

ABSTRACT

Nonretroviral RNA viruses (NRVs) are viruses with an RNA-based genome that, despite not coding for reverse-transcriptase and integrase, have been found integrated in the genome of several eukaryotes. These viral integrated sequences are called nonretroviral integrated RNA virus sequences, NIRVS.

Currently, the precise molecular mechanisms that explain which and under which circumstances NRVs integrate and become endogenous are poorly characterized. NRVs comprise highly epidemiologically-relevant viruses, including arboviruses (arthropod-borne viruses), which are biologically-transmitted by arthropod vectors, such as mosquitoes, ticks and sand flies.

The Asian tiger mosquito, *Aedes albopictus*, is one of the most important arboviral vectors, being able to transmit 26 arboviruses belonging to the *Flaviviridae* (e.g. Dengue, Zika, Japanese encephalitis and Yellow fever viruses) and *Togaviridae* (e.g. Chikungunya virus) families. We previously identified a total of 72 loci harboring sequences of viral origin in the genome of *Ae. albopictus*. These viral integrations map primarily in piRNA clusters and produce piRNAs, which are antisense to the incoming viral mRNA. These results suggest NIRVS may contribute to mosquito immunity. A molecular validation of the bioinformatically-characterized NIRVS is essential to further study their potential role in mosquito immunity and to estimate their variability across natural populations.

On this basis, in this thesis I combined molecular and bioinformatics approaches to validate the presence and the sequences of the bioinformatically-characterized NIRVS.

For the molecular work, I focused my attention on NIRVS with similarities to viruses of the *Flavivirus* genus due to their high public health relevance. Combining PCRs and Southern blotting, I was able to validate 13 NIRVS. The application of molecular techniques was limited by the repetitive nature of the genomic context where the majority of NIRVS are located. As a consequence, I developed a bioinformatic pipeline to further validate NIRVS. WGS data from mosquitoes of the reference Foshan strain and from strains that were recently derived from 4 geographic locations, including the *Ae. albopictus* native home range and both old and newly colonized regions, were used to test presence/absence of NIRVS based on read coverage and to identify NIRVS variability (i.e. presence of indels and SNPs). This approach proved complementary to the application of molecular techniques.

My results provide the first insights on NIRVS widespread in wild *Ae. albopictus* mosquitoes and NIRVS variability.

Interestingly, NIRVS variability across geographic strains was consistent with the history of *Ae. albopictus* colonization. NIRVS with highest variability were conserved across strains and the Foshan reference strains and included the first-ever described viral integration.

Despite their high polymorphism, which would make NIRVS ideal genetic markers, the observed difficulties in identifying NIRVS using simple and cheap molecular techniques like PCR limits their usefulness in population genetics applications.