, I., Weeks, A. R., and Hoffmann, A. A. (2014). Genome-wide SNPs lead to strong signals of geographic structure and relatedness patterns in the major arbovirus vector, Aedes aegypti. *BMC genomics*, 15(1):275.

[Rosen et al., 1985] Rosen, L., Roseboom, L. E., Gubler, D. J., Lien, J. C., and Chaniotis, B. N. (1985). Comparative Susceptibility of Mosquito Species and Strains to Oral and Parenteral Infection with Dengue and Japanese Encephalitis Viruses. *American Journal of Tropical Medicine and Hygiene*, 34(3):603–615.

[Rozen-Gagnon et al., 2014] Rozen-Gagnon, K., Stapleford, K. A., Mongelli, V., Blanc, H., Failloux, A.-B., Saleh, M.-C., and Vignuzzi, M. (2014). Alphavirus mutator variants present host-specific defects and attenuation in mammalian and insect models. *PLoS Pathogens*, 10(1):e1003877.

[Sessions et al., 2015] Sessions, O. M., Wilm, A., Kamaraj, U. S., Choy, M. M., Chow, A., Chong, Y., Ong, X. M., Nagarajan, N., Cook, A. R., and Ooi, E. E. (2015). Analysis of Dengue Virus Genetic Diversity during Human and Mosquito Infection Reveals Genetic Constraints. *PLoS neglected tropical diseases*, 9(9):e0004044.

[Sim et al., 2015] Sim, S., Aw, P. P. K., Wilm, A., Teoh, G., Hue, K. D. T., Nguyen, N. M., Nagarajan, N., Simmons, C. P., and Hibberd, M. L. (2015). Tracking Dengue Virus Intra-host Genetic Diversity during Human-to-Mosquito Transmission. *PLoS neglected tropical diseases*, 9(9):e0004052.

[Stapleford et al., 2014] Stapleford, K. A., Coffey, L. L., Lay, S., Bordería, A. V., Duong, V., Isakov, O., Rozen-Gagnon, K., Arias-Goeta, C., Blanc, H., Beaucourt, S., Haliloğlu, T., Schmitt, C., Bonne, I., Ben-Tal, N., Shomron, N., Failloux, A.-B., Buchy, P., and Vignuzzi, M. (2014). Emergence and transmission of arbovirus evolutionary intermediates with epidemic potential. *Cell host & microbe*, 15(6):706–716.

[Tardieux et al., 1990] Tardieux, I., Poupel, O., Lapchin, L., and Rodhain, F. (1990). Variation Among Strains of Aedes-Aegypti in Susceptibility to Oral Infection with Dengue Virus Type-2. *American Journal of Tropical Medicine and Hygiene*, 43(3):308–313.

[Thai et al., 2012] Thai, K. T. D., Henn, M. R., Zody, M. C., Tricou, V., Nguyet, N. M., Charlebois, P., Lennon, N. J., Green, L., de Vries, P. J., Hien, T. T.,

Farrar, J., van Doorn, H. R., de Jong, M. D., Birren, B. W., Holmes, E. C., and Simmons, C. P. (2012). High-resolution analysis of intrahost genetic diversity in dengue virus serotype 1 infection identifies mixed infections. *Journal of Virology*, 86(2):835–843.

[Vazeille-Falcoz et al., 1999] Vazeille-Falcoz, M., Mousson, L., Rodhain, F., Chungue, E., and Failloux, A.-B. (1999). Variation in oral susceptibility to dengue type 2 virus of populations of Aedes aegypti from the islands of Tahiti and Moorea, French Polynesia. *American Journal of Tropical Medicine and Hygiene*, 60(2):292–299.

[Vignuzzi et al., 2006] Vignuzzi, M., Stone, J. K., Arnold, J. J., Cameron, C. E., and Andino, R. (2006). Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature*, 439(7074):344–348.

[Wang et al., 2002] Wang, W.-K., Lin, S. R., Lee, C. M., King, C. C., and Chang, S. C. (2002). Dengue type 3 virus in plasma is a population of closely related genomes: Quasispecies. *Journal of Virology*, 76(9):4662–4665.

[Wickham, 2007] Wickham, H. (2007). Reshaping data with the reshape package. *Journal of Statistical Software*.

[Wickham, 2009] Wickham, H. (2009). *ggplot2*. Elegant Graphics for Data Analysis. Springer Science & Business Media, New York, NY.

[Wickham, 2011] Wickham, H. (2011). The Split-Apply-Combine Strategy for Data Analysis. *Journal of Statistical Software*, 40(1):1.

[Wickham, 2015] Wickham, H. (2015). Simple, Consistent Wrappers for Common String Operations.

[Wilm et al., 2012] Wilm, A., Aw, P. P. K., Bertrand, D., Yeo, G. H. T., Ong, S. H., Wong, C. H., Khor, C. C., Petric, R., Hibberd, M. L., and Nagarajan, N. (2012). LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Research*, 40(22):11189–11201.

[Wright, 1931] Wright, S. (1931). Evolution in Mendelian Populations. *Genetics*, 16(2):97–159.

Supporting information

Day post-exposure	Sample	d_N	d_S	d_N/d_S
0	Input	5.503035×10^{-6}	2.480305×10^{-5}	0.2218693
4	Line A	2.151092×10^{-4}	9.700066×10^{-4}	0.2217606
	Line B	1.033578×10^{-4}	4.659567×10^{-4}	0.2218186
	Line CD	7.363042×10^{-5}	3.319168×10^{-4}	0.2218340
7	Line A	1.086174×10^{-4}	4.896736×10^{-4}	0.2218158
	Line B	4.520744×10^{-5}	2.037760×10^{-4}	0.2218487
	Line CD	1.538954×10^{-5}	6.936470×10^{-5}	0.2218642
14	Line CD	1.174938×10^{-4}	5.297019×10^{-4}	0.2218112

Table S1 – d_N/d_S ratios using the conservative marker set.

Number	Name	Oligo Sequence 5' - 3'	Modification	Purification	Synthesis scale
		P1			
1	TCGA_NlaIII_P1.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCGACATG	none	standard desalt	100nmole
2	GTAGT_NlaIII_P1.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTAGTCATG	none	standard desalt	100nmole
3	GCCAAT_NlaIII_P1.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCCAATCATG	none	standard desalt	100nmole
4	TAGTTT_NlaIII_P1.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAGTTTCATG	none	standard desalt	100nmole
5	CTTGG_NlaIII_P1.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTTGGCATG	none	standard desalt	100nmole
6	CGATGTA_NlaIII_P1.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGATGTACATG	none	standard desalt	100nmole
7	AGTCAA_NlaIII_P1.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGTCAACATG	none	standard desalt	100nmole
8	AACCA_NlaIII_P1.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAACCACATG	none	standard desalt	100nmole
9	TGCAT_NlaIII_P1.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGCATCATG	none	standard desalt	100nmole
10	ACTTC_NlaIII_P1.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACTTCCATG	none	standard desalt	100nmole
11	CGTAC_NlaIII_P1.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGTACCATG	none	standard desalt	100nmole
12	GGTTG_NlaIII_P1.1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGTTGCATG	none	standard desalt	100nmole
1	TCGA_NlaIII_P1.1	/5Phos/TCGAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	5'Phosphorylation	standard desalt	100nmole
2	GTAGT_NlaIII_P1.1	/5Phos/ACTACAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	5'Phosphorylation	standard desalt	100nmole
3	GCCAAT_NlaIII_P1.1	/5Phos/ATTGGCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	5'Phosphorylation	standard desalt	100nmole
4	TAGTTT_NlaIII_P1.1	/5Phos/AAACTAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	5'Phosphorylation	standard desalt	100nmole
5	CTTGG_NlaIII_P1.1	/5Phos/CCAAGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	5'Phosphorylation	standard desalt	100nmole
6	CGATGTA_NlaIII_P1.1	/5Phos/TACATCGGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	5'Phosphorylation	standard desalt	100nmole
7	AGTCAA_NlaIII_P1.1	/5Phos/TTGACTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	5'Phosphorylation	standard desalt	100nmole
8	AACCA_NlaIII_P1.1	/5Phos/TGGTTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	5'Phosphorylation	standard desalt	100nmole
9	TGCAT_NlaIII_P1.2	/5Phos/ATGCAAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	5'Phosphorylation	standard desalt	100nmole
10	ACTTC_NlaIII_P1.2	/5Phos/GAAGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	5'Phosphorylation	standard desalt	100nmole
11	CGTAC_NlaIII_P1.2	/5Phos/GTACGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	5'Phosphorylation	standard desalt	100nmole
12	GGTTG_NlaIII_P1.2	/5Phos/CAACCAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	5'Phosphorylation	standard desalt	100nmole
		P2			
1	GCTGA_MluCI_P2.1	GTGACTGGAGTTCAGACGTGT**GCTCTTCCGATCT**GCTGA	none	standard desalt	100nmole
2	ATGAG_MluCI_P2.1	GTGACTGGAGTTCAGACGTGT**GCTCTTCCGATCT**ATGAG	none	standard desalt	100nmole
3	TGGAA_MluCI_P2.1	GTGACTGGAGTTCAGACGTGT**GCTCTTCCGATCT**TGGAA	none	standard desalt	100nmole
4	CATAT_MluCI_P2.1	GTGACTGGAGTTCAGACGTGT**GCTCTTCCGATCT**CATAT	none	standard desalt	100nmole
5	ATTAC_MluCI_P2.1	GTGACTGGAGTTCAGACGTGT**GCTCTTCCGATCT**ATTAC	none	standard desalt	100nmole
1	GCTGA_MluCI_P2.2	/5Phos/AATTTCAGC**AGATCGGAAGAGC**GAGAACAA	5'Phosphorylation	standard desalt	100nmole
2	ATGAG_MluCI_P2.2	/5Phos/AATTCTCAT**AGATCGGAAGAGC**GAGAACAA	5'Phosphorylation	standard desalt	100nmole
3	TGGAA_MluCI_P2.2	/5Phos/AATTTTCCA**AGATCGGAAGAGC**GAGAACAA	5'Phosphorylation	standard desalt	100nmole
4	CATAT_MluCI_P2.2	/5Phos/AATTATATG**AGATCGGAAGAGC**GAGAACAA	5'Phosphorylation	standard desalt	100nmole
5	ATTAC_MluCI_P2.2	/5Phos/AATTGTAAT**AGATCGGAAGAGC**GAGAACAA	5'Phosphorylation	standard desalt	100nmole
	PCR 1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG			
	PCR 2	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGC			

File S1 – Primers used for mosquito RAD-sequencing.

	Primer name	Sequence	Adapted from	DENV genome fragment	Fragment size	Position on the genome	Annealing temperature	Comment
Forward	d1s1C	AGTTGTTAGTCTACGTGGAC	Christenbury et al., 2010	1	1570	1-1570	50	
Reverse	D1-3R	CCATTGTTTGTGGACGAGCC	Fansiri et al., 2013					
Forward	D1-4F	TGCATTGAAGCCAAAATATCAAA	Fansiri et al., 2013	2	1448	1112-2559	50	
Reverse	d1a17	CCAATGGCYGCTGAYAGTCT	Christenbury et al., 2010					
Forward	D1-5F	GCATGGGACTTCGGCTCTATAGG	Fansiri et al., 2013	3	1294	2189-3482	55	
Reverse	D1-5R	CTGACCCTGCAGAGACCATTGA	Fansiri et al., 2013					
Forward	D1-LAB-6Fm	AGGAGAAGATGGGTGCTGGTACGG	Fansiri et al., 2013	4	1295	3391-4685	50	
Reverse	D1-LAB-6R	CTCCCCTGGTGACGTGCCACATT	Fansiri et al., 2013					
Forward	D1-8F	CAAAGAGGACTGTTGGGCAGG	Fansiri et al., 2013	5	1273	4598-5870	55	
Reverse	D1-8R	TCACTGGCATCGGTCCGGCTA	Fansiri et al., 2013					
Forward	D1-10F	GGTAATAGACCCAAGGCGGTG	Fansiri et al., 2013	6	1172	5785-6956	60	
Reverse	D1-10R	CTGCATAGAGAGTCCAGGCTGAAG	Fansiri et al., 2013					
Forward	D1-12Fm	CACAAAGAAAGATTTAGGGATTG	Fansiri et al., 2013	7	1317	6850-8166	50	
Reverse	D1-12R	CTTTGCATTTGCTCCAGAGTTTC	Fansiri et al., 2013					
Forward	D1-14Fm	TACGCGTCCTAAAGATGGTGGAACC	Fansiri et al., 2013	8	1053	8058-9110	55	
Reverse	D1-14Rm	TGTGGAGTCCTTCTCCTTCCACTC	Fansiri et al., 2013					
Forward	D1-16F	GGAAAGGCAAAAGGAAGTCGTG	Fansiri et al., 2013	9	1000	8966-9965	50	
Reverse	D1-16R	CATGGATTGACCAGGTTGTG	Fansiri et al., 2013					
Forward	D1-18F	AGCTGATGTACTTCCACAGGAGA	Fansiri et al., 2013	10	878	9858-10735	60	
Reverse	d1a5B	AGAACCTGTTGATTCAACRGC	Christenbury et al., 2010					Used for RT step
Forward	qPCR-F	GGAAGGAGAAGGACTCCACA	Kong et al., 2006		105	9091-9195	60	Used for qPCR
Reverse	qPCR-R	ATCCTTGTATCCCATCCGGCT	Kong et al., 2006					

References

Kong YY, Thay CH, Tin TC, Devi S. *Rapid detection, serotyping and quantitation of dengue viruses by TaqMan real-time one-step RT-PCR.* 2006 ***J Virol Methods*** 138(1-2):123-30

Christenbury JG, Aw PP, Ong SH, Schreiber MJ, Chow A, Gubler DJ, Vasudevan SG, Ooi EE, Hibberd ML. *A method for full genome sequencing of all four serotypes of the dengue virus.* 2010 ***J Virol Methods*** 169(1):202-6. doi: 10.1016/j.jviromet.2010.06.013.

Fansiri T, Fontaine A, Diancourt L, Caro V, Thaisomboonsuk B, Richardson JH, Jarman RG, Ponlawat A, Lambrechts L. *Genetic mapping of specific interactions between Aedes aegypti mosquitoes and dengue viruses.* 2013 ***PLoS Genet.*** 9(8):e1003621. doi: 10.1371/journal.pgen.1003621.

File S2 – Primers used for virus deep sequencing.

Code S1 – R script used for bottleneck simulations.

```
#PREAMBULE####
library(ggplot2)
library(plyr)

#Create data

set.seed(1)
SNV_frequency=rexp(100, rate=100)
plot(SNV_frequency)
Viral_pop=round(rnorm(100, mean=2000, sd=200))
Bottleneck = round(rnorm(100, mean=28, sd=5))
sample = 1:100

#Run simulation

results2 <- data.frame()

for (i in sample){
  variants = length(SNV_frequency)
  result.variant <- data.frame()
  for(j in 1:variants){
    nA = rep("A",Viral_pop[i]*SNV_frequency[j])
    nB = rep("B",Viral_pop[i]*(1-SNV_frequency[j]))
    pop = as.factor(c(nA, nB))
    ini=data.frame(Variant=j, Sample=i,Frequency=SNV_frequency[j], Statut ="
      Before")

    sampling = sample(pop, size=Bottleneck[i], replace=FALSE)
    New.frequency = length(sampling[sampling=="A"]) / Bottleneck[i]
    result.b <- data.frame(Variant=j, Sample=i,Frequency=New.frequency, Statut
      ="After")
    result.variant <- rbind.fill(result.variant,result.b,ini)
  }
  results2 = rbind.fill(results2, result.variant)
}
results2$Statut <- factor(results2$Statut, levels = c("Before","After"))

#Parse results

```

```
results2.2 = subset(results2, Frequency > 0.01)

data.Sn <- ddply(results2.2, .(Statut, Sample), mutate, Sn.site = -(Frequency*
  log(Frequency)+(1-Frequency)*log(1-Frequency))/log(4))
data.Sn[is.na(data.Sn)] <- 0
data.Sn2 <- ddply(data.Sn, .(Statut, Sample), summarize, mean.Sn=sum(Sn.site)/
  11000)

data.pi <- ddply(results2.2, .(Statut, Sample), mutate, Pi.site = (1-((
  Frequency^2)+((Frequency-1)^2))))
data.pi[is.na(data.pi)] <- 0
data.pi2 <- ddply(data.pi, .(Statut, Sample), summarize, mean.pi=sum(Pi.site)/
  11000)

data.Sn.omni <- ddply(results2, .(Statut, Sample), mutate, Sn.site = -(
  Frequency*log(Frequency)+(1-Frequency)*log(1-Frequency))/log(4))
data.Sn.omni[is.na(data.Sn.omni)] <- 0
data.Sn2.omni <- ddply(data.Sn.omni, .(Statut, Sample), summarize, mean.Sn=sum(
  Sn.site)/11000)

data.pi.omni <- ddply(results2, .(Statut, Sample), mutate, Pi.site = (1-((
  Frequency^2)+((Frequency-1)^2))))
data.pi.omni <- na.omit(data.pi.omni)

data.pi2.omni <- ddply(data.pi.omni, .(Statut, Sample), summarize, mean.pi=sum(
  Pi.site)/11000, N=length(Pi.site))

#Display results

data.Sn2 <- data.Sn2[100:200,]
p1 <- ggplot(data=data.Sn2, aes(x=as.factor(Statut), y=mean.Sn))
p1 <- p1 + geom_boxplot(outlier.shape=NA, width=0.5, position=position_dodge
  (0.5)) + geom_point(position=position_jitter())
p1 <- p1 + theme_bw() + ylab(expression(atop("Averaged Sn per site over", paste
  ("all positions per sample")))) + xlab(label="bottleneck") + theme(text =
  element_text(size=15))
print(p1)

data.pi2 <- data.pi2[100:200,]

```

```
67 p2 <- ggplot(data=data.pi2, aes(x=as.factor(Statut), y=mean.pi))
68 p2 <- p2 + geom_boxplot(outlier.shape=NA, width=0.5, position=position_dodge
     (0.5)) + geom_point(position=position_jitter())
69 p2 <- p2 + theme_bw() + ylab(expression(atop(paste("Averaged ",pi," over all
     positions"), paste("per sample")))) + xlab(label="bottleneck") + theme(text =
      element_text(size=15))
70 print(p2)
71
72 data.Sn2.omni <- data.Sn2.omni[100:200,]
73
74 p3 <- ggplot(data=data.Sn2.omni, aes(x=as.factor(Statut), y=mean.Sn))
75 p3 <- p3 + geom_boxplot(outlier.shape=NA, width=0.5, position=position_dodge
     (0.5)) + geom_point(position=position_jitter())
76 p3 <- p3 + theme_bw() + ylab(expression(atop("Averaged Sn per site over", paste
     ("all positions per sample")))) + xlab(label="bottleneck") + theme(text =
     element_text(size=15))
77 print(p3)
78
79 data.pi2.omni <- data.pi2.omni[100:200,]
80
81 p4 <- ggplot(data=data.pi2.omni, aes(x=as.factor(Statut), y=mean.pi))
82 p4 <- p4 + geom_boxplot(outlier.shape=NA, width=0.5, position=position_dodge
     (0.5)) + geom_point(position=position_jitter())
83 p4 <- p4 + theme_bw() + ylab(expression(atop(paste("Averaged ",pi," over all
     positions"), paste("per sample")))) + xlab(label="bottleneck") + theme(text =
      element_text(size=15))
84 print(p4)
85
86 #Statistical tests
87 wilcox.test(data.pi2.omni[2:101,3], mu=data.pi2.omni[1,3])
88 wilcox.test(data.Sn2.omni[2:101,3], mu=data.Sn2.omni[1,3])
89 wilcox.test(data.pi2[2:101,3], mu=data.pi2[1,3])
90 wilcox.test(data.Sn2[2:101,3], mu=data.Sn2[1,3])
```

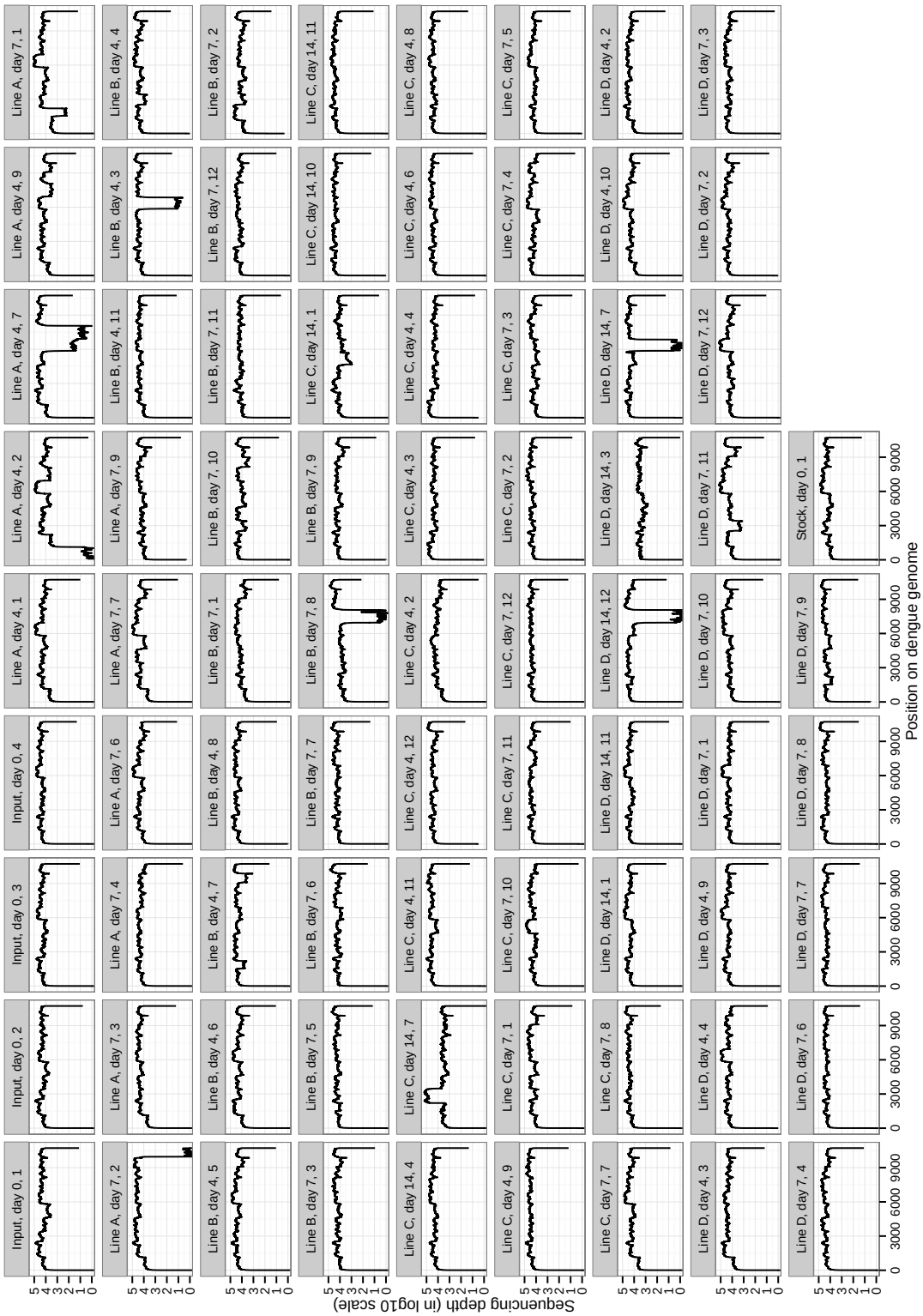

Figure S1 – Sequencing coverage and depth by sample.

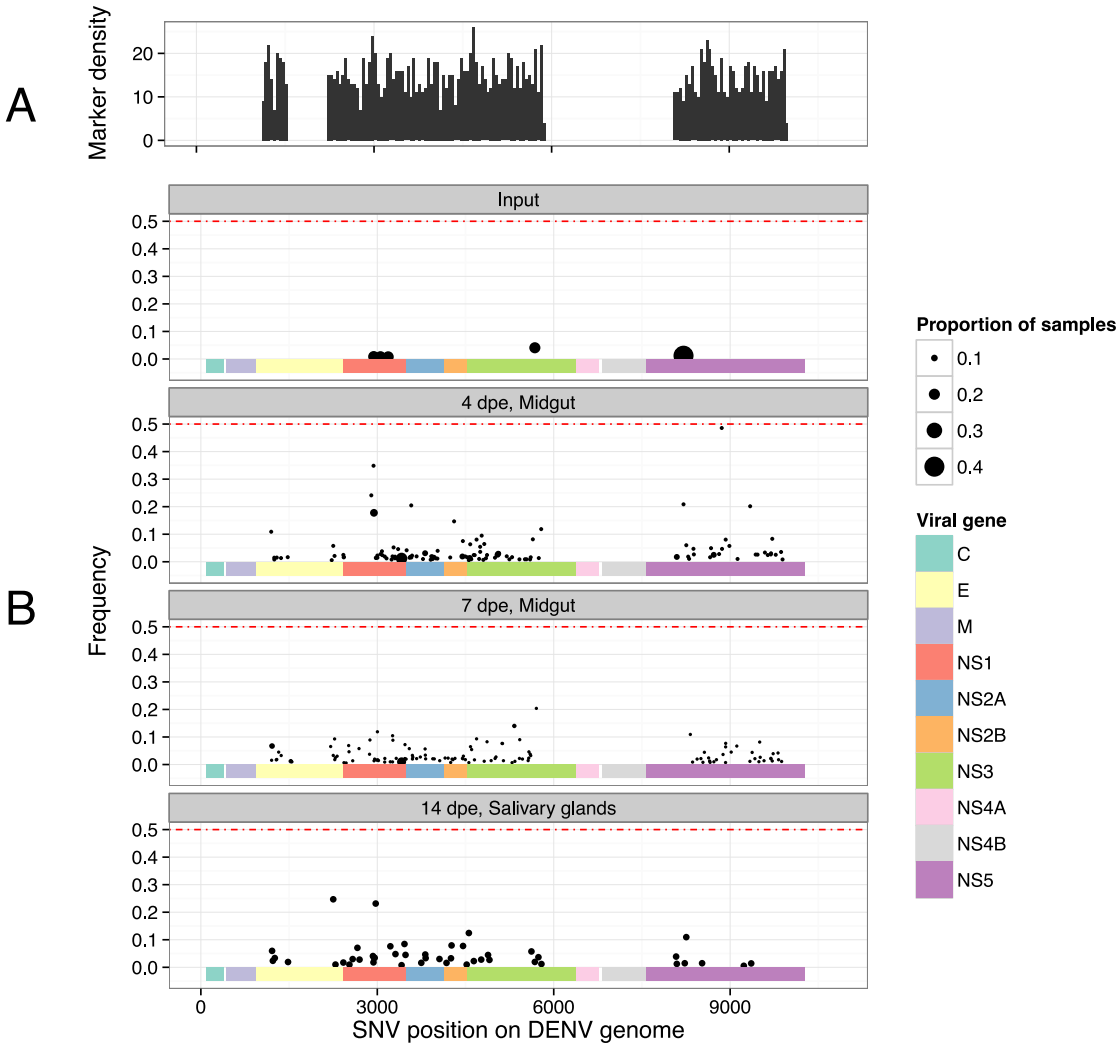

Figure S2 – Distribution of SNV positions and their mean detected frequencies in the conservative marker set. (A) Bars represent the density of markers retained in the conservative marker set for diversity and natural selection analyses along the DENV reference genome indicated on the x-axis. (B) Each dot represents the minor allele frequency of a single SNV along the DENV reference genome indicated on the x-axis, averaged over all samples from the same time point in which the SNV was detected. Dot size corresponds to the proportion of samples from the same time point in which the SNV was detected. The horizontal red dashed line represents a frequency of 0.5 over which a new variant becomes the consensus sequence. SNV distributions are stratified by time point. C=capsid protein, E=envelope glycoprotein, M=membrane glycoprotein, NS1=non-structural glycoprotein 1; NS2A=non-structural protein 2A; NS2B= non-structural protein 2B; NS3=non-structural protein 3 (protease/helicase); NS4A=non-structural protein 4A; NS4B=non-structural protein 4B; NS5=non-structural protein 5 (RNA-dependent-RNA polymerase).

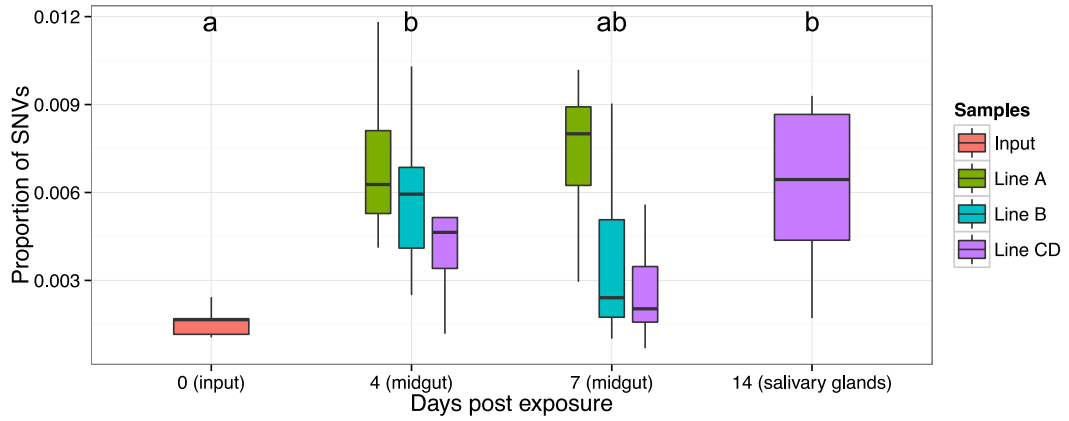

Figure S3 – Proportion of variable sites detected in the full marker set. Letters indicate statistically significant pairwise differences between time points.

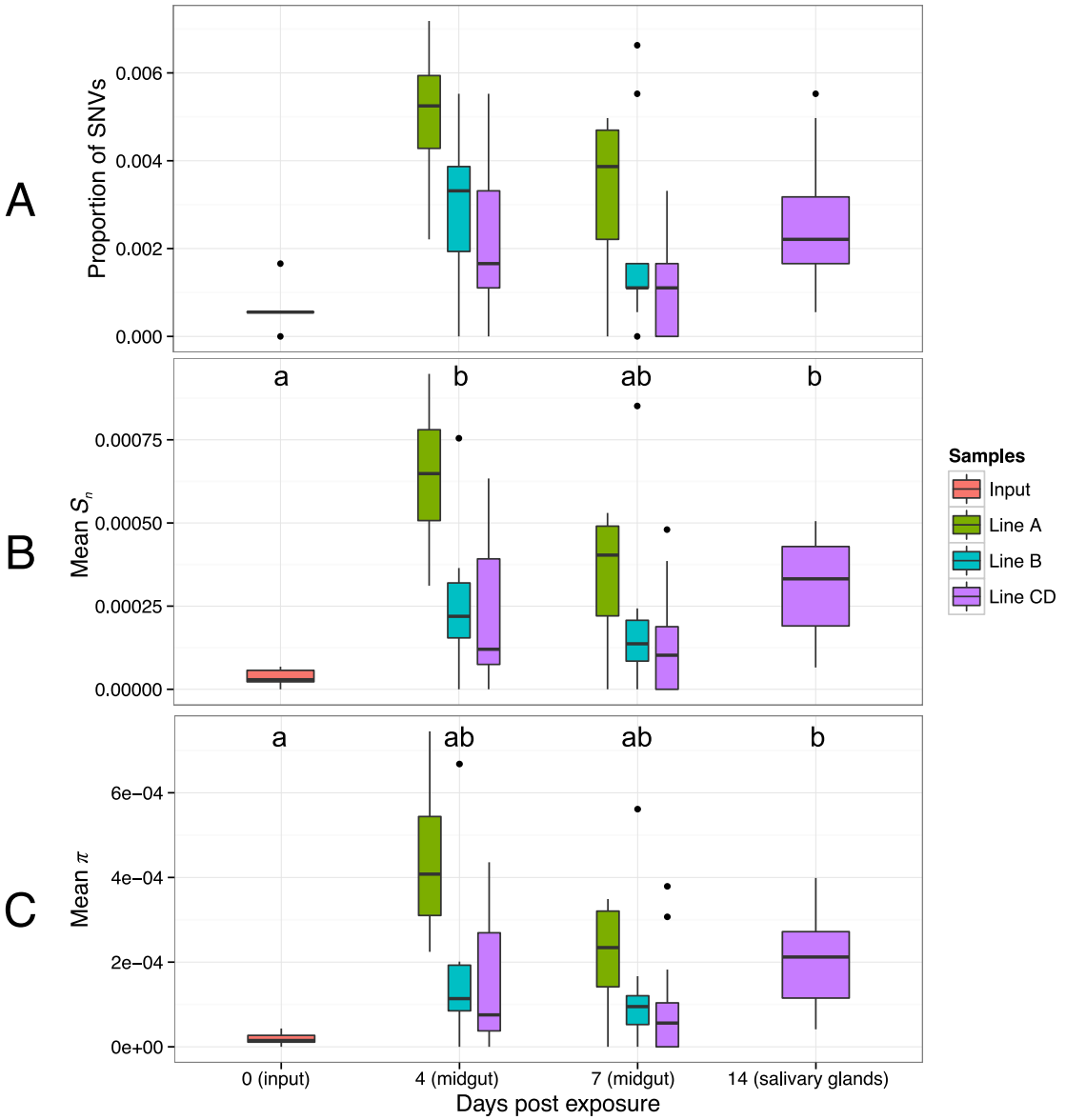

Figure S4 – Observed levels of DENV intra-host genetic diversity using the conservative marker set. (A) Proportion of variable sites detected. (B) Averaged Shannon entropy (S_n) per site over all positions per sample. (C) Averaged nucleotide diversity (π) over all positions per sample.

C Vertical transmission of arboviruses in mosquitoes: a historical perspective

Sebastian Lequime & Louis Lambrechts. Vertical transmission of arboviruses in mosquitoes: a historical perspective. *Infection, Genetics and Evolution*, 28:681-90, December 2014. doi:10.1016/j.meegid.2014.07.025

Abstract

Arthropod-borne viruses (arboviruses) are mainly transmitted horizontally among vertebrate hosts by blood-feeding invertebrate vectors, but can also be transmitted vertically in the vector from an infected female to its offspring. Vertical transmission (VT) is considered a possible mechanism for the persistence of arboviruses during periods unfavorable for horizontal transmission, but the extent and epidemiological significance of this phenomenon have remained controversial. To help resolve this question, we reviewed over a century of published literature on VT to analyze historical trends of scientific investigations on experimental and natural occurrence of VT in mosquitoes. Our synthesis highlights the influence of major events of public health significance in arbovirology on the number of VT publications. Epidemiological landmarks such as emergence events have significantly stimulated VT research. Our analysis also reveals the association between the evolution of virological assays and the probability of VT detection. Increased sensitivity and higher-throughput of modern laboratory assays resulted in enhanced VT detection. In general, VT contribution to arbovirus persistence is likely modest because vertically infected mosquitoes are rarely observed

in nature. Taken together, however, our results call for caution when interpreting VT studies because their conclusions are context- and method-dependent.

1 Introduction

According to the United States Center for Disease Control and Prevention (CDC) Arbovirus Catalog [Centers for Disease Control and Prevention, 2010], there are currently at least 530 identified arthropod-borne viruses (arboviruses), of which about a hundred cause human disease. Among them, four major viral genera account for the majority of arboviral diseases: *Flavivirus* (e.g., dengue, West Nile, Japanese encephalitis, and yellow fever viruses), *Alphavirus* (e.g., chikungunya, Eastern equine encephalomyelitis, Western equine encephalomyelitis and Venezuelan equine encephalitis viruses), *Orthobunyavirus* (e.g., California encephalitis and LaCrosse viruses) and *Phlebovirus* (e.g., Rift Valley fever and sandfly fever viruses).

During the past few decades, several arboviruses have emerged globally and are now considered among the most important public health concerns for the 21st century [Gubler, 2002]. Dengue, for example, has become the most prevalent arthropod-borne viral disease of humans over the last few decadesMessina:2014jo; it has recently been estimated that there are 390 million human dengue infections each year [Bhatt et al., 2013]. With only a few licensed vaccines and virtually no therapeutics available, antivectorial measures are often the only way to prevent arboviral diseases. Historically, however, the implementation of vector control measures has generally been difficult to sustain.

Arboviruses are naturally maintained in a transmission cycle between vertebrate and arthropod hosts [Gubler, 2001]. The majority of arthropod hosts, generally referred to as vectors, are blood-feeding mosquitoes. Rather than a simple alternation within a single host-vector pair, arbovirus transmission often occurs through highly complex transmission networks that include various hosts and vectors [Diaz et al., 2012]. Humans in particular, are not necessarily at the center of the transmission network and may only be incidental hosts (e.g., West Nile virus). Whereas some host or vector species are central to epidemic arbovirus transmission, others can be part of alternative transmission pathways, participating in the maintenance of the virus in nature during inter-epidemic periods. As an example, for some authors, fox squirrels may contribute to alternative transmission of West Nile virus in suburban

communities [Root et al., 2006, Root et al., 2007]. In many regions of the world, climatic conditions do not allow mosquito reproductive activity all year long. During the dry season in tropical areas or the cold season in temperate regions, the absence or low density of adult mosquitoes is unlikely to support continuous host-to-vector (horizontal) transmission [Leake, 1984]. Survival of mosquitoes during dry and cold seasons involves different physiological and/or behavioral mechanisms that may impact virus transmission differently. Besides, arboviral infections in vertebrate hosts typically produce a short-lived viremic period that eventually results in immunization of the host. A high level of herd immunity in the host community may thus prevent transmission above the minimum level required for sustained horizontal transmission. The existence of reservoir host species, alternative transmission mechanism or virus re-introduction, have been proposed to explain the maintenance of arboviruses during unfavorable periods or when herd immunity is high (for review, see [Reeves, 2004]).

One popular hypothesis to explain the persistence of arboviruses during unfavorable periods is the occurrence of vertical transmission (VT) in the arthropod vector. In this article, we define VT as the transmission of an arbovirus from an infected female mosquito to its offspring, regardless of the underlying mechanism. VT may occur through two main mechanisms. Transovarial transmission (TOT) occurs when the virus infects the germinal tissues of the female mosquito, whereas trans-egg VT takes place during oviposition in the fully formed egg [Rosen, 1988]. TOT typically achieves a higher efficiency of VT than trans-egg mechanisms especially when the germ cells are permanently infected so that most of the offspring are infected in the following generation [Tesh, 1984].

Under the VT scenario, the arbovirus present in the mosquito eggs, larvae or adults, including nulliparous females entering diapause, may survive throughout the unfavorable period without the need for a vertebrate host. Such a mixed-mode transmission (i.e., both horizontal and VT) is widespread among symbionts across taxa [Ebert, 2013]. Here symbiosis is defined as any type of persistent biological interaction, which includes mutualistic, commensalistic and antagonistic relationships. Although the infection cost is often modest to the vector, arboviruses are considered parasites of mosquitoes [Lambrechts and Scott, 2009]. Combining horizontal with VT enlarges considerably the range of ecological conditions in which a symbiont can per-

sist. In host species with diapause or discrete generations, VT may allow the symbiont to endure periods when horizontal transmission is not possible. Trade-offs between the two modes of transmission have been documented but are not universal [Ebert, 2013]. In addition to ecological factors that may favor horizontal over VT (e.g., climate), theory suggests that VT should be reduced in arboviruses with complex transmission networks because horizontal transmission among genetically disparate hosts hinders co-adaptation between vertically transmitted viruses and their hosts.

Both the very existence and the epidemiological significance of arbovirus VT have remained controversial since it was first suggested in the scientific literature over a century ago, at the onset of arbovirology. Carlos Finlay, who first introduced the idea of vectorial transmission of yellow fever virus by mosquitoes in 1881, extended his theory in 1899, suggesting that the yellow fever agent could be transmitted by an infected mosquito to its progeny [Finlay, 1899]. During their investigations in Cuba, the Yellow Fever U.S. Army Commission proved Carlos Finlay's original theory right in 1901 and experimentally tested, albeit unsuccessfully, the possibility of yellow fever virus VT in mosquitoes [Reed, 1901]. The same year, the Cuban physician Juan Guiteras also failed to succeed in demonstrating yellow fever virus VT [Guiteras, 1901].

Between 1901 and 1905, a group of French scientists from the Pasteur Institute carried out studies on yellow fever in Rio de Janeiro, Brazil. Among them, Paul-Louis Simond and Emile Marchoux finally demonstrated the VT hypothesis. They stated, however, that in their opinion the phenomenon was certainly infrequent [Marchoux and Simond, 1905, Marchoux and Simond, 1906]. Following this discovery, many have tried to replicate the experiment, but none have succeeded [Rosenau and Goldberger, 1906, Davis and Shannon, 1930, Hindle, 1930, Frobisher Jr et al., 1931].

The possibility of VT of other arboviruses was also investigated during the first half of the 20th century. Whereas early studies on dengue virus failed to provide evidence of VT [Siler et al., 1926, Simmons et al., 1931], a Japanese team demonstrated VT of Japanese encephalitis virus [Mitamura et al., 1939], but their work, published in German in a Japanese journal on the eve of World War II, went unnoticed by the scientific community. Studies on VT resumed in the 1950s and 1960s with a team from the Communicable Disease Center (ancestor of CDC) that was investigating viruses responsible for encephalitis (alphaviruses, such as Eastern equine encephalomyelitis, Western equine encephalomyelitis or Venezuelan equine encephalitis viruses;

flaviviruses, such as St. Louis encephalitis virus). Their conclusions, however, were inconsistent with both positive [Kissling et al., 1956, Chamberlain et al., 1956b, Kissling et al., 1957] and negative results [Chamberlain et al., 1956a, Chamberlain and Sudia, 1957, Chamberlain et al., 1959]. The outcome of these studies did not seem to depend on the virus under consideration.

In 1972, Robert B. Tesh and colleagues provided a clear demonstration of vesicular stomatitis virus VT in *Lutzomyia* sandflies [Tesh et al., 1972]. The next year, another team led by Douglas M. Watts, published evidence of VT for LaCrosse virus (*Orthobunyavirus*) in experimentally infected *Aedes triseriatus* mosquitoes [Watts et al., 1973]. Research on arbovirus VT by mosquito vectors has been vigorous ever since (Fig. 1), although a closer look reveals considerable heterogeneity associated with historical events and the evolution of laboratory assays used to detect VT experimentally or in a natural setting.

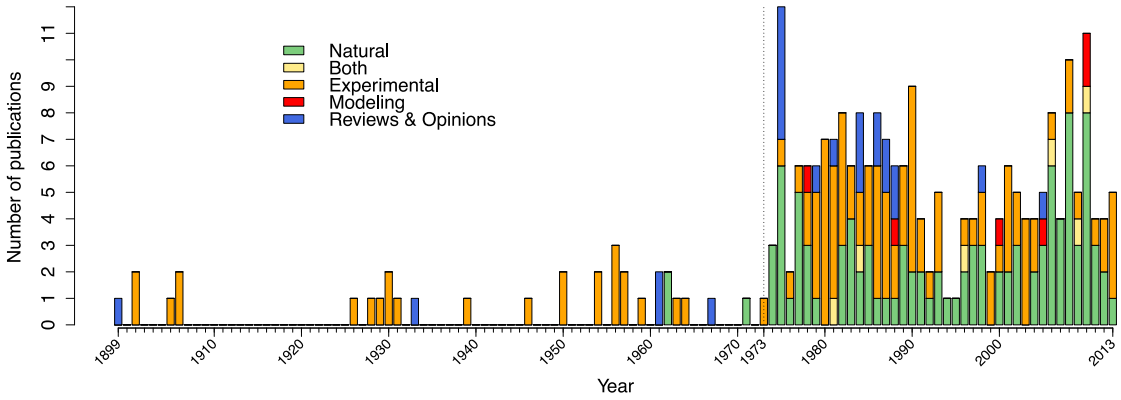

Figure 1 – Evolution of the yearly number of publications on arbovirus VT by mosquitoes. Bars show the yearly number of publications in each of five broad categories represented by different colors. The dashed vertical line indicates the publication of an influential Science article on LaCrosse virus VT by [Watts et al., 1973].

In the present study, we conducted a systematic review of over a century of published literature on arbovirus VT to analyze quantitatively historical trends of research on experimental and natural occurrence of VT in mosquitoes.

2 Material and methods

2.1 Literature search

Between 22 June 2011 and 25 September 2013, a systematic literature search was conducted in NCBI PubMed, ISI Web of Science, Armed Forces Pest Management Board Literature Retrieval System and Pasteur Institute Media Library. Citations in the identified articles were also examined individually in order to recover additional references. When the article was not found using the databases mentioned above, the corresponding authors and/or journal were contacted to obtain a copy of the publication. Older publications were found through Internet Archive (http://www.archive.org/).

Publications were searched regardless of their language, including English, French, German, Japanese and Chinese. Japanese and Chinese publications without a full abstract in English were excluded for practical reasons.

The review focused on arboviruses transmitted by mosquitoes and therefore articles dealing with VT in ticks or other arthropods were excluded. Likewise, publications about insect-specific viruses were excluded. Within mosquito-borne arboviruses, the review was restricted to VT in three main arboviral families: *Bunyaviridae*, *Flaviviridae* and *Togaviridae*.

2.2 Databases

Three databases were created in MySQL using the Sequel Pro© software. Contents of the three databases are described below.

Historical trends

To analyze historical trends in VT research, a first database (database # 1) was built that contains all publications related to VT according to the inclusion criteria indicated above. For each individual publication contained in this database, basic bibliographical data, virus taxonomy and study type were recorded. Study type consisted of five categories: 'Experimental' studies looked for evidence of VT based on laboratory experiments under controlled environments; 'Natural' studies attempted to demonstrate VT in nature, by collecting immature, male or overwintering female

mosquitoes; 'Both' are studies that conducted both experimental and natural investigations; 'Modeling' studies used mathematical models to evaluate the potential role of VT in arbovirus epidemiology; 'Reviews & Opinion' articles synthesized earlier work and/or commented the possibility of VT.

Evolution of virological assays

In order to examine how the evolution of laboratory assays in virology influenced the outcome of VT studies, two additional databases were assembled. Database # 2 consists of 'experimental' studies and database # 3 consists of 'natural' studies. Both were derived from database # 1, but were restricted to the 'experimental', 'natural', and 'both' types of studies. Additionally, inclusion criteria were more stringent for databases # 2 and # 3 than for database # 1. Databases # 2 and # 3 only included publications that specified mosquito and virus species tested, sample size and detection technique (the full list of studies included in these databases is provided in Supplementary Table S1). A total of 32 and 16 factors were included in databases # 2 and # 3, respectively, and each unique combination of factors was considered a different entry (the full list of factors recorded is provided in Supplementary Table **??**). Laboratory assays used for virus detection were divided into four broad categories. 'Animal' assays refer to the detection of pathological effects (including death) in laboratory animals, usually suckling mice, following inoculation with mosquito extracts; 'Cellular' assays refer to the detection of cytopathological effects in cell culture in vitro, following inoculation with mosquito extracts; 'Immunological' assays rely on the detection of viral antigens by antibodies (e.g., immunofluorescence assays) with or without previous amplification in cell culture or animal tissues; 'Molecular' assays refer to detection of viral nucleic acids, generally by reverse transcription (RT)-PCR. Whereas animal/cellular assays detect infectious virus and molecular assays detect viral nucleic acids (non-infectious), immunological assays include both detection of viral antigens in primary samples (non-infectious) and detection of viral antigens following amplification in vivo (e.g., intra-thoracic inoculation of Toxorhynchites mosquitoes) or in cell culture (infectious). Whether the use of infectious or non-infectious assays influenced the outcome of VT studies was also evaluated.

2.3 Statistical analyses

All statistical analyses were performed in the statistical environment R, version 3.0.2 (http://www.r-project.org/).

Historical trends

To normalize the overall scientific production about VT of dengue, West Nile and chikungunya viruses, the yearly number of publications between 1950 and 2013 was obtained through queries to the ISI Web of Science version 15.13.1 database, using the keywords 'dengue', 'West Nile virus' and 'chikungunya', respectively. For each virus, the normalized yearly number of publication was calculated as the ratio of the yearly number of publications about VT of this virus over the total yearly number of publications about the virus. The normalized yearly number of publications was analyzed using non-parametric Wilcoxon tests.

Evolution of virological assays

The evolution of laboratory assays used for VT detection in databases # 2 and # 3 was visualized by plotting the yearly proportion of each assay category (i.e., animal, cellular, immunological and molecular) over time. To weight each category according to the sample size, the yearly proportion of each assay category was calculated relative to the total number of individual mosquitoes tested for VT (i.e., the relative likelihood of each category over time). Smoothed curves were obtained by kernel density estimation using a Gaussian kernel and the bandwidth selection method proposed by [Sheather and Jones, 1991].

The association between assay category and VT detection probability was further explored with a subset of databases # 2 and # 3. The subset consisted of the *Aedes-Flavivirus* pair, which is the most frequent vector-virus association in the databases (49.7% and 31.8% of entries in databases # 2 and # 3, respectively). Analysis of the relationship between assay category and VT detection probability was restricted to the *Aedes-Flavivirus* pair to rule out the potential confounding effect of a differential VT probability among arbovirus families. Indeed, some arbovirus families may be vertically transmitted more efficiently than others [Turell, 1988]. If VT of arboviruses with higher VT efficiency were predominantly examined using one particular virological

assay, then a spurious association would be detected between this assay and higher VT detection probability. In experimental studies, only three assay categories were compared for the *Aedes-Flavivirus* pair because the molecular assay category was not represented in database # 2. VT occurrence was defined as the proportion of database entries that found evidence of VT. Evolution of VT detection over time for the *Aedes-Flavivirus* pair was represented as the relative likelihood of yearly VT occurrence. The relative likelihood function was estimated by kernel density estimation as described above.

To disentangle the respective effects of assay and sample size, VT occurrence was analyzed with a generalized linear model that included the $\log_{10}$-transformed sample size, the assay category and their interaction. As VT occurrence is a binary variable (0=VT undetected, 1=VT detected), the model was fitted with a binomial error distribution and a logit link function. Statistical significance of the effects was assessed by analysis of deviance [Hastie and Pregibon, 1991]. The same model was used to evaluate the influence of infectious compared to non-infectious assays.

To compare sample sizes between assay categories, distributions of $\log_{10}$-transformed sample sizes for each assay category were compared with non-parametric Kruskall-Wallis tests. When significant, pairwise Wilcoxon tests were subsequently performed with a Bonferroni correction of the p-values.

3 Results and discussion

3.1 Summary description of databases

The primary database (database # 1) includes 257 articles related to arbovirus VT published between 1899 and 2013 (the full list of publications is provided in Supplementary Table 1). Two additional databases were derived from database # 1 that were restricted to 'experimental' and 'natural' types of studies (database # 2 and # 3, respectively) and had more stringent inclusion criteria (see material and methods).

Database # 2 includes a total of 1,119 entries from 94 distinct publications that examined VT in an experimental setting. Overall, VT occurrence was 60.2% in database # 2 (i.e., evidence of VT was found in 60.2% of entries). Virological assays consisted of 58.5% immunological, 23.6% cellular, 16.8% animal and 1.1% molecular assays.

Appendix C. Vertical transmission of arboviruses in mosquitoes: a historical perspective

The most frequent vector-virus pairs in database # 2 were *Aedes-Flavivirus*, *Culex-Flavivirus*, *Aedes-Orthobunyavirus* and *Aedes-Alphavirus* that are represented by 49.7%, 22.6% 18.9% and 5.1% of entries, respectively. The remaining 3.7% consist of 11 less studied vector-virus pairs. The median VT prevalence, estimated as the number of positive pools over the total number of mosquitoes tested, was 0.15%, 0.0%, 13.5% and 0.0% in *Aedes-Flavivirus*, Culex-*Flavivirus*, *Aedes-Orthobunyavirus* and *Aedes-Alphavirus* pairs, respectively.

Database # 3 includes a total of 368 entries from 97 distinct publications that examined VT in a natural setting. Overall, VT occurrence was 37.5% in database # 3 (i.e., evidence of VT was found in 37.5% of the entries). Virological assays consisted of 29.6% animal, 28.5% immunological, 25.3% cellular and 16.3% molecular assays. *Aedes-Flavivirus*, *Culex-Flavivirus*, *Aedes-Orthobunyavirus* and *Aedes-Alphavirus* pairs are represented by 31.8%, 21.2% 16.8% and 2.7% of database # 3 entries, repectively. The remaining 27.5% consist of 19 less studied vector-virus pairs. The median VT prevalence was 0.03%, 0.0%, 0.0% and 0.06% in *Aedes-Flavivirus*, *Culex-Flavivirus*, *Aedes-Orthobunyavirus* and *Aedes-Alphavirus* pairs, respectively. The 10- to 1000-fold lower estimates of VT prevalence observed in natural studies likely reflect two different processes. First, the infection prevalence in the previous adult generation is generally close to 100% in experimental studies because mothers are experimentally exposed to the arbovirus before VT is assayed in the their progeny. The prevalence of arbovirus infection in wild mosquito populations, by contrast, is typically low and the prevalence of VT infection observed in a given generation actually measures the product of adult population prevalence and VT efficiency. For example, only 0.1% of field-caught *Ae. aegypti* were found to be infected by dengue virus in an endemic region of Thailand [Yoon et al., 2012]. Second, experimental VT studies possibly overestimate VT efficiencies because they tend to focus on naturally efficient combinations of vector-virus strains. VT efficiency may be further enhanced in experimental studies due to the unnatural mode of infection commonly used in the parental generation (i.e., intrathoracic inoculation). Moreover, VT efficiency can be quickly selected in the laboratory. For example, dengue virus VT efficiency increased more than 4-fold within two *Ae. aegypti* generations [Joshi et al., 2002]. 'Stabilized' infections have also been described for arboviruses of the California serogroup (*Bunyaviridae*: *Orthobunyavirus*), such as California encephalitis virus in *Ae. dorsalis* [Turell et al., 1982] and

San Angelo virus in *Ae. albopictus* [Tesh and Shroyer, 1980]. Whereas TOT occurred in 10-20% of the progeny of horizontally infected females, transovarially infected daughters in the following generations consistently transmitted virus to over 90% of their progeny. The proposed mechanism is that following a systemic infection by horizontal transmission, only a small fraction of the developing oocytes of a female become infected transovarially. However, if the germinal cells of a female become infected, then nearly 100% of her progeny are infected. It has been argued that most experimental studies of TOT have in fact underestimated VT efficiency because they considered 'non-stabilized' infections [Tesh, 1984]. An additional complicating factor is that TOT may not occur during the first gonotrophic cycles of orally infected females because the ovaries only become infected during the subsequent gonotrophic cycles. Only considering the first batch of eggs, usually free of virus, would also underestimate VT efficiency in this case [Tesh, 1984].

3.2 Historical trends

We first examined whether epidemiological landmarks, such as emergence events, large outbreaks or a significant change in geographic distribution of an arbovirus, influenced VT research. We considered three major epidemiological events in the recent history of arbovirology: the 1981 dengue hemorrhagic fever (DHF) epidemic in the Americas, the 1999 West Nile virus emergence in North America, and the 2005 chikungunya epidemic in the Indian Ocean region. Whereas dengue was already a growing problem in the tropics in the early 1980s [Messina et al., 2014], the emergence of West Nile virus in 1999 and that of chikungunya in 2005 were sudden and largely unexpected.

DHF in the Americas

Dengue viruses (*Flaviviridae*: *Flavivirus*) consist of four serotypes (DENV-1 to -4) that are currently estimated to collectively cause about 96 million symptomatic cases of dengue fever each year, including the severe form DHF [Bhatt et al., 2013]. Dengue viruses are primarily transmitted by the mosquito *Ae. aegypti*, which was largely eliminated from Central and South America during the 1950s and 1960s by the eradication program of the Rockefeller Foundation. Approximately ten years after the end of the eradication program, however, dengue epidemics started to occur in the

Americas, quickly followed by hyperendemicity and the emergence of DHF [Bennett et al., 2010]. The first cases of DHF were reported in Cuba during a DENV-2 epidemic in 1981 [Guzmán et al., 1999].

This epidemiological event did not appear to have an obvious impact on dengue research overall, as shown by a query with the keyword 'dengue' to the ISI Web of Science database (Fig. 2). By contrast, research on dengue virus VT markedly increased after 1981 (Fig. 2). The normalized yearly number of VT publications was significantly larger during the period 1982-2013 than during the period 1950-1981 (Wilcoxon test, $p = 2.0 \times 10^{-8}$). We cannot exclude the existence of a concomitant factor, however, because this effect was no longer significant when considering a period spanning 5 years before and 5 years after the event (Wilcoxon test, $p = 1.2 \times 10^{-1}$). The burst of VT research after the early 1980s may have been driven by studies investigating the possibility of dengue virus maintenance in areas where *Ae. aegypti* re-established, following 'rediscovery' of arbovirus VT in the previous decade (i.e., a delayed consequence of the [Watts et al., 1973] publication). Despite a lack of evidence in the early studies [Siler et al., 1926, Simmons et al., 1931], the VT hypothesis has been particularly popular in the case of dengue. Although a sylvatic transmission cycle exists whereby sylvatic dengue strains circulate between arboreal mosquitoes and non-human primates, there is no identified animal reservoir of dengue virus strains responsible for human epidemics [Vasilakis and Weaver, 2008]. This probably explains why VT has often been suggested to explain dengue virus maintenance during inter-epidemic periods.

West Nile virus emergence in North America

West Nile virus (*Flaviviridae*: *Flavivirus*) is primarily transmitted between birds and ornithophilic mosquitoes but can occasionally infect humans and horses, which are dead-ends for the virus [World Health Organization, 2011]. Following its introduction in North America, it is estimated that between 1999 and 2010, West Nile virus infected 1.8 million people, resulting in 1,308 deaths [Kilpatrick, 2011]. Research on this virus clearly increased after the North-American emergence (Fig. 3) but the normalized yearly number of studies on West Nile virus VT also increased significantly (Fig. 3; Wilcoxon test, $p = 6.1 \times 10^{-9}$). The effect was also significant when considering a period spanning 5 years before and 5 years after the event (Wilcoxon test, $p = 3.7 \times 10^{-3}$). More frequent studies on West Nile virus VT in Culex mosquitoes

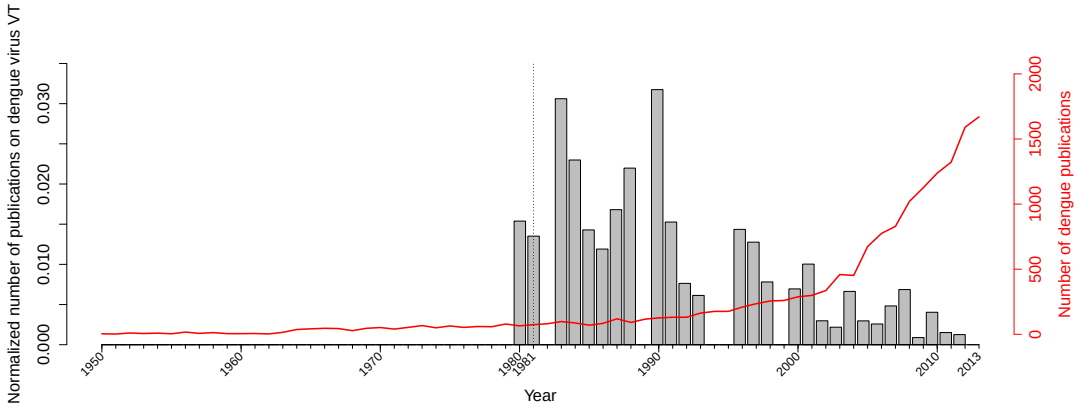

Figure 2 – Evolution of the yearly number of publications on dengue virus VT. Bars represent the normalized number of yearly publications on dengue virus VT. The red curve shows the overall number of dengue publications used for normalization. The vertical dashed line indicates the start of the 1981 DHF epidemic in the Americas.

after the 1999 epidemic have probably been stimulated by the search for a possible arthropod reservoir. Adult female mosquitoes are not active all year long in temperate regions of North America, yet overwintering persistence of West Nile virus transmission has been observed for over a decade [Reisen, 2013]. The role of diapausing female Culex mosquitoes as a West Nile virus reservoir is one of the hypotheses to explain West Nile virus overwintering in North America [Reisen, 2013]. Indeed, infected diapausing females have been reported in nature [Fechter-Leggett et al., 2012]. VT is a prerequisite for this mechanism to operate, because Culex females do not blood feed prior to diapause [Mitchell and Briegel, 1989, Nelms et al., 2013]. In warmer regions of North America, however, female *Culex* mosquitoes do not necessarily enter diapause and may instead become quiescent, sometimes after taking a blood meal [Eldridge, 1966, Nelms et al., 2013].

Chikungunya emergence in the Indian Ocean region

Chikungunya virus (*Togaviridae*: *Alphavirus*) is an arbovirus originating in Africa that primarily circulates between non-human primates and arboreal mosquitoes. In 2004, it emerged in the human population on the coast of Kenya and was subsequently responsible for a vigorous epidemic in the Indian Ocean region between 2005 and 2007. On La Réunion Island, 34.4% of the population was infected overall, while 1.4 million cases were reported in 2006 in India [Pialoux et al., 2007]. The global emer-

Appendix C. Vertical transmission of arboviruses in mosquitoes: a historical perspective

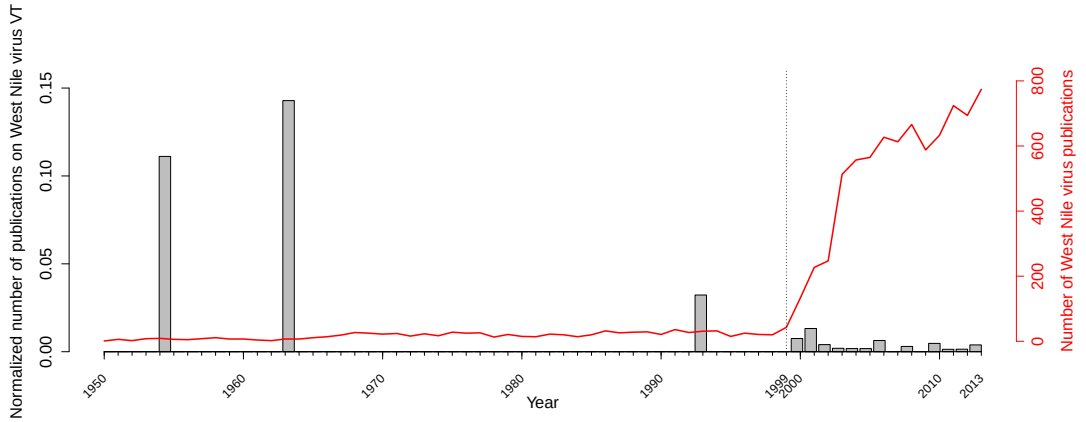

Figure 3 – Evolution of the yearly number of publications on West Nile virus VT. Bars represent the normalized number of yearly publications on West Nile virus VT. The red curve shows the overall number of West Nile virus publications used for normalization. The vertical dashed line indicates the start of the 1999 West Nile virus emergence in North America.

gence of chikungunya that followed the 2005 outbreak triggered a renewed scientific interest for chikungunya virus, as shown by a query with the keyword 'chikungunya' to the ISI Web of Science database (Fig. 4). Although the normalized yearly number of publications on chikungunya virus VT after 2005 remains relatively low (due to the dramatic increase in the total number of chikungunya publications used for normalization), the frequency of VT publications significantly increased (Fig. 4; Wilcoxon test, $p = 2.9 \times 10^{-3}$). The effect was also significant when considering a period spanning 5 years before and 5 years after the event (Wilcoxon test, $p = 3.6 \times 10^{-2}$). Increase of VT research after the 2005 epidemic was, however, less dramatic than in the cases of the 1981 DHF epidemic in the Americas and the 1999 emergence of West Nile virus in North America. This may be a consequence of earlier studies, which concluded that Alphavirus VT rarely occurs and is unlikely to be epidemiologically significant when it does [Turell, 1988]. Thus, the modest increase of research on chikungunya virus VT after 2005 may reflect a general lack of interest and/or a significant proportion of negative results, a well known bias in systematic reviews [Littell et al., 2008]. The risk of chikungunya virus emergence in Europe, as illustrated by the Italian outbreak in 2007 [Rezza et al., 2007] and the current emergence in the Americas [Leparc-Goffart et al., 2014] may stimulate additional research on chikungunya virus VT.

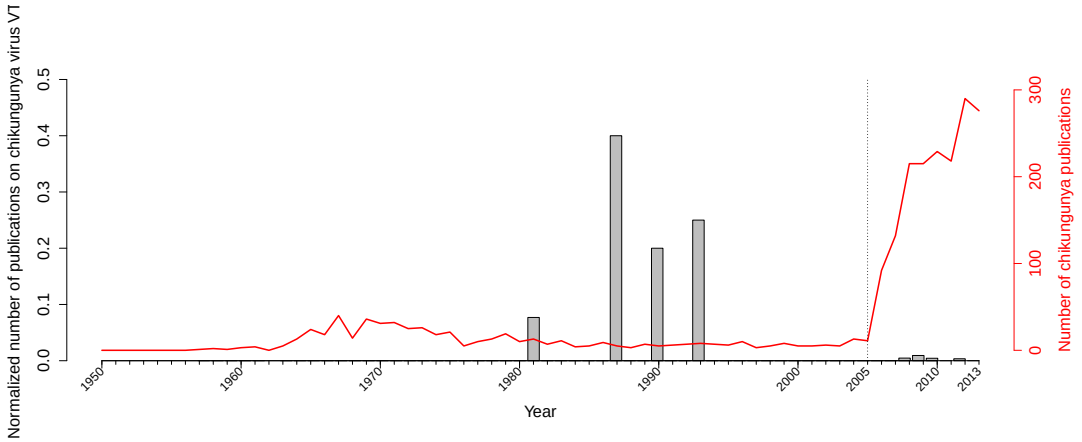

Figure 4 – Evolution of the yearly number of publications on chikungunya virus VT. Bars represent the normalized number of yearly publications on chikungunya virus VT. The red curve shows the overall number of chikungunya publications used for normalization. The vertical dashed line indicates the start of the 2005 chikungunya epidemic in the Indian Ocean region.

3.3 Evolution of virological assays

For some authors [Bocquet and Molero, 1996], the expansion of arbovirus VT research in the 1970s may be related to the development of new investigation methods, notably with the introduction of immunological reactions in laboratory assays. Overall, virological assays used for VT detection changed dramatically over the last century (Fig. 5). In the 1970s, traditional animal models were replaced by immunological methods that became progressively more common through the 1990s. Immunological assays were then replaced by cellular and molecular assays in the 2000s. Although the use of animal, cellular and immunological assays followed a relatively similar evolution in both natural and experimental studies of arbovirus VT, molecular assays have seldom been used in experimental studies to date (Fig. 5-A).

The advent of novel virological methods such as PCR-based molecular assays may have facilitated VT detection due to improved sensitivity and higher-throughput capacity. It is worth noting that molecular assays only detect viral RNA and therefore do not provide evidence of infectious virus unless they are confirmed by virus isolation. This is also true for immunological assays that detect viral antigens without direct evidence for infectivity. VT being a relatively rare event [Bocquet and Molero, 1996], one might expect that the likelihood of VT detection would have increased with the

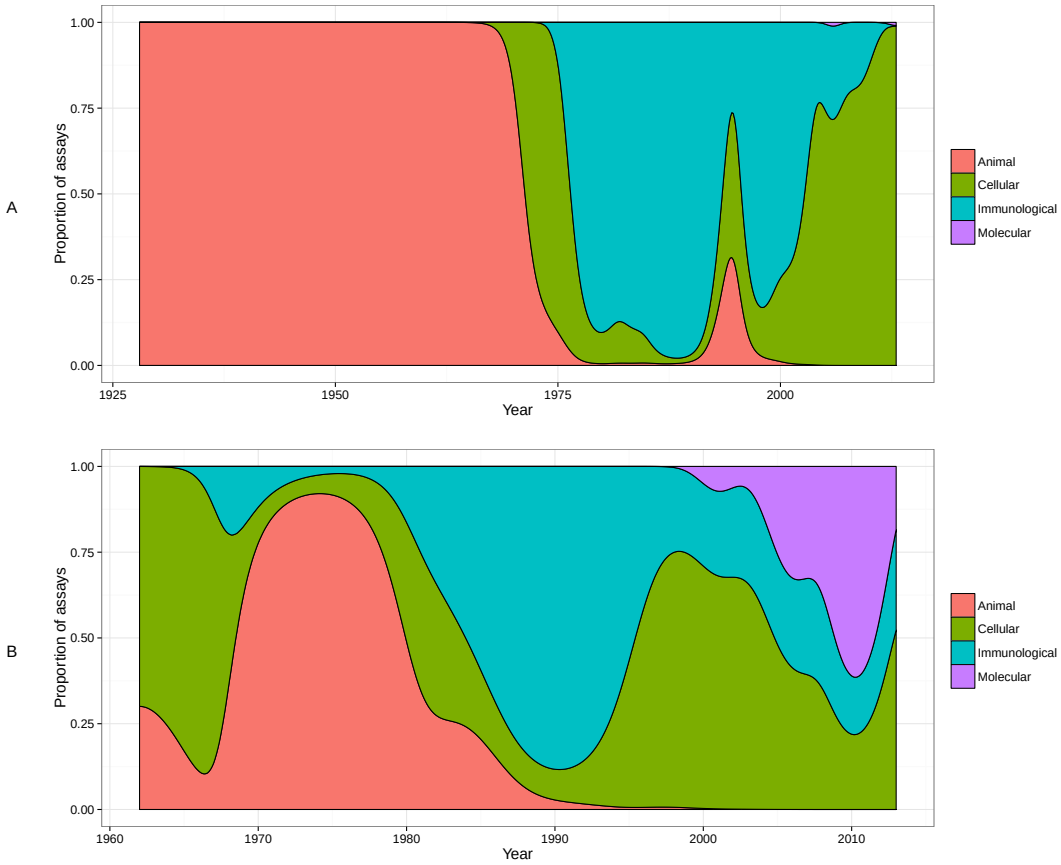

Figure 5 – Evolution of virological assays used for arbovirus VT detection in (A) experimental and (B) natural studies. In each panel, colored areas represent the relative likelihood of each assay category at a given time-point on the x-axis. The yearly proportion of each assay is calculated relative to the total number of individual mosquitoes tested for VT.

evolution of assay performance. We tested this hypothesis in the subset of studies that examined *Flavivirus* VT in *Aedes* mosquitoes. The *Aedes-Flavivirus* association is the most widely represented in the published literature and has been investigated with a variety of laboratory assays. The probability of VT detection in *Aedes-Flavivirus* studies significantly increased over time, both in experimental (Fig. 6-A) and natural studies (Fig. 6-B). Although a causal relationship cannot be conclusively inferred, these increasing trends coincide with the advent of cellular and immunological assays in experimental studies (Fig. 5-A) and of cellular and molecular assays in natural studies (Fig. 5-B). Thus, our analyses support the hypothesis that the probability of VT detection has been influenced by the evolution of virological assays.

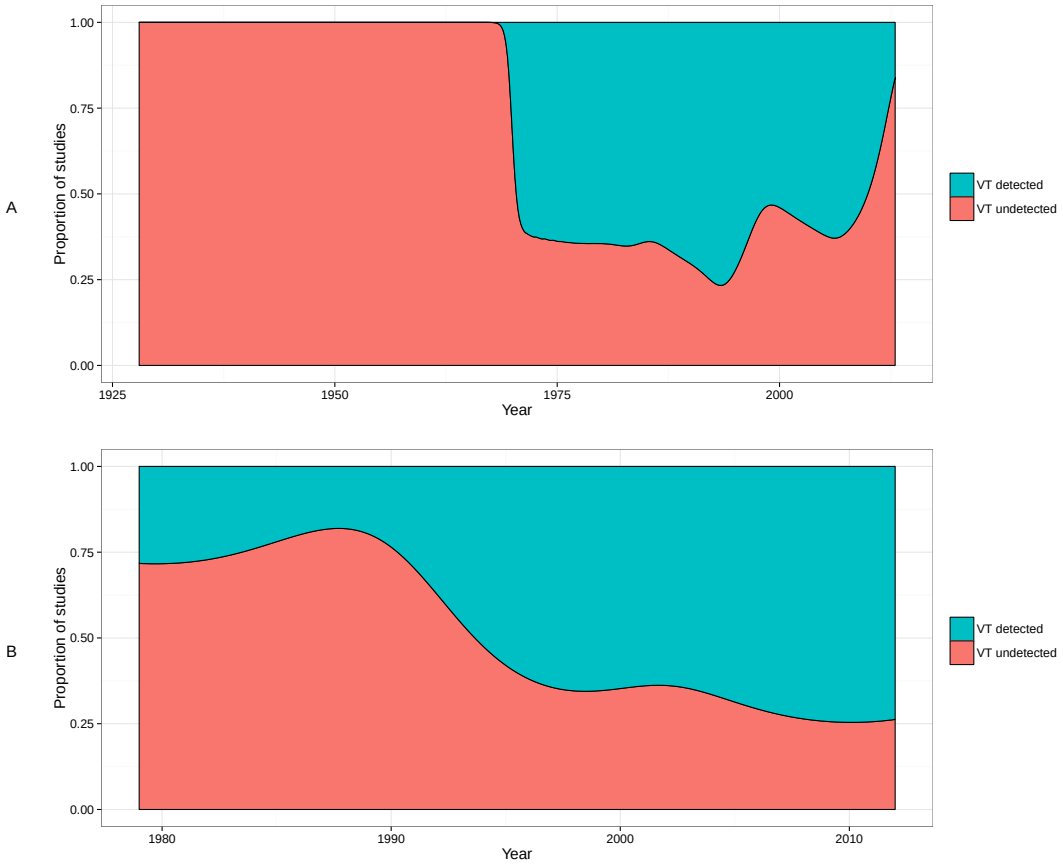

Figure 6 – Evolution of VT detection in (A) experimental and (B) natural studies of the *Aedes-Flavivirus* pair. In each panel, the curve represents the relative likelihood of VT occurrence at a given time-point on the x-axis. The yearly detection rate is calculated relative to the total number of database entries.

The probability of VT detection could have been influenced by the sensitivity of the assay and/or the ability to process a larger number of samples by modern assays. More recent detection methods such as cellular or immunological assays used in experimental studies are indeed associated with larger sample sizes (Fig. 7-A; Kruskal-Wallis test, $p = 5.2 \times 10^{-5}$; pairwise Wilcoxon tests: $p = 6.4 \times 10^{-3}$ and $p = 3.7 \times 10^{-5}$ for animal vs. cellular and animal vs. immunological assays, respectively, $p = 6.3 \times 10^{-1}$ for cellular vs. immunological assays). Similarly, in natural studies, significantly larger sample sizes are found in studies based on molecular assays (Fig. 7-B; Kruskal-Wallis test, $p = 2.9 \times 10^{-2}$; pairwise Wilcoxon tests: $p = 2.6 \times 10^{-2}$ for animal vs. molecular assays, $p = 9.5 \times 10^{-1}$ for animal vs. cellular assays, $p = 2.3 \times 10^{-1}$ for animal vs. immunological assays, $p = 1.0$ for immunological vs. cellular assays, $p = 2.5 \times 10^{-1}$

for molecular vs. cellular assays, and $p = 9.6 \times 10^{-1}$ for molecular vs. immunological assays). To determine the respective contributions of sample size and detection assay to changes in VT detection probability, VT occurrence in the databases was analyzed with a model that incorporated both factors and their interaction. In experimental studies, we found that VT occurrence was statistically significantly influenced by both the $\log_{10}$-transformed sample size ($p = 2.5 \times 10^{-7}$) and the detection assay ($p = 3.7 \times 10^{-2}$), but not by their interaction ($p = 4.7x10^{-1}$). Therefore, the increase in VT detection probability did not simply result from a higher-throughput capacity of assays but also from improved assay sensitivity. In natural studies, VT occurrence was significantly influenced by the $\log_{10}$-transformed sample size ($p = 3.9 \times 10^{-4}$), but not by the detection assay ($p = 2.2 \times 10^{-1}$) or their interaction ($p = 7.2 \times 10^{-2}$). In this case, the observed increase in VT detection probability primarily resulted from a significant increase in the number of samples tested.

We also evaluated whether the use of infectious or non-infectious assays influenced the probability of VT detection. Both animal and cellular assays detect infectious virus, whereas molecular assays detect viral nucleic acids (non-infectious). Immunological assays include both detection of viral antigens in primary samples (non-infectious) and detection of viral antigens following amplification in vivo or in cell culture (infectious). We did not find a statistically significant effect of whether the assay was infectious or non-infectious, both in experimental studies ($p = 1.8 \times 10^{-1}$) and in natural studies ($p = 1.8 \times 10^{-1}$), neither did we find an interaction with the $\log_{10}$-transformed sample size ($p = 1.5 \times 10^{-1}$ and $p = 8.6 \times 10^{-2}$, respectively).

4 Conclusions

Arbovirus VT in mosquitoes has rarely taken center stage in arbovirology, but has remained a debated topic since it was initially discovered more than a century ago. In the present study, fortunately, controversy about the extent and epidemiological importance of VT for arbovirus epidemiology has likely alleviated publication bias, a well-known pitfall of meta-analytical studies due to the under reporting of negative results. Despite their intrinsic value, negative results often remain unpublished because they are abandoned by authors and/or rejected by editors. Controversial topics like arbovirus VT do not entirely escape publication bias, but are thought to minimize it because negative results are more frequently published (e.g., [Vazeille

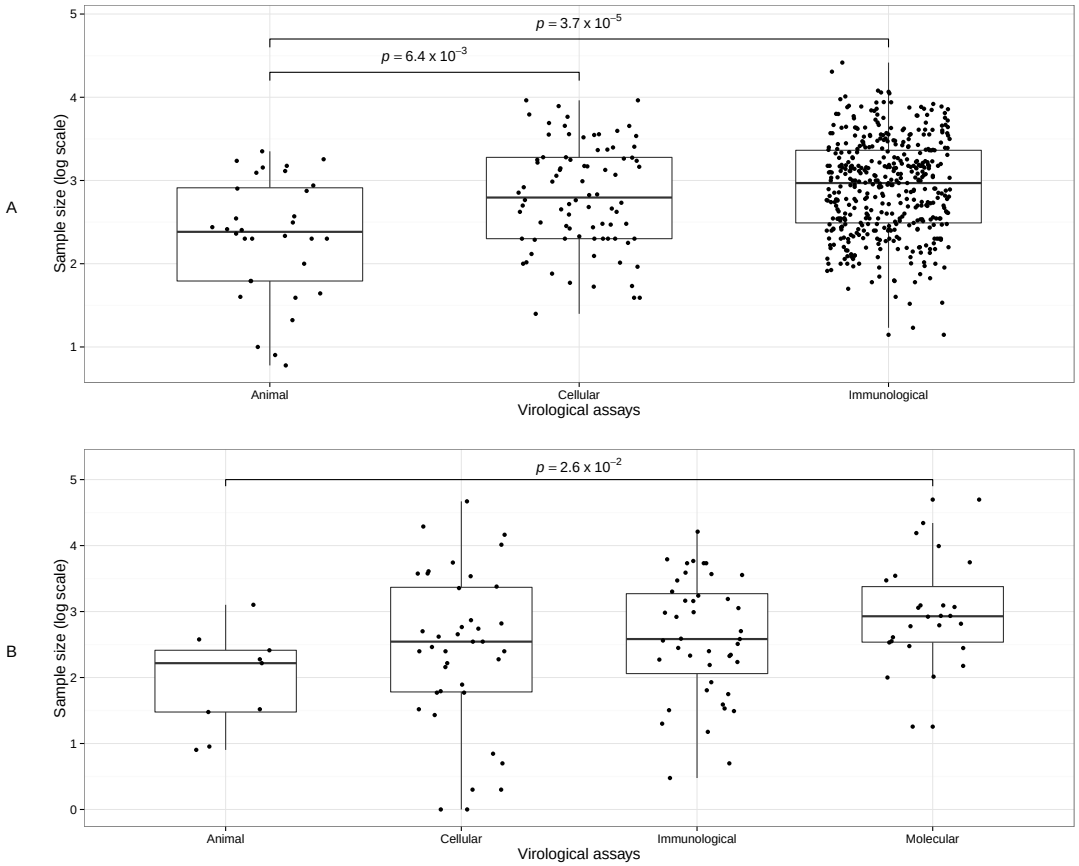

Figure 7 – Distribution of sample sizes by assay category in (A) experimental and (B) natural studies of arbovirus VT. Sample size was log-transformed prior to analysis and graphical representation. Boxplots represent the 75th percentile, median and 25th percentile; whiskers extend to the highest and lowest value in the 1.5 × interquartile range.

et al., 2009, Watts et al., 1985]. By nature, however, publication bias is difficult to quantify. Therefore caution must be used to interpret the results, keeping in mind that under reporting of negative results leads to overestimating VT.

Although the existence of VT is unequivocal for several mosquito-arbovirus pairs of public health significance, our databases showed that VT is collectively rare. Studies that estimated VT prevalence in a natural situation typically found that <0.1% of mosquitoes are vertically infected, regardless of the virus family or mosquito species. VT efficiency per se, however, cannot be inferred in field studies because infection prevalence in the adult population is unknown. VT prevalence in experimental studies is 10 to 1000-fold higher, which reflects more closely VT efficiency because the

parental generation is usually 100% infected. Even when it is relatively efficient, VT is unlikely to explain the persistence of arboviruses in the prolonged absence of horizontal transmission. A recent mathematical model on dengue concluded that VT efficiency would have to be 5- to 30-fold higher than the experimentally measured VT efficiency to support long-term persistence of the virus [Adams and Boots, 2010]. It has been argued that TOT alone could maintain the virus indefinitely in a mosquito subpopulation whose germinal tissues are permanently infected in 'stabilized' infections [Turell et al., 1982] even if the prevalence in the overall population is low [Tesh, 1984]. Unless VT efficiency is 100%, however, occasional horizontal amplification is predicted to be necessary for the long-term maintenance of arboviruses [Fine, 1975].

Our historical perspective highlighted two important aspects of VT research in arbovirology. First, scientific production on arbovirus VT in mosquitoes has been extremely uneven over time. In particular, epidemiological landmarks such as emergence events have significantly stimulated VT research. Second, the evolution of virological assays used in VT studies has profoundly influenced the ability to detect it. Increased sensitivity and higher-throughput of more recent laboratory assays are associated with higher estimates of VT rates. Taken together, our results call for caution when interpreting VT studies because their conclusions are context- and method-dependent.

Acknowledgements

The authors thank members of the Lambrechts lab for insightful discussions, Richard Paul for critical reading of the manuscript, and two anonymous reviewers for useful comments. S.L. was supported by a doctoral fellowship from University Pierre and Marie Curie. L.L. received funding from the French Government's Investissement d'Avenir program, Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases (grant ANR-10-LABX-62-IBEID).

References

[Adams and Boots, 2010] Adams, B. and Boots, M. (2010). How important is vertical transmission in mosquitoes for the persistence of dengue? Insights from a mathematical model. *Epidemics*, 2(1):1–10.

[Bennett et al., 2010] Bennett, S. N., Drummond,

A. J., Kapan, D. D., Suchard, M. A., Muñoz-Jordán, J. L., Pybus, O. G., Holmes, E. C., and Gubler, D. J. (2010). Epidemic dynamics revealed in dengue evolution. *Molecular biology and evolution*, 27(4):811–818.

[Bhatt et al., 2013] Bhatt, S., Gething, P. W., Brady, O. J., Messina, J. P., Farlow, A. W., Moyes, C. L., Drake, J. M., Brownstein, J. S., Hoen, A. G., Sankoh, O., Myers, M. F., George, D. B., Jaenisch, T., Wint, G. R. W., Simmons, C. P., Scott, T. W., Farrar, J. J., and Hay, S. I. (2013). The global distribution and burden of dengue. *Nature*, 496(7446):504–507.

[Bocquet and Molero, 1996] Bocquet, J. P. and Molero, M. (1996). Une hypothèse originale: la transmission transovarienne des arbovirus de Finlay à nos jours. In Waast, R. and Moulin, A. M., editors, *Les sciences hors d'Occident au 20ème siècle*, pages 83–88. Paris.

[Centers for Disease Control and Prevention, 2010] Centers for Disease Control and Prevention (2010). Arbovirus Catalog.

[Chamberlain et al., 1956a] Chamberlain, R. W., Kissling, R. E., Stamm, D. D., Nelson, D. B., and Sikes, R. K. (1956a). Venezuelan equine encephalomyelitis in wild birds. *American journal of hygiene*, 63(3):261–273.

[Chamberlain et al., 1956b] Chamberlain, R. W., Sikes, R. K., and Nelson, D. B. (1956b). Infection of Mansonia perturbans and Psorophora ferox mosquitoes with Venezuelan equine encephalomyelitis virus. *Proceedings of the Society for Experimental Biology and Medicine. Society for Experimental Biology and Medicine (New York, N.Y.)*, 91(2):215–216.

[Chamberlain and Sudia, 1957] Chamberlain, R. W. and Sudia, W. D. (1957). The North American arthropod-borne encephalitis viruses in Culex tarsalis Coquillett. *American journal of hygiene*, 66(2):151–159.

[Chamberlain et al., 1959] Chamberlain, R. W., Sudia, W. D., and Gillett, J. D. (1959). St. Louis encephalitis virus in mosquitoes. *American journal of hygiene*, 70:221–236.

[Davis and Shannon, 1930] Davis, N. C. and Shannon, R. C. (1930). The location of yellow fever virus in infected mosquitoes and the possibility of hereditary transmission. *American Journal of Epidemiology*, 11(2):335–344.

[Diaz et al., 2012] Diaz, L. A., Flores, F. S., Quaglia, A., and Contigiani, M. S. (2012). Intertwined arbovirus transmission activity: reassessing the transmission cycle paradigm. *Frontiers in physiology*, 3:493.

[Ebert, 2013] Ebert, D. (2013). The Epidemiology and Evolution of Symbionts with Mixed-Mode Transmission. *Annual Review of Ecology, Evolution, and Systematics*, 44(1):623–643.

[Eldridge, 1966] Eldridge, B. F. (1966). Environmental Control of Ovarian Development in Mosquitoes of the Culex pipiens Complex. *Science*, 151(3712):826–828.

[Fechter-Leggett et al., 2012] Fechter-Leggett, E., Nelms, B. M., Barker, C. M., and Reisen, W. K. (2012). West Nile virus cluster analysis and vertical transmission in Culex pipiens complex mosquitoes in Sacramento and Yolo Counties, California, 2011. *Journal of vector ecology : journal of the Society for Vector Ecology*, 37(2):442–449.

[Fine, 1975] Fine, P. E. (1975). Vectors and vertical transmission: an epidemiologic perspective. *Annals of the New York Academy of Sciences*, 266:173–194.

Appendix C. Vertical transmission of arboviruses in mosquitoes: a historical perspective

[Finlay, 1899] Finlay, C. J. (1899). Mosquitoes considered as transmitters of yellow fever and malaria. *Medical Record*, 55(21):737–739.

[Frobisher Jr et al., 1931] Frobisher Jr, M., Davis, N. C., and Shannon, R. C. (1931). On the failure of yellow fever virus to persist in a colony of Aedes aegypti. *American Journal of Epidemiology*, 14(1):142–146.

[Gubler, 2001] Gubler, D. J. (2001). Human arbovirus infections worldwide. *Annals of the New York Academy of Sciences*, 951:13–24.

[Gubler, 2002] Gubler, D. J. (2002). The global emergence/resurgence of arboviral diseases as public health problems. *Archives of medical research*, 33(4):330–342.

[Guiteras, 1901] Guiteras, J. (1901). Experimental yellow fever at the inoculation station of the Samtary Department of Havana with a view to producing immunization. *American Medicine*, 2:809–817.

[Guzmán et al., 1999] Guzmán, M. G., Kouri, G., Vazquez, S., and Rosario, D. (1999). DHF epidemics in Cuba, 1981 and 1997: some interesting observations. *Dengue Bulletin*, 23:39–43.

[Hastie and Pregibon, 1991] Hastie, J. T. and Pregibon, D. (1991). Generalized linear models. In *Statistical models in S*. Wadsworth & Brooks/-Cole.

[Hindle, 1930] Hindle, E. (1930). The transmission of yellow fever. *The Lancet*, 216(5590):835–842.

[Joshi et al., 2002] Joshi, V., Mourya, D. T., and Sharma, R. C. (2002). Persistence of dengue-3 virus through transovarial transmission passage in successive generations of Aedes aegypti mosquitoes. *American Journal of Tropical Medicine and Hygiene*, 67(2):158–161.

[Kilpatrick, 2011] Kilpatrick, A. M. (2011). Globalization, land use, and the invasion of West Nile virus. *Science*, 334(6054):323–327.

[Kissling et al., 1956] Kissling, R. E., Chamberlain, R. W., Nelson, D. B., and Stamm, D. D. (1956). Venezuelan equine encephalomyelitis in horses. *American journal of hygiene*, 63(3):274–287.

[Kissling et al., 1957] Kissling, R. E., Chamberlain, R. W., Sudia, W. D., and Stamm, D. D. (1957). Western equine encephalitis in wild birds. *American journal of hygiene*, 66(1):48–55.

[Lambrechts and Scott, 2009] Lambrechts, L. and Scott, T. W. (2009). Mode of transmission and the evolution of arbovirus virulence in mosquito vectors. *Proceedings of the Royal Society B-Biological Sciences*, 276(1660):1369–1378.

[Leake, 1984] Leake, C. J. (1984). Transovarial Transmission of Arboviruses by Mosquitoes. In *Vectors in virus biology*, pages 63–91. Academic Press, New-York & London.

[Leparc-Goffart et al., 2014] Leparc-Goffart, I., Nougairede, A., Cassadou, S., Prat, C., and De Lamballerie, X. (2014). Chikungunya in the Americas. *Lancet (London, England)*, 383(9916):514.

[Littell et al., 2008] Littell, J. H., Corcoran, J., and Pillai, V. (2008). *Systematic Reviews and Meta-Analysis*. Oxford Univ Press, New-York.

[Marchoux and Simond, 1905] Marchoux, É. and Simond, P.-L. (1905). La transmission héréditaire du virus de la fièvre jaune chez le Stegomyia fasciata. *Comptes Rendus des Séances et Mémoires de la Société de Biologie*, pages 259–260.

[Marchoux and Simond, 1906] Marchoux, É. and Simond, P.-L. (1906). Etudes sur la fièvre jaune -

Deuxième mémoire. *Annales de l'Institut Pasteur*, 20(1):1–25.

[Messina et al., 2014] Messina, J. P., Brady, O. J., Scott, T. W., Zou, C., Pigott, D. M., Duda, K. A., Bhatt, S., Katzelnick, L., Howes, R. E., Battle, K. E., Simmons, C. P., and Hay, S. I. (2014). Global spread of dengue virus types: mapping the 70 year history. *Trends in microbiology*, 22(3):138–146.

[Mitamura et al., 1939] Mitamura, T., Kitaoka, M., Watanabe, S., Hosoi, T., Tenjin, S., Seki, O., Nagahata, K., Jo, K., and Shimizu, M. (1939). Weitere Untersuchungen über die Übertragung der japanischen epidemischen enzephalitis durch Mücken. *Transaction of the Japanese Society of Pathology*, 29:92–105.

[Mitchell and Briegel, 1989] Mitchell, C. J. and Briegel, H. (1989). Fate of the blood meal in force-fed, diapausing Culex pipiens (Diptera: Culicidae). *Journal of medical entomology*, 26(4):332–341.

[Nelms et al., 2013] Nelms, B. M., Macedo, P. A., Kothera, L., Savage, H. M., and Reisen, W. K. (2013). Overwintering biology of Culex (Diptera: Culicidae) mosquitoes in the Sacramento Valley of California. *Journal of medical entomology*, 50(4):773–790.

[Pialoux et al., 2007] Pialoux, G., Gaüzère, B.-A., Jauréguiberry, S., and Strobel, M. (2007). Chikungunya, an epidemic arbovirosis. *The Lancet. Infectious diseases*, 7(5):319–327.

[Reed, 1901] Reed, W. (1901). The propagation of yellow fever; observations based on recent researches. *Medical Record*, 60:201–209.

[Reeves, 2004] Reeves, W. C. (2004). Overwintering of arboviruses. In *Epidemiology and Control of Mosquito-borne Arboviruses in California*, pages 357–382. Calif Mosq Vector Control Assoc, Sacramento, California.

[Reisen, 2013] Reisen, W. K. (2013). Ecology of West Nile virus in North America. *Viruses*, 5(9):2079–2105.

[Rezza et al., 2007] Rezza, G., Nicoletti, L., Angelini, R., Romi, R., Finarelli, A. C., Panning, M., Cordioli, P., Fortuna, C., Boros, S., Magurano, F., Silvi, G., Angelini, P., Dottori, M., Ciufolini, M. G., Majori, G. C., Cassone, A., and CHIKV study group (2007). Infection with chikungunya virus in Italy: an outbreak in a temperate region. *Lancet (London, England)*, 370(9602):1840–1846.

[Root et al., 2006] Root, J. J., Oesterle, P. T., Nemeth, N. M., Klenk, K., Gould, D. H., McLean, R. G., Clark, L., and Hall, J. S. (2006). Experimental infection of fox squirrels (Sciurus niger) with West Nile virus. *American Journal of Tropical Medicine and Hygiene*, 75(4):697–701.

[Root et al., 2007] Root, J. J., Oesterle, P. T., Sullivan, H. J., Hall, J. S., Marlenee, N. L., McLean, R. G., Montenieri, J. A., and Clark, L. (2007). Fox squirrel (Sciurus niger) associations with West Nile virus. *American Journal of Tropical Medicine and Hygiene*, 76(4):782–784.

[Rosen, 1988] Rosen, L. (1988). Further observations on the mechanism of vertical transmission of flaviviruses by Aedes mosquitoes. *American Journal of Tropical Medicine and Hygiene*, 39(1):123–126.

[Rosenau and Goldberger, 1906] Rosenau, M. J. and Goldberger, J. (1906). Attemps to grow the yellow fever parasite - The hereditary transmission of the yellow fever parasite in the mosquito. *Yellow Fever Institute, Bulletin*, 15:103–115.

[Sheather and Jones, 1991] Sheather, S. J. and Jones, M. C. (1991). A reliable data-based band-

width selection method for kernel density estimation. *Journal of the Royal Statistical Society Series B (Methodological)*, 53(3):683–690.

[Siler et al., 1926] Siler, J. F., Hall, M. W., and Hitchens, A. P. (1926). Dengue: its history, epidemiology, mechanism of transmission, etiology, clinical manifestations, immunity, and prevention. *The Philippine Journal of Science*, 29(1-2):1–322.

[Simmons et al., 1931] Simmons, J. S., St John, J. H., and Reynolds, F. H. K. (1931). Experimental studies of dengue. *The Philippine Journal of Science*, 44(1-2):1–255.

[Tesh, 1984] Tesh, R. B. (1984). Transovarial Transmission of Arboviruses in Their Invertebrate Vectors. In *Current topics in vector research*, pages 57–76. New-York, NY: Praeger.

[Tesh et al., 1972] Tesh, R. B., Chaniotis, B. N., and Johnson, K. M. (1972). Vesicular stomatitis virus (Indiana serotype): transovarial transmission by phlebotomine sandlies. *Science*, 175(4029):1477–1479.

[Tesh and Shroyer, 1980] Tesh, R. B. and Shroyer, D. A. (1980). The mechanism of arbovirus transovarial transmission in mosquitoes: San Angelo virus in Aedes albopictus. *American Journal of Tropical Medicine and Hygiene*, 29(6):1394.

[Turell, 1988] Turell, M. J. (1988). Horizontal and vertical transmission of viruses by insect and tick vectors. In *The Arboviruses*, pages 127–152. CRC PressI Llc, Boca Raton, Florida.

[Turell et al., 1982] Turell, M. J., Hardy, J. L., and Reeves, W. C. (1982). Stabilized Infection of California Encephalitis Virus in Aedes dorsalis, and Its Implications for Viral Maintenance in Nature. *American Journal of Tropical Medicine and Hygiene*, 31(6):1252–1259.

[Vasilakis and Weaver, 2008] Vasilakis, N. and Weaver, S. C. (2008). The history and evolution of human dengue emergence. *Advances in virus research*, 72:1–76.

[Vazeille et al., 2009] Vazeille, M., Mousson, L., and Failloux, A.-B. (2009). Failure to demonstrate experimental vertical transmission of the epidemic strain of Chikungunya virus in Aedes albopictus from La Réunion Island, Indian Ocean. *Memórias do Instituto Oswaldo Cruz*, 104(4):632–635.

[Watts et al., 1985] Watts, D. M., Harrison, B. A., Pantuwatana, S., Klein, T. A., and Burke, D. S. (1985). Failure to detect natural transovarial transmission of dengue viruses by Aedes aegypti and Aedes albopictus (Diptera: Culicidae). *Journal of medical entomology*, 22(3):261–265.

[Watts et al., 1973] Watts, D. M., Pantuwatana, S., DeFoliart, G. R., Yuill, T. M., and Thompson, W. H. (1973). Transovarial transmission of LaCrosse virus (California encephalitis group) in the mosquito, Aedes triseriatus. *Science*, 182(4117):1140–1141.

[World Health Organization, 2011] World Health Organization (2011). West Nile virus.

[Yoon et al., 2012] Yoon, I.-K., Getis, A., Aldstadt, J., Rothman, A. L., Tannitisupawong, D., Koenraadt, C. J. M., Fansiri, T., Jones, J. W., Morrison, A. C., Jarman, R. G., Nisalak, A., Mammen, M. P., Thammapalo, S., Srikiatkhachorn, A., Green, S., Libraty, D. H., Gibbons, R. V., Endy, T., Pimgate, C., and Scott, T. W. (2012). Fine scale spatiotemporal clustering of dengue virus transmission in children and Aedes aegypti in rural Thai villages. *PLoS neglected tropical diseases*, 6(7):e1730.

Supporting information

Table S1 – Full list of studies included in databases 1, 2 and 3.

First author and Year	Title	Journal or book	Study type	Inclusion in databases 2 and/or 3
Adams 2010	How important is vertical transmission in mosquitoes for the persistence of dengue? Insights from a mathematical model.	Epidemics	Modelisation	
Ahmad 1997	Detection of dengue virus from field Aedes aegypti and Aedes albopictus adults and larvae	Southeast Asian J Trop Med Pub Health	Natural	+
Aitken 1979	Transovarial transmission of yellow fever virus by mosquitoes (Aedes aegypti)	Am J Trop Med Hyg	Experimental	+
Aitken 1988	Yellow fever: evolution of ideas concerned with demonstrating the natural occurrence of transovarial transmission of virus in mosquitoes.	Bull Soc Vector Ecol	Review	
Akbar 2008	PCR detection of dengue transovarial transmissibility in Aedes aegypti in Bandung, Indonesia.	Proc ASEAN Congr Trop Med Parasitol	Natural	+
Anderson 2006a	West Nile virus from female and male mosquitoes (Diptera: Culicidae) in subterranean, ground, and canopy habitats in Connecticut	J Med Entomol	Natural	+
Anderson 2006b	Importance of vertical and horizontal transmission of west nile virus by culex pipiens in the Northeastern United States	J Infect Dis	Natural	+
Anderson 2008	Extrinsic incubation periods for horinzontal and vertical transmission of West Nile Virus by Culex pipiens pipiens (Diptera: Culicidae)	J Med Entomol	Experimental	+
Anderson 2012	Horizontal and vertical transmission of West Nile virus genotype NY99 by Culex salinarius and genotypes NY99 and WN02 by Culex tarsalis	Am J Trop Med Hyg	Experimental	+
Andreadis 2010	Studies on hibernating populations of Culex pipiens from a West Nile virus endemic focus in New York City: parity rates and isolation of West Nile virus	J Am Mosq Control Asso	Natural	+
Andrews 1977	Isolation of trivittatus virus from larvae and adults reared from field-collected larvae of Aedes trivittatus (Diptera: Culicidae)	J Med Entomol	Natural	+
Angel 2008a	Association of ovarian proteins with transovarial transmission of dengue viruses by Aedes mosquitoes in Rajasthan, India.	Indian J Med Res	Natural	
Angel 2008b	Distribution and seasonality of vertically transmitted dengue viruses in Aedes mosquitoes in arid and semi-arid areas of Rajasthan, India	J Vector Borne Dis	Natural	+
Arunachalam 2002	Vertical transmission of Japanese encephalitis virus in Mansonia species, in an epidemic-prone area of southern India	Ann Trop Med Parasitol	Natural	+
Arunachalam 2008	Natural vertical transmission of dengue viruses by Aedes aegypti in Chennai, Tamil Nadu, India	Indian J Med Res	Natural	+
Bailey 1978	Isolation of St. Louis encephalitis virus from overwintering Culex pipiens mosquitoes	Science	Natural	+
Balfour 1975	Isolates of California encephalitis (LaCrosse) virus from field collected eggs and larvae of Aedes triseriatus: identification of the overwintering site of California encephalitis	J Infect Dis	Natural	+
Baqar 1993	Vertical transmission of West Nile virus by Culex and Aedes species mosquitoes	Am J Trop Med Hyg	Experimental	+
Bardos 1975	Isolation of Tahyna virus from field collected Culiseta annulata (Schrk.) larvae	Acta Virol	Natural	+
Bardos 1978	Virological examination of mosquito larvae from southern Moravia	Folia Parasitol	Natural	+
Beaty 1975	Emergence of La Crosse virus from endemic foci	Am J Trop Med Hyg	Natural	+
Beaty 1980	Transovarial transmission of yellow fever virus in Stegomyia mosquitoes	Am J Trop Med Hyg	Experimental	+
Beck 2009	Patterns of variation in the inhibitor of apoptosis 1 gene of Aedes triseriatus, a transovarial vector of La Crosse virus	J Mol Evol	Natural	
Bellini 2012	Impact of Chikungunya virus on Aedes albopictus females and possibility of vertical transmission using the actors of the 2007 outbreak in Italy	PLoS One	Experimental	+
Belloncik 1982	Activity of California encephalitis group viruses in Entrelacs (province of Quebec, Canada)	Can J Microbiol	Natural	+
Berry 1974	Isolation of LaCrosse virus (California encephalitis group) from field collected Aedes triseriatus (Say) larvae in Ohio (Diptera: Culicidae)	Mosq News	Natural	+

Follows on the next page...

Appendix C. Vertical transmission of arboviruses in mosquitoes: a historical perspective

Berry 1977	Evidence for transovarial transmission of Jamestown canyon virus, in Ohio	Mosq News	Natural	+
Bina 2008	Natural vertical transmission of dengue virus in peak summer collections of Aedes aegypti (Diptera: Culicidae) from Urban Areas of Jaipur (Rajasthan) and Delhi	J Commun Dis	Natural	+
Boromisa 1986	Virus-vector-host relationships of Aedes stimulans and Jamestown Canyon virus in a northern Indiana enzootic focus.	Am J Trop Med Hyg	Experimental	
Borucki 1999	Bunyavirus superinfection and segment reassortment in transovarially infected mosquitoes.	J Gen Virol	Experimental	
Bosio 1992	Variation in the efficiency of vertical transmission of dengue-1 virus by strains of Aedes albopictus (Diptera: Culicidae).	J Med Entomol	Experimental	+
Broom 1995	Two possible mechanisms for survival and initiation of Murray Valley encephalitis virus activity in the Kimberley region of western Australia	Am J Trop Med Hyg	Natural	+
Bugbee 2004	The discovery of West Nile virus in overwintering Culex pipiens (Diptera: Culicidae) mosquitoes in Lehigh County, Pennsylvania	J Am Mosq Control Asso	Natural	+
Burgdorfer 1967	Trans-stadial and transovarial development of disease agents in arthropods	Annu Rev Entomol	Review	
Busenberg 1988	The population dynamics of two vertically transmitted infections.	Theor Popul Biol	Modelisation	
Campbell 1991	Isolation of Jamestown Canyon virus from boreal Aedes mosquitoes from the Sierra Nevada of California	Am J Trop Med Hyg	Natural	+
Cecilio 2009	Natural vertical transmission by Stegomyia albopicta as dengue vector in Brazil	Braz J Biol	Natural	+
Chamberlain 1956a	Venezuelan equine encephalomyelitis in wild birds.	Am J Hyg	Experimental	
Chamberlain 1956b	Infection of Mansonia perturbans and Psorophora ferox mosquitoes with Venezuelan equine encephalomyelitis virus.	Proc Soc Exp Biol Med	Experimental	
Chamberlain 1957	The North American arthropod-borne encephalitis viruses in Culex tarsalis Coquillett.	Am J Hyg	Experimental	+
Chamberlain 1959	St. Louis encephalitis virus in mosquitoes	Am J Hyg	Experimental	+
Chamberlain 1961	Mechanism of transmission of viruses by mosquitoes.	Annu Rev Entomol	Review	
Chamberlain 1964	Studies on transovarial transmission of St. Louis encephalitis virus by Culex quiquefasciatus Say.	Am J Hyg	Experimental	+
Chandler 1990	Heterologous reassortment of bunyaviruses in Aedes triseriatus mosquitoes and transovarial and oral transmission of newly evolved genotypes.	J Gen Virol	Experimental	
Chandler 1998	La Crosse virus infection of Aedes triseriatus (Diptera: Culicidae) ovaries before dissemination of virus from the midgut.	J Med Entomol	Experimental	
Chen 1990	A study on transovarial transmission of dengue type I virus in Aedes aegypti	Chinese J Microbiol Immunol	Experimental	+
Chen 2010	Screening of dengue virus in field-caught Aedes aegypti and Aedes albopictus (Diptera: Culicidae) by one step SYBR Green-based reverse transcriptase-polymerase chain reaction assay during 2004-2007 in Southern Taiwan	Vector Borne Zoonotic Dis	Natural	+
Chow 1998	Monitoring of dengue viruses in field-caught Aedes aegypti and Aedes albopictus mosquitoes by a type-specific polymerase chain reaction and cycle sequencing.	Am J Trop Med Hyg	Natural	
Christensen 1978	Laboratory studies of transovarial transmission of trivittatus virus by Aedes trivitattus.	Am J Trop Med Hyg	Experimental	+
Clark 1982	Lacrosse virus activity in Illinois detected by ovitraps	Mosq News	Natural	+
Clark 1983	Persistence of La Crosse virus (California encephalitis serogroup) in north-central Illinois	Am J Trop Med Hyg	Natural	+
Clark 1985	Absence of eastern equine encephalitis (EEE) virus in immature Coquillettidia perturbans associated with equine cases of EEE	J Am Mosq Control Asso	Natural	+
Cordellier 1983	Circulation selvatique du virus Dengue 2 en 1980, dans les savanes sub-soudaniennes de Côte d'Ivoire.	Cah ORSTOM, sér Ent méd Parasitol	Natural	
Corner 1980	Cache Valley virus: experimental infection in Culiseta inornata.	Can J Microbiol	Experimental	+
Cornet 1979	Une poussée épizootique de Fièvre jaune selvatique au Sénégal oriental. Isolement du virus de lots de Moustiques adultes mâles et femelles	Med Mal Infect	Natural	+
Cornet 1984	Dengue 2 au Sénégal oriental : une poussée épizootique en milieu selvatique ; isolements du virus à partir de moustiques et d'un singe et considérations épidémiologiques	Cah ORSTOM, sÈr Ent mÉd Parasitol	Natural	+
Coz 1976	Transmission transovarienne d'un Flavivirus, le virus Koutango chez Aedes aegypti L.	C R Acad Sci Hebd Seances Acad Sci D	Experimental	

Follows on the next page...

Crane 1977	Transovarial transmission of California encephalitis virus in the mosquito Aedes dorsalis at Blue Lake, Utah	Mosq News	Natural	+
Danielova 1979	Laboratory demonstration of transovarial transmission of Tahyna virus in Aedes vexans and the role of this mechanism in overwintering of this arbovirus	Folia Parasitol	Experimental	+
Davies 1954a	Observations on the biology of West Nile virus, with special reference to its behavior in the mosquito Aedes aegypti	Ann Trop Med Parasitol	Experimental	+
Davies 1954b	The transmission of Semliki Forest virus by Aedes aegypti.	J Trop Med Hyg	Experimental	
Davis 1930	The location of yellow fever virus in infected mosquitoes and the possibility of hereditary transmission	Am J Epidemiol	Experimental	+
de Castro 2004	Dengue virus detection using reverse transcription-polymerase chain reaction in salive and progeny of experimentally infected Aedes albopictus from Brazil	Mem Inst Oswaldo Cruz	Experimental	+
de Souza 1991	Vertical transmission of dengue 1 virus by Haemagogus equinus mosquitoes	J Am Mosq Control Asso	Experimental	+
DeFoliart 1986	Changing patterns in mosquito-borne arboviruses.	J Am Mosq Control Asso	Review	
DeFoliart 1987	Advances in mosquito-borne arbovirus/vector research.	Annu Rev Entomol	Review	
Delatte 2008	in Aedes albopictus, vecteur des virus du chikungunya et de la dengue à la Réunion : biologie et contrôle	Parasite	Natural	+
Dhanda 1989	Japanese encephalitis virus infection in mosquitoes reared from field-collected immatures and in wild-caught males	Am J Trop Med Hyg	Natural	+
Dhileepan 1996	Evidence of Vertical Transmission of Ross River and Sindbis Viruses (Togaviridae: Alpha virus) by Mosquitoes (Diptera: Culicidae) in Southeastern Australia	J Med Entomol	Natural	+
Diallo 2000	Vertical transmission of the yellow fever virus by Aedes aegypti (Diptera, Culicidae): dynamics of infection in F1 adult progeny of orally infected females	Am J Trop Med Hyg	Experimental	+
Dohm 2002	Experimental vertical transmission of West Nile virus by Culex pipiens (Diptera: Culicidae)	J Med Entomol	Experimental	+
Dutary 1981	Transovarial transmission of yellow fever virus by a sylvatic vector, Haemagogus equinus.	Trsn Roy Soc Trop Med Hyg	Experimental	+
Eastwood 2011	West Nile virus vector competency of Culex quiquefasciatus mosquitoes in the Galapagos Islands	Am J Trop Med Hyg	Experimental	+
Esteva 2000	Influence of vertical and mechanical transmission on the dynamics of dengue disease.	Math Biosci	Modelisation	
Fan 2010	The impact of maturation delay of mosquitoes on the transmission of West Nile virus.	Math Biosci	Modelisation	
Farajollahi 2005	Detection of West Nile viral RNA from an overwintering pool of Culex pipiens pipiens (Diptera: Culicidae) in New Jersey, 2003	J Med Entomol	Natural	+
Fauran 1990	Etude sur la transmission verticale des virus de la dengue dans le Pacifique Sud	Bull Soc Path Ex	Natural	+
Fechter-Leggett 2012	West Nile virus cluster analysis and vertical transmission in Culex pipiens complex mosquitoes in Sacramento and Yolo Counties, California, 2011.	J Vector Ecol	Natural	
Fine 1975	Vectors and vertical transmission: an epidemiologic perspective.	Ann N Y Acad Sci	Review	
Fine 1978	Towards a quantitative understanding of the epidemiology of Keystone virus in the eastern United States.	Am J Trop Med Hyg	Modelisation	
Finlay 1899	Mosquitoes considered as transmitters of yellow fever and malaria	Medical Record	Review	
Flores 2010	Vertical transmission of St. Louis encephalitis virus in Culex quiquefasciatus (Diptera:Culicidae) in Cordoba, Argentina	Vector Borne Zoonotic Dis	Both	+
Fontenille 1997	First evidence of natural vertical transmission of yellow fever virus in Aedes aegypti, its epidemic vector	Trsn Roy Soc Trop Med Hyg	Natural	+
Fontenille 1998	La transmission verticale du virus amaril et ses consÈquences	International Seminar on Yellow Fever in Africa	Natural	+
Fouque 1996	Aedes aegypti en Guyane française : quelques aspects de l'histoire, de l'écologie générale et de la transmission verticale des virus de la dengue	Bull Soc Path Ex	Natural	+
Fouque 2004	Epidemiological and entomological surveillance of the co-circulation of DEN-1, DEN-2 and DEN-4 viruses in French Guiana	Trop Med Int Health	Natural	+
Francy 1981	Transovarial transmission of St. Louis encephalitis virus by Culex pipiens complex mosquitoes	Am J Trop Med Hyg	Experimental	+
Freier 1984	Oral and transovarial transmission of La Crosse virus by Aedes atropalpus	Am J Trop Med Hyg	Experimental	+
Freier 1987	Vertical transmission of dengue virus by mosquitoes of the Aedes scutellaris complex mosquitoes.	Am J Trop Med Hyg	Experimental	+
Freier 1988	Vertical transmission of dengue viruses by Aedes mediovitattus	Am J Trop Med Hyg	Experimental	+
Fulhorst 1994	Natural vertical transmission of western equine encephalomyelitis virus in mosquitoes	Science	Natural	+

Follows on the next page...

Appendix C. Vertical transmission of arboviruses in mosquitoes: a historical perspective

Gargan 1988	Panveld oviposition sites of floodwater Aedes mosquitoes and attempts to detect transovarial transmission of Rift Valley fever virus in South Africa	Med Vet Entomol	Natural	+
Gillett 1950	Experiments to test the possibility of transovarial transmission of yellow fever virus in the mosquito Aedes (Stegomyia) africanus Theobald	Ann Trop Med Parasitol	Experimental	+
Glass 2005	Ecological mechanisms that promote arbovirus survival: a mathematical model of Ross River virus transmission.	Trsn Roy Soc Trop Med Hyg	Modelisation	
Goddard 2003	Vertical transmission of West Nile virus by three California Culex (Diptera: Culicidae) species	J Med Entomol	Experimental	+
Gokhale 2001	Vertical transmission of dengue-2 through Aedes albopictus mosquitoes	J Commun Dis	Experimental	+
Gottfried 2002	Temporal abundance, parity, surivaval rates and arbovirus isolation of field-collected container-inhabiting mosquitoes in eastern Tennessee	J Am Mosq Control Asso	Natural	+
Graham 1999	Selection of refractory and persmissive strains of Aedes triseriatus (Diptera: Culicidae) for transovarial transmission of La Crosse virus	J Med Entomol	Experimental	+
Graham 2003	Quantitative trait loci conditioning transovarial transmission of La Crosse virus in the eastern treehole mosquito, Ochlerotatus triseriatus.	Insect Mol Biol	Experimental	
Guedes 2010	Patient-based dengue virus surveillance in Aedes aegypti from Recife, Brazil	J Vector Borne Dis	Natural	+
Guiteras 1901	Experimental yellow fever at the inoculation station of the Samtary Department of Havana with a view to producing immunization.	American Medicine	Experimental	
G¸nther 2007	Evidence of vertical transmission of dengue virus in two endemic localities in the state of Oaxaca, Mexico.	Intervirol	Natural	+
Hardy 1980	Effect of rearing temperature on transovarial transmission of St. Louis encephalitis virus in mosquitoes.	Am J Trop Med Hyg	Experimental	+
Hardy 1984	Experimental transovarial transmission of St. Louis encephalitis virus by Culex and Aedes mosquitoes	Am J Trop Med Hyg	Both	+
Hartanti 2010	Dengue virus transovarial transmission by Aedes aegypti	Univ Med	Natural	+
Hayes 1962	Detection of eastern encephalitis virus and antibody in wild and domestic birds in Massachusett	Am J Hyg	Natural	+
Hindle 1930	The transmission of yellow fever.	The Lancet	Experimental	
Hinman 1933	Hereditary transmisson of infections through arthropods.	Am J Trop Med Hyg	Review	
Hughes 2006	Comparative potential of Aedes triseriatus, Aedes albopictus, and Aedes aegypti (Diptera: Culicidae) to transovarially transmit La Crosse virus.	J Med Entomol	Experimental	+
Hull 1984	Natural transovarial transmission of dengue 4 virus in Aedes aegypti in Trinidad	Am J Trop Med Hyg	Natural	+
Hutamai 2007	A survey of dengue viral infection in Aedes aegypti and Aedes albopictus from re-epidemic areas in the North of Thailand using nucleic acid sequence based amplification assay.	Southeast Asian J Trop Med Pub Health	Natural	+
Ibanez-Bernal 1997	First record in America of Aedes albopictus naturally infected with dengue virus during the 1995 outbreak at Reynosa, Mexico	Med Vet Entomol	Natural	+
Ilkal 1991	Entomological investigations during outbreaks of dengue fever in certain villages in Maharashra state	Indian J Med Res	Natural	+
Joshi 1996	Transovarial transmission of dengue 3 virus by Aedes aegypti	Trsn Roy Soc Trop Med Hyg	Both	+
Joshi 2002	Persistence of Dengue-3 virus through transovarial transmission passage in successive generations of Aedes aegypti mosquitoes	Am J Trop Med Hyg	Experimental	+
Joshi 2006	Importance of socioeconomic status and tree holes in distribution of Aedes mosquitoes (Diptera: Culicidae) in Jodhpur, Rajasthan, India.	J Med Entomol	Natural	
Jousset 1981	Geographic Aedes aegypti strains and dengue-2 virus: susceptibility, ability to transmit to vertebrate and transovarial transmission	Ann Virol (Inst Pasteur)	Experimental	+
Jupp 1981	Laboratory vector studies on six mosquito and on tick species with chikungunya virus.	Trsn Roy Soc Trop Med Hyg	Experimental	+
Jupp 1990	Ae. Furcifer and other mosquitoes as vectors of Chikungunya virus at Mica, northeastern Transvaal, South Africa	J Am Mosq Control Asso	Natural	+
Kappus 1982	La Crosse virus infection and disease in Western North Carolina	Am J Trop Med Hyg	Natural	+
Kay 1980	Transovarial transmission of Murray Valley encephalitis virus by Aedes aegypti (L).	Aust J Exp Biol Med Sci	Experimental	+
Kay 1982	Three modes of transmission of Ross River virus vy Aedes vigilax (Skuse).	Aust J Exp Biol Med Sci	Experimental	+
Khin 1983	Transovarial transmission of dengue 2 virus by Aedes aegypti in nature	Am J Trop Med Hyg	Natural	+

Follows on the next page...

Kissling 1956	Venezuelan equine encephalomyelitis in horses.	Am J Hyg	Experimental	
Kissling 1957	Western equine encephalitis in wild birds.	Am J Hyg	Experimental	
Kow 2001	Detection of dengue viruses in field caught male Aedes aegypti and Aedes albopictus (Diptera: Culicidae) in Singapore by type-specific PCR	J Med Entomol	Natural	+
Kumari 2013	First indigenous transmission of Japanese Encephalitis in urban areas of National Capital Territory of Delhi, India	Trop Med Int Health	Natural	+
Kuno 2005	Biological Transmission of Arboviruses: Reexamination of and New Insights into Components, Mechanisms, and Unique Traits as Well as Their Evolutionary Trends.	Clin Microbiol Rev	Review	
Labuda 1983	Experimental model of transovarial transmission of Tahyna virus in Aedes aegypti	Acta Virol	Experimental	+
Le Goff 2011	Natural vertical transmission of dengue viruses by Aedes aegypti in Bolivia	Parasite	Natural	+
Leake 1984	Transovarial Transmission of Arboviruses by Mosquitoes.	Vectors in Virus Biology	Review	
LeDuc 1975	Ecology of California encephalitis viruses on the Del Mar Va Peninsula. II. Demonstration of transovarial transmission	Am J Trop Med Hyg	Natural	+
LeDuc 1979	The ecology of California groupviruses.	J Med Entomol	Review	
Lee 1997	Does transovarial transmission of dengue virus occur in Malaysian Aedes aegypti and Aedes albopictus?	Southeast Asian J Trop Med Pub Health	Experimental	+
Lee 2005	Transovarial transmission of Dengue virus in Aedes aegypti and Aedes albopictus in relation to dengue outbreak in an Urban area in Malaysia	Dengue Bull	Natural	+
Lindsay 1993	Ross River virus isolations from mosquitoes in Arid Regions of Western Australia: Implication of vertical transmission as a means of persistence of the virus	Am J Trop Med Hyg	Natural	+
Linthicum 1985	Rift Valley fever virus (family Bunyaviridae, genus Phlebovirus). Isolations from Diptera collected during an inter-epizootic period in Kenya	J Hyg	Natural	+
Lisitza 1977	Prevalence rates of LaCrosse virus (California encephalitis group) in larvae from overwintered eggs of Aedes triseriatus	Mosq News	Natural	+
Manore 2013	Inter-Epidemic and Between-Season Persistence of Rift Valley Fever: Vertical Transmission or Cryptic Cycling?	Transbound Emerg Dis	Modelisation	
Marchoux 1905	La transmission hÈrÈditaire du virus de la fiËvre jaune chez le Stegomyia fasciata.	Annales de l'Institut Pasteur	Experimental	
Marchoux 1906	Etudes sur la fiËvre jaune	Annales de l'Institut Pasteur	Experimental	
McAbee 2008	Identification of Culex pipiens complex mosquitoes in a Hybrid zone of West Nile virus transmission in Fresno county, California	Am J Trop Med Hyg	Natural	+
McLean 1975	California encephalitis virus prevalence throughout the Yukon Territory, 1971-1974.	Am J Trop Med Hyg	Natural	
McLean 1975	Mosquito-borne arboviruses in arctic america.	Med Biol	Review	
McLean 1977	Natural foci of California encephalitis virus activity in the Yukon territory	Can J Public Health	Natural	+
McLintock 1976	Isolation of snowshoe hare virus from Aedes implicatus larvae in Saskatchewan	Mosq News	Natural	+
Micieli 2013	Vector competence of Argentine mosquitoes (Diptera: Culicidae) for West Nile virus (Flaviviridae: Flavivirus).	J Med Entomol	Experimental	+
Miller 1977	Vertical transmission of La Crosse virus (California encephalitis group): transovarial and filial infection rates in Aedes triseriatus (Diptera: Culicidae)	J Med Entomol	Experimental	+
Miller 1979	Aedes triseriatus and La Crosse virus: lack of infection in eggs of the first ovarian cycle following oral infection of females	Am J Trop Med Hyg	Experimental	+
Miller 1982	Variation of La Crosse virus filial infection rates in geographic strains of Aedes triseriatus (Diptera: Culicidae)	J Med Entomol	Experimental	+
Miller 2000	First field evidence for natural vertical transmission of West Nile virus in Culex univittatus complex mosquitoes from Rift Valley province, Kenya	Am J Trop Med Hyg	Natural	+
Mishra 2001	Transovarial transmission of West Nile virus in Culex vishnui mosquito.	Indian J Med Res	Experimental	+
Mitamura 1939	Weitere Untersuchungen über die Übertragung der japanischen epidemischen enzephalitis durch Mücken.	Trans Soc Pathol Jpn	Experimental	
Mitamura 1950	Seasonal occurrence of mosquito in Okayama 1946 and infectivity of the mosquito with Japanese B encephalitis virus; trans ovary infection of the virus in mosquito	Jap Med J	Experimental	+
Mitchell 1990a	Vector competence of Aedes albopictus for a newly recognized Bunyavirus from mosquitoes collected in Potosi, Missouri.	J Am Mosq Control Asso	Experimental	+

Follows on the next page...

Appendix C. Vertical transmission of arboviruses in mosquitoes: a historical perspective

Mitchell 1990b	Vertical transmission of dengue viruses by strains of Aedes albopictus recently introduced into Brazil.	J Am Mosq Control Asso	Experimental	+
Mondet 2002	Isolation of yellow fever virus from nulliparous Haemagogus janthinomys in Eastern Amazonia	Vector Borne Zoonotic Dis	Natural	+
Morris 1978	An Evaluation of the Hypothesis of Transovarial transmission of Eastern Equine encephalomyelitis virus by Culiseta melanura	Am J Trop Med Hyg	Natural	+
Mourya 1987a	Experimental transmission of Chikungunya virus by Aedes vittatus mosquitoes.	Indian J Med Res	Experimental	+
Mourya 1987b	Absence of transovarial transmission of chikungunya virus in Aedes aegypti & Ae. albopictus mosquitoes.	Indian J Med Res	Experimental	+
Mourya 2001	Horizontal and vertical transmission of dengue virus type 2 in highly and lowly susceptible strains of Aedes aegypti mosquitoes	Acta Virol	Experimental	+
Mulyatno 2012	Vertical transmission of dengue virus in Aedes aegypti collected in Surabaya, Indonesia, during 2008-2011	Jpn J Infect Dis	Natural	+
Muul 1975	Ecological studies of Culiseta melanura (Diptera:Culicidae) in relation to eastern and western equine encephalomyelitis viruses on the eastern shore of Maryland	J Med Entomol	Natural	+
Nasci 2001	West Nile virus in overwintering Culex mosquitoes, New York City 2000	Emerg Infect Dis	Natural	+
Nayar 1986	Experimental vertical transmission of Saint Louis encephalitis virus by Florida mosquitoes.	Am J Trop Med Hyg	Experimental	+
Nelms 2013a	Experimental and natural vertical transmission of West Nile virus by California Culex (Diptera: Culicidae) mosquitoes	J Med Entomol	Experimental	+
Nelms 2013b	Phenotypic variation among Culex pipiens complex (Diptera: Culicidae) populations from the Sacramento Valley, California: horizontal and vertical transmission of West Nile virus, diapause potential, autogeny, and host selection.	Am J Trop Med Hyg	Experimental	+
Nir 1963	Failure to obtain experimental transovarian transmission of West Nile virus by Aedes aegypti	Ann Trop Med Parasitol	Experimental	+
Niyas 2010	Molecular characterization of Chikungunya virus isolates from clinical samples and adult Aedes albopictus mosquitoes emerged from larvae from Kerala, South India.	Virol J	Natural	
Pantuwatana 1974	Isolation of La Crosse virus from field collected Aedes triseriatus larvae	Am J Trop Med Hyg	Natural	+
Paramasivan 2006	Serological and entomological investigation of an outbreak of dengue fever at Nagercoil, Kanyakumari district, Tamil Nadu.	Indian J Med Res	Natural	
Patrican 1985	Lack of adverse effect of transovarially acquired La Crosse virus infection on the reproductive capacity of Aedes triseriatus (Diptera: Culicidae).	J Med Entomol	Experimental	
Patrican 1985	La Crosse viremias in juvenile, subadult and adult chipmunks (Tamias striatus) following feeding by transovarially-infected Aedes triseriatus.	Am J Trop Med Hyg	Experimental	
Paulson 1989	Replication and dissemination of La Crosse virus in the competent vector Aedes triseriatus and the incompetent vector Aedes hendersoni and evidence for transovarial transmission by Aedes hendersoni (Diptera: Culicidae)	J Med Entomol	Experimental	+
Pelz 1990	Vertical transmission of St Louis encephalitis virus to autogenously developed eggs of Aedes atroplapus mosquitoes	J Am Mosq Control Asso	Experimental	+
Pessoa Martins 2012	Occurrence of Natural Vertical transmission of Dengue-2 and Dengue-3 viruses in Aedes aegypti and Aedes albopictus in Fortaleza, Ceara, Brazil	PLoS One	Natural	+
Philip 1929	Possibility of hereditary transmission of yellow fever virus by Aedes aegypti (Linn.)	J Exp Med	Experimental	+
Philipps 2006	Field-caught Culex erythrothorax larvae found naturally infected with West Nile virus in Grand county, utah.	J Am Mosq Control Asso	Natural	+
Pinger 1983	Isolation of La Crosse and other arboviruses from Indiana mosquitoes	Mosq News	Natural	+
Pinheiro 2005	Detection of dengue virus serotype 3 by reverse transcription polymerase chain reaction in Aedes aegypti (Diptera: Culicidae) captured in Manaus, Amazonas	Mem Inst Oswaldo Cruz	Natural	+
Ramalingam 1986	Does transovarial transmission of dengue virus occur in Malaysia	Trop Biomed	Natural	+
Reed 1901	Experimental yellow fever.	American Medicine	Experimental	
Reese 2009	Aedes triseriatus females transovarially-infected with La Crosse virus mat more efficiently than uninfected mosquitoes.	J Med Entomol	Both	
Reese 2010	Identification of super-infected Aedes triseriatus mosquitoes collected as eggs from the field and partial characterization of the infecting La Crosse viruses	Virol J	Natural	+

Follows on the next page...

Reeves 1946	Laboratory transmission of Japanese B encephalitis virus by seven species (three genera) of North America mosquitoes	J Exp Med	Experimental	+
Reeves 1961	Overwintering of arthropod-borne viruses.	Prog Med Virol	Review	
Reisen 2006	Overwintering of West Nile virus in Southern California	J Med Entomol	Both	+
Rodhain 1998	La notion de rÈservoir naturel en arbovirologie.	Bull Soc Path Ex	Review	
Rohani 2007	Detection of transovarial dengue virus from field-caught Aedes aegypti and Ae albopictus larvae using C6/36 cell culture and reverse transcriptase polymerase chain reaction (RT-PCR) techniques	Dengue Bull	Natural	+
Rohani 2008	Persistency of transovarial dengue virus in Aedes aegypti (Linn.)	Southeast Asian J Trop Med Pub Health	Experimental	+
Romero-Vivas 1998	Determination of dengue virus serotypes in individual Aedes aegypti mosquitoes in Colombia	Med Vet Entomol	Natural	+
Romoser 2011	Rift Valley fever virus-infected mosquito ova and associated pathology: possible implications for endemic maintenance.	Res Rep Trop Med	Natural	
Rosen 1978	Transovarial transmission of Japanese encephalitis virus by mosquitoes	Science	Experimental	+
Rosen 1980	Transovarial transmission of Japanese encephalitis virus by Culex tritaeniorhynchus mosquitoes.	Am J Trop Med Hyg	Experimental	+
Rosen 1981	Transmission transovarienne des arbovirus par les moustiques.	Med Trop (Mars)	Review	
Rosen 1983	Transovarial transmission of dengue virus by mosquitoes: Aedes albopictus and Aedes aegypti	Am J Trop Med Hyg	Experimental	+
Rosen 1984	Ovarian Infection and Transovarial Transmission of Viruses in Insects	Concepts in Viral Pathogenesis	Review	
Rosen 1987	Mechanism of vertical transmission of the dengue virus in mosquitoes	C R Acad Sciences	Experimental	+
Rosen 1987	Overwintering Mechanisms of Mosquito-Borne Arboviruses in Temperate Climates.	Am J Trop Med Hyg	Review	
Rosen 1988	Further observation on the mechanism of vertical transmission of flaviviruses by Aedes mosquitoes	Am J Trop Med Hyg	Experimental	+
Rosen 1989a	A longitudinal study of the prevalence of Japanese encephalitis virus in adult and larval Culex tritaeniorhynchus mosquitoes in northern Taiwan	Am J Trop Med Hyg	Natural	+
Rosen 1989b	Experimental vertical transmision of Japanese encephalitis virus by Culex tritaeniorhynchus and other mosquitoes	Am J Trop Med Hyg	Experimental	+
Rosenau 1906	Hereditary transmission of the yellow fever parasite in the mosquito.	Yellow Fever Institute, Bulletin	Experimental	
Russell 1989	New South Wales mosquito and arbovirus surveillance: the program after 5 years.	Arbovirus research in Australia - 5th Symposium	Natural	
Russell 1992	The surveillance of arbovirus activity in N.S.W. 1989-1992.	Arbovirus research in Australia - 6th Symposium	Natural	
Scherer 1986	Vector incompetency: its implication in the disappearance of epizootic Venezuelan equine encephalomyelitis virus from Middle America	J Med Entomol	Experimental	+
Schopen 1991	Vertical and veneral transmission of California group viruses by Aedes triseriatus and Culiseta inornata mosquitoes	Acta Virol	Experimental	+
Scott 1984	The distribution and development of eastern equine encephalitis virus in its enzootic mosquito vector, Culiseta melanura.	Am J Trop Med Hyg	Experimental	
Scott 1990	Susceptibility of Aedes albopictus to infection with eastern equine encephalomyelitis virus	J Am Mosq Control Asso	Experimental	+
Serufo 1993	Isolation of dengue virus type 1 from larvae from Aedes albopictus in Campos Altos city, state of Minas Gerais, Brazil	Mem Inst Oswaldo Cruz	Natural	+
Shroyer 1986a	Transovarial maintenance of San Angelo virus in sequential generations of Aedes albopictus	Am J Trop Med Hyg	Experimental	+
Shroyer 1986b	Aedes albopictus and arboviruses: a concise review of the literature.	J Am Mosq Control Asso	Review	
Shroyer 1990	Vertical maintenance of dengue-1 virus in sequential generations of Aedes albopictus	J Am Mosq Control Asso	Experimental	+
Siler 1926	Dengue: Its history, epidemiology, mechanism of transmission, etiology, clinical manifestations, immunity and prevention	Philipp J Sci	Experimental	
Simmons 1931	Experimental studies of dengue.	Philipp J Sci	Experimental	
Soman 1985	Transovarial transmission of Japanese encephalitis virus in Culex bitaeniorhynchus mosquitoes	Indian J Med Res	Experimental	+
Soman 1986	Transovarial transmission of Japanese encephalitis virus in Culex vishnui mosquitoes.	Indian J Med Res	Experimental	
Spielman 1975	Inherited infection in the epidemiology of diptera-borne disease: perspectives and an introduction.	Ann N Y Acad Sci	Review	

Follows on the next page...

Appendix C. Vertical transmission of arboviruses in mosquitoes: a historical perspective

Sprance 1981	Experimental evidence against the transovarial transmission of eastern equine encephalitis virus in Culiseta melanura	Mosq News	Both	+
Stamm 1962	Arbovirus studies in south Alabama, 1957-1958	Am J Hyg	Natural	+
Stockes 1928	Experimental transmission of yellow fever to laboratory animals	Am J Trop Med Hyg	Experimental	+
Sudeep 2013	Preliminary findings on Bagaza virus (Flavivirus: Flaviviridae) growth kinetics, transmission potential & transovarial transmission in three species of mosquitoes.	Indian J Med Res	Experimental	+
Takashima 1989	Horizontal and Vertical transmission of Japanese Encephalitis Virus by Aedes japonicus (Diptera: Culicidae)	J Med Entomol	Experimental	+
Taylor 1971	California group arboviruses in Florida. Host-vector relations.	Am J Trop Med Hyg	Natural	
Tesh 1975	Laboratory studies of transovarial transmission of La Crosse and other arboviruses by Aedes albopictus and Culex fatigans	Am J Trop Med Hyg	Experimental	+
Tesh 1979	Studies of transovarial transmission of yellow fever and Japanese encephalitis viruses in Aedes mosquitoes and their implications for the epidemiology of dengue.	Dengue in the Caribbean	Experimental	
Tesh 1980a	Experimental studies on the transovarial transmission of Kunjin and San Angelo viruses in mosquitoes	Am J Trop Med Hyg	Experimental	+
Tesh 1980b	The mechanism of arbovirus transovarial transmission in mosquitoes: San Angelo virus in Aedes albopictus	Am J Trop Med Hyg	Experimental	+
Tesh 1981	The location of San Angelo virus in developing ovaries of transovarially infected Aedes albopictus mosquitoes as revealed by fluorescent antibody technique.	Am J Trop Med Hyg	Experimental	
Tesh 1984	Transovarial transmission of arboviruses in their invertebrate vectors.	Current topics in vector research	Review	
Thavara 2009	Outbreak of chikungunya fever in Thailand and virus detection in field population of vector mosquitoes, Aedes aegypti (L.) and Aedes albopictus Skuse (Diptera: Culicidae).	Southeast Asian J Trop Med Pub Health	Natural	+
Thenmozhi 2000	Natural vertical transmission of dengue viruses in Aedes aegypti in southern India	Trsn Roy Soc Trop Med Hyg	Natural	+
Thenmozhi 2006	Long-Term study of Japanese encephalitis virus infection in Anopheles subpictus in Cuddalore district, Tamil Nadu, South India	Trop Med Int Health	Natural	+
Thenmozhi 2007	Natural vertical transmission of dengue virus in Aedes albopictus (Diptera: Culicidae) in Kerala, a southern Indian state	Jpn J Infect Dis	Natural	+
Thongrungkiat 2011	Prospective field study of transovarial dengue virus transmission by two different forms of Aedes aegypti in an urban area of Bangkok, Thailand	J Vector Ecol	Natural	+
Turell 1982a	Transovarial and trans-stadial transmission of California encephalitis virus in Aedes dorsalis and Aedes melanimon	Am J Trop Med Hyg	Experimental	+
Turell 1982b	Stabilized infection of California encephalitis virus in Aedes dorsalis, and its implications for viral maintenance in nature	Am J Trop Med Hyg	Experimental	+
Turell 1982c	Evaluation of the efficiency of transovarial transmission of California encephalitis viral strains in Aedes dorsalis and Aedes melanimon	Am J Trop Med Hyg	Experimental	+
Turell 1988	Horizontal an vertical transmission of viruses by insect and tick vectors.	The Arboviruses: Epidemiology and Ecology	Review	
Turell 2001	Vector competence of North American mosquitoes (Diptera: Culicidae) for West Nile virus	J Med Entomol	Experimental	+
Unlu 2010	Evidence of vertical transmission of West Nile virus in field-collected mosquitoes	J Vector Ecol	Natural	+
van den Hurk 2003	Vector competence of Australian mosquitoes (Diptera: Culicidae) for Japanese encephalitis virus	J Med Entomol	Experimental	+
Vazeille 2009	Failure to demonstrate experimental vertical transmission of the epidemic strain of Chikungunya virus in Aedes albopictus from La Réunion Island, Indian Ocean.	Mem Inst Oswaldo Cruz	Experimental	+
Vilela 2010	Dengue virus 3 genotype I in Aedes aegypti mosquitoes and eggs, Brazil, 2005-2006	Emerg Infect Dis	Natural	+
Wasinpiyamongkoi 2003	Susceptibility and transovarial transmission of dengue virus in Aedes aegypti: a preliminary study of morphological variations.	Southeast Asian J Trop Med Pub Health	Experimental	+
Watts 1973	Transovarial transmission of La Crosse virus (California encephalitis group) in the mosquito, Aedes triseriatus	Science	Experimental	+
Watts 1974	Overwintering of La Crosse virus in Aedes triseriatus	Am J Trop Med Hyg	Natural	+
Watts 1975a	Transovarial transmission of arboviruses by mosquitoes: a review.	Med Biol	Review	
Watts 1975b	Transovarial transmission of LaCrosse virus in Aedes triseriatus.	Ann N Y Acad Sci	Review	
Watts 1985	Failure to detect natural transovarial transmission of dengue viruses by Aedes aegypti and Aedes albopictus (Diptera: Culicidae)	J Med Entomol	Natural	+
Watts 1987	Ecological evidence against vertical transmission of eastern equine encephalitis virus by mosquitoes (Diptera: Culicidae) on the Delmarva Peninsula, USA	J Med Entomol	Natural	+

Follows on the next page...

Woodring 1998	Short report: Diapause, transovarial transmission, and filial infection rates in geographic strains of La Crosse virus-infected Aedes triseriatus	Am J Trop Med Hyg	Experimental	+
Zeidler 2008	Dengue virus in Aedes aegypti larvae and infestation dynamics in Roraima, Brazil	Rev Saude Publica	Natural	+
Zhang 1993	Transovarial transmission of Chikungunya virus in Aedes albopictus and Aedes aegypti mosquitoes	Chin J Virol	Experimental	+
Zhang 1996	Transovarial transmission of Dengue viruses in Aedes albopictus and Aedes aegypti mosquitoes	Virol Sinica	Experimental	+
Zuo 2004	Isolation, identification, and phylogenetic analysis of a dengue virus strain from Aedes albopictus collected in Mawei town in Guizhou Province, China	Chin Med J (Engl)	Experimental	
Zytoon 1993	Transovarial transmission of chikungunya virus by Aedes albopictus mosquitoes ingesting microfilariae of Dirofilaria immitis under laboratory conditions	Microbiol Immunol	Experimental	+

Appendix C. Vertical transmission of arboviruses in mosquitoes: a historical perspective

Factor category	Factors
	Publication ID
Virus identification	Virus specie
	Virus strain
	Virus isolation
Mosquito identification	Mosquito genus
	Mosquito sub-genus
	Mosquito specie
	Mosquito strain
	Mosquito-Virus pair
Passage history of the virus	Vertebrate cell passages
	Number of vertebrate cell passages
	Arthropod cell passages
	Number of arthropod cell passages
	Alternate passage
	VT history
	Host type of the last passage
	Complete passage history
	Infection method of the parental mosquitoes
Was the mosquito infected by a parasite?	Parasite (Yes or No)
	Parasite specie
Rearing conditions	Temperature
	Humidity
	Generation of tested offspring
	Ovarian cycle of tested offspring
	Egg laying day after infection for tested offspring
Detection and identification of VT	Mosquito development stage when test for VT
	Previous cell culture before detection
	Detection assay
	Identification assay
	VT (Yes or No)
	Sample size
	Minimum Infection Rate

(a) Database # 2

Factor category	Factors
	Publication ID
Virus identification	Virus specie
Mosquito identification	Mosquito genus
	Mosquito sub-genus
	Mosquito specie
	Mosquito-Virus pair
Geographic and epidemiological contexte	Study site
	Epidemiological contexte
	Mosquito capture development stage
Detection and identification of VT	Mosquito development stage when test for VT
	Previous cell culture before detection
	Detection assay
	Identification assay
	VT (Yes or No)
	Sample size
	Minimum Infection Rate

(b) Database # 3

Table S2 – Full list of factors for databases # 2 and # 3.

D Determinants of arbovirus vertical transmission in mosquitoes

Sebastian Lequime, Richard E. Paul & Louis Lambrechts. Determinants of arbovirus vertical transmission in mosquitoes. *PLoS Pathogens*, 12(5): e1005548, May 2016. doi:10.1371/journal.ppat.1005548 *In press.*

Abstract

Vertical transmission (VT) and horizontal transmission (HT) of pathogens refer to parental and non-parental chains of host-to-host transmission. Combining HT with VT enlarges considerably the range of ecological conditions in which a pathogen can persist, but the factors governing the relative frequency of each transmission mode are poorly understood for pathogens with mixed-mode transmission. Elucidating these factors is particularly important for understanding the epidemiology of arthropod-borne viruses (arboviruses) of public health significance. Arboviruses are primarily maintained by HT between arthropod vectors and vertebrate hosts in nature, but are occasionally transmitted vertically in the vector population from an infected female to her offspring, which is a proposed maintenance mechanism during adverse conditions for HT. Here, we review over a century of published primary literature on natural and experimental VT, which we previously assembled into large databases, to identify biological factors associated with the efficiency of arbovirus VT in mosquito vectors. Using a robust statistical framework, we highlight a suite of environmental, taxonomic, and physiological predictors of arbovirus VT. These novel insights contribute to refine our understanding of strategies employed by arboviruses to persist in the environment

and cause substantial public health concern. They also provide hypotheses on the biological processes underlying the relative VT frequency for pathogens with mixed-mode transmission that can be tested empirically.

1 Introduction

Host-to-host transmission of pathogens is usually categorized as either vertical or horizontal, irrespective of the physical route of transmission [Fine, 1975]. Vertical transmission (VT), also called hereditary transmission, refers to transmission of a pathogen from parent to offspring. Horizontal transmission (HT) encompasses all other modes of non-parental transmission, including sexual and vector-borne transmission. VT and HT are not mutually exclusive and a combination of both, known as mixed-mode transmission, is common across taxa of hosts and pathogens, including eukaryotes, bacteria, and viruses [Ebert, 2013]. Combining both modes of transmission allows pathogens to persist under conditions that would otherwise lead to extinction. Indeed, demographic and epidemiological changes of host populations lead to different opportunities for transmission by one mode or the other. When the density of susceptible hosts is low, for example, VT becomes an essential link in the transmission chain [Lipsitch et al., 1996]. Elucidating the biological factors governing the relative frequency of HT and VT is essential to understanding the ecology and evolution of pathogens with mixed-mode transmission [Ebert, 2013]. This is particularly critical for infectious agents of public health relevance because their epidemiology and control will largely depend on their mode of transmission. Rare VT can be important for persistence of pathogens that would otherwise go extinct during the winter [Mink, 1993]. Such knowledge can be used to prevent VT and consequently reduce the likelihood of future epidemics.

Arthropod-borne viruses (arboviruses) that are pathogenic for humans mainly belong to four distinct genera of RNA viruses: *Alphavirus* (e.g., chikungunya and Eastern equine encephalomyelitis viruses), *Flavivirus* (e.g., Zika, yellow fever, dengue, and West Nile viruses), *Orthobunyavirus* (e.g., California encephalitis virus) and *Phlebovirus* (e.g., Rift Valley fever virus). There are more than 530 described arboviruses, of which about a hundred are pathogenic to humans [Centers for Disease Control and Prevention, 2010]. Epidemiological cycles of arboviruses often consist of complex transmission networks that involve a variety of vertebrate hosts and hematophagous

arthropods, usually referred to as vectors, such as mosquitoes or ticks [Diaz et al., 2012]. Humans are not necessarily at the centre of the transmission network and may only be incidental hosts (e.g., West Nile virus).

Arboviruses are primarily maintained by cross-species transmission between arthropod vectors and vertebrate hosts. Vectors become infected during blood feeding on a viremic vertebrate and, after a period of development within the vector, the virus can infect a new vertebrate host during a subsequent blood meal. This mode of transmission qualifies as HT because it involves unrelated hosts [Fine, 1975]. Although HT largely determines arbovirus epidemiology, some epidemiological features remain unexplained. In particular, the maintenance of arboviruses in endemic areas during adverse conditions for vector activity has long puzzled researchers. Dry seasons in tropical areas, cold seasons in temperate regions, or insecticide spraying campaigns can drastically reduce vector density and thus opportunities for HT [Leake, 1984]. In addition, arbovirus infections in vertebrates usually result in long-lasting protective immunity so that high levels of herd immunity will prevent HT following epidemics. Several hypotheses have been suggested to explain the maintenance of arboviruses during inter-epidemic periods, such as virus re-introduction, circulation in unknown host species (i.e., reservoir), and alternative transmission mechanisms [Reeves, 2004]. Phylogeographic studies of dengue virus genomic sequences from isolates collected prior to, during, and post epidemics found that despite a degree of mutation, the virus was the same within each site over time, but different between sites [Dupont-Rouzeyrol et al., 2014, Cao-Lormeau et al., 2011]. This would suggest that the virus was thus maintained locally despite the lack of conditions permissive to HT.

Arbovirus VT in the arthropod vector population is a proposed maintenance mechanism during adverse conditions for HT. VT is well documented in mosquitoes for several arboviruses across the four aforementioned major viral genera [Lequime and Lambrechts, 2014]. Arbovirus VT in mosquitoes is generally maternal and occurs through two main mechanisms: transovarial transmission (TOT), whereby the virus infects the germinal tissues of the female mosquitoes, and trans-egg transmission, whereby the virus infects the egg during oviposition [Rosen, 1988]. Arboviruses may persist during unfavourable periods through the infection of eggs, larvae, or adults (including diapausing individuals) without the need of HT. The epidemiological significance of arbovirus VT, however, has remained controversial for several arboviruses

of public health concern [Lequime and Lambrechts, 2014]. Controversy primarily stems from the inability of estimated rates of VT to explain long-term arbovirus maintenance, combined with discrepancies between field observations and laboratory studies, as well as inconsistencies among virological techniques [Adams and Boots, 2010, Lequime and Lambrechts, 2014, Grunnill and Boots, 2015]. Here, we used the largest contemporary databases on natural and experimental VT occurrence records to provide a comprehensive list of positive and negative predictors of arbovirus VT in mosquitoes.

Following Clements [Clements, 2012], we defined VT rate as the proportion of vertically infected offspring from a population of infected females. Infection status of mothers is typically unknown in natural settings and rarely assessed in published laboratory experiments. Clements introduced the effective vertical transmission (eVT) rate to describe the proportion of vertically infected offspring, irrespective of the infection status of their mothers [Clements, 2012]. Therefore, eVT is the product of VT rate and infection prevalence in the population of females under consideration. Only when 100% of mothers are infected, such as in most laboratory experiments, does eVT rate equal VT rate. It is worth noting that VT and eVT definitions do not make any assumption about the underlying mechanism (i.e., TOT or trans-egg VT).

2 Overview of the Primary Literature

Literature search and database assembly are described elsewhere [Lequime and Lambrechts, 2014]. Briefly, we previously conducted a systematic review of the literature published prior to 25 September 2013 from various online and physical sources (the full list of publication is provided in Table S1). We subsequently conducted a "resilience" test to evaluate the impact of publications that may have been missed during the literature search (see below). Literature search was restricted to the three main arboviral families (i.e., *Bunyaviridae*, *Flaviviridae*, and *Togaviridae*). We compiled two databases. Database # 1 includes "experimental" VT studies typically conducted in a laboratory setting. Database # 2 includes "natural" studies that investigated arbovirus VT in nature by collecting and testing wild, immature male or nulliparous female mosquitoes. Both databases only included publications that specified at least the mosquito and virus species tested, sample size, and detection technique. The full list of variables in databases # 1 and # 2 can be found in Table S2.

In natural VT studies, mosquitoes were collected from the field and brought back to the laboratory. Field collections usually consisted of adult males or immature stages (eggs, larvae, or pupae) that were reared in the laboratory to later developmental stages. These were then processed in pools to detect the presence of virus. In experimental VT studies, female mosquitoes from a parental generation were artificially infected (orally, by intra-thoracic [IT] inoculation, or vertically) and subsequently allowed to lay eggs. Eggs were hatched and the offspring sampled at various developmental stages to determine the presence of virus, usually in pools of individuals.

Virological detection techniques belonged to four main categories [Lequime and Lambrechts, 2014]: animal, cellular, immunological, and molecular assays. In "animal" assays, mosquito extracts were inoculated in vivo into susceptible laboratory animals that were subsequently checked for pathological effects. In "cellular" assays, cytopathological effects were monitored following inoculation of mosquito extracts onto cell cultures in vitro. In "immunological" assays, viral antigens were detected by antibodies with or without previous amplification in cell culture or animal tissues. In "molecular" assays, viral RNA was detected by reverse transcription (RT)-PCR. Although a few studies tested the progeny individually, individual mosquitoes were most often tested in pools. Therefore, the proportion of infected individuals was usually estimated as the minimum infection rate based on the assumption that only one individual was infected in the pool. This assumption is generally reasonable because observed eVT rates are low for arboviruses (see below), but will underestimate VT rate when it is efficient.

In both databases, the *Aedes–Flavivirus* pair was the most represented vector–virus combination (56.7% and 24.6% of all mosquitoes tested in databases # 1 and # 2, respectively), followed by *Culex–Flavivirus* (34.1% and 43.7% of all mosquitoes tested, respectively). Within the *Aedes–Flavivirus* pair in database # 1, dengue viruses (DENV1, DENV2, DENV3, and DENV4) were the most represented viruses (41% of all *Aedes* mosquitoes tested for flavivirus infection and 23% of all mosquitoes tested). Other vector–virus pairs included *Aedes–Orthobunyavirus* (19.5% and 6.1% of all mosquitoes in databases # 1 and # 2, respectively) and *Aedes–Alphavirus* (4.6% and 1.4% of all mosquitoes tested, respectively).

3 Rates of Effective Vertical Transmission (eVT)

In experimental studies, estimates of eVT rates were significantly higher for the orthobunyaviruses California encephalitis virus (CEV) and LaCrosse virus (LACV) than for flaviviruses or alphaviruses (Figure 1-A). CEV and LACV had weighted mean eVT rates of 16% and 28%, respectively, whereas weighted mean eVT rates were 1‰ for yellow fever virus (YFV), 4‰ for Japanese encephalitis virus (JEV), 1‰ for West Nile virus (WNV), 2%–6‰ for DENV1-4, and 1‰ for chikungunya virus (CHIKV). eVT rate estimates were up to several orders of magnitude smaller in natural studies than in experimental studies (Figure 1-B). Weighted mean eVT rates in natural studies were 0.1‰ for CEV, 5‰ for LACV, 8‰ for YFV, 0.02‰ for JEV, 0.4‰ for WNV, 2‰ for DENV1-4, and 0.8‰ for CHIKV.

Mosquito and virus taxonomic groups were strong predictors of the observed level of VT in experimental studies (Figure 2). Mosquitoes in the *Culex* genus were associated with significantly lower eVT rates than mosquitoes in the *Aedes* genus (odds ratio [OR] = 0.32, $p < 0.001$, 95% CI [0.28–0.36]). Using flaviviruses as the reference, viruses of the Orthobunyavirus genus had significantly higher eVT rates (OR = 45.71, $p < 0.001$, 95% CI [38.34–54.49]), whereas members of the *Alphavirus* genus had significantly lower eVT rates (OR = 0.08, $p < 0.01$, 95% CI [0.14–0.45]). The method used to infect mothers significantly influenced eVT rate estimates (Figure 2). Both IT inoculated and vertically infected mothers resulted in higher eVT rates (OR = 1.80, $p < 0.001$, 95% CI [1.63–2.00] and OR = 4.71, $p < 0.001$, 95% CI [3.94–5.64], respectively) compared to orally infected mothers. eVT rate estimates also depended on virological assays (Figure 2). Using cellular assays as the methodological reference, only immunological assays were associated with higher eVT rates (OR = 3.27, $p < 0.001$, 95% CI [2.70–3.98]). Offspring produced during the second or later gonotrophic cycles displayed higher eVT rates than offspring produced during the first gonotrophic cycle (OR = 1.66, $p < 0.001$, 95% CI [1.55–1.77]). Lower eVT rates were found when the virus was detected at the adult stage as opposed to immature stages of the progeny (OR = 0.60, $p < 0.001$, 95% CI [0.54–0.67]).

The *Aedes* genus is divided into several subgenera [Wilkerson et al., 2015], among which Stegomyia (including, for example, *Aedes aegypti* and *Aedes albopictus*) and *Ochlerotatus* (including, for example, *Aedes dorsalis*) are the most represented in the

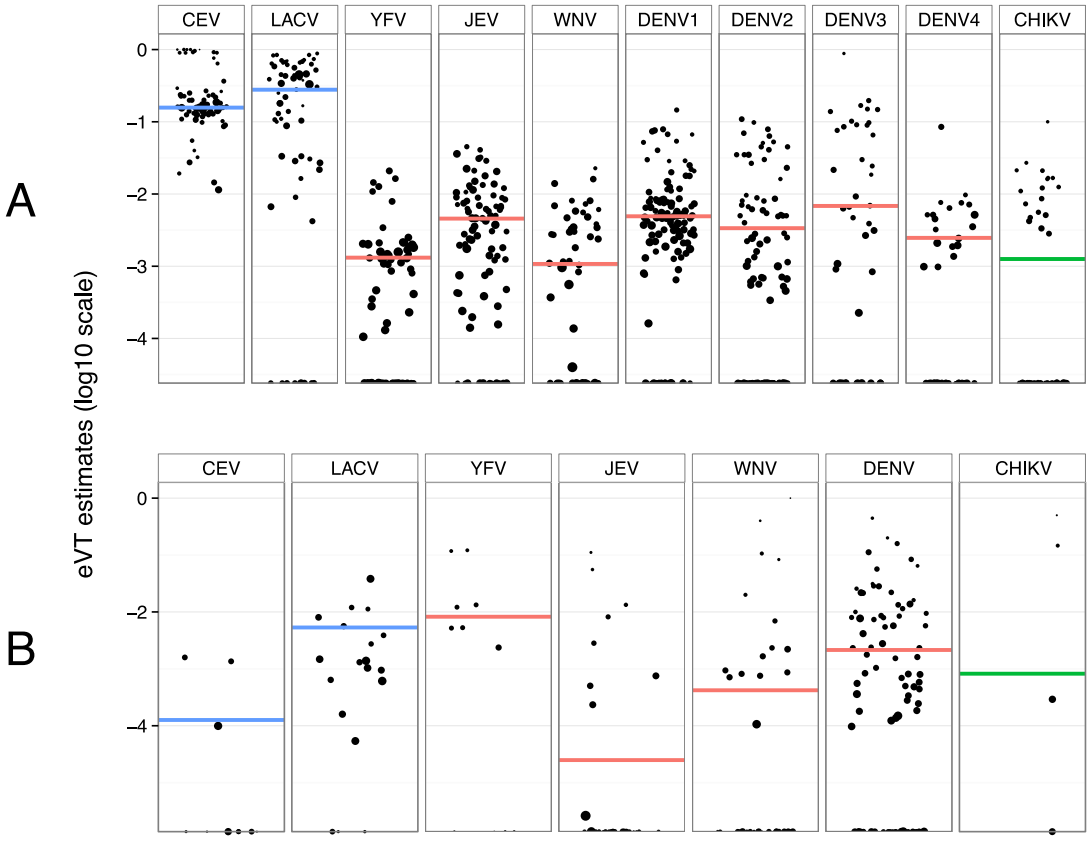

Figure 1 – Distributions of eVT estimates for the most represented arboviruses in (A) experimental and (B) natural studies. Each data point represents the log10-transformed eVT rate obtained from a single database entry. Dot size is proportional to log10-transformed sample size. A horizontal coloured line shows the weighted mean eVT for each virus (blue: orthobunyaviruses; red: flaviviruses; green: alphaviruses). CEV: California encephalitis virus; LACV: LaCrosse virus; YFV: yellow fever virus; JEV: Japanese encephalitis virus; WNV: West Nile virus; DENV1-4: dengue viruses; CHIKV: chikungunya virus.

VT literature. Because *Stegomyia* mosquitoes are typically associated with flaviviruses, whereas *Ochlerotatus* mosquitoes are typically associated with orthobunyaviruses, the strong effect of the viral genus observed (Figure 2) might be confounded by the mosquito subgenera within the *Aedes* genus. A separate analysis focused on the two *Aedes* subgenera and the two viral genera, *Orthobunyavirus* and *Flavivirus*, showed that their interaction did not significantly influence eVT rates ($p = 0.09$), confirming that orthobunyaviruses had higher eVT rates independently of the mosquito subgenera.

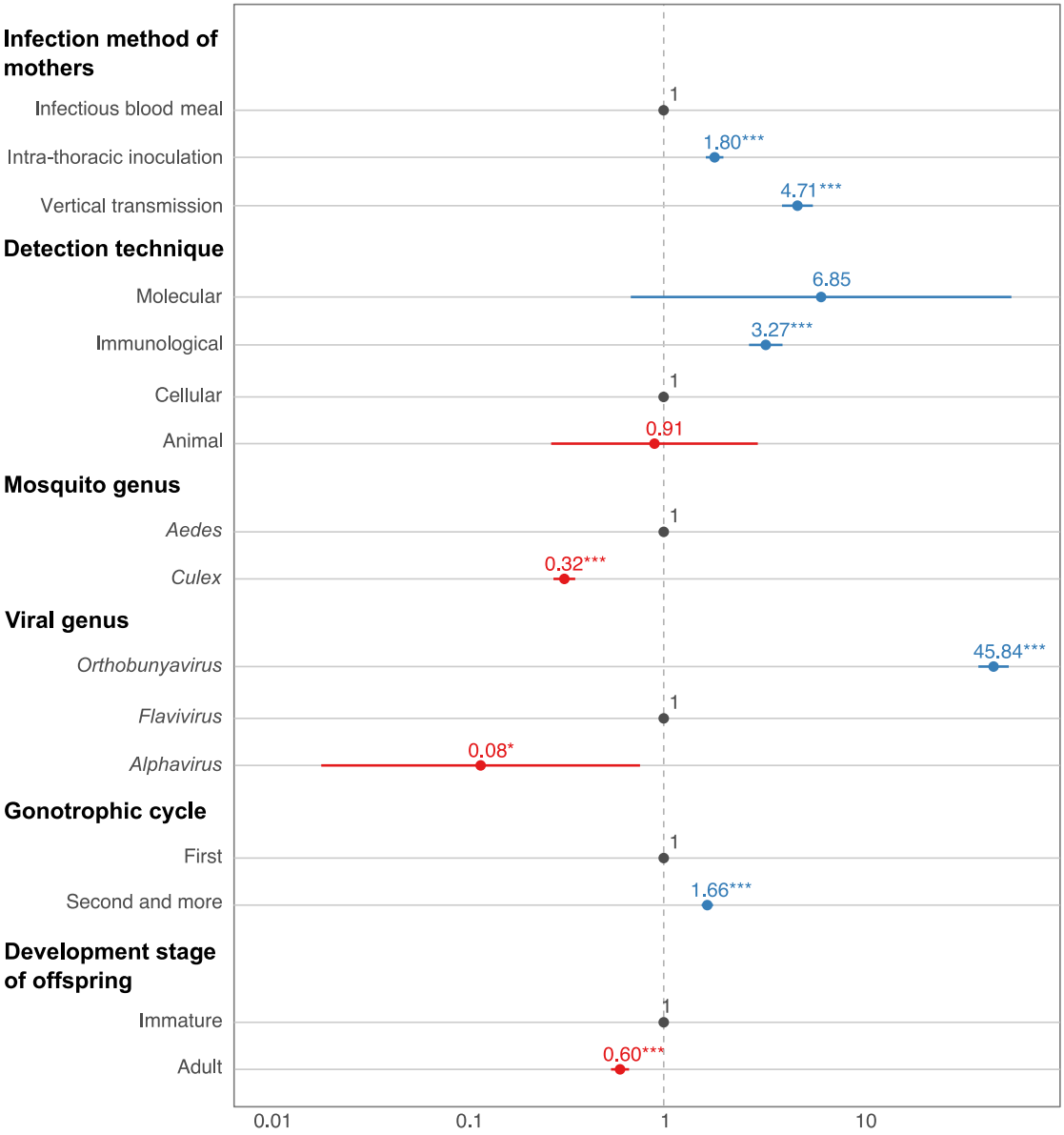

Figure 2 – Significant predictors of arbovirus eVT rates in experimental studies. Odds ratios (ORs) and their 95% confidence intervals are shown on a log10 scale for statistically significant factors. ORs were calculated from a marginal logistic regression based on a generalized linear mixed model that included the random effect of the study and fixed effects of other covariates. Reference level is shown in grey, positive effects are shown in blue, and negative effects are shown in red. Stars represent statistical significance levels: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Because the systematic review process to assemble our databases may have missed some publications, we conducted a "resilience" test to evaluate the impact of missing primary data on the outcome of the analyses. We examined the effect of in-

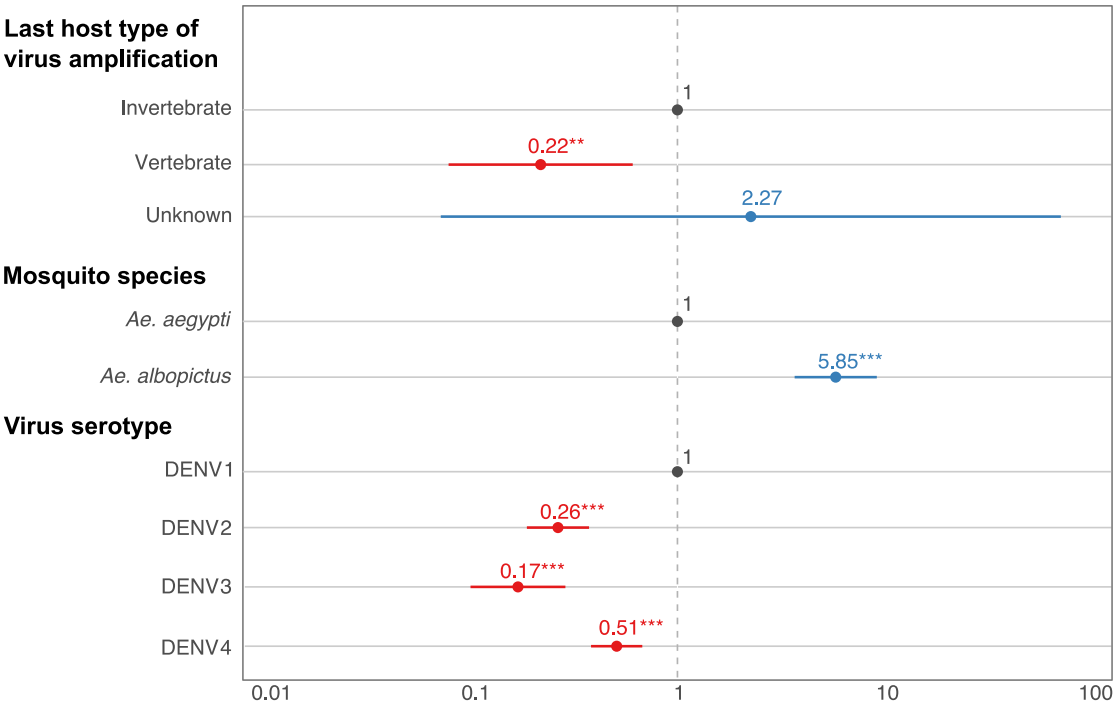

Figure 3 – Odds ratios (ORs) and their 95% confidence intervals are shown on a log10 scale for statistically significant factors. ORs were calculated from a marginal logistic regression based on a generalized linear mixed model that included the random effect of the study and fixed effects of other covariates. Reference level is shown in grey, positive effects are shown in blue, and negative effects are shown in red. Stars represent statistical significance levels: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

cluding four publications that we subsequently identified as missing from the original databases [Kramer et al., 1992, Kramer et al., 1993a, Kramer et al., 1993b, Kramer et al., 1998]. Running the same statistical analyses with the four additional publications did not significantly alter the ORs, their CIs, or associated p-values (Table S3). Likewise, it did not change the conclusions from the separate analysis that focused on the two *Aedes* subgenera and the two viral genera, *Orthobunyavirus* and *Flavivirus* ($p = 0.25$). According to this resilience test, a few additional publications do not lead to meaningful differences in our conclusions, owing to the large size of the databases.

4 Vertical Transmission of Dengue Viruses

DENV1-4 were the most represented viruses in database # 1, which allowed more in-depth analyses of the two main DENV vectors worldwide, *Ae. aegypti* and *Ae. albopictus*. DENV isolates amplified in vertebrate cells (*in vivo* or *in vitro*) resulted in

lower eVT rates than virus isolates amplified in invertebrate cells (OR = 0.21, $p < 0.001$, 95% CI [0.11–0.38]) (Figure 3). Significantly higher eVT rates were observed in Ae. albopictus compared to Ae. aegypti mosquitoes (OR = 5.65, $p < 0.001$, 95% CI [3.57–8.93]). DENV2, DENV3, and DENV4 had significantly lower eVT rates than DENV1 (OR = 0.28, $p < 0.001$, 95% CI [0.20–0.40]; OR = 0.18, $p < 0.001$, 95% CI [0.11–0.31]; and OR = 0.55, $p < 0.001$, 95% CI [0.42–0.73], respectively). DENV4 had significantly higher eVT rates than DENV2 and DENV3 (OR = 0.51, $p < 0.01$, 95% CI [0.34–0.79]; OR = 0.33, $p < 0.001$, 95% CI [0.18–0.58]).

5 Climate and Vertical Transmission

Climatic information (main climate, precipitation, and temperature) about each study site was obtained from published data [Rubel and Kottek, 2010] using the Köppen-Geiger climate classification. Study sites from database # 2 were geo-localized with Google Earth [Google, 2009] and layered with climatic data for the 1976–2000 period (available online http://koeppen-geiger.vu-wien.ac.at/shifts.htm), which corresponded to the majority of publications.

Most of the natural studies were conducted in Asia, North America, and South America, whereas only a few studies were conducted in Europe and Africa. Figure 4 shows the geographical distribution of VT studies in natural populations overlaid with the main climatic regions defined according to the Köppen-Geiger classification. Overall, statistical power to detect differences was low in database # 2 due to exceedingly small eVT rates. Moreover, data structure was highly clustered because of multicollinearity between climate classification and several other variables. For instance, viruses in the *Orthobunyavirus* genus were exclusively associated with warm temperate and continental climate types. Therefore, statistical analyses were performed on selected subsets of data to minimize confounding effects. A subset of the database focusing on arbovirus VT in *Culex* mosquitoes did not reveal any significant predictor. Another subset focusing on arbovirus VT in *Aedes* mosquitoes showed higher eVT rates in mosquito–virus pairs found under arid climate (OR = 28.40, $p < 0.001$, 95% CI [5.52–146.25]) and lower eVT rates under warm temperate climate (OR = 0.13, $p < 0.01$, 95% CI [0.03–0.58]) with equatorial climate as the reference (Figure 5). Note that statistically significant associations between climate type and eVT rates represent overall trends and many counterexamples exist. A final subset focusing

on *Flavivirus* VT found a significantly lower eVT rate in adults compared to when immature stages were tested (OR = 0.60, $p < 0.05$, 95% CI [0.40–0.91]).

6 Discussion

We reviewed the abundant published literature on arbovirus VT in mosquitoes to identify determinants of VT rates in a case study of pathogens with mixed-mode transmission. We used the largest contemporary databases of natural and experimental VT occurrence records to identify several biological factors associated with arbovirus VT in mosquitoes. Our previous study, in which we assembled the databases, focused on historical and technological aspects of arbovirus VT research [Lequime and Lambrechts, 2014]. In particular, we found that the probability of VT detection changed with the evolution of virological assays, with enhanced eVT detection associated with increased assay sensitivity and larger sample sizes [Lequime and Lambrechts, 2014]. In the present review, we used the databases to examine a broader array of biological factors associated with arbovirus VT in mosquitoes. Our analyses provide new insights into the biology of arboviruses and, more generally, into the epidemiology of pathogens with mixed-mode transmission.

Our analysis found striking differences in eVT rates across virus and mosquito taxa. Overall, *Aedes* mosquitoes displayed higher eVT rates than *Culex* mosquitoes. *Aedes* eggs are generally more resistant to desiccation than *Culex* eggs [Clements, 2000], which may confer a selective advantage to vertically transmitted viruses. In addition, *Aedes* mosquitoes had higher eVT rates in nature under arid climatic conditions compared to equatorial or warm temperate climatic conditions. This supports the hypothesis that VT could be a maintenance mechanism during the adverse season. In agreement with earlier observations [Tesh, 1984, Leake, 1984, Turell, 1988], orthobunyaviruses such as LACV and CEV were vertically transmitted better than flaviviruses and alphaviruses. Orthobunyaviruses are assumed to achieve higher VT rates because they are vertically transmitted by TOT, a more efficient VT mechanism than trans-egg VT of flaviviruses [Rosen, 1988]. It is noteworthy that the *Flavivirus* genus includes several insect-specific flaviviruses that are primarily maintained by VT in nature [Blitvich and Firth, 2015]. Unlike dual-host flaviviruses [Rosen, 1988], insect-specific flaviviruses display high rates of VT that are assumed to result from TOT [Saiyasombat et al., 2011], suggesting that different VT mechanisms may have

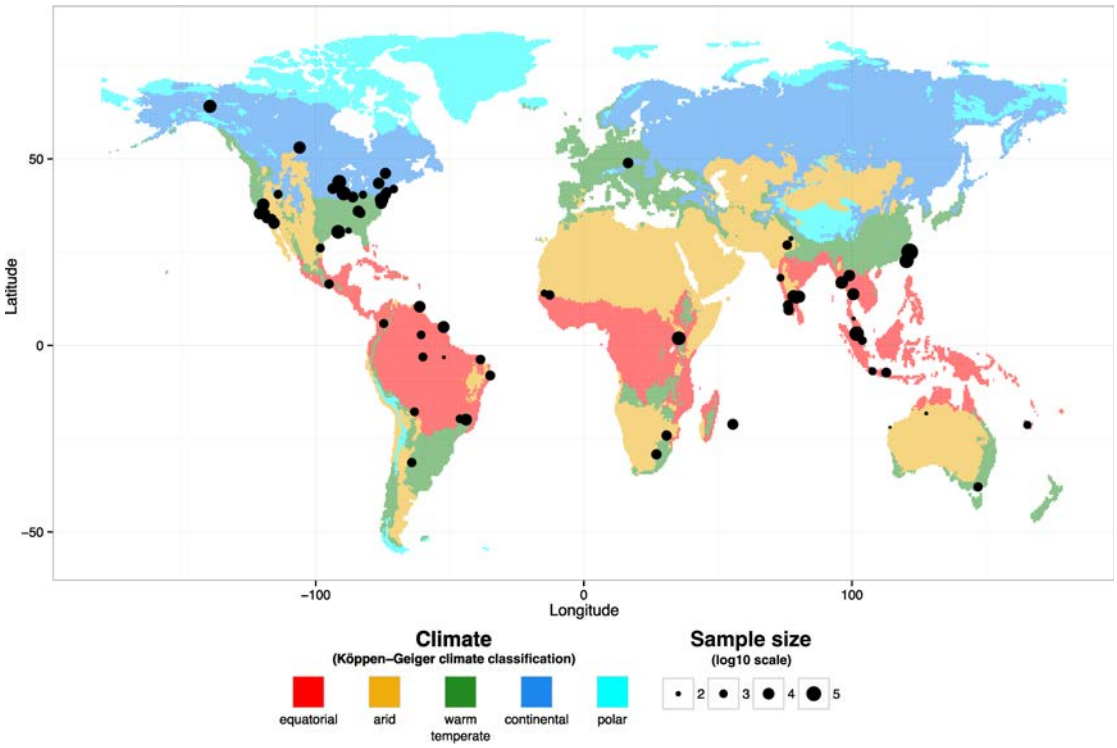

Figure 4 – Geographical distribution of main climate types and study sites for natural studies of arbovirus VT in mosquitoes. Climatic data were obtained from Rubel and Kottek (2010) for the 1975–2000 period. Dots represent geographic location of study sites. Dot size is proportional to the log10-transformed number of mosquitoes tested in the corresponding study.

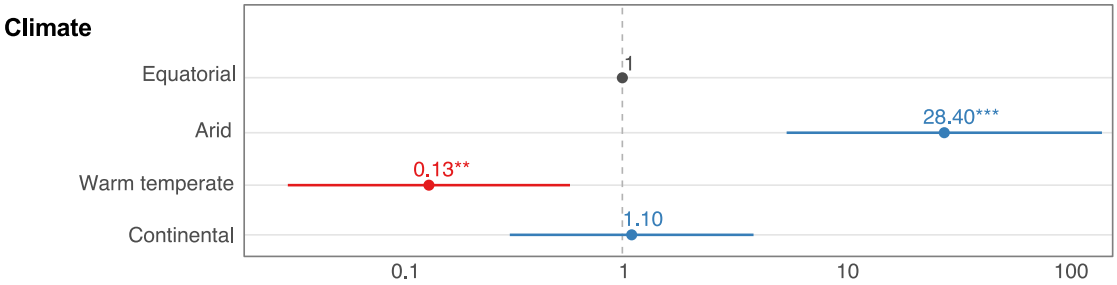

Figure 5 – Significant predictors of arbovirus eVT rates in Aedes mosquitoes in natural studies. Odds ratios (ORs) and their 95% confidence intervals are shown on a log10 scale for statistically significant factors. ORs were calculated from a marginal logistic regression based on a generalized linear mixed model that included the random effect of the study and fixed effects of other covariates. Reference level is shown in grey, positive effects are shown in blue, and negative effects are shown in red. Stars represent statistical significance levels: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

evolved in the same viral genus. TOT relies on viral infection of developing oocytes, which results in close to 100% infection in the subsequent generations [Tesh, 1984]. Such "stabilized" infections have been described for several members of the Orthobunyavirus genus such as CEV in *Ae. dorsalis* [Turell et al., 1982] and San Angelo virus in *Ae. albopictus* [Tesh and Shroyer, 1980]. Note that even when TOT rates are close to 100%, prevalence in the vector population is expected to remain low unless the virus confers an evolutionary advantage to the infected subpopulation. TOT may not be required for trans-generational persistence of arboviral infection in the vector, as patterns consistent with stabilized infections were also reported for the flaviviruses DENV1 in *Ae. albopictus* [Shroyer, 1990] and DENV3 in *Ae. aegypti* [Joshi et al., 2002]. A fourfold increase in DENV3 eVT rate indicated that selection could be involved [Joshi et al., 2002]. Interestingly, we found that prior DENV amplification in invertebrate host cells enhanced subsequent eVT, supporting the idea that stabilized infections require adaptation. Because systemic dissemination of the virus acquired by HT during blood feeding is a prerequisite for VT, stabilized infections are also more likely to be established during later gonotrophic cycles. VT efficiency could be underestimated in studies that focus on the first gonotrophic cycle and therefore do not account for stabilized infections [Tesh, 1984]. When a stabilized infection is established, however, VT is expected to occur as soon as the first gonotrophic cycle.

Our analyses support the idea that arbovirus VT efficiency largely depends on the interplay between gonotrophic cycle and viral infection dynamics. Offspring produced during the second or later gonotrophic cycles displayed higher eVT rates than offspring produced during the first gonotrophic cycle following the infectious blood meal. This can be attributed to the lack of virus dissemination to the ovaries before the first batch of eggs is produced and laid, and/or increased permeability to virus of the ovaries during oogenesis [Agarwal et al., 2014]. In addition, the experimental method to infect females was a significant predictor of eVT rates in their progeny. IT inoculations, which bypass the natural infection route through the gut, resulted in higher eVT rates than oral infections. This is presumably because IT-inoculated mothers more quickly develop a systemic viral infection and reach higher viral titers than orally infected females [Smith et al., 2007]. Vertically infected mothers also had higher eVT rates, which could be a consequence of a stabilized infection of the germ line [Tesh, 1984]. Surprisingly, eVT rates measured in immature developmental stages were higher than

in the corresponding adults, although the underlying mechanism is unclear. *Aedes aegypti* vertically infected with YFV [Beaty et al., 1980], Kunjin virus, and JEV [Tesh, 1980] have delayed development and vertically infected larvae may also suffer lower survival, hence leading to lower infection prevalence in adults.

Some of our findings have important implications for arbovirus–mosquito pairs of public health significance. For instance, DENV had significantly higher eVT rates in *Ae. albopictus* than in *Ae. aegypti* mosquitoes. Although this has been previously suggested in the literature [Bosio et al., 1992, Rosen et al., 1983], reports have been conflicting [Lee et al., 1997]. This finding raises important concerns for potential maintenance of arboviruses in areas where *Ae. albopictus* has recently expanded geographically, including temperate zones where VT would enable the virus to overwinter [Kraemer et al., 2015]. There is currently a critical lack of field and laboratory studies that have examined DENV VT in European *Ae. albopictus* populations. Our analysis also revealed that eVT rates in *Aedes* mosquitoes varied considerably among DENV serotypes, as was recently hypothesized [Grunnill and Boots, 2015]. Specifically, DENV1 was significantly better transmitted vertically than the other DENV serotypes. Although there is no obvious explanation for this phenomenon, accounting for such a serotype-specific feature could improve dengue epidemiological models. Such genetic variation could fuel adaptive evolution of VT under a scenario where vectors move into more seasonally hostile areas.

Evolution and maintenance of VT as part of a mixed-mode transmission strategy is expected to be under strongest selective pressure when the horizontal route is limiting. For arboviruses, this will occur when there is periodicity in host (arthropod or vertebrate) abundance, whether due to seasonal climate forcing, herd immunity, or boom-bust cycles of population densities. Imposed seasonality in vector population densities will inevitably select for VT irrespective of the number of competent vector species and thus VT would be expected to be under increasing selection along a latitudinal gradient. Depending on the extent of population genetic structure of vector species and the distribution of the virus, an intra-specific latitudinal cline of VT frequency might occur, and thus VT would not necessarily be a fixed characteristic of a virus–vector couple. Likewise, the lack of sufficient susceptible hosts may be alleviated when viruses have multiple vertebrate host species. Thus, VT might be expected less likely to occur in generalists than specialists. Finally, the selective advantage of a mixed

mode of transmission will be limited by the duration of infection and life expectancy of the vertebrates and vectors. Maintaining a reservoir of infection in one or the other is the key, and thus evolution towards persistent infection in the vertebrate host would offer a viable alternative to the risky VT route where egg mortality due to abiotic factors will likely be high. The complete gamut of possibilities may not be available for all pathogens because of the specific physiologies of their hosts; thus, identifying the ecological factors selecting for VT may be best approached within a phylogenetic framework. VT is expected to favour co-divergence of hosts and pathogens [Jackson and Charleston, 2004]. Co-phylogenetic relationships have been suggested between *Aedes* mosquitoes of the *Ochlerotatus* subgenus and orthobunyaviruses in North America [Eldridge, 1990]. However, this was not supported by our analyses, which failed to find a statistically significant effect of the *Orthobunyavirus–Ochlerotatus* pairings on experimental eVT rates.

In conclusion, our review uncovered a variety of environmental (e.g., climate), taxonomic (e.g., viral genus), and physiological (e.g., gonotrophic cycle) predictors of arbovirus eVT rate. The influence of experimental factors such as the infection route or the mosquito developmental stage tested calls for caution in interpreting results generated from different experimental designs. Our results emphasize the fact that arbovirus VT efficiency is a dynamic process that may vary within and between mosquito generations. Further studies are needed to determine whether permanent germ line infection in so-called stabilized infections could contribute to long-term arbovirus maintenance in a vector subpopulation [Tesh, 1984, Turell et al., 1982]. Ultimately, this knowledge could help to prevent VT and reduce the risk of arbovirus transmission. More generally, our study provides empirically testable hypotheses to investigate the biological processes underlying the relative frequency of each transmission mode for pathogens with mixed-mode transmission.

7 Methods

Databases # 1 and # 2 are provided as available upon request (in supplementary tables S4 and S5 in original manuscript). All quantitative analyses were performed in the statistical environment R, version 3.2.0 (http://www.r-project.org/), using the following packages: plyr [Wickham, 2011], ggplot2 [Wickham, 2009], stringr [Wickham, 2015], gridExtra [Auguie, 2015], car [46], lme4 [47], multcomp [Hothorn et al., 2008],

sjPlot [Lüdecke, 2015], and arm [Gelman and Su, 2015].

We computed the mean eVT rate for the most represented arboviruses in each database based on eVT rate estimates for each entry in the database weighted by its sample size. Prevalence is usually close to 100% in experimental studies and, therefore, eVT rate is a reasonable proxy for VT rate. Variation in eVT rate estimates was analysed using generalized linear mixed models (GLMMs). Publication was included as random-effect variable in the GLMMs to account for the collective effect of several potentially confounding variables (experimenter, laboratory conditions, etc.). All other variables were considered as fixed-effect variables. For each entry in the database, the numbers of vertically infected and uninfected individuals were calculated from the eVT rate estimate and the sample size. These numbers were analysed as a contingency table by fitting a model with a binomial error distribution, a logit link function, and a bobyqa optimizer with a maximum of 100,000 function evaluations.

Multicollinearity between model variables was evaluated using the variance inflation factor (VIF) and the condition number (CN). A VIF < 4 and a CN < 30 were interpreted as a low level of multicollinearity [O'Brien, 2007, Dormann et al., 2013]. Validity of the models was checked using quantile-quantile (Q-Q) plots of random effects and binned residuals average against expected values plots. The Ω_0^2 measure of explained variation for mixed-effects models [Xu, 2003] was calculated as a simple goodness-of-fit metric. Following model validation, statistically significant effects ($p < 0.05$) were determined by analysis of deviance and type II Wald χ^2 statistics. Statistically insignificant variables were removed from the model in a stepwise fashion, repeating model validation at each step. The minimum adequate model (MAM) was obtained when all variables had a statistically significant effect ($p < 0.05$) according to the type II Wald χ^2 test. Odds ratios (ORs) were calculated based on estimated regression coefficients of the MAM. OR confidence intervals and p-values were computed using estimated regression coefficients and their standard errors.

Acknowledgments

The authors thank all members of the Lambrechts lab for insightful discussions and two anonymous reviewers whose comments helped improve an earlier version of the manuscript.

References

[Adams and Boots, 2010] Adams, B. and Boots, M. (2010). How important is vertical transmission in mosquitoes for the persistence of dengue? Insights from a mathematical model. *Epidemics*, 2(1):1–10.

[Agarwal et al., 2014] Agarwal, A., Dash, P. K., Singh, A. K., Sharma, S., Gopalan, N., Rao, P. V. L., Parida, M. M., and Reiter, P. (2014). Evidence of experimental vertical transmission of emerging novel ECSA genotype of Chikungunya Virus in Aedes aegypti. *PLoS neglected tropical diseases*, 8(7):e2990.

[Auguie, 2015] Auguie, B. (2015). Miscellaneous Functions for "Grid" Graphics.

[Beaty et al., 1980] Beaty, B. J., Tesh, R. B., and Aitken, T. H. (1980). Transovarial transmission of yellow fever virus in Stegomyia mosquitoes. *American Journal of Tropical Medicine and Hygiene*, 29(1):125–132.

[Blitvich and Firth, 2015] Blitvich, B. J. and Firth, A. E. (2015). Insect-specific flaviviruses: a systematic review of their discovery, host range, mode of transmission, superinfection exclusion potential and genomic organization. *Viruses*, 7(4):1927–1959.

[Bosio et al., 1992] Bosio, C. F., Thomas, R. E., Grimstad, P. R., and Rai, K. S. (1992). Variation in the efficiency of vertical transmission of dengue-1 virus by strains of Aedes albopictus (Diptera: Culicidae). *Journal of medical entomology*, 29(6):985–989.

[Cao-Lormeau et al., 2011] Cao-Lormeau, V.-M., Roche, C., Aubry, M., Teissier, A., Lastere, S., Daudens, E., Mallet, H.-P., Musso, D., and Aaskov, J. (2011). Recent Emergence of Dengue Virus Serotype 4 in French Polynesia Results from Multiple Introductions from Other South Pacific Islands. *Plos One*, 6(12).

[Centers for Disease Control and Prevention, 2010] Centers for Disease Control and Prevention (2010). Arbovirus Catalog.

[Clements, 2000] Clements, A. N. (2000). The egg shell. In *The Biology of Mosquitoes*, pages 63–73. CABI.

[Clements, 2012] Clements, A. N. (2012). *The Biology of Mosquitoes*, volume 3 of *Transmissions of viruses and interactions with bacteria*. CABI.

[Diaz et al., 2012] Diaz, L. A., Flores, F. S., Quaglia, A., and Contigiani, M. S. (2012). Intertwined arbovirus transmission activity: reassessing the transmission cycle paradigm. *Frontiers in physiology*, 3:493.

[Dormann et al., 2013] Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Garcia Marquez, J. R., Gruber, B., Lafourcade, B., Leitão, P. J., Muenkemueller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., and Lautenbach, S. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1):27–46.

[Dupont-Rouzeyrol et al., 2014] Dupont-Rouzeyrol, M., Aubry, M., O'Connor, O., Roche, C., Gourinat, A.-C., Guigon, A., Pyke, A., Grangeon, J.-P., Nilles, E., Chanteau, S., Aaskov, J., and Cao-Lormeau, V.-M. (2014). Epidemiological and molecular features of dengue virus type-1 in New Caledonia, South Pacific, 2001-2013. *Virology journal*, 11(1):61.

[Ebert, 2013] Ebert, D. (2013). The Epidemiology and Evolution of Symbionts with Mixed-Mode Transmission. *Annual Review of Ecology, Evolution, and Systematics*, 44(1):623–643.

[Eldridge, 1990] Eldridge, B. F. (1990). Evolutionary relationships among California serogroup viruses (Bunyaviridae) and Aedes mosquitoes (Diptera: Culicidae). *Journal of medical entomology*, 27(5):738–749.

[Fine, 1975] Fine, P. E. (1975). Vectors and vertical transmission: an epidemiologic perspective. *Annals of the New York Academy of Sciences*, 266:173–194.

[Gelman and Su, 2015] Gelman, A. and Su, Y.-S. (2015). Data Analysis Using Regression and Multilevel/Hierarchical Models.

[Google, 2009] Google (2009). Google Earth.

[Grunnill and Boots, 2015] Grunnill, M. and Boots, M. (2015). How Important is Vertical Transmission of Dengue Viruses by Mosquitoes (Diptera: Culicidae)? *Journal of medical entomology*, page tjv168.

[Hothorn et al., 2008] Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, 50(3):346–363.

[Jackson and Charleston, 2004] Jackson, A. P. and Charleston, M. A. (2004). A cophylogenetic perspective of RNA-virus evolution. *Molecular biology and evolution*, 21(1):45–57.

[Joshi et al., 2002] Joshi, V., Mourya, D. T., and Sharma, R. C. (2002). Persistence of dengue-3 virus through transovarial transmission passage in successive generations of Aedes aegypti mosquitoes. *American Journal of Tropical Medicine and Hygiene*, 67(2):158–161.

[Kraemer et al., 2015] Kraemer, M. U. G., Sinka, M. E., Duda, K. A., Mylne, A. Q. N., Shearer, F. M., Barker, C. M., Moore, C. G., Carvalho, R. G., Coelho, G. E., Van Bortel, W., Hendrickx, G., Schaffner, F., Elyazar, I. R. F., Teng, H.-J., Brady, O. J., Messina, J. P., Pigott, D. M., Scott, T. W., Smith, D. L., Wint, G. R. W., Golding, N., and Hay, S. I. (2015). The global distribution of the arbovirus vectors Aedes aegypti and Ae. albopictus. *eLife*, 4:e08347.

[Kramer et al., 1993a] Kramer, L. D., Bowen, M. D., Hardy, J. L., Reeves, W. C., Presser, S. B., and Eldridge, B. F. (1993a). Vector competence of alpine, Central Valley, and coastal mosquitoes (Diptera: Culicidae) from California for Jamestown Canyon virus. *Journal of medical entomology*, 30(2):398–406.

[Kramer et al., 1993b] Kramer, L. D., Hardy, J. L., Reeves, W. C., Presser, S. B., Bowen, M. D., and Eldridge, B. F. (1993b). Vector competence of selected mosquito species (Diptera: Culicidae) for California strains of Northway virus (Bunyaviridae: Bunyavirus). *Journal of medical entomology*, 30(3):607–613.

[Kramer et al., 1992] Kramer, L. D., Reeves, W. C., Hardy, J. L., Presser, S. B., Eldridge, B. F., and Bowen, M. D. (1992). Vector competence of California mosquitoes for California encephalitis and California encephalitis-like viruses. *American Journal of Tropical Medicine and Hygiene*, 47(5):562–573.

[Kramer et al., 1998] Kramer, L. D., Reisen, W. K., and Chiles, R. E. (1998). Vector competence of Aedes dorsalis (Diptera: Culicidae) from Morro Bay, California, for western equine encephalomyelitis virus. *Journal of medical entomology*, 35(6):1020–1024.

[Leake, 1984] Leake, C. J. (1984). Transovarial Transmission of Arboviruses by Mosquitoes. In

Vectors in virus biology, pages 63–91. Academic Press, New-York & London.

[Lee et al., 1997] Lee, H. L., Mustafakamal, I., and Rohani, A. (1997). Does transovarial transmission of dengue virus occur in Malaysian Aedes aegypti and Aedes albopictus? *The Southeast Asian journal of tropical medicine and public health*, 28(1):230–232.

[Lequime and Lambrechts, 2014] Lequime, S. and Lambrechts, L. (2014). Vertical transmission of arboviruses in mosquitoes: a historical perspective. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 28:681–690.

[Lipsitch et al., 1996] Lipsitch, M., Siller, S., and Nowak, M. A. (1996). The evolution of virulence in pathogens with vertical and horizontal transmission. *Evolution*, 50(5):1729–1741.

[Lüdecke, 2015] Lüdecke, D. (2015). sjPlot: Data Visualization for Statistics in Social Science.

[Mink, 1993] Mink, G. I. (1993). Pollen-Transmitted and Seed-Transmitted Viruses and Viroids. *Annual Review of Phytopathology*, 31:375–402.

[O'Brien, 2007] O'Brien, R. M. (2007). A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Quality & Quantity*, 41(5):673–690.

[Reeves, 2004] Reeves, W. C. (2004). Overwintering of arboviruses. In *Epidemiology and Control of Mosquito-borne Arboviruses in California*, pages 357–382. Calif Mosq Vector Control Assoc, Sacramento, California.

[Rosen, 1988] Rosen, L. (1988). Further observations on the mechanism of vertical transmission of flaviviruses by Aedes mosquitoes. *American Journal of Tropical Medicine and Hygiene*, 39(1):123–126.

[Rosen et al., 1983] Rosen, L., Shroyer, D. A., Tesh, R. B., Freier, J. E., and Lien, J. C. (1983). Transovarial Transmission of Dengue Viruses by Mosquitos: Aedes albopictus and Aedes aegypti. *American Journal of Tropical Medicine and Hygiene*, 32(5):1108–1119.

[Rubel and Kottek, 2010] Rubel, F. and Kottek, M. (2010). Observed and projected climate shifts 1901-2100 depicted by world maps of the Koppen-Geiger climate classification. *Meteorologische Zeitschrift*, 19(2):135–141.

[Saiyasombat et al., 2011] Saiyasombat, R., Bolling, B. G., Brault, A. C., Bartholomay, L. C., and Blitvich, B. J. (2011). Evidence of efficient transovarial transmission of Culex flavivirus by Culex pipiens (Diptera: Culicidae). *Journal of medical entomology*, 48(5):1031–1038.

[Shroyer, 1990] Shroyer, D. A. (1990). Vertical Maintenance of Dengue-1 Virus in Sequential Generations of Aedes-Albopictus. *Journal of the American Mosquito Control Association*, 6(2):312–314.

[Smith et al., 2007] Smith, D. R., Arrigo, N. C., Leal, G., Muehlberger, L. E., and Weaver, S. C. (2007). Infection and dissemination of Venezuelan equine encephalitis virus in the epidemic mosquito vector, Aedes taeniorhynchus. *American Journal of Tropical Medicine and Hygiene*, 77(1):176–187.

[Tesh, 1980] Tesh, R. B. (1980). Experimental studies on the transovarial transmission of Kunjin and San Angelo viruses in mosquitoes. *American Journal of Tropical Medicine and Hygiene*, 29(4):657–666.

[Tesh, 1984] Tesh, R. B. (1984). Transovarial Transmission of Arboviruses in Their Invertebrate Vectors. In *Current topics in vector research*, pages 57–76. New-York, NY: Praeger.

[Tesh and Shroyer, 1980] Tesh, R. B. and Shroyer, D. A. (1980). The mechanism of arbovirus transovarial transmission in mosquitoes: San Angelo virus in Aedes albopictus. *American Journal of Tropical Medicine and Hygiene*, 29(6):1394.

[Turell, 1988] Turell, M. J. (1988). Horizontal and vertical transmission of viruses by insect and tick vectors. In *The Arboviruses*, pages 127–152. CRC PressI Llc, Boca Raton, Florida.

[Turell et al., 1982] Turell, M. J., Hardy, J. L., and Reeves, W. C. (1982). Stabilized Infection of California Encephalitis Virus in Aedes dorsalis, and Its Implications for Viral Maintenance in Nature. *American Journal of Tropical Medicine and Hygiene*, 31(6):1252–1259.

[Wickham, 2009] Wickham, H. (2009). *ggplot2*. Elegant Graphics for Data Analysis. Springer Science & Business Media, New York, NY.

[Wickham, 2011] Wickham, H. (2011). The Split-Apply-Combine Strategy for Data Analysis. *Journal of Statistical Software*, 40(1):1.

[Wickham, 2015] Wickham, H. (2015). Simple, Consistent Wrappers for Common String Operations.

[Wilkerson et al., 2015] Wilkerson, R. C., Linton, Y.-M., Fonseca, D. M., Schultz, T. R., Price, D. C., and Strickman, D. A. (2015). Making Mosquito Taxonomy Useful: A Stable Classification of Tribe Aedini that Balances Utility with Current Knowledge of Evolutionary Relationships. *Plos One*, 10(7):e0133602.

[Xu, 2003] Xu, R. (2003). Measuring explained variation in linear mixed effects models. *Statistics in medicine*, 22(22):3527–3541.

Supporting information

Table S1 – Full list of publications included in databases # 1 and # 2. Bold font indicates studies that tested mosquitoes individually for VT.

First author and Year	Title	Journal or book	Study type
Ahmad 1997	Detection of dengue virus from field Aedes aegypti and Aedes albopictus adults and larvae	Southeast Asian J Trop Med Pub Health	Natural
Aitken 1979	Transovarial transmission of yellow fever virus by mosquitoes (Aedes aegypti)	Am J Trop Med Hyg	Experimental
Akbar 2008	PCR detection of dengue transovarial transmissibility in Aedes aegypti in Bandung, Indonesia.	Proc ASEAN Congr Trop Med Parasitol	Natural
Anderson 2006a	West Nile virus from female and male mosquitoes (Diptera: Culicidae) in subterranean, ground, and canopy habitats in Connecticut	J Med Entomol	Natural
Anderson 2006b	Importance of vertical and horizontal transmission of west nile virus by culex pipiens in the Northeastern United States	J Infect Dis	Natural
Anderson 2008	Extrinsic incubation periods for horinzontal and vertical transmission of West Nile Virus by Culex pipiens pipiens (Diptera: Culicidae)	J Med Entomol	Experimental
Anderson 2012	Horizontal and vertical transmission of West Nile virus genotype NY99 by Culex salinarius and genotypes NY99 and WN02 by Culex tarsalis	Am J Trop Med Hyg	Experimental
Andreadis 2010	Studies on hibernating populations of Culex pipiens from a West Nile virus endemic focus in New York City: parity rates and isolation of West Nile virus	J Am Mosq Control Asso	Natural
Andrews 1977	Isolation of trivittatus virus from larvae and adults reared from field-collected larvae of Aedes trivittatus (Diptera: Culicidae)	J Med Entomol	Natural

Follows on the next page...

Angel 2008b	Distribution and seasonality of vertically transmitted dengue viruses in Aedes mosquitoes in arid and semi-arid areas of Rajasthan, India	J Vector Borne Dis	Natural
Arunachalam 2002	Vertical transmission of Japanese encephalitis virus in Mansonia species, in an epidemic-prone area of southern India	Ann Trop Med Parasitol	Natural
Arunachalam 2008	Natural vertical transmission of dengue viruses by Aedes aegypti in Chennai, Tamil Nadu, India	Indian J Med Res	Natural
Bailey 1978	Isolation of St. Louis encephalitis virus from overwintering Culex pipiens mosquitoes	Science	Natural
Balfour 1975	Isolates of California encephalitis (LaCrosse) virus from field collected eggs and larvae of Aedes triseriatus: identification of the overwintering site of California encephalitis	J Infect Dis	Natural
Baqar 1993	Vertical transmission of West Nile virus by Culex and Aedes species mosquitoes	Am J Trop Med Hyg	Experimental
Bardos 1975	Isolation of Tahyna virus from field collected Culiseta annulata (Schrk.) larvae	Acta Virol	Natural
Bardos 1978	Virological examination of mosquito larvae from southern Moravia	Folia Parasitol	Natural
Beaty 1975	Emergence of La Crosse virus from endemic foci	Am J Trop Med Hyg	Natural
Beaty 1980	Transovarial transmission of yellow fever virus in Stegomyia mosquitoes	Am J Trop Med Hyg	Experimental
Bellini 2012	Impact of Chikungunya virus on Aedes albopictus females and possibility of vertical transmission using the actors of the 2007 outbreak in Italy	PLoS One	Experimental
Belloncik 1982	Activity of California encephalitis group viruses in Entrelacs (province of Quebec, Canada)	Can J Microbiol	Natural
Berry 1974	Isolation of LaCrosse virus (California encephalitis group) from field collected Aedes triseriatus (Say) larvae in Ohio (Diptera: Culicidae)	Mosq News	Natural
Berry 1977	Evidence for transovarial transmission of Jamestown canyon virus, in Ohio	Mosq News	Natural
Bina 2008	Natural vertical transmission of dengue virus in peak summer collections of Aedes aegypti (Diptera: Culicidae) from Urban Areas of Jaipur (Rajasthan) and Delhi	J Commun Dis	Natural
Bosio 1992	**Variation in the efficiency of vertical transmission of dengue-1 virus by strains of Aedes albopictus (Diptera: Culicidae).**	**J Med Entomol**	**Experimental**
Broom 1995	Two possible mechanisms for survival and initiation of Murray Valley encephalitis virus activity in the Kimberley region of western Australia	Am J Trop Med Hyg	Natural
Bugbee 2004	The discovery of West Nile virus in overwintering Culex pipiens (Diptera: Culicidae) mosquitoes in Lehigh County, Pennsylvania	J Am Mosq Control Asso	Natural
Campbell 1991	Isolation of Jamestown Canyon virus from boreal Aedes mosquitoes from the Sierra Nevada of California	Am J Trop Med Hyg	Natural
Cecilio 2009	Natural vertical transmission by Stegomyia albopicta as dengue vector in Brazil	Braz J Biol	Natural
Chamberlain 1957	The North American arthropod-borne encephalitis viruses in Culex tarsalis Coquillett.	Am J Hyg	Experimental
Chamberlain 1959	St. Louis encephalitis virus in mosquitoes	Am J Hyg	Experimental
Chamberlain 1964	Studies on transovarial transmission of St. Louis encephalitis virus by Culex quiquefasciatus Say.	Am J Hyg	Experimental
Chen 1990	A study on transovarial transmission of dengue type I virus in Aedes aegypti	Chinese J Microbiol Immunol	Experimental
Chen 2010	Screening of dengue virus in field-caught Aedes aegypti and Aedes albopictus (Diptera: Culicidae) by one step SYBR Green-based reverse transcriptase-polymerase chain reaction assay during 2004-2007 in Southern Taiwan	Vector Borne Zoonotic Dis	Natural
Christensen 1978	**Laboratory studies of transovarial transmission of trivittatus virus by Aedes trivitattus.**	**Am J Trop Med Hyg**	**Experimental**
Clark 1982	Lacrosse virus activity in Illinois detected by ovitraps	Mosq News	Natural
Clark 1983	Persistence of La Crosse virus (California encephalitis serogroup) in north-central Illinois	Am J Trop Med Hyg	Natural
Clark 1985	Absence of eastern equine encephalitis (EEE) virus in immature Coquillettidia perturbans associated with equine cases of EEE	J Am Mosq Control Asso	Natural
Corner 1980	Cache Valley virus: experimental infection in Culiseta inornata.	Can J Microbiol	Experimental
Cornet 1979	Une poussée épizootique de Fièvre jaune selvatique au Sénégal oriental. Isolement du virus de lots de Moustiques adultes m,les et femelles	Med Mal Infect	Natural

Follows on the next page...

Cornet 1984	Dengue 2 au Sénégal oriental : une poussée épizootique en milieu selvatique ; isolements du virus à partir de moustiques et d'un singe et considérations épidémiologiques	Cah ORSTOM, sér Ent méd Parasitol	Natural
Crane 1977	Transovarial transmission of California encephalitis virus in the mosquito Aedes dorsalis at Blue Lake, Utah	Mosq News	Natural
Danielova 1979	Laboratory demonstration of transovarial transmission of Tahyna virus in Aedes vexans and the role of this mechanism in overwintering of this arbovirus	Folia Parasitol	Experimental
Davies 1954a	Observations on the biology of West Nile virus, with special reference to its behavior in the mosquito Aedes aegypti	Ann Trop Med Parasitol	Experimental
Davis 1930	The location of yellow fever virus in infected mosquitoes and the possibility of hereditary transmission	Am J Epidemiol	Experimental
de Castro 2004	Dengue virus detection using reverse transcription-polymerase chain reaction in salive and progeny of experimentally infected Aedes albopictus from Brazil	Mem Inst Oswaldo Cruz	Experimental
de Souza 1991	Vertical transmission of dengue 1 virus by Haemagogus equinus mosquitoes	J Am Mosq Control Asso	Experimental
Delatte 2008	in Aedes albopictus, vecteur des virus du chikungunya et de la dengue à la Réunion : biologie et contrôle	Parasite	Natural
Dhanda 1989	Japanese encephalitis virus infection in mosquitoes reared from field-collected immatures and in wild-caught males	Am J Trop Med Hyg	Natural
Dhileepan 1996	Evidence of Vertical Transmission of Ross River and Sindbis Viruses (Togaviridae: Alpha virus) by Mosquitoes (Diptera: Culicidae) in Southeastern Australia	J Med Entomol	Natural
Diallo 2000	Vertical transmission of the yellow fever virus by Aedes aegypti (Diptera, Culicidae): dynamics of infection in F1 adult progeny of orally infected females	Am J Trop Med Hyg	Experimental
Dohm 2002	Experimental vertical transmission of West Nile virus by Culex pipiens (Diptera: Culicidae)	J Med Entomol	Experimental
Dutary 1981	Transovarial transmission of yellow fever virus by a sylvatic vector, Haemagogus equinus.	Trsn Roy Soc Trop Med Hyg	Experimental
Eastwood 2011	West Nile virus vector competency of Culex quiquefasciatus mosquitoes in the Galapagos Islands	Am J Trop Med Hyg	Experimental
Farajollahi 2005	Detection of West Nile viral RNA from an overwintering pool of Culex pipiens pipiens (Diptera: Culicidae) in New Jersey, 2003	J Med Entomol	Natural
Fauran 1990	Etude sur la transmission verticale des virus de la dengue dans le Pacifique Sud	Bull Soc Path Ex	Natural
Flores 2010	Vertical transmission of St. Louis encephalitis virus in Culex quiquefasciatus (Diptera:Culicidae) in Cordoba, Argentina	Vector Borne Zoonotic Dis	Both
Fontenille 1997	First evidence of natural vertical transmission of yellow fever virus in Aedes aegypti, its epidemic vector	Trsn Roy Soc Trop Med Hyg	Natural
Fontenille 1998	La transmission verticale du virus amaril et ses consÈquences	International Seminar on Yellow Fever in Africa	Natural
Fouque 1996	Aedes aegypti en Guyane française : quelques aspects de l'histoire, de l'écologie générale et de la transmission verticale des virus de la dengue	Bull Soc Path Ex	Natural
Fouque 2004	Epidemiological and entomological surveillance of the co-circulation of DEN-1, DEN-2 and DEN-4 viruses in French Guiana	Trop Med Int Health	Natural
Francy 1981	Transovarial transmission of St. Louis encephalitis virus by Culex pipiens complex mosquitoes	Am J Trop Med Hyg	Experimental
Freier 1984	Oral and transovarial transmission of La Crosse virus by Aedes atropalpus	Am J Trop Med Hyg	Experimental
Freier 1987	Vertical transmission of dengue virus by mosquitoes of the Aedes scutellaris complex mosquitoes.	Am J Trop Med Hyg	Experimental
Freier 1988	Vertical transmission of dengue viruses by Aedes mediovitattus	Am J Trop Med Hyg	Experimental
Fulhorst 1994	Natural vertical transmission of western equine encephalomyelitis virus in mosquitoes	Science	Natural
Gargan 1988	Panveld oviposition sites of floodwater Aedes mosquitoes and attempts to detect transovarial transmission of Rift Valley fever virus in South Africa	Med Vet Entomol	Natural
Gillett 1950	Experiments to test the possibility of transovarial transmission of yellow fever virus in the mosquito Aedes (Stegomyia) africanus Theobald	Ann Trop Med Parasitol	Experimental
Goddard 2003	Vertical transmission of West Nile virus by three California Culex (Diptera: Culicidae) species	J Med Entomol	Experimental
Gokhale 2001	Vertical transmission of dengue-2 through Aedes albopictus mosquitoes	J Commun Dis	Experimental

Follows on the next page...

Gottfried 2002	Temporal abundance, parity, surivaval rates and arbovirus isolation of field-collected container-inhabiting mosquitoes in eastern Tennessee	J Am Mosq Control Asso	Natural
Graham 1999	Selection of refractory and persmissive strains of Aedes triseriatus (Diptera: Culicidae) for transovarial transmission of La Crosse virus	J Med Entomol	Experimental
Guedes 2010	Patient-based dengue virus surveillance in Aedes aegypti from Recife, Brazil	J Vector Borne Dis	Natural
Günther 2007	Evidence of vertical transmission of dengue virus in two endemic localities in the state of Oaxaca, Mexico.	Intervirol	Natural
Hardy 1980	Effect of rearing temperature on transovarial transmission of St. Louis encephalitis virus in mosquitoes.	Am J Trop Med Hyg	Experimental
Hardy 1984	Experimental transovarial transmission of St. Louis encephalitis virus by Culex and Aedes mosquitoes	Am J Trop Med Hyg	Both
Hartanti 2010	Dengue virus transovarial transmission by Aedes aegypti	Univ Med	Natural
Hayes 1962	Detection of eastern encephalitis virus and antibody in wild and domestic birds in Massachusett	Am J Hyg	Natural
Hughes 2006	Comparative potential of Aedes triseriatus, Aedes albopictus, and Aedes aegypti (Diptera: Culicidae) to transovarially transmit La Crosse virus.	J Med Entomol	Experimental
Hull 1984	Natural transovarial transmission of dengue 4 virus in Aedes aegypti in Trinidad	Am J Trop Med Hyg	Natural
Hutamai 2007	A survey of dengue viral infection in Aedes aegypti and Aedes albopictus from re-epidemic areas in the North of Thailand using nucleic acid sequence based amplification assay.	Southeast Asian J Trop Med Pub Health	Natural
Ibanez-Bernal 1997	First record in America of Aedes albopictus naturally infected with dengue virus during the 1995 outbreak at Reynosa, Mexico	Med Vet Entomol	Natural
Ilkal 1991	Entomological investigations during outbreaks of dengue fever in certain villages in Maharashra state	Indian J Med Res	Natural
Joshi 1996	**Transovarial transmission of dengue 3 virus by Aedes aegypti**	**Trsn Roy Soc Trop Med Hyg**	**Both**
Joshi 2002	**Persistence of Dengue-3 virus through transovarial transmission passage in successive generations of Aedes aegypti mosquitoes**	**Am J Trop Med Hyg**	**Experimental**
Jousset 1981	Geographic Aedes aegypti strains and dengue-2 virus: susceptibility, ability to transmit to vertebrate and transovarial transmission	Ann Virol (Inst Pasteur)	Experimental
Jupp 1981	Laboratory vector studies on six mosquito and on tick species with chikungunya virus.	Trsn Roy Soc Trop Med Hyg	Experimental
Jupp 1990	Ae. Furcifer and other mosquitoes as vectors of Chikungunya virus at Mica, northeastern Transvaal, South Africa	J Am Mosq Control Asso	Natural
Kappus 1982	La Crosse virus infection and disease in Western North Carolina	Am J Trop Med Hyg	Natural
Kay 1980	**Transovarial transmission of Murray Valley encephalitis virus by Aedes aegypti (L).**	**Aust J Exp Biol Med Sci**	**Experimental**
Kay 1982	**Three modes of transmission of Ross River virus vy Aedes vigilax (Skuse).**	**Aust J Exp Biol Med Sci**	**Experimental**
Khin 1983	Transovarial transmission of dengue 2 virus by Aedes aegypti in nature	Am J Trop Med Hyg	Natural
Kow 2001	**Detection of dengue viruses in field caught male Aedes aegypti and Aedes albopictus (Diptera: Culicidae) in Singapore by type-specific PCR**	**J Med Entomol**	**Natural**
Kumari 2013	First indigenous transmission of Japanese Encephalitis in urban areas of National Capital Territory of Delhi, India	Trop Med Int Health	Natural
Labuda 1983	Experimental model of transovarial transmission of Tahyna virus in Aedes aegypti	Acta Virol	Experimental
Le Goff 2011	Natural vertical transmission of dengue viruses by Aedes aegypti in Bolivia	Parasite	Natural
LeDuc 1975	Ecology of California encephalitis viruses on the Del Mar Va Peninsula. II. Demonstration of transovarial transmission	Am J Trop Med Hyg	Natural
Lee 1997	Does transovarial transmission of dengue virus occur in Malaysian Aedes aegypti and Aedes albopictus?	Southeast Asian J Trop Med Pub Health	Experimental
Lee 2005	Transovarial transmission of Dengue virus in Aedes aegypti and Aedes albopictus in relation to dengue outbreak in an Urban area in Malaysia	Dengue Bull	Natural
Lindsay 1993	Ross River virus isolations from mosquitoes in Arid Regions of Western Australia: Implication of vertical transmission as a means of persistence of the virus	Am J Trop Med Hyg	Natural
Linthicum 1985	Rift Valley fever virus (family Bunyaviridae, genus Phlebovirus). Isolations from Diptera collected during an inter-epizootic period in Kenya	J Hyg	Natural

Follows on the next page...

Lisitza 1977	Prevalence rates of LaCrosse virus (California encephalitis group) in larvae from overwintered eggs of Aedes triseriatus	Mosq News	Natural
McAbee 2008	**Identification of Culex pipiens complex mosquitoes in a Hybrid zone of West Nile virus transmission in Fresno county, California**	**Am J Trop Med Hyg**	**Natural**
McLean 1977	Natural foci of California encephalitis virus activity in the Yukon territory	Can J Public Health	Natural
McLintock 1976	Isolation of snowshoe hare virus from Aedes implicatus larvae in Saskatchewan	Mosq News	Natural
Micieli 2013	Vector competence of Argentine mosquitoes (Diptera: Culicidae) for West Nile virus (Flaviviridae: Flavivirus).	J Med Entomol	Experimental
Miller 1977	**Vertical transmission of La Crosse virus (California encephalitis group): transovarial and filial infection rates in Aedes triseriatus (Diptera: Culicidae)**	**J Med Entomol**	**Experimental**
Miller 1979	Aedes triseriatus and La Crosse virus: lack of infection in eggs of the first ovarian cycle following oral infection of females	Am J Trop Med Hyg	Experimental
Miller 1982	**Variation of La Crosse virus filial infection rates in geographic strains of Aedes triseriatus (Diptera: Culicidae)**	**J Med Entomol**	**Experimental**
Miller 2000	First field evidence for natural vertical transmission of West Nile virus in Culex univittatus complex mosquitoes from Rift Valley province, Kenya	Am J Trop Med Hyg	Natural
Mishra 2001	Transovarial transmission of West Nile virus in Culex vishnui mosquito.	Indian J Med Res	Experimental
Mitamura 1950	Seasonal occurrence of mosquito in Okayama 1946 and infectivity of the mosquito with Japanese B encephalitis virus; trans ovary infection of the virus in mosquito	Jap Med J	Experimental
Mitchell 1990a	Vector competence of Aedes albopictus for a newly recognized Bunyavirus from mosquitoes collected in Potosi, Missouri.	J Am Mosq Control Asso	Experimental
Mitchell 1990b	Vertical transmission of dengue viruses by strains of Aedes albopictus recently introduced into Brazil.	J Am Mosq Control Asso	Experimental
Mondet 2002	Isolation of yellow fever virus from nulliparous Haemagogus janthinomys in Eastern Amazonia	Vector Borne Zoonotic Dis	Natural
Morris 1978	An Evaluation of the Hypothesis of Transovarial transmission of Eastern Equine encephalomyelitis virus by Culiseta melanura	Am J Trop Med Hyg	Natural
Mourya 1987a	Experimental transmission of Chikungunya virus by Aedes vittatus mosquitoes.	Indian J Med Res	Experimental
Mourya 1987b	Absence of transovarial transmission of chikungunya virus in Aedes aegypti & Ae. albopictus mosquitoes.	Indian J Med Res	Experimental
Mourya 2001	Horizontal and vertical transmission of dengue virus type 2 in highly and lowly susceptible strains of Aedes aegypti mosquitoes	Acta Virol	Experimental
Mulyatno 2012	Vertical transmission of dengue virus in Aedes aegypti collected in Surabaya, Indonesia, during 2008-2011	Jpn J Infect Dis	Natural
Muul 1975	Ecological studies of Culiseta melanura (Diptera:Culicidae) in relation to eastern and western equine encephalomyelitis viruses on the eastern shore of Maryland	J Med Entomol	Natural
Nasci 2001	West Nile virus in overwintering Culex mosquitoes, New York City 2000	Emerg Infect Dis	Natural
Nayar 1986	Experimental vertical transmission of Saint Louis encephalitis virus by Florida mosquitoes.	Am J Trop Med Hyg	Experimental
Nelms 2013a	Experimental and natural vertical transmission of West Nile virus by California Culex (Diptera: Culicidae) mosquitoes	J Med Entomol	Experimental
Nelms 2013b	Phenotypic variation among Culex pipiens complex (Diptera: Culicidae) populations from the Sacramento Valley, California: horizontal and vertical transmission of West Nile virus, diapause potential, autogeny, and host selection.	Am J Trop Med Hyg	Experimental
Nir 1963	Failure to obtain experimental transovarian transmission of West Nile virus by Aedes aegypti	Ann Trop Med Parasitol	Experimental
Pantuwatana 1974	Isolation of La Crosse virus from field collected Aedes triseriatus larvae	Am J Trop Med Hyg	Natural
Paulson 1989	**Replication and dissemination of La Crosse virus in the competent vector Aedes triseriatus and the incompetent vector Aedes hendersoni and evidence for transovarial transmission by Aedes hendersoni (Diptera: Culicidae)**	**J Med Entomol**	**Experimental**
Pelz 1990	Vertical transmission of St Louis encephalitis virus to autogenously developed eggs of Aedes atroplapus mosquitoes	J Am Mosq Control Asso	Experimental
Pessoa Martins 2012	Occurrence of Natural Vertical transmission of Dengue-2 and Dengue-3 viruses in Aedes aegypti and Aedes albopictus in Fortaleza, Ceara, Brazil	PLoS One	Natural

Follows on the next page...

Philip 1929	Possibility of hereditary transmission of yellow fever virus by Aedes aegypti (Linn.)	J Exp Med	Experimental
Philipps 2006	Field-caught Culex erythrothorax larvae found naturally infected with West Nile virus in Grand county, utah.	J Am Mosq Control Asso	Natural
Pinger 1983	Isolation of La Crosse and other arboviruses from Indiana mosquitoes	Mosq News	Natural
Pinheiro 2005	Detection of dengue virus serotype 3 by reverse transcription polymerase chain reaction in Aedes aegypti (Diptera: Culicidae) captured in Manaus, Amazonas	Mem Inst Oswaldo Cruz	Natural
Ramalingam 1986	Does transovarial transmission of dengue virus occur in Malaysia	Trop Biomed	Natural
Reese 2010	Identification of super-infected Aedes triseriatus mosquitoes collected as eggs from the field and partial characterization of the infecting La Crosse viruses	Virol J	Natural
Reeves 1946	Laboratory transmission of Japanese B encephalitis virus by seven species (three genera) of North America mosquitoes	J Exp Med	Experimental
Reisen 2006	Overwintering of West Nile virus in Southern California	J Med Entomol	Both
Rohani 2007	Detection of transovarial dengue virus from field-caught Aedes aegypti and Ae albopictus larvae using C6/36 cell culture and reverse transcriptase polymerase chain reaction (RT-PCR) techniques	Dengue Bull	Natural
Rohani 2008	Persistency of transovarial dengue virus in Aedes aegypti (Linn.)	Southeast Asian J Trop Med Pub Health	Experimental
Romero-Vivas 1998	Determination of dengue virus serotypes in individual Aedes aegypti mosquitoes in Colombia	Med Vet Entomol	Natural
Rosen 1978	Transovarial transmission of Japanese encephalitis virus by mosquitoes	Science	Experimental
Rosen 1980	Transovarial transmission of Japanese encephalitis virus by Culex tritaeniorhynchus mosquitoes.	Am J Trop Med Hyg	Experimental
Rosen 1983	Transovarial transmission of dengue virus by mosquitoes: Aedes albopictus and Aedes aegypti	Am J Trop Med Hyg	Experimental
Rosen 1987	Mechanism of vertical transmission of the dengue virus in mosquitoes	C R Acad Sciences	Experimental
Rosen 1988	Further observation on the mechanism of vertical transmission of flaviviruses by Aedes mosquitoes	Am J Trop Med Hyg	Experimental
Rosen 1989a	A longitudinal study of the prevalence of Japanese encephalitis virus in adult and larval Culex tritaeniorhynchus mosquitoes in northern Taiwan	Am J Trop Med Hyg	Natural
Rosen 1989b	Experimental vertical transmision of Japanese encephalitis virus by Culex tritaeniorhynchus and other mosquitoes	Am J Trop Med Hyg	Experimental
Scherer 1986	Vector incompetency: its implication in the disappearance of epizootic Venezuelan equine encephalomyelitis virus from Middle America	J Med Entomol	Experimental
Schopen 1991	Vertical and veneral transmission of California group viruses by Aedes triseriatus and Culiseta inornata mosquitoes	Acta Virol	Experimental
Scott 1990	Susceptibility of Aedes albopictus to infection with eastern equine encephalomyelitis virus	J Am Mosq Control Asso	Experimental
Serufo 1993	Isolation of dengue virus type 1 from larvae from Aedes albopictus in Campos Altos city, state of Minas Gerais, Brazil	Mem Inst Oswaldo Cruz	Natural
Shroyer 1986a	Transovarial maintenance of San Angelo virus in sequential generations of Aedes albopictus	Am J Trop Med Hyg	Experimental
Shroyer 1990	Vertical maintenance of dengue-1 virus in sequential generations of Aedes albopictus	J Am Mosq Control Asso	Experimental
Soman 1985	Transovarial transmission of Japanese encephalitis virus in Culex bitaeniorhynchus mosquitoes	Indian J Med Res	Experimental
Sprance 1981	Experimental evidence against the transovarial transmission of eastern equine encephalitis virus in Culiseta melanura	Mosq News	Both
Stamm 1962	Arbovirus studies in south Alabama, 1957-1958	Am J Hyg	Natural
Stockes 1928	Experimental transmission of yellow fever to laboratory animals	Am J Trop Med Hyg	Experimental
Sudeep 2013	Preliminary findings on Bagaza virus (Flavivirus: Flaviviridae) growth kinetics, transmission potential & transovarial transmission in three species of mosquitoes.	Indian J Med Res	Experimental
Takashima 1989	Horizontal and Vertical transmission of Japanese Encephalitis Virus by Aedes japonicus (Diptera: Culicidae)	J Med Entomol	Experimental
Tesh 1975	**Laboratory studies of transovarial transmission of La Crosse and other arboviruses by Aedes albopictus and Culex fatigans**	**Am J Trop Med Hyg**	**Experimental**
Tesh 1980a	**Experimental studies on the transovarial transmission of Kunjin and San Angelo viruses in mosquitoes**	**Am J Trop Med Hyg**	**Experimental**

Follows on the next page...

Appendix D. Determinants of arbovirus vertical transmission in mosquitoes

Tesh 1980b	**The mechanism of arbovirus transovarial transmission in mosquitoes: San Angelo virus in Aedes albopictus**	**Am J Trop Med Hyg**	**Experimental**
Thavara 2009	Outbreak of chikungunya fever in Thailand and virus detection in field population of vector mosquitoes, Aedes aegypti (L.) and Aedes albopictus Skuse (Diptera: Culicidae).	Southeast Asian J Trop Med Pub Health	Natural
Thenmozhi 2000	Natural vertical transmission of dengue viruses in Aedes aegypti in southern India	Trsn Roy Soc Trop Med Hyg	Natural
Thenmozhi 2006	Long-Term study of Japanese encephalitis virus infection in Anopheles subpictus in Cuddalore district, Tamil Nadu, South India	Trop Med Int Health	Natural
Thenmozhi 2007	Natural vertical transmission of dengue virus in Aedes albopictus (Diptera: Culicidae) in Kerala, a southern Indian state	Jpn J Infect Dis	Natural
Thongrungkiat 2011	Prospective field study of transovarial dengue virus transmission by two different forms of Aedes aegypti in an urban area of Bangkok, Thailand	J Vector Ecol	Natural
Turell 1982a	Transovarial and trans-stadial transmission of California encephalitis virus in Aedes dorsalis and Aedes melanimon	Am J Trop Med Hyg	Experimental
Turell 1982b	Stabilized infection of California encephalitis virus in Aedes dorsalis, and its implications for viral maintenance in nature	Am J Trop Med Hyg	Experimental
Turell 1982c	Evaluation of the efficiency of transovarial transmission of California encephalitis viral strains in Aedes dorsalis and Aedes melanimon	Am J Trop Med Hyg	Experimental
Turell 2001	Vector competence of North American mosquitoes (Diptera: Culicidae) for West Nile virus	J Med Entomol	Experimental
Unlu 2010	Evidence of vertical transmission of West Nile virus in field-collected mosquitoes	J Vector Ecol	Natural
van den Hurk 2003	Vector competence of Australian mosquitoes (Diptera: Culicidae) for Japanese encephalitis virus	J Med Entomol	Experimental
Vazeille 2009	Failure to demonstrate experimental vertical transmission of the epidemic strain of Chikungunya virus in Aedes albopictus from La Réunion Island, Indian Ocean.	Mem Inst Oswaldo Cruz	Experimental
Vilela 2010	Dengue virus 3 genotype I in Aedes aegypti mosquitoes and eggs, Brazil, 2005-2006	Emerg Infect Dis	Natural
Wasinpiyamongko 2003	**Susceptibility and transovarial transmission of dengue virus in Aedes aegypti: a preliminary study of morphological variations.**	**Southeast Asian J Trop Med Pub Health**	**Experimental**
Watts 1973	**Transovarial transmission of La Crosse virus (California encephalitis group) in the mosquito, Aedes triseriatus**	**Science**	**Experimental**
Watts 1974	Overwintering of La Crosse virus in Aedes triseriatus	Am J Trop Med Hyg	Natural
Watts 1985	Failure to detect natural transovarial transmission of dengue viruses by Aedes aegypti and Aedes albopictus (Diptera: Culicidae)	J Med Entomol	Natural
Watts 1987	Ecological evidence against vertical transmission of eastern equine encephalitis virus by mosquitoes (Diptera: Culicidae) on the Delmarva Peninsula, USA	J Med Entomol	Natural
Woodring 1998	**Short report: Diapause, transovarial transmission, and filial infection rates in geographic strains of La Crosse virus-infected Aedes triseriatus**	**Am J Trop Med Hyg**	**Experimental**
Zeidler 2008	Dengue virus in Aedes aegypti larvae and infestation dynamics in Roraima, Brazil	Rev Saude Publica	Natural
Zhang 1993	Transovarial transmission of Chikungunya virus in Aedes albopictus and Aedes aegypti mosquitoes	Chin J Virol	Experimental
Zhang 1996	Transovarial transmission of Dengue viruses in Aedes albopictus and Aedes aegypti mosquitoes	Virol Sinica	Experimental
Zytoon 1993	Transovarial transmission of chikungunya virus by Aedes albopictus mosquitoes ingesting microfilariae of Dirofilaria immitis under laboratory conditions	Microbiol Immunol	Experimental

Factor category	Factors
	Publication ID
Virus identification	Virus specie
	Virus strain
	Virus isolation
Mosquito identification	Mosquito genus
	Mosquito sub-genus
	Mosquito specie
	Mosquito strain
	Mosquito-Virus pair
Passage history of the virus	Vertebrate cell passages
	Number of vertebrate cell passages
	Arthropod cell passages
	Number of arthropod cell passages
	Alternate passage
	VT history
	Host type of the last passage
	Complete passage history
	Infection method of the parental mosquitoes
Was the mosquito infected by a parasite?	Parasite (Yes or No)
	Parasite specie
Rearing conditions	Temperature
	Humidity
	Generation of tested offspring
	Ovarian cycle of tested offspring
	Egg laying day after infection for tested offspring
Detection and identification of VT	Mosquito development stage when test for VT
	Previous cell culture before detection
	Detection assay
	Identification assay
	VT (Yes or No)
	Sample size
	Minimum Infection Rate

(a) Database # 2

Factor category	Factors
	Publication ID
Virus identification	Virus specie
Mosquito identification	Mosquito genus
	Mosquito sub-genus
	Mosquito specie
	Mosquito-Virus pair
Geographic and epidemiological contexte	Study site
	Epidemiological contexte
	Mosquito capture development stage
Detection and identification of VT	Mosquito development stage when test for VT
	Previous cell culture before detection
	Detection assay
	Identification assay
	VT (Yes or No)
	Sample size
	Minimum Infection Rate

(b) Database # 3

Table S2 – Full list of factors for databases # 2 and # 3.

	Before (original database)				After (addition of 4 missing publications)			
	OR	lower CI	upper CI	*p*-value	OR	lower CI	upper CI	*p*-value
Infection method of mothers								
Infectious blood meal	1				1			
Intra-thoracic inoculation	1.80409471908522	1.62906683087477	1.99792770544808	< 2e-16	1.82399774717766	1.6481773798541	2.01857386369644	< 2e-16
Vertical transmission	4.71292367467743	3.94050918348373	5.6367460470421	< 2e-16	4.7239804147911	3.95132847256638	5.64771851144932	< 2e-16
Detection technique								
Molecular	6.8484194056505	0.780046071085632	60.1257414070633	0.082590	8.36856806246735	0.831286952101107	84.2463979966699	0.071367
Immunological	3.27447056459551	2.69706225403153	3.97549498991918	< 2e-16	3.33442041722097	2.74615889568899	4.04869490117052	< 2e-16
Cellular	1				1			
Animal	0.90507964045962	0.276367874385298	2.96405346459506	0.869121	1.00012522928223	0.281046480308544	3.55902152963711	0.999846
Mosquito genus								
Aedes	1				1			
Culex	0.315591189875811	0.278192312231082	0.358017798293789	< 2e-16	0.320006714469033	0.282070645002086	0.363044858157814	< 2e-16
Virus genus								
Orthobunyavirus	45.7088751112907	38.34215405052 93	54.4909725516773	< 2e-16	45.0547568357332	37.8163259995692	53.6786972259063	< 2e-16
Flavivirus	1				1			
Alphavirus	0.078076467686153 9	0.013656070985515	0.446390093666984	0.004148	0.050804973011765 7	0.008071757372621 66	0.3197748846467	0.001500
Gonotrophic cycle								
First	1				1			
Second or more	1.65807806145314	1.55171185513587	1.7717354216073	< 2e-16	1.66601830849933	1.5590614806298	1.78031273220459	< 2e-16
Development stage of offspring								
Immature	1				1			
Adult	0.602880594111 85	0.542920333282554	0.669462881522801	< 2e-16	0.688722946937029	0.625542228961504	0.758285013667426	3.05e-14

Table S3 – Resilience test. Comparison of odds ratios, confidence intervals and *p*-values before and after addition of 4 missing publications.

LEQUIME SEBASTIAN

Interactions flavivirus-moustiques : diversité et transmission

Les infections humaines dues aux virus du genre *Flavivirus* constituent depuis longtemps un problème de santé publique majeur à travers le monde, en particulier dans les zones à climat tropical. Ces virus à ARN sont des arbovirus qui infectent alternativement un hôte vertébré et un arthropode « vecteur », dont majoritairement des moustiques de la sous-famille des Culicinae. D'autres flavivirus, en revanche, sont incapables d'infecter les cellules de vertébrés et sont qualifiés de flavivirus spécifiques d'insectes (FSI). L'interaction entre les vecteurs et les flavivirus est centrale dans leur biologie, par l'influence qu'elle a sur leur diversité génétique, leur évolution et leur transmission. Cependant, après plus d'un siècle de recherches scientifiques, certains points de ces aspects fondamentaux restent méconnus, malgré une abondance accrue de données. Les approches basées sur les « mégadonnées » (*big data*) ont été au cœur du travail de cette thèse, qu'elles aient été générées par des technologies modernes ou par compilation de travaux plus anciens.

Dans une première partie, nous avons exploré des génomes de moustiques anophèles disponibles dans les bases de données publiques, à la recherche de traces d'éléments viraux endogènes (EVEs) d'origine flavivirale. Nous avons réussi à identifier *in silico*, puis à confirmer *in vivo*, la présence d'EVEs proches des FSI exogènes, chez les espèces *Anopheles sinensis* et *An. minimus*. Ces résultats suggèrent l'existence de FSI chez les anophèles, habituellement non associés aux flavivirus, et mettent en lumière la diversité du genre *Flavivirus*, loin d'être restreinte aux seuls arbovirus.

Dans une deuxième partie, nous avons généré et analysé un jeu de données complexe basé sur le séquençage haut-débit d'un flavivirus à l'intérieur de son vecteur. Cette étude nous a permis d'explorer la fine interaction entre le génotype du moustique *Aedes aegypti* et la diversité virale intra-hôte du virus de la dengue-1. En effet, comme tous les virus à ARN, les flavivirus existent sous la forme d'une population de variants génétiques apparentés, considérée comme critique pour leur *fitness* et leur potentiel adaptatif. Nos résultats ont mis en évidence : (i) un fort effet de la dérive génétique liée à un goulot d'étranglement démographique lors de l'infection initiale du tube digestif, diminuant l'importance relative de la sélection naturelle, et (ii) une modulation de la diversité génétique intra-hôte du virus par le génotype du moustique, indiquant que la diversité génétique du virus, et donc sa *fitness* et son évolution, est inextricablement liée à la variation génétique de l'hôte.

Enfin, nous avons compilé de manière systématique la riche littérature disponible sur la transmission verticale des arbovirus chez le vecteur moustique, c'est-à-dire de la femelle infectée à sa descendance, afin d'identifier des facteurs techniques, environnementaux, taxonomiques et physiologiques sous-jacents. Nos résultats, étayés par une analyse statistique robuste, éclairent d'un jour nouveau ce mode de transmission complémentaire à la transmission horizontale, entre le vecteur et l'hôte vertébré. Ils permettent d'affiner notre compréhension des stratégies employées par les arbovirus pour persister dans leur environnement, tout en fournissant des hypothèses testables sur les processus biologiques impliqués.

Collectivement, nos résultats ont mis en évidence de nouveaux aspects de la complexité des relations entre flavivirus et moustiques. Au-delà, ils soulignent l'intérêt d'étudier les « mégadonnées » générées au cours de plus d'un siècle de recherche, qu'elles soient issues de l'accumulation historique d'études descriptives ou produites par les technologies récentes de séquençage haut-débit. Ces analyses inscrivent résolument l'étude du système flavivirus-moustique dans l'ère du *big data*.

MOTS-CLÉS : arbovirus, *Aedes*, *Anopheles*, *Flavivirus* spécifique d'insectes, évolution intra-hôte, transmission verticale