UNIVERSITY OF TRENTO

International PhD Program in Biomolecular Sciences

Cycle 33rd

Genome-scaled molecular clock studies of invasive mosquitoes and other organisms of societal relevance

Tutor: Prof. Nicola Segata, CIBIO, University of Trento

Advisor: Prof. Omar Rota-Stabelli, C3A, University of Trento

Co-Advisor: Prof. Annapaola Rizzoli, Fondazione Edmund Mach

PhD Thesis of

Nicola Zadra

Department of Cellular, Computational and Integrative Biology

University of Trento

Academic year 2019-2020

I, Nicola Zadra, confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

TABLE OF CONTENTS

Abstract

Molecular dating (or molecular clock) is a powerful technique that uses the mutation rate of biomolecules to estimate divergence times among organisms. In the last two decades, the theory behind the molecular clock has been intensively developed, and it is now possible to employ sophisticated evolutionary models on genome-scaled datasets in a Bayesian framework. The molecular clock has been successfully applied to virtually all types of organisms and molecules to estimate timing of speciation, timing of gene duplications, and generation times: this knowledge allows contextualizing past and present events in the light of (paleo)ecological scenarios. Molecular clock studies are routinely used in evolutionary and ecological studies, but their use in applied fields such as agricultural and medical entomology is still scarce in particular because of a paucity of genome data. Genome-scaled clocks have been successfully applied, for example, to various model organisms such as *Anopheles* and *Drosophila*, as well as to invasive mosquitoes *Aedes aegypti* and *Aedes albopictus.* Many other invasive pests are emerging worldwide aided by global trade, increased connectivity among countries, lack of prevention, and flawed invasive species management. Among them, there is *Aedes koreicus* and *Aedes japonicus,* two invasive mosquito species which are monitored for public health concerns because of their harboured human pathogenic viruses. For these, as well as for other insects of societal relevance, such as the parasitoid *Trissolcus japonicus*, there is a paucity of gene markers and no genome data for large scale molecular clock studies.

Invasive pests are typically studied using microevolutionary approaches that tackle events at an intraspecific level: these approaches provide important information for the pest management, for example, by revealing invasion routes and insecticide resistances. Approaches that tackle the deep-time evolution of the pest, such as the molecular clock, are instead less used in pest science. Many important traits associated with invasiveness have evolved by speciation over a long time frame: the molecular clock can reveal the paleo-ecological conditions that favoured these traits helping a better understanding of pest biology. Molecular clock, when coupled with phylogenomics, can further identify genes and patterns that characterize the pest: this knowledge can be used to enhance management practices. Although this is a data-driven thesis, its major aim is to

provide new results to demonstrate the utility of the molecular clock in pest science. This has been done by systematically apply the molecular clock to various neglected organisms of medical and agricultural relevance. To this aim, I generated new genome data and/or assembled the largest genome-scaled data to date.

I studied the molecular clock in mosquitoes, focusing on the Aedini radiation (Chapter 2) and identified a strong incongruence between the mitochondrial and nuclear phylogeny for what concerns their molecular clock. This result highlighted the importance of employing genome scaled data for these species to exclude stochastic effects due to poor/inaccurate sampling in clock studies. To tackle the absence of data, I further assembled the whole mitogenome of emerging invasive species *Aedes koreicus* and *Aedes japonicus* with the aim of producing useful data for molecular typing and of inferring divergence estimates using whole mitogenomes (Chapter 3). Dated phylogenies point toward more recent diversification of Aedini and Culicini compared to estimates from previous works, addressing the issue of taxon sampling sensitivity in dated phylogeny. Although it is possible to perform molecular clock studies on single/few gene markers, the current trend is to couple this methodology with genome-scaled datasets to reduce the stochastic effect of using few genes. For this reason, I sequenced the draft genome of *A. koreicus* and *A. japonicus* (Chapter 4). The assemblies were extremely fragmented, highlighting the problem of sequencing large genomes using short reads. The assemblies provided, however enough information for genome skimming allowing extraction of BUSCO genes for downstream analyses, whole mitogenome assemblies (used in Chapter 3), and characterisation of the associated metagenome. These data need to be integrated by long reads; it provides, however a first framework to investigate the genome evolution of these species. I further sequenced and assembled the genome of *Trissolcus japonicus*, the parasitoid wasp of the invasive pest *Halyomorpha halys*. To elucidate its divergence, estimate and define an intraspecific typing system to differentiate strains for biocontrol strategies, I reconstructed the mitochondrial genomes of two populations: the mitogenomes were surprisingly identical, suggesting that they belong to the same *de facto* population. I further provide a detailed clock investigation of Zika, a virus harboured and transmitted by some *Aedes* species (Chapter 5). Using the largest

set of genomes to date, I could set the origin of ZIKV in the middle age and its first diversification in the mid-19th century. From a methodological point of view, the clocking of this virus highlighted the importance of checking for recombination and for cell-passages to obtain correct divergence estimates. I finally show my contributions to molecular clock studies of three other invasive species (Chapter 6): I helped disentangle the divergence times of *Bactrocera*, a genus of invasive fruit files pest of agriculture; I contributed in performing a phylogenomics study of opsin genes in Diptera; I used chloroplast and nuclear genome data to reconstruct the divergences of the invasive reed *Arundo.*

In the various Chapters of my thesis, I highlighted the limits and the problems of current molecular clock methodologies and identified the best practices for different types of organisms in order to develop a cross-discipline understanding of the molecular clock techniques. The various results presented in this thesis further demonstrate the utility of the molecular clock approach in pest studies.

Thesis outline

Chapter 1 – Study invasive species using the molecular clock

This thesis uses the molecular clock and other macroevolutionary approaches to study invasive pests. I first introduce in detail the theoretical and practical aspects of this methodology. I then present a brief introduction to invasive species focusing on *Aedes* mosquitoes, the main model organisms of the thesis, followed by a short description of other model organisms studied: viruses and pests of agriculture. I discuss that macroevolutionary approaches such as clock are neglected in pest science, and I introduce the general aim of this thesis which is to provide new genome data and results to support the hypothesis that macroevolutionary studies are beneficial for invasive pest studies. A first author review article is in preparation on the "Utility of molecular clock in pest science".

Chapter 2 - Chronological Incongruences between Mitochondrial and Nuclear Phylogenies of *Aedes* Mosquitoes

Synopsis: The aim was to explore state of the art for molecular clock of *Aedes* species to test 1) if there is a congruent signal between the two genomes of the cell (nuclear versus mitochondrial); 2) if a PCR based dataset is enough to establish mosquito divergence times. Results revealed discrepancies between nuclear and mitochondrial clock, a methodological issue that has been used to guide the research in Chapter 3. Published first author article (Zadra, Rizzoli, Rota Stabelli, 2021 Life).

Chapter 3 - The mitogenomes of invasive mosquitoes *Aedes koreicus* and *Aedes japonicus* and an updated molecular clock analysis of mosquito mitogenomes

I assembled the whole mitogenome of two Italian invasive species *A. koreicus* and *A. japonicus*. The general aim was to produce data useful for molecular typing in the field for two species of medical relevance and to build a taxon-enriched mitogenomic dataset to generate an updated timetree of mosquito evolution. Results were used to test if some of the chronological discrepancies observed in Chapter 2 are the result of stochastic error. A first author article in preparation.

Chapter 4 - Short-reads genome sequencing of pest and biocontrol species

Part A. New tigers on the block: Genome-skimming of the invasive mosquitoes *Aedes koreicus* and *Aedes japonicus*

I present data on the genome assemblies of *A. koreicus* and *A. japonicus* using short reads. As expected for this type of sequencing and species, the assemblies are very fragmented. The raw data and the assemblies allowed, however, a genome skimming approach for the reconstruction of the microbiota present in the raw data and the extraction of genes for downstream clock and phylogenomic analyses of these species. This data is a useful source of information for demographic and genetic studies of these two invasive species. A first author article is in preparation.

Part B. The short-read genome of two populations of the parasitoid wasp *Trissolcus japonicus* from Italy: male haploidy, metagenomic screening, identical mitogenomes

Synopsis: *Trissolcus japonicus* is a parasitoid wasp used for the classical biological control of the invasive bug *Halyomorpha halys*. Here we used short-read sequencing to generate high-quality genome data from a lab-reared population used for biocontrol studies and from a recently established Trentino population. No differences were detected in the mitochondrial nucleotide sequences suggesting that the mtDNA is not suitable for distinguishing the two populations. These genomes indicate that it is possible to obtain good quality assemblies even from short reads. This genome is a source of data for a variety of evolutionary and applied studies. Sheared first author article in preparation for Biological Control.

Chapter 5 - Extensive phylogenomic analysis of Zika virus provides an updated scenario of its origin and evolution

One peculiar aspect of *Aedes* is that they can harbour and transmit a series of viruses. Here I applied molecular clocks to explore the evolution of a virus typically transmitted by *Aedes* mosquitoes, the Zika virus. By assembling the largest genome scaled dataset so far, I provide an updated picture of the

molecular evolution of these viruses, focusing on recent recombination and the first steps of the Asian colonization of the virus. Results are relevant for understanding the early epidemiological patterns of arboviruses. First author article in preparation.

Chapter 6 - Other studies of species of agricultural importance

Phylogenomics of opsin genes in Diptera reveals lineage-specific events and contrasting evolutionary dynamics in *Anopheles* and *Drosophila*

Role of candidate: manual extraction and curation of mosquito orthologs, construction of phylogenetic trees, figure preparation. Published co-author article (Feuda et al. 2021 GBE).

Phylogenomic proof of Recurrent Demipolyploidization and Evolutionary Stalling of the "Triploid Bridge" in *Arundo* (Poaceae)

Role of candidate: experimental design and analyses of the chloroplast clock and the nuclear clock employing a genomic dataset. Figure preparation. Published co-author article (Jike et al. 2020 IJMS).

The impact of fast radiation on the phylogeny of *Bactrocera* fruit flies

Role of candidate: experimental design and guidance for assemblies, clock analyses, and species tree reconstruction. In review co-author article (Valerio et al.).

CHAPTER 1 - STUDY INVASIVE SPECIES USING THE MOLECULAR CLOCK

1.1 The molecular clock hypothesis

The molecular clock is an idiom used to describe the hypothesis that molecules evolve at a constant pace through time. It is nowadays often used as a synonymous for inferring divergence times from molecules. This idea of a molecular clock started to thrive around the 1960s from the first comparison of protein sequences. Cutting-edge studies in the following decades have pushed the molecular clock theory into the scientific community (Zuckerkandl and Pauling 1965; Doolittle and Blombäck 1964; Margoliash 1963). Emile Zuckerkandl and Linus Pauling were the first to introduce the term 'molecular evolutionary clock' to the scientific community in a seminal paper in 1965 (Zuckerkandl and Pauling 1965). Their new idea arose from comparing the haemoglobin amino acid sequences of many animals: they observed that the amino acid diversity between two different species was positively correlated with the age of their supposed common ancestor inferred from fossil data. The observation led to the linear evolutionary rate assumption and to the possibility of applying this estimate to measure the timespan among homologous haemoglobin sequences (Morgan 1998). Zuckerkandl and Pauling's studies on haemoglobin led to surprisingly accurate estimations, such as the divergence between humans and gorillas set at 11 Ma (million years ago); this result was later confirmed by cytochrome C and fibrinopeptides (Doolittle and Blombäck 1964; Margoliash 1963), eventually reinforcing the power and the applicability of the molecular clock hypothesis. Nevertheless, the molecular clock hypothesis received a cold reception in its initial phase, even from eminent biologists such as Lewontin (Stebbins and Lewontin 1971) and Ernst Mayr (Morgan 1998). Initially, the molecular clock hypothesis was integrated into Kimura's neutral theory of molecular evolution (Motoo Kimura 1968; 1969). The molecular clock was thought to fit Kimura's neutral theory, for which the mutations are neutrally occurring at a nearly constant rate per generation. Many lineages, however, showed a slower or faster substitution rate that correlates mainly with species' generation time (C. I. Wu and Li 1985; Kikuno, Hayashida, and Miyata 1985; Kohne 1970). The species-specific generation time was insufficient to explain the large rate variability displayed in the analyzed protein sequences: the substitution rate variability was much greater than expected, contradicting the assumption of a neutral (strict)

molecular clock (Ohta and Gillespie 1996; Gillespie 1989; Langley and Fitch 1974) (Figure 1).

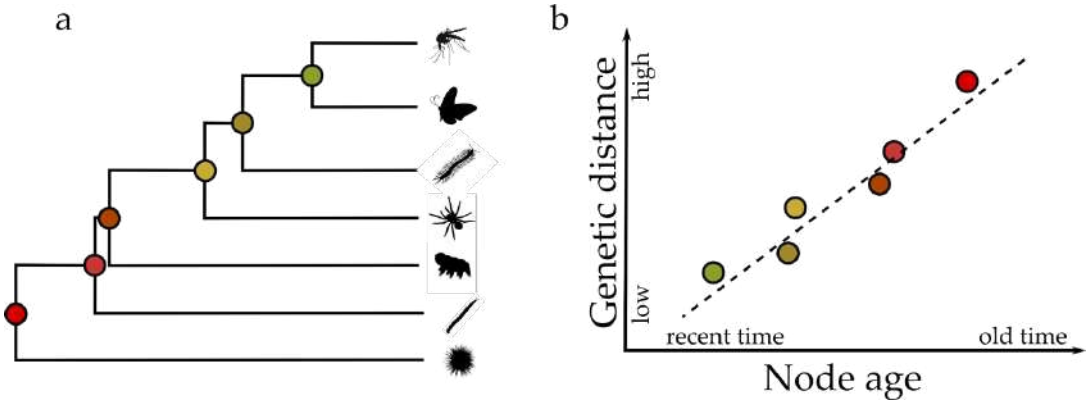


Figure 1. Graphical representation of the molecular clock. Panel *a* represent a reduced schematic phylogeny of animal; colored dots at nodes indicate the point in the phylogeny where the species coalesce. Panel *b* depicts the genetic distances between two species subtending each of the nodes in panel A (y-axis) plotted against the age of that node inferred from paleontological data (x-axis). Genetic distances correlate with the age of the assumed common ancestors; the slope of the linear regression is the rate at which the genetic material is evolving.

Although the molecular clock's rate variation was a serious limitation of the theory, efforts have been made to understand why the clock behaves so erratically and how to mathematically model the rate variation. In the first decade of 2000, it has become clear that the clock is, in general, not strict (identical/similar rates in different lineages) but is rather relaxed (different rates in different lineages): this fundamental advance in molecular dating has led to more sophisticated methods and new statistically robust tools to estimate the phylogenetic timescale. Nowadays, the molecular clock is a useful tool for investigating the rate and timescale of molecular evolution in many biological fields such as epidemiology, ecology, population genetics, genomics, and pest management.

1.1.1 Why and how does the clock rate vary?

Many causes may contribute to the variability of the evolutionary rate along lineages (Figure 2). The answers can be found in the biology of a species. This variation can have biochemical explanations, or it can be a physical constraint. Considerable rate variation among species makes it challenging to provide

reliable estimations in molecular dating, but understanding this pattern of variation offers the chance to enhance the evolutionary clock models. To better understand the clock, we have to clarify the differences between a mutation and substitution, how they occur in the DNA and how they influence the molecular clock.

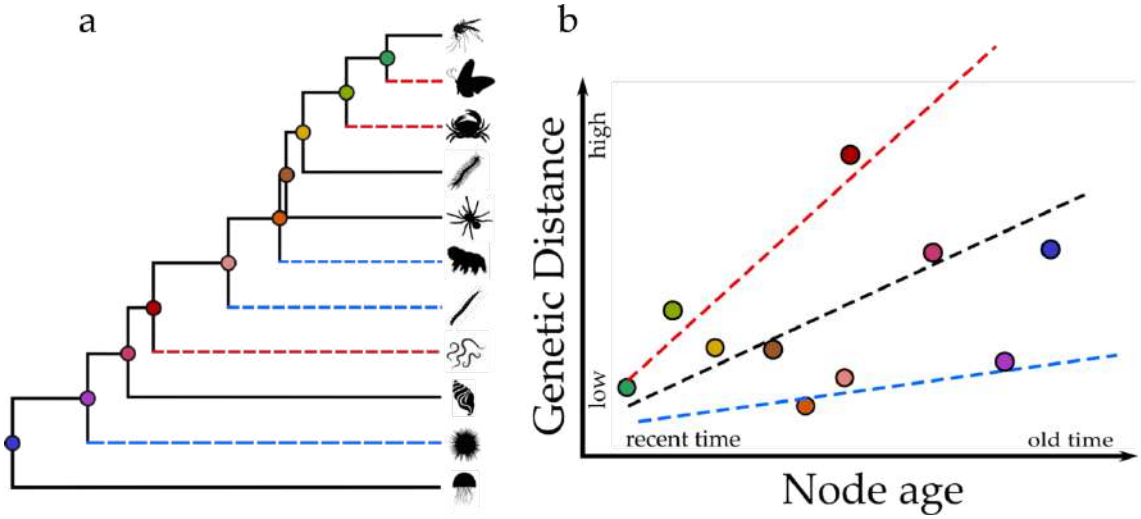


Figure 2. The molecular clock is relaxed. The scientific community has reached a consensus that, in most cases, the evolutionary dynamics do not show a constant rate of evolution among branches. Clades evolve at different rates, with some clades showing a slower rate of evolution (blue dotted line), while other clades are characterised by faster evolutionary rate (red dotted line). In panel *a*, the hypothetical animal phylogeny is represented. In panel *b*, different clades are characterised by different line interpolations and, therefore, different evolutionary rates. There is no universal rate to describe the evolution in every lineage, rate changes along with the biology, ecology, and population dynamics of the organism.

1.1.1.1 Mutation rate

Mutations are changes in DNA sequences. Large scale mutations such as whole gene duplications and chromosomal rearrangements are rare events. Small-scaled mutations such as single base-pair mutations (one base is replaced by another one) are more common and are generally caused by errors during the DNA duplication process. If mutations happen in the germ line and they are viable for the cell and the individual, they may be retained, creating a heritable variation. The mutation rate is an approximation of the number of these mutations that occurs in a certain genomic region (it could be genes, non-coding DNA, or the entire genome) per unit of time. The unit of time can be represented by both DNA replication events (generations) or by years.

The rate of mutation is influenced by several mechanisms; the primary mutation source is a copy error, influenced by the DNA polymerase, and the second is the efficiency of its proofreading activity. The mutation rate can differ among species because of a trade-off of cost and benefits. A more accurate DNA copy leads to increased metabolic costs, while a less accurate copy can increase the deleterious mutation burden (Kunkel 2009). The mutation occurrences during the DNA copying process are mainly a physiochemical phenomenon; however, life-history traits can affect the mutation rate. Biological species can shape the DNA repair efficiency and replication fidelity accordingly to their fitness. This can change throughout populations, niches, and over time according to the fluctuation of the fitness landscape (Lanfear, Kokko, and Eyre-Walker 2014; Sniegowski et al. 2000).

Mutation rate has been estimated using whole-genome sequencing in *D. melanogaster* by counting the average number of mutations passed in a generation between a parent and its offspring and the differences among siblings (Keightley et al. 2009). The instantaneous mutation rate has been calculated for *Heliconius molpomene* butterflies (Keightley et al. 2015). *Drosophila melanogaster* and *H. molpomene* spontaneous mutation rate has shown to be very similar; both are estimated to be ~3×10^{-9} mutation/generation. In the last decades, many mutation rates were estimated and proposed for a vast plethora of organisms, particularly model organisms such as yeast and *Escherichia coli* (Williams 2014; H. Liu and Zhang 2019).

1.1.1.2 Substitution rate

Distinguishing between mutation and substitution rates is crucial to properly understand the molecular clock. Once a mutation occurs, it could be inherited by the following generations or lost throughout generations. The substitutions are the mutations that get fixed in a population. The substitution rate can be defined as the number of new mutations in each generation (Nu) multiplied by the probability of each new mutation reaching fixation (1/N). Even though the substitution rate is correlated with the mutation rate, other mechanisms can contribute to the substitution rate variation. Two major forces contribute to the fixation of mutation in a population: genetic drift and selection. Moreover, these two forces are influenced by the effective population size (N_e), which can be a

game-changing factor that can alter, substantially, the fate of the mutations. In the case of neutral mutations, the mutation and the replacement rate tend to be equal.

1.1.1.3 Life traits affect mutation and substitution rate

There are some biological properties of species that may influence the mutation and/or the substitution rates.

- Generation time. All new mutations in the somatic cells of an individual disappear with the death of the individual, while only mutations that occur in the germline can be passed to the progeny and enter the population. One obvious factor influencing the mutation rate is, therefore, the generation time - the average time between two consecutive generations. For example, an insect with up to 10 generations per year has many more chances of introducing new mutations in the population than a mammal with one or fewer generations per year. The mutation rate has been experimentally correlated with the generation time of the organism (Mooto Kimura 1987). Mutation rates are indeed often expressed as the number of mutations per site per generation. It is, however, common to scale the mutation rate per unit of time, not per generation (Mooers and Harvey 1994; J. A. Thomas et al. 2010). Rates per generation are poorly informative for molecular clock studies because they should be scaled to time to be used as calibrations. Overall, the correlation between generation time and the mutation rate is not completely straightforward, and it can be influenced by other life-history factors.

- Fecundity. Fecundity - *sensu* the number of offspring (descendants) per lifespan - of a species can affect the mutation rate. More offspring means more chances to pass a mutation in the germline into the population. For a high fecundity organism, the DNA copy fidelity is less important than for an organism that produces few offspring (Bromham and Leys 2005). Consequently, high fecundity is correlated with a higher mutation rate. (S. Duffy, Shackelton, and Holmes 2008; Belshaw et al. 2008).

- Body size. Surprisingly, the body size of an animal or a plant is correlated with a slower mutation rate (Lanfear, Kokko, and Eyre-Walker 2014; Gillooly et al. 2005; Barrera-Redondo, Ramírez-Barahona, and Eguiarte 2018; Lanfear et al.

2013). Body size affects the mutation rate for two reasons. First, it is linked with life-history traits such as metabolic rate, population size and generation time. Secondly, the mutation rate is lower in the bigger animal because of the higher number of cells and more genome copies. The mutation load is not controlled only in the germline but even in the somatic cell, and it's connected to the fitness of an individual. More cells mean more chances to develop unviable mutations in a cell line (e.g. cancer) (Nunney 2018; 1999); hence bigger body sizes tend to select more efficient genome replication, reducing the average error per site compared to species with fewer cells and slower cell turnover. For this reason, the mutation rate is not selected only on the germline, but the somatic cells could drive the selection toward a more efficient or less efficient mutation rate. Eventually, the body size selects the mutation rate due to a trade-off between the energetic cost of replication fidelity and the mutation burden tolerated by the organism.

- Selection. Selection can shape the frequencies of variants (from new mutations) within a population, eventually promoting the fixation of advantageous variants. Mutations are heritable, so the mutation that provides an advantage in the offspring has a higher chance to get selected and get fixed in the population. For the same reason, deleterious mutations are prone to be erased from the population. Things are often not that straightforward balancing selection can favour the existence of more than one variant (allele) in a population. The interactions among environment, mutation genotype, and phenotype define if a mutation is neutral, advantageous, or deleterious (selective coefficient). The species are in a dynamic equilibrium with their environment; hence, the selective coefficient of a mutation can change through time, space and niches. For this reason, clock studies should be ideally conducted on neutrally evolving sites (Lanfear, Kokko, and Eyre-Walker 2014).

- Genetic drift and population size. Population size may be the most crucial factor shaping the substitution patterns because it can deeply affect genetic drift and selection (Charlesworth 2009; Lynch et al. 2016). The population size becomes important in light of the nearly neutral theory exposed by Otha and Kimura (Motoo Kimura and Ohta 1971), stating that a slightly advantageous or slightly deleterious mutation can behave as a neutral mutation, and they have the same chance of getting lost or fixed as a neutral mutation. In the nearly neutral

scenario, the population size determines how many mutations escape the selection and are affected by drift. In a small population, a slightly deleterious mutation could be easily get fixed into a population by drift, whereas a slightly advantageous mutation could get lost for the same reason (Figure 3). The population size can therefore strongly modulate the effect of the selection and drift; usually, a small population fixes more mutations than a large population, resulting in an higher substitution rate in the small populations (Ohta and Gillespie 1996). In large populations, purifying selection is more efficient to remove even the slightly deleterious mutations (T. Ohta 1973).

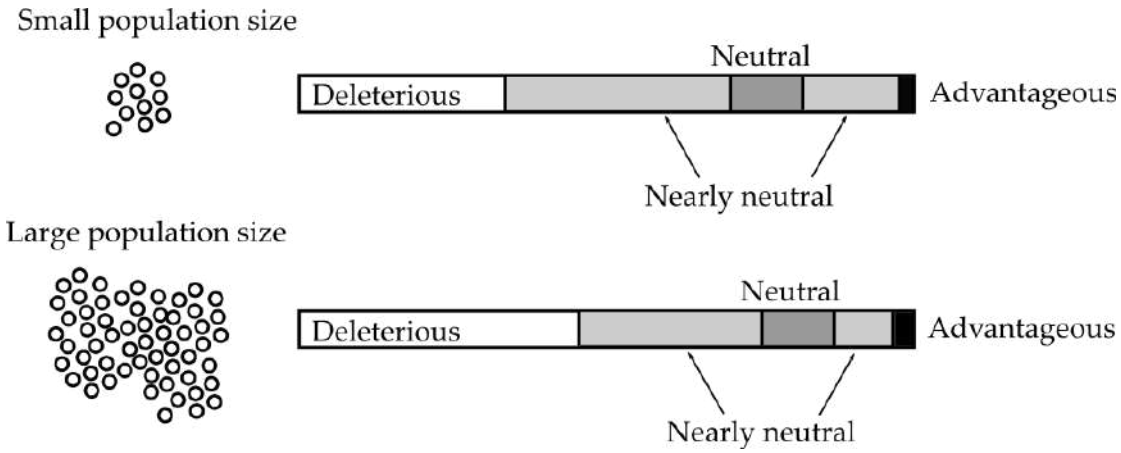


Figure 3. Nearly neutral theory and the effect of population size on shaping the selection coefficient between two identical mutation ranges (modified from Otha 2002). The bars displayed in the figure represent the same ideal set of mutations. We can appreciate how the proportion of neutral and nearly neutral mutation changes depending on the population size. Smaller population sizes do not allow the nearly deleterious or nearly advantageous to be seen by natural selection. On the other head, larger population size allows the selection to act even on mutations that show a nearly neglectable effect on the fitness of the individuals.

In the last decades, we have clarified most of the causes behind among-lineages variation of the molecular clock; despite this knowledge, reconstructing the evolutionary history of the species is still challenging. This is because most of the forces outlined above work at the population level, whereas the molecular clock is generally applied to species phylogeny without taking the population into account. In addition to the phenomena described above, other events can affect the rate of evolution, such as bottlenecks, the founder effect, and genetic hitchhiking. In general, when estimating divergence times, we approximate the rate of substitution per unit of time on branches by calibrating and modelling it in a Bayesian framework. (Bromham et al. 2018; S. Y. W. Ho et al. 2011). In the

next section, I will explain how a clock analysis can be modelled and how we can estimate the divergence time between species, focusing on Bayesian analyses. I will further show that estimating divergence time is challenging for various other parametric reasons, in particular, model misspecification.

1.1.2 Bayesian inference for molecular dating

In the last two decades, the phylogenetics reconstruction relied mainly on two statistical inferences, Maximum likelihood (ML) and Bayesian. Both are suitable for phylogenetic analysis and are widely applied in phylogenetic investigations such as molecular dating, phylogenomics, and population genetics. The maximum likelihood approach was initially theorized for estimating phylogenies from genetic data (Cavalli-Sforza and Edwards 1967), and it was made more applicable to sequence alignments by Felsenstein in 1981 (Felsenstein 1981). ML expresses the probability of the data given the model; the model refers to the set of parameters to search for the best tree that fits the data. Thus, more precisely, the ML phylogeny is the probability of the data given the model and the tree. The ML approach has been widely used for inferring phylogeny, but it has some computational limitations that are unsuitable for many big datasets and analyses with many parameters, such as dating a phylogeny. Moreover, ML inference has issues with missing data and partitioned datasets (Simmons 2012), which pose a limitation to applying this method in many datasets. Partitioned datasets can be easier described to capture the actual phylogenetic dynamic; they better explore the relations among species and individuals because different genes, loci and positions can describe different evolutionary histories.

Bayesian, unlike ML, expresses the probability of the tree and the model given the data. In this view, the Bayesian approach may fit better the purpose of a phylogenetic framework. Bayesian deals better than ML with the uncertainty associated with priors and estimated parameters such as rate, branch length, and population size. The parameters in a Bayesian inference framework are easily expressed as statistical distributions, whereas in a ML framework, they are treated as fixed (although estimated) constants. In a Bayesian framework, a specific distribution (the prior, see below) is assigned to each of the free parameters that describe the model. During the analysis, all priors are combined

with the likelihood of the data, thus generating the posterior distribution (Z. Yang and Rannala 2012). In contrast to ML, Bayesian does not provide a single optimal result but summarizes what has been observed in exploring the tree space, providing a more comprehensive view of the variability of the trees with the highest posterior probability values.

Bayesian inference was introduced in the 1990s for phylogenetics (Rannala and Yang 1996). The first Bayesian software to infer phylogeny was released in 1999 by Larget in the package BAMBE (Larget and Simon 1999). Since then, the Bayesian tools face increased in popularity in the phylogenetic field, providing several programs to the phylogenies that keep updating to perform better the new theoretical needs of the scientific community and keep up with the times of the post-genomic era.

The Bayesian approach strongly relies on an approximation algorithm called Markov Chain Monte Carlo (MCMC). MCMC takes samples walking through the tree and the parameters space, calculating the respective posterior probability of the hypothesis. After the analysis has run enough to collect an amount of tree states that approximates the posterior distribution, it is necessary to summarize all the results. The result is summarized from a subsample of steps made by the MCMC, and the sampled results will be in proportion to the highest posterior probability regions.

Bayesian inference provides the probability of the tree and the model given the data. The inference combines the prior information and the likelihood to obtain the posterior probability of the tree.

$$P(model|data) = \frac{P(data|model)P(model)}{P(data)}$$

P(model) = prior

P(data|model) = likelihood

P(data) = data probability (model evidence)

P(model|data) = posterior

The prior is the probability density of a parameter which is combined with the likelihood to calculate the posterior probability of the model. The prior is a range of values/parameters in which we can sample; it is the "a priori" knowledge that conditions the posterior probability. It is possible to make prior assumptions about how the molecules/traits have evolved. The priors depict the belief of the parameters involved in the phylogenetic reconstruction, ignoring the information that comes from the data, so estimating a prior distribution does not depend on the data. In a phylogenetic framework in a Bayesian space, the hyperpriors need to be mentioned; the hyperprior is the prior information that describes a hyperparameter, a parameter that influences a prior distribution. Estimating a prior distribution can be hierarchical because a prior may require other priors for being estimated.

1.1.3 The priors: models, parameters, calibrations

1.1.3.1 Substitution model

Like any other Likelihood-based phylogenetic method, Bayesian dating requires a substitution model, a set of priors that allow us to estimate the number of changes observed in the alignment and that occurred along each of the branches of a phylogenetic tree. Different substitution models use different assumptions to describe the molecular substitution process. Commonly applied substitution models for DNA are JC (Jukes-Cantor), HKY (Hasegawa, Kishino and Yano), TN93 (Tamura and Nei 1993) and GTR (Generalized time-reversible) (Jukes, Cantor, and others 1969; Hasegawa, Kishino, and Yano 1985; Tavaré and others 1986; Tamura and Nei 1993). The differences among these models are the expectation of observing each possible nucleotide exchange. JC assumes that all the substitutions (both transition and transversion) are equally likely, and the nucleotide frequencies are always equal. In HKY, all the transitions have the same replacement rate, and the same occurs for the transversion rate: nucleotide frequencies are different for all the nucleotides. The TN93 distinguishes between the two transitions (A to G and C to T), but the transversions are assumed to occur at the same pace. This model allows for unequal base frequencies. These substitution models are particular cases of the GTR model, which allows a different rate for all the six possible substitutions and unequal nucleotide

frequencies. The GTR is composed of two sub-models: a set of four nucleotide frequencies (P) and a 4x4 rate matrix (R): together, they describe the pattern of nucleotide replacement. The GTR model is called reversible because the R matrix is specular. As with the nucleotides, models are available for amino acid and codons (Z. Yang, Nielsen, and Hasegawa 1998; Adachi and Hasegawa 1996; Goldman and Yang 1994). However, they are less employed because they demand high computational power and many parameters. A nucleotide time-reversible replacement rate matrix has only six parameters to estimate, whereas an amino acid matrix has 190 rates to estimate which can affect the speed of the analysis. For this reason, it is common to use a precompiled empirical matrix calculated on a large aminoacidic alignment such as in Dayhoff (Dayhoff, Schwartz, and Orcutt 1978).

Choosing the right substitutional model is not the only issue; both the rate and the pattern of evolution may vary across sites. Gamma distribution categories were applied successfully to overcome the issue of among-site rate variation: sites are assigned to a rate category and similarly evolving sites are clustered together (Z. Yang 1994). Commonly used models such as GTR do not, however, account for the among-site heterogeneity of the pattern of replacement. The GTR model is among-sites homogeneous because every type of replacement (e.g.: A→G) is modelled in the same way at every site. Efforts have been made to relax this assumption. Among-site heterogeneous models of replacement such as CAT have been shown to be capable of inferring more correct phylogenies and reducing systematic errors due to model violation (Drummond et al. 2006; Lartillot and Philippe 2004).

1.1.3.2 Tree priors

A central problem of Bayesian analysis is how to assign prior probabilities to the inferred trees. Each tree has to explain the observed data (the alignment), but an evolutionary history can have many variables, node heights, branch length, and topology, making this challenging. However, we can assume branching process to assign a prior to the trees that does not rely on specific evolutionary events. Assigning a branching model means constraining the tree's appearance, assuming evolutionary dynamics increase the prior probability of a tree with a

given shape. But let's see what the main characteristics of two of the widely used branching model are.

Birth-death (BD) process. The birth and death process represents an oversimplified model to describe speciation and the extinction events (Kendall 1948; Rannala and Yang 1996). The birth-and-death process is an umbrella term to describe tree priors that model speciation-extinction events in the model. The model has two main parameters to estimate speciation (birth process) and extinction (death process) which means a lineage could go extinct or be born throughout evolution (Yule 1925). The Yule process is the simplest BD process that considers only speciation events, which means the model does not allow for extinction, which tends to produce younger node ages. The pure birth and death model considers events, the extinction and the speciation, leading to longer branches in deep time. Through the years, evidence was collected to evaluate the effect of Yule and BD processes on phylogeny. Some studies highlighted the small impact on the results (C. S. P. Foster and Ho 2017), but some found an important incongruence between the two models (S. Y. W. Ho et al. 2005; Heled and Drummond 2015). It is important to select priors that are best suited to each dataset, otherwise, it can affect the accuracy of estimates. Another source of bias in a BD analysis is the taxon sampling. Incomplete extant taxon sampling results in artificially long terminal branches. In some BD models, the proportion of the missing taxa can be specified; however, the analysis could be strongly biased by an incorrect sampling fraction(Heath, Huelsenbeck, and Stadler 2014). A more recent application of the BD process is the fossilized birth and death process that allows the Bayesian analysis to consider the fossil evidence jointly with the molecular evidence. Moreover, this new model can be useful to understand the fossilization rate of a lineage and to calculate the extinction rate (Heath, Huelsenbeck, and Stadler 2014).

Coalescent tree prior. The coalescent theory traces back the evolutionary history from the tip to the root, connecting the sequences to their parent node in the previous generation. This model of evolution describes the branching dynamics within the population and not among species (Drummond et al. 2002). A coalescent tree reconstructs the evolutionary history of a population where each organism has the same chance to contribute with offspring to the next generation. The model assumes that generations do not overlap and are panmittic (Fisher

1930). The coalescent prior is sensitive to the population size, under the simplest model, population size is constant through time. However, a constant assumption is not describing the natural dynamics of many populations. The population can grow exponentially or change over time. The skyline coalescent describes population variation through time; this assumes discrete intervals of different population sizes during the evolution of the population.

1.1.3.3 The Clock priors

One of the most attractive innovations of the last two decades is the plethora of clock models developed and implemented in Bayesian programs for inferring dated phylogenies, such as BEAST, BEAST2, RevBayes, Phylobayes, and MrBayes (Drummond and Rambaut 2007; R. Bouckaert et al. 2014; Höhna et al. 2016; Lartillot, Lepage, and Blanquart 2009; Huelsenbeck and Ronquist 2001). The clock model (a.k.a. branch models) indicates the number of rates adopted in the analysis compared to the number of branches in the phylogeny; more rates represent a more complex model. The rational to choose a model over another is a matter of both our knowledge of the organism biology and of hypothesis testing. Different software implements different clocks; here, I describe the widely used clock models implemented in Bayesian analysis.

The simplest model assumes a constant substitution rate (strict clock) through all the branches in the phylogeny (M. Dos Reis, Donoghue, and Yang 2016). The constant rate model is theoretically sound only when the rate variance through all the branches is negligible. This can be easily inspected in BEAST by inspecting the coefficient of variation, which explores the rate variance through the tree. The coefficient of variation helps to understand if the strict clock assumption would fit the dataset or not.

The strict clock is usually assumed to properly describe shallow time evolutionary processes, as in the case of population studies. The life-history traits influencing the mutation rate do not change much in a narrow time window. Even though the constant evolution rate seems unreliable, as I discussed above, the mutation and substitution rate vary a lot for many reasons (Paragraph 1.2.1), it has many applications. Basically, the strict clock model has fewer parameters to infer than the other clock model, which is extremely advantageous in terms of

computational power, especially with big datasets (e.g., many taxa, partitions, sites). Moreover, a non-strict clock assumption can increase the background noise and make the analysis fail to converge, making the analysis unsuitable for further discussion. In fact, in a Bayesian dating analysis with poor posterior convergence for many parameters, it is better to use a strict clock model than a relaxed clock (Vilaça et al. 2021).

Local clock, or discrete-rate models are models that allow for several rate categories higher than one but lower than the branches of the phylogeny. The local clock assumes that some branches may have the same evolutionary rate, the same rate can be assigned to different branches within the phylogeny, or the rate categories can describe a group of branches that cluster together (Yoder and Yang 2000). There are different methods to assign these categories into a phylogeny based on prior knowledge about the rate variation pattern, such as defining categories according to the host infected by a virus/parasite or according to geographic and ecological previous knowledge. However, a stochastic rate assignment to the branches is preferred in a local clock analysis.

In BEAST, the model applied is called the 'random local clock' to stress the stochastic distribution of rate categories through the phylogeny. A random local clock tests whether a branch shares the same rate category with the parent branch (Drummond and Suchard 2010). The child branch inherits the parent rate unless the data shows evidence for a rate category jump. A Poisson prior distribution describes the number of rates categories, the discrete number of rates can be calculated during the analysis; in this approach, the rate categories are treated as a variable (Heath, Holder, and Huelsenbeck 2012).

Relaxed molecular clock. The relaxed molecular clock model of evolution is the most widely applied model in Bayesian dating analyses. It allows each branch in the tree to independently have its own rate. The first formalized relaxed clock model was the nonparametric rate smoothing (NPRS) method developed by Sanderson (Sanderson 1997; 2002), then a vast plethora of Bayesian models arose at the beginning of 2000 (Thorne, Kishino, and Painter 1998; Huelsenbeck, Larget, and Swofford 2000; Kishino, Thorne, and Bruno 2001; Aris-Brosou and Yang 2003; 2002), Drummond formalizes the uncorrelated relaxed clock method and it is implemented in BEAST (Drummond et al. 2006; Drummond and Rambaut 2007).

In the relaxed autocorrelated clock, the rates can be related in adjacent branches (Thorne, Kishino, and Painter 1998): if a branch has a low rate in the uncorrelated clock, the surrounding branches do not have a higher probability of having a low rate, but the probability of having a low or high rate is equal. The uncorrelated clock assigns the evolutionary rate from a parametric distribution and assigns it to a branch; this parametric distribution is usually lognormal or exponential (Rannala and Yang, n.d.; 1996). In BEAST, the distribution of the clock means and standard deviation can be set; this allows the use of prior estimates from different studies. A more complex relaxed clock model allows the rate to change through the tree, meaning that long branches are more prone to face rate variation events through time (Huelsenbeck, Larget, and Swofford 2000).

1.1.3.4 Calibrations: the fundamental prior

The tree priors discussed in the previous paragraphs provide information about the tree topology and the branching dynamics (tree priors), the distribution of rates on the tree (clock prior), and the replacement dynamics. These priors do not bear intrinsic information about the absolute timing of evolution. Without a rate, even the molecular clock priors do not provide intrinsic information about the evolutionary timescale. How to cope with the lack of temporal information in dating analysis? The answer is calibration, which is probably the most fundamental and widely known prior in Bayesian molecular dating. Calibration priors are used to transform the observed number of substitutions into rates by using prior knowledge of the timing of evolution.

The most common type of calibration is node calibration, where we provide a probability density for a specific tree bifurcation. The information to determine how to calibrate a node may come from biogeographic data or, more commonly, from paleontological evidence. For example, an organism with the characteristics of clade A has been present in the fossil record since 100 Ma (million years ago): we can assume that the current diversity of clade A (the stem of clade A in paleontological terms) is older than 100 Ma. Thus, 100 Ma is the calibration for the diversification of clade A. Because we do not know the exact position of clade A on the current phylogenetic tree, 100 Ma should be used as an upper bound (minimum, a.k.a. more recent bound) and a lower bound (maximum a.k.a older

bound). In general, a fossil provides a reliable upper bound because it proves that the lineage to which it belongs existed before its first fossils occurrence. On the contrary, fossils cannot provide a lower bound. Misplacing fossil evidence and placing unjustified maximum bound can affect the result of the analysis. This hard bound to the phylogeny can overwhelm the signal of any other age constrain employed in the phylogeny (Battistuzzi et al. 2015). To overcome the problem of placing maximum bounds, a common approach is to apply a softbound into a uniform distribution, allowing the analysis to sample node age even outside the given range (this is the default in Phylobayes). Another approach is to use a lognormal distribution to describe the fossil calibration range. The fossil evidence describes the minimum age of a given node and the decreasing probability of the long tail representing its older age (S. Y. W. Ho and Phillips 2009).

Cautious implementation of fossils and biogeographic information is needed to avoid sources of error in placing calibrations. For instance, misplacement of a fossil calibration to the crown node instead of the stem node would provide artificial younger nodes (Biffin, Hill, and Lowe 2010). Unreliable taxonomic assignment of a fossil specimen could provide uncertainty in placing a calibration in a phylogeny. Another source of error is an incorrect age estimation of fossils resulting in an inaccurate calibration signal.

In most Bayesian calibrated analyses, the date and rates are estimated mainly by implemented calibration distribution at least for one node. Biogeographic and geological calibration shed light on the evolutionary history of species and populations: in insects, the Hawaiian volcanoes provided calibrations that could disentangle *Drosophila* evolutionary history (Obbard et al. 2012), whereas the morphogenesis of Aegean Islands provided a useful calibration for estimating the evolution of endemic Coleoptera (Papadopoulou, Anastasiou, and Vogler 2010). Historical events provided minimum evolutionary bounds to virus evolution: the end of the slave trade, for instance, was taken as a minimum calibration point for the introduction of Yellow fever virus (YFV) in the Americas (Bryant, Holmes, and Barrett 2007), and Beringian glaciation was used as internal node calibration to estimate the complex Flavivirus evolution (Pettersson and Fiz-Palacios 2014). Fossil calibration is widespread in vertebrate animals and other large animals with external calcareous or chitinous exoskeletons; this is

because hard body parts are more prone to get fossilized. Soft-bodied animals and small insects do not provide a fruitful source of fossils. This lack of calibrations makes it challenging to set up a Bayesian dating analysis for these taxa.

Calibrations can be applied to nodes and even to tips; this can be done with viruses (see Chapter 5) in a tip dating analysis where the collection date provides a calibration probability distribution. The extraordinary short generation time of the virus allows calibrating with samples which are only months old. It is possible to use tip dating also for bacterial or eukaryotic studies. From a methodological point of view, the tip dating approach works in the same way in every taxon where it is applied. The only difference between virus tip date and other tip date is in their time range: viruses accumulate enough mutations to provide a temporal signal in a human-scaled timespan, whereas, for all other organisms, including bacteria, a human-scale timespan evolution does not leave any appreciable time signal in genomes. Ancient DNA studies have thrived in the last years because old samples can provide both DNA and a valuable calibration for inferring phylogeny. Ancient bacterial genomes provide an interesting calibration point for bacterial evolution. The exploitation of ancient DNA provided, for example, insight into the evolutionary history of many human pathogens such as *Yersinia pestis, Mycobacterium tuberculosis,* and human immunodeficiency viruses (HIV) (Bos et al. 2011; Menardo et al. 2019; Gryseels et al. 2020). The molecular clock has been successfully applied to reconstruct the human microbiome's evolutionary history using ancient bacterial DNA as tip calibrations (Wibowo et al. 2021; Tett et al. 2019).

In conclusion, calibration is one of the most relevant prior in a dating analysis and it inevitably bears some degree of error and uncertainty. In analyses involving many calibration points, a good practice is to exclude some calibration priors to identify possible over-constraints that may affect the entire dating (Rota-Stabelli, Daley, and Pisani 2013). It is also useful to conduct Bayesian estimates by sampling only the prior information without analyzing the actual data (the alignment): this would provide a tree that reflects only the prior information. If this tree is too similar to the tree obtained using the data, it indicates that the prior is influencing too strongly the results.

1.1.4 Model selection: how to choose the model that suits better the dataset

The choice of clock model and other priors can affect the output of a Bayesian dating analysis. Therefore, selecting an appropriate model is necessary to provide a reliable answer to any evolutionary question. A model can be chosen from our knowledge of the evolutionary dynamics and life-history traits of the species under scrutiny. If, for example, we study intra-specific population dynamics, a coalescent tree prior suits better the analysis than a Birth and Death prior that is thought to describe speciation. If we investigate RNA viruses, we should apply a broad clock rate inferred from previous knowledge on RNA virus evolution (S. Duffy, Shackelton, and Holmes 2008).

In a Bayesian framework, the model fit can be however weighted by the marginal likelihood. The marginal likelihood ratio compares the models using the Bayesian factor. The Bayesian factor provides which models are more likely to have generated the data (Kass and Raftery 1995) and it is employed in phylogenetics for testing hypotheses, such as different calibrations, tree priors, tree models and topology (Baele et al. 2013; Bergsten, Nilsson, and Ronquist 2013). BEAST implements (or has implemented) various methods to evaluate the marginal likelihood. The two simplest methods are the harmonic-mean, which calculates the marginal likelihood from the posterior, and the arithmetic mean, which employs the prior (Newton and Raftery 1994). These two methods are no longer recommended to calculate Bayes factors due to their infinite variance (Baele et al. 2013; Baele, Lemey, and Suchard 2016). A more reliable approach is path sampling (PS) (Lartillot and Philippe 2006), stepping stone (SS) (Xie et al. 2011), and nested sampling (NS) (Russel et al. 2019); all of these can be calculated in BEAST2 by means of a series of MCMC-like runs.

1.1.5 Partitioning data in clock studies

An important notion to better understand the molecular clock is that rates do not only vary among branches but also among. Some genes can be under relaxed selection, some others can be ruled by purifying selection, and some genomic regions can face genomic hitchhiking: different genes might provide a different rate and pattern of substitution rate (Gillespie 1991). Even distinct codon positions have independent evolutionary rates. All of these variables can affect

the rate through the genome, meaning that calculating a unique clock on a genomic scale can be inaccurate, and it is required to calculate multiple branch rate models to account for the rate diversity across the genome (Duchêne and Ho 2014).

Should we divide the datasets into partitions in which each one has its own evolutionary rate? The answer in our opinion is negative. Too many parameters can lead to overparameterization issues that can alter the final result. However, to avoid this issue, statistical methods help to reduce the partition scheme and the number of clock models involved in the analysis. It is possible to identify the number of rate variation patterns in a multi-locus analysis and assign each locus to a rate variation class. Tools like clockstaR and other approaches can identify the optimal partitioning scheme for a clock analysis (Duchêne, Molak, and Ho 2014; Snir 2014). The partition can reflect clusters of loci evolving at the same pace or the codon position of an in-frame set of genes. In fact, in BEAST it is possible to split an in-frame alignment in a codon position scheme automatically. Different codon positions have independent substitutional rates under different degrees of selection; in fact, the third codon position is supposed to evolve neutrally. Employing a substitutional clock prior only to one partition is extremely useful when we have prior clock information only on a subset of our dataset: in this way, we can infer the clock of all the other partitions involved in the analysis. It only works if all the partitions share the same tree topology.

1.2 Invasive species

1.2.1 What is an invasive species?

Invasive species are non-native organisms that have entered new territories and are expanding their areal in these territories. They directly affect wildlife, agriculture, and human health, impacting native species and ecosystems and posing socio-economical threats. Invasive species are considered one of the main factors of global changes (Vitousek et al. 1996), along with other ecosystem threatening phenomena such as exploitation, climate change and pollution. Indeed, invasive species are involved in biodiversity loss, ecosystem disruption, and species extinction.

Invasive species are mainly introduced by trade and other human activities; they can affect human agriculture, damage crops and reduce or destroy the harvest having a severe impact on the local economy. Many invasive species are known to have disrupted the entire ecosystem after their introduction and taken over native species. Hybridization and introgression are evolutionary consequences of invasive species in new ecosystems and, moreover, they can promote new selective pressure on native species (Mooney and Cleland 2001).

Many invasive species were introduced intentionally as decorative plats, domesticated animals, biological control, or wild animals for hunting purposes. But lately, human connectivity and intensified trade have introduced new threats to agriculture and human health. The economic impact of invasive species has been estimated for many countries (Pimentel, Zuniga, and Morrison 2005). However, the impact on humans of pathogens carried by invasive vectors has recently emerged in local or wider epidemics. Among those invasive vectors, mosquitoes are the most important spreader of new and old human diseases all over the world.

1.2.2 An overview of European *Aedes* invasive mosquitoes

Mosquitoes (Culicidae) are among the most successful Diptera lineages: they include more than 3,600 species classified in two subfamilies and 44 genera (Pombi and Montarsi 2020; Wilkerson et al. 2015; Reinert, Harbach, and Kitching 2004). They are vectors of various diseases that make them the more significant indirect cause of mortality among humans. They also cause serious nuisance to humans due to their aggressive biting activity to the point that they are affecting the quality of life in various countries (Halasa et al. 2014).

Aedini tribe is the broader taxonomic group within the mosquito family, counting more than 1.200 species divided into 10 genera (Pombi and Montarsi 2020). Aedini mosquitoes can be vectors of many zoonoses that affect humans and animals, including filarial nematodes (Taylor, Hoerauf, and Bockarie 2010) and many arboviruses such as Chikungunya, Dengue, Zika, Yellow Fever, West Nile, Japanese Encephalitis viruses (Pfeffer and Dobler 2010; Kilpatrick and Randolph 2012; Kraemer et al. 2015). Aedini mosquitoes are globally distributed; some *Aedes* vectors of diseases have become invasive in non-native environments

in recent years, mainly favoured by climate change and increased human connectivity (Couper et al. 2021; Fischer et al. 2011; Liu-Helmersson et al. 2019). In particular, increasing temperatures are enhancing the risk of arboviral tropical diseases' introduction and transmission in temperate regions (Brady et al. 2013; Scott et al. 2000; William K Reisen, Fang, and Martinez 2014).

In Europe, we count several recent Aedini mosquitoes introductions: *Aedes albopictus*, *Aedes koreicus*, *Aedes japonicus*, *Aedes atropalpus*. Except for *A. albopictus*, the other invasive mosquitoes have not been investigated intensively, even though their invasiveness and potential health risk have been known since the 1990s in North America and the 2000s in Europe.

Aedes albopictus is a daytime biting invasive mosquito endemic to South-East Asia. It has experienced a dramatic global expansion in the last decades; its spreading is mainly favoured by the trades of tyres and live plants (such as 'lucky bamboo'). *Aedes albopictus* was first reported out of its native range in the 18th century in Hawaii (Rai 1991). After that, it was first reported in Albania in 1979 and in the continental United States in the 1980s (Benedict et al. 2007), and in South and Central America before the 90s (Rai 1991).

Aedes albopictus spread has been intensified in the last three decades; indeed, in 1990 was reported in Italy, probably introduced by tyre trade in Genova harbour (Scholte and Schaffner 2007). Since its introduction, *A. albopictus* has settled in all the Italian regions, mainly in the urban areas. Outside Italy, *A. albopictus* was detected in France in Cost Azure and in Belgium in 2000 (Francis Schaffner, Van Bortel, and Coosemans 2004). *Aedes albopictus* spread in Greece, Switzerland and Balkan countries, and Spain along the Mediterranean seaside (Scholte and Schaffner 2007). Recently, this highly invasive species has settled out of the Mediterranean basin, establishing in Germany and England (Werner and Kampen 2015; Jolyon M Medlock et al. 2017)

Aedes albopictus is rapidly adapting to the European environment; the diapause eggs of the European population are already more tolerant to a colder temperature than the eggs laid by *A. albopictus* in their native range (S. M. Thomas et al. 2012). Rapid local adaptation of invasive species, usually by means of fixation of genetic variants that are rare in the native range, is a key feature that makes them successful in spreading and surviving the new environment (Moran

and Alexander 2014). *Aedes albopictus* populations in Europe are the results of different strain admixtures resulting from a complex and articulated invasive pattern (Manni et al. 2017). This species' rapid adaptation to colder climates can be explained by the introgression of populations from Asian temperate regions, such as Japan.

Aedes koreicus and ***Aedes japonicus*** were introduced in Europe in the 2000s (Francis Schaffner, Chouin, and Guilloteau 2003; Versteirt et al. 2009; 2012). These two new invasive species occupy different ecological niches than *A. albopictus*, preferring colder climates and finding a suitable ecosystem in the Alps and continental Europe. North-East Italy is the region most affected by the invasive mosquitoes phenomena. Indeed, together with southern Bavaria, it is the only area in Europa colonized by the major three invasive *Aedes* mosquitoes in Europe *A. albopictus*, *A. koreicus* and *A. japonicus* (Figure 4).

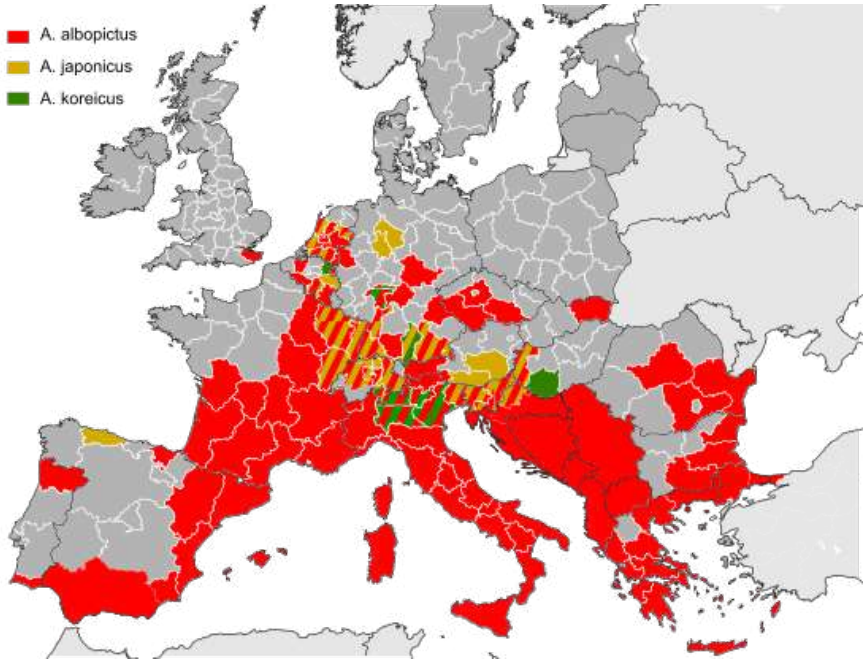


Figure 4. The actual distribution of *A. albopictus*, *A. japonicus*, and *A. koreicus*. The map shows that North-East Italy is characterised by the presence of all these three invasive species. *Aedes albopictus* is the most spread species in Europe. It has colonized all the countries in the Mediterranean basin, and in recent years, it has spread to the central and northern European countries. *Aedes koreicus* and *A. japonicus* show a restricted occupied area, mainly in central Europe and the Alps. Their distribution is expected to increase in the following seasons.

Aedes koreicus is endemic to Japan, North-eastern China, the Republic of Korea (South Korea), and Russia. *Aedes koreicus* first record in Europe occurred in 2008 and was reported again in the following years from the same area in Belgium (Versteirt et al. 2012). Subsequently, new collections of *A. koreicus* have been reported, starting from 2011 in Belluno and in the North-East Alps region of Italy (Capelli et al. 2011; Montarsi et al. 2013). Since then, its presence has been quickly escalating: in 2012, and 2013 it was found close to Swiss-Italian border (Suter et al. 2015), in 2013 in Slovenia, in 2015 in southern Germany (Werner, Zielke, and Kampen 2016), and in 2018 in Austria (Fuehrer et al. 2020). *Aedes koreicus* larval habitat preference is for man-made small containers and is usually found in association with human dwellings, while the adult stage prefers suburban and rural habitats. This species is a candidate vector of Japanese encephalitis virus (JEV) and *Dirofilaria immitis;* however, there are no reported cases of its involvement in JEV epidemics (J. M. Medlock et al. 2015).

Aedes japonicus strongly resembles *A. koreicus,* and it is endemic to Russia, Korea, Japan, and Southern China (J. M. Medlock et al. 2015). Its geographical expansion has been facilitated by human activities such as international trade. *Aedes japonicus* has spread initially to North America in the 1990s, and in 2002 its presence was reported in Europe (Peyton et al. 1999; Versteirt et al. 2009); in 2008, stable populations were confirmed in Switzerland (Francis Schaffner et al. 2009). Currently, *A. japonicus* is reported in Belgium, France, Germany (Werner and Kampen 2013), Austria, Slovenia (B Seidel et al. 2012), Croatia (Klobučar et al. 2019), Hungary (Bernhard Seidel, Nowotny, et al. 2016), Italy (Montarsi et al. 2019) and Spain (Eritja et al. 2019). *Aedes japonicus* has a high freezing tolerance; in the native range, the eggs overwinter in regions where the temperature can reach -20 °C (J. M. Medlock et al. 2015). *Aedes japonicus* prefers rural and sylvan habitats, but it can tolerate suburban and urban settings; the eggs develop in natural and artificial containers (Kaufman and Fonseca 2014).

Aedes japonicus is known to feed on mammals, and no record of bird blood meal was taken. Among mammals, human blood meal seems to have a high incidence along with domestic animals like cows (Damiens et al. 2014; Molaei et al. 2009). *Aedes japonicus* opportunistic feeding behaviour has caused public health concerns in the introduced countries. This species bites humans and a vast plethora of domesticated and wild mammals, increasing the chance of possible

cross-species viruses transmission (Kaufman and Fonseca 2014). This species can be involved in local arboviral disease transmission in newly colonized areas when the contact probability between domestic animals and human arboviruses is increased (Sardelis and Turell 2001; Francis Schaffner et al. 2011).

Mosquitoes have consistently raised concern in the last centuries because of their involvement in vectoring many pathogens such as *Plasmodium (*by Anophelini*)*, arboviruses and filariasis (by Aedini). In the Aedini clade are some of the mosquitoes facing the most rapid expansion and have been involved in several outbreaks outside their original range. While *A. aegypti* and *A. albopictus* have been studied intensively in many aspects of their biology, there is still poor and fragmented knowledge of *A. japonicus* and *A. koreicus* biology. In this thesis, I contribute to shedding some light on the evolution of these new invasive species by producing new genome data and by conducting molecular clock studies (see 1.2) of the Aedini in Chapters 2-3. I further contributed in showing the utility of phylogenomics (see 1.3) in Diptera, the order to which Aedini belongs (Chapter 6).

1.2.3 Arboviruses and invasive mosquitoes

Arbovirus is not a taxonomically defined group; it indicates any virus transmitted by arthropod vectors, including ticks and mosquitoes. Mosquitoes are competent vectors of many human pathogenetic viruses belonging to the *Togaviridae*, *Bunyaviridae* and *Flaviviridae* families, leading to the emergence of human diseases such as haemorrhagic fever, encephalitis, and meningitis (Cheng et al. 2016). Arboviruses infect humans via hematophagous arthropods bite. The infection can be spread only by arthropods that are competent for the virus, which means the virus needs to complete its lifecycle in the arthropod vectors to be transmitted.

In the case of mosquitoes, we can define three infection cycle scenarios. The most common scenario occurs by accidental spillover when a mosquito first bites an infected wild animal and then bites a human. In this case, mosquitoes act as a bridge for cross-species infection, as with West Nile virus (WNV) and Usutu virus (USUV) that were imported from Africa, probably by birds, and then have caused local human outbreaks in Europe.

The second scenario happens when the vector transmits the virus to a domesticated animal. Then the virus undergoes a secondary amplification circulating within domestic animals and subsequently infects humans, as in the case of the Japanese Encephalitis virus (JEV). In the first infection cycle described, there is a direct cross-species infection between wild animals and humans, whereas the second cycle involves domesticated animals as the reservoir of the infection. In both cases, humans are probably a dead-end host.

The third scenario is the most dangerous; once the cross-species event occurs from animals to humans (spillover) and the virus spreads among humans, skipping the intermediate animal host step. In this case, the virus can spread rapidly across humans with high efficiency in urban environments (Weaver 2018; P. Wu et al. 2019). This effect has been becoming more pronounced with the increasing population and the enhanced global connectivity of humans and goods. Humans are exposed to arboviruses when people invade rural and wild environments or when new bridge vectors can transmit viruses into peridomestic areas, increasing the spillover chance. In recent years various arboviruses experienced a geographical expansion in America and Europe: WNV has been spread mainly by bird hosts while JEV by domestic animals, in both cases using mosquitoes of the genus *Culex* as vectors (Erlanger et al. 2009; Muñoz et al. 2012). Urbanization and the globalized society are playing a pivotal role in facilitating the spread of arbovirus mediated disease among humans (E. Gould et al. 2017; K. F. Smith et al. 2007).

Arboviruses expand their range when there is availability of intermediate hosts and the concurrent presence of autochthonous competent mosquitoes vectors as in the cases of WNV and JEV. Arboviruses may further expand their natural range in the presence of competent invasive mosquito vectors in non-native territories. This is the case of various arboviruses vectored by *A. aegypti* and *A. albopictus*, which have caused outbreaks in every continent (excluding Antarctica). The most common outbreaks were caused by the Yellow fever virus (YFV), Dengue virus (DENV), Chikungunya virus (CHIKV) and Zika virus (ZIKV); these viruses have spread globally, and they have claimed many lives all over the world (Tomori 2004; N. I. O. Silva et al. 2020; Harapan et al. 2020; Young 2018; Chibueze et al. 2017; Higuera and Ramírez 2019).

Chikungunya virus (CHIKV) is responsible for acute infections. The main complications are chronic joint pain, severe organ dysfunction and encephalitis. It was first described in 1952 during an outbreak on the Tanzanian coast and other African regions (Lumsden 1955), Uganda and East Africa. Infections were reported in Asia between the 1960s and the 1970s, causing several epidemics and eventually, it was reported in the Americas (Leparc-Goffart et al. 2014). In Europe, CHIKV has already manifested a short autochthonous outbreak with *A. albopictus* as a vector in Ravenna, Italy, in 2007 and in Rome and in other municipalities in Central and South Italy (Rezza 2018). Autochthonous CHIKV infections were reported in southern France as well (Grandadam et al. 2011). Out of Africa, CHIKV outbreaks are caused by *A. aegypti* and *A. albopictus* (Higuera and Ramírez 2019).

DENV is a Flavivirus transmitted by Aedini species, mainly *A. aegypti* and *A. albopictus*, which causes what is known as Dengue fever; the symptomatic infection includes high fever, headache, retroorbital pain, myalgia. DENV has four serotypes, and it is considered the most important mosquito-borne disease, distributed in all the tropical and subtropical regions (Higuera and Ramírez 2019). Dengue has been reported in more than 100 countries, and the global spread is thought to date back centuries ago (Badii et al. 2007), and its infections are estimated at around 360 million per year (estimated in 2013) (Bhatt et al. 2013). Sporadic outbreaks have been registered in South USA states since the 1970s, with a contained number of infections (Beaumier, Garcia, and Murray 2014). Local DENV infections are reported yearly in Europe, mostly due to infections acquired by tourists that have visited countries where DENV is endemic, but lately, DENV has been involved in local outbreaks (Tomasello and Schlagenhauf 2013). *Aedes albopictus* has been involved as the primary vector in the DENV limited outbreak, where *A. aegypti* is either not present nor significant.

Yellow fever virus (YFV) is endemic to Africa and Central and South America in tropical regions, but, interestingly, it has never become endemic in Asia, although the conditions seem favourable, as climate and vector distribution. YFV originated in Africa and then was brought to South America through the trade of slaves between the 1600s and 1800s (Kuno 2020). YFV infections have risen in the South Americas since the 1990s; in 2016, Brazil registered the last large outbreak. YFV is an example of virus local adaptation to new vectors. In the sylvatic

environment, YFV is vectored especially by *Hemagogus* and *Sabethes* species, which occupy a non-anthropic niche, whereas, in urban outbreaks, the main vector is the invasive *A. aegypti* (L. H. Chen and Wilson 2020).

Zika virus (ZIKV) originated in Africa, and the first human case of ZIKV was reported in 1954 in Nigeria (MacNamara 1954). It was involved in a few local outbreaks, mainly in the Pacific Islands, such as Yap Island 2007 (M. R. Duffy et al. 2009), French Polynesia 2013 and New Caledonia (Hayes 2009; Cao-Lormeau 2014). However, in the 2016 ZIKV outbreak in the Americas, this virus gained attention for its unforeseen expansion and infection. ZIKV infection usually causes mild effects that last less than seven days. However, it may cause microcephaly and birth defects in infected pregnant women (Ventura et al. 2016). Only imported cases were registered in Europe. The epidemic hit heavily in South and Central America, reaching US southern states, in particular, Florida (Metsky et al. 2017).

Emerging and re-emerging arboviruses, especially Flavivirus, are becoming more common in every continent. *Aedes* vectors are not only responsible for the re-emergence of arboviruses in their Asian and African native range but are also responsible for novel vectoring in Europe and North America, where they have been recently introduced. It is therefore essential to study arboviruses' evolution and molecular epidemiology in detail to understand past outbreaks and protect from future ones. The molecular clock is a powerful tool to understand past and present epidemiological patterns of arboviruses. This technique has the potential to inform when a virus has moved from one place to another and if it has adapted to new hosts, vectors, or environmental conditions (Beaver et al. 2018; Pettersson, Bohlin, et al. 2018; Bertrand et al. 2012; Campos et al. 2018). In this thesis, I provide my contribution to show the utility of molecular clocks in understanding arboviruses epidemiology by studying ZIKV in Chapter 5.

1.2.4 Other agricultural invasive species

While invasive mosquitoes have a direct threat to human health, there are a plethora of other invasive species that pose ecological and/or economic concerns, for example, the insect pests of agriculture (Hulme 2009). One notable example is the vinegar fly *Drosophila suzukii*, whose recent outbreak has been responsible

for up to 80% of small fruits yield loss in certain European regions and a cost in the orders of billions of euros worldwide (Rota-Stabelli, Blaxter, and Anfora 2013; Asplen et al. 2015). As in the case of invasive mosquitoes, the introduction and spread of pest species outside their native range are eased by the global economy and its associated movement of goods (Hulme 2009). Once introduced, pests are extremely difficult to eradicate because they often occupy empty ecological niches and because scientists have a fragmented knowledge of their biology. In this thesis, along with studies of *Aedes* and one of their vectored arboviruses, I will present the results of evolutionary analyses of other three species of relevance for biodiversity and agriculture.

The first model I study is *Trissolcus japonicus*, the parasitoid of *Halyomorpha halys* an invasive hemipteran insect from South-East Asia (Hoebeke 2003). This species belongs to the Pentadomidae, a large family that includes more than 4100 species with a wide world distribution. Many species of this family may act as crop pests like *Palomena prasiana, Nezara viridula,* and *Aelia rostrata*. *Halyomorpha halys* first report outside its natural range was in the United States in Allentown, PA, in 1996 (Hoebeke 2003). Since then, it has been introduced in many other countries worldwide, including Italy (Maistrello et al. 2016). The peculiar combination of biological traits such as cold resistance, high polyphagy, strong dispersal capacity and lack of natural enemies (see below) have made *H. halys* an extremely aggressive invasive species (Maistrello et al. 2016). In its original environment, *H. halys* are naturally controlled by oophagous parasitoids, and various studies identify the parasitoid wasps *Trissolcus mitsukurii* and especially *Trissolcus japonicus* as candidate with the highest potential as an agent for classical *H. halys* biological control programs in newly invaded regions (J. Zhang et al. 2017). The ineffectiveness of nets and chemical methods to fight *H. halys* has promoted a classical biological control programme in Italy that consists of releasing and monitoring *T. japonicus* individuals from a lab-reared control population named CREA. From an applied point of view, it is important to assess the success of the CREA population releases: this is done by performing a recapturing experiment and by typing the captured specimens using genetic markers in order to discriminate them from local *T. japonicus* populations which have been independently introduced in Italy recently. To help in the definition of markers

for this typing, I sequenced and analyzed the genome of CREA and local *T. japonicus* individuals (Chapter 4).

Second model I studied is the plant *Arundo donax*, one of the several species known as reeds. It is one of the most aggressive invasive plants in invading warm regions worldwide, threatening the native biodiversity in its ecological niche. The genus *Arundo* is known to have gone through demipolyploidization events (Hardion et al. 2014). The evolutionary history of this genus is not, however, clear. In this thesis, I will show how molecular clock reconstructions based on transcriptome data can help in defining the timing and, therefore, the paleo-ecological context of the key demipolyploidization event in the genus *Arundo* and in *A. don*ax in particular (Chapter 6).

The third model is *Bactrocera*, an economically important genus comprising at least 50 different pest species (White and Elson-Harris 1992). In the Mediterranean basin, *Bactrocera oleae* is the most damaging insect pest in olives orchards. It is possible to use phylogenomics to identify genetic traits that can explain their invasiveness and may increase control strategies (Attardo et al. 2019) section 1.3). Phylogenomics relies on a robust dated phylogeny, which is still debated for some important *Bactrocera* pests (Choo et al. 2019). In Chapter 6 I contributed in dating the speciation events within the economically relevant *Bactrocera* genus.

1.3 The importance of molecular clocks and phylogenomics in pest science

Pest science generally focuses on the direct management of invasive species, chasing down the best practice to remove them, contain their damage, and prevent future invasions. At least for insects, in the absence of prior knowledge of the pest's biology, the most common, non-specific, management practice is the use of chemical pesticides (Pettis et al. 2013). The environmental and health cost of pesticides are fostering the application of alternative Integrated Pest Management (IPM) techniques, which make use of less harmful tools such as semio-chemicals for sexual disruption and of natural bio-controllers such as parasitoids and *Wolbachia.* Modern management agendas, in particular the IPM ones, are often based on the pest life-history traits for example generation time,

genetic variability, optimal environmental conditions. This information is based on a fine understanding of pest biology, a complex task that is normally achieved only after years of field, ecological, and physiological studies. Evolutionary studies can quickly provide insights into pest biology and, therefore, into its beneficial management: for this reason, a general call for more evolutionary studies in the field of pest science has been proposed (Thrall et al. 2011; Leftwich, Bolton, and Chapman 2016). One of the most interesting aspects of the evolutionary approach is that it is relatively cost-effective and rapid in providing results; this is because evolutionary analyses are generally computational and rely on sequences rather than on lab experiments. Speed is a key factor when designing management and governmental policies.

A large array of evolutionary tools to study invasive pests exists, which are now gaining a certain degree of methodological maturity. When applied to insect management, evolutionary approaches are, however, still mostly restricted at elucidating present events that are occurring within the pest population. The most widely used evolutionary approach is population genetics, the study of allele variation at the intra-specific level. Population genetics can easily provide fundamental demographic information, as in *A. albopictus* and *A. aegypti*, for which it has been possible to trace the invasion route (Gloria-Soria et al. 2016; Manni et al. 2017). Thanks to the reduction of sequencing costs, it is currently possible to perform population genetics at the genome level. Successful examples include *D. suzukii* and *A. albopictus*, whose large scale population genomic studies could describe with unprecedented precision their invasion routes as well as events of admixtures between different populations (Lewald et al. 2021). Population genomics can reveal unexpected high genetic differences among populations worldwide: if these differences are mirrored by different life-history traits, they may impair the global effectiveness of management practices (Rota-Stabelli et al. 2020). Another type of microevolutionary approach in pest science is the intraspecific genetic/genomic survey aimed at identifying pesticide resistance. This has been done for many agricultural pests, such as *Plutella xylostella* (Troczka et al. 2012) and *Myzus persicae* (Bass et al. 2014), as well as for invasive *Aedes* mosquitoes (L. B. Smith, Kasai, and Scott 2016; Tancredi et al. 2020). These studies are crucial because resistance to a pesticide should be detected as early as possible when the allele frequencies are still low in the

population: this may allow the definition of compensatory management practices such as pesticides rotation.

Population genetics and genomics are microevolutionary approaches: they tackle events and processes at an intraspecific scale. An aspect so far neglected in pest science is instead the study of the deep-time evolution of the pest. This approach can be referred to as macroevolutionary (Hautmann 2020) because it studies the evolution of the pest at the interspecific level. It is becoming clear that many important traits (and therefore genes) associated with invasiveness have evolved by speciation(s) over a long time frame (Thrall et al. 2011). Macroevolutionary approaches such as the molecular clock and phylogenomics have therefore the potential of revealing the genetic bases of these traits as well as the evolutionary pattern associated to their emergence. Knowledge of these traits may have direct consequences for pest management. Here I would like to present some evidence from the available literature in support of the utility of macroevolutionary studies in pest science.

1.3.1 Phylogenomics application in pest science.

Phylogenomics has at least two meanings: it may refer to 1) genome-scaled phylogenetics and 2) the evolution of gene families on the species tree. The two meanings are often coupled because the evolution of gene families is studied on a genome-scaled phylogenetic tree. Phylogenomics is a macroevolutionary approach because it investigates evolutionary events that have led to speciation. It can provide insight into gene functional prediction by revealing which genes (or gene families) are peculiar to a certain species: these genes may be at least partially responsible for the invasiveness of the pest.

One example of successful phylogenomic study of a pest is that of *Drosophila suzukii*: this pest is coincidentally from one of the most important genus models in genetics, for which many genomes are available. This allowed studying of a taxon-rich phylogenetic framework, some important gene families that regulate olfaction and taste, two senses that play a fundamental role in regulating insect behaviour and niche specialization (Sánchez-Gracia, Vieira, and Rozas 2009). For this reason, Odorant and Gustatory Receptors genes (ORs and GRs) have been targeted by phylogenomics studies in D. *suzukii.* Phylogenomics showed a

substantial increase of ORs and GRs in the lineage leading to *D. suzukii* (Crava et al. 2016; Ramasamy et al. 2016). This can be explained by selective forces imposed by environmental changes that forced *D. suzukii* to oviposition on ripe fruits instead of rotten fruits usually used by other *Drosophila* species. Some of these receptors are used by *D. suzukii* to detect fresh fruits and are being studied to define their ligand, which shall be used as a lure in traps for ameliorating its management.

Other successful examples of phylogenomic studies in pest science are those of insects of medical relevance for which many genomes have been produced. The analysis of the genomes of 16 closely related *Anopheles* species could define species-specific gene novelties thus describing the biology of each of the 16 species with high detail (Neafsey et al. 2015). In a similar approach, the genome of 5 *Glossina* species (tze-tze flies) has been analyzed. Results highlighted the unique features of each of the species: this data is relevant form management because it shall guide studies aiming at ameliorating traps and defining novel control strategies (Attardo et al. 2019).

Overall, the phylogenomics approach in pest science has the potential of identifying genes that correlate to phenotypical novelties responsible for invasive traits. In Chapter 6 I show the utility of the phylogenomic approach in identifying genes in flies and mosquitoes that may become the target for downstream analyses aimed at improving their management.

1.3.2 Molecular clock application in pest science

The molecular clock can reveal when in time, a pest differentiated from other species that do not show invasive or deleterious characteristics. Linking these evolutionary events with paleoecological scenarios may provide a solid background to understand why certain pest traits emerged as past adaptations to changing environmental conditions.

Pioneering work in this direction has been done to understand the unusual ecology of the invasive fruit fly *D. suzukii*. Clock studies provided evidence of its Asian origin circa 7 Ma; by crossing this information with the actual geographical distribution of the pest, it was possible to hypothesise a long history of adaptation to temperate climates due to the Tibet upraise during the late Miocene (Ometto

et al. 2013). This has direct management significance: it indicates that *D. suzukii* is pre-adapted to the European temperate climate because it originated by adaptation in a similarly temperate climate in Asia. Clock studies further revealed a reduced substitution rate in the branch leading to *D. suzukii.* This indicates that this species has been characterised during its evolutionary history by a reduced number of generations per year compared to other *Drosophila* species. Winter diapauses decreased the number of generations per year and consequently decreased the mutation rate. The case of *D. suzukii* taught us that paleoecological preadaptation to temperate (colder) environments could have favoured its invasive potential in similarly temperate regions (Ometto et al. 2013). This clock study provided a biological background for field applications that aim at containing *D. suzukii* in the field. It instructed us that cold is not a limiting factor for this pest which does not need overwintering shelters in the field: this information could be confirmed by subsequent monitoring experiments (Rossi-Stacconi et al. 2016).

Molecular clock studies, when coupled with comparative (phylogenomics) approaches, are particularly effective in describing the evolution of traits. It is, for example, possible to reconstruct past ecological behaviour of pests and better understand its current ecology. This has been done, for example, in Aedini mosquitoes whose habitat specialization of laying eggs in containers had emerged multiple times independently in this group: this trait is shared by most of the invasive *Aedes,* providing a phylogenetic indication of possible future new invasive species (Soghigian, Andreadis, and Livdahl 2017). Another successful example of deep time molecular clock and ecological niche modelling is that used to study *Leishmania brazilensis* species complex. Results show a diversification in the Late Pleistocene, dating the hybridization event associated with the habitat contraction driven by climate change that occurred 150.000 years ago in Amazonia (Van den Broeck et al. 2020).

1.4 Aims of the dissertation

Macroevolutionary studies may be useful for pest science studies. The timing of pest speciation can reveal the paleoecological context that has contributed in shaping the key features that we currently observe in the pest. Inference of substitution rates in the branch leading to the pest can inform about the number of generations per year, providing useful information to understand its demography. Molecular clocks coupled with phylogenomics may reveal which are the genetic novelties that characterize the pest in comparison to closely related species that are not pests providing with genes for downstream applied studies.

The general aim of this thesis is to provide new results to further demonstrate the utility of macroevolutionary studies, particularly molecular clock, in pest science. I will do this by applying state of the art molecular clock methodologies to various invasive species for which macroevolutionary studies have been conducted fragmentary because of a lack of available genome data. I will present new genome data and clock studies from different model organisms ranging from mosquitoes of biomedical importance and their arboviruses (Chapter 2, 3, 5) to insect species of agriculture interest (Chapter 4, 5, 6) and invasive plants (Chapter 6). The outcome of these studies will be used to improve a review paper on the utility of the molecular clock in pest science (Chapter 1.3).

Studying different model organisms allows me to further define objectives aimed at answering specific scientific questions.

- How robust is the currently available time frame of Aedini evolution (Soghigian, Andreadis, and Livdahl 2017)? To answer this question, I employed, for the first time in *Aedes*, enhanced models of replacement and data type comparison in Chapter 2.
- What is the evolutionary history of emerging invasive *A. koreicus* and *A. japonicus*? To answer this question, I generated genome data for these two species and performed clock studies of these species using whole mitogenomes in Chapters 3 and 4A.
- What are the evolutionary patterns of the ecologically important Opsin genes family in dipteran pests? To answer this question, I performed a large phylogenomic study of this gene family in Diptera (Chapter 6.2).

- What are the poorly studied patterns of Zika virus early evolution? To answer this question in Chapter 5, I analysed the largest genome scaled dataset to date for this arbovirus.
- Can we exploit divergence estimates to manage the release of the biocontrol agent *T. japonicas*? To answer this question, I sequenced the genome and compared the mitogenomes of two different populations in Chapter 4B.
- What is the timing of the evolution of two important invasive genera, the *Arundo* plant and the *Bactrocera* fruit flies? To better understand the emergence of pest-specific novel traits, I contributed in performing up to date clock analyses of these genera in Chapter 6.

Macroevolutionary studies such as those proposed in this thesis could provide novel insights into the paleoecology of pests, help understanding which evolutionary events and novelties are responsible for their invasiveness, and contribute in ameliorating their management.

CHAPTER 2 - CHRONOLOGICAL INCONGRUENCES BETWEEN MITOCHONDRIAL AND NUCLEAR PHYLOGENIES OF *AEDES* MOSQUITOES

2.1 Introduction to the Chapter

This Chapter presents the results of the first topic I investigated during my PhD training. It represents an initial approach to investigate *Aedes* mosquitoes' phylogeny and clock. First, I created two different Diptera datasets, one nuclear and one mitochondrial, and I further concatenated the two datasets. I was able to identify four genes per dataset that were suitable for phylogenetic analysis. The lack of molecular data in the Aedini clade makes the phylogenetic investigation challenging; moreover, the molecular data available are mainly mitochondrial, and many nuclear DNA data are missing for many important and representative species. I employed maximum-likelihood (RAxML), and a Bayesian (PhyloBayes) approach to infer phylogenies and a Bayesian (BEAST2) approach to infer divergences. Results show phylogenetic and chronological incongruences between nuclear and mitochondrial data. In this work, I provide a first attempt to date some important divergences, such as the split between *Aedes albopictus* and its sister species *Aedes flavopictus* and the split between *Aedes koreicus* and *Aedes japonicus*. I set a probable origin of mosquitoes within the Angiosperm radiation, and I estimated a more recent origin for most clades in comparison to previous work. Results revealed some methodological issues which have been used to guide further research during my PhD studies (see Chapter 3).

The paper is published on MPDI Life:

https://www.mdpi.com/2075-1729/11/3/181/htm#app1-life-11-00181

Contribution. I carried out all the analysis of this study along with the experimental design. I was guided by my supervisor Omar Rota-Stabelli. Omar Rota-Stabelli supervised the data interpretation and manuscript writing.

Chronological Incongruences between Mitochondrial and Nuclear Phylogenies of *Aedes* Mosquitoes

Nicola Zadra[1,2], Annapaola Rizzoli[2], Omar Rota Stabelli [1,2,4]

[1]CIBIO department, University of Trento; [2] Fondazione Edmund Mach; [3 4] Center C3A, University of Trento.

2.2 Abstract

One-third of all mosquitoes belong to the Aedini, a tribe comprising common vectors of viral zoonoses such as *Aedes aegypti* and *Aedes albopictus*. To improve our understanding of their evolution, we present an updated multigene estimate of Aedini phylogeny and divergence, focusing on the disentanglement between nuclear and mitochondrial phylogenetic signals. We first show that there are some phylogenetic discrepancies between nuclear and mitochondrial markers, which may be caused by wrong taxa assignment in samples collections or by some stochastic effect due to small gene samples. We indeed show that the concatenated dataset is model and framework dependent, indicating a general paucity of signal. Our Bayesian calibrated divergence estimates point toward a mosquito radiation in the mid-Jurassic and an *Aedes* radiation from the mid-Cretaceous on. We observe, however a strong chronological incongruence between mitochondrial and nuclear data, the latter providing divergence times within the Aedini significantly younger than the former. We show that this incongruence is consistent over different datasets and taxon sampling and that may be explained by either peculiar evolutionary events such as different levels of saturation in certain lineages or a past history of hybridization throughout the genus. Overall, our updated picture of Aedini phylogeny reveal a strong nuclear-mitochondrial incongruence which may be of help in setting the research agenda for future phylogenomic studies of Aedini mosquitoes.

Keywords: divergence; mtDNA; Diptera; phylogeny; saturation; rates

2.3 Introduction

Mosquitoes (Culicidae) are one of the most successful Diptera radiations. They include more than 3600 species classified in two subfamilies and 44 genera and 145 subgenera (Reinert, Harbach, and Kitching 2004; Pombi and Montarsi 2020; Wilkerson et al. 2015). Because they vector a variety of disease, mosquitoes are still the largest indirect cause of mortality among humans than any other group of organisms. Approximately one-third of mosquito species belong to the tribe Aedini, including 1261 species classified in 10 genera (Wilkerson et al. 2015). Aedini species are globally distributed and are vectors of many zoonosis of human and animals including filarial nematodes (Taylor, Hoerauf, and Bockarie 2010) and many arboviruses such as Chikungunya, Dengue, Zika, Yellow Fever, West Nile (Pfeffer and Dobler 2010; Silverj and Rota-Stabelli 2020; Kraemer et al. 2015). Aedini species include some of the most invasive and medically relevant mosquitoes: *Aedes aegypti* and *Aedes albopictus* (Jolyon M Medlock et al. 2012; E. C. Cameron et al. 2010; J. M. Medlock et al. 2015; Grard et al. 2010; Francis Schaffner, Chouin, and Guilloteau 2003). *Aedes aegypti* has mainly spread outside its original African range, although it does not seem capable of settling stable populations in continental climates, such as the European one. *Aedes albopictus*, originally from South East Asia, is instead now reported from every continent and has quickly settled in Europe, China, and other temperate zones (Kraemer et al. 2015; N. R. Faria et al. 2017). Genome resources exist for only these two species of *Aedes* (X.-G. Chen et al. 2015; Matthews et al. 2018; Dritsou et al. 2015), while whole genome data for other invasive *Aedes* is still lacking. These include *Aedes japonicus* and *Aedes koreicus*, which are quickly invading and establishing, respectively, in central Europe (Bernhard Seidel, Nowotny, et al. 2016) and North Italy (Montarsi et al. 2015; Capelli et al. 2011) showing competence for the transmission of many arboviruses such as West Nile virus and Zika virus (Jolyon M Medlock et al. 2012; Capelli et al. 2011; Huber, Jansen, et al. 2014).

Knowledge of the reciprocal affinities of these and other invasive *Aedes* species and the timing of their evolution is important for various reasons. First, a robust phylogeny is essential to polarize key behavioural and ecological traits, as recently shown by (Soghigian, Andreadis, and Livdahl 2017). In particular, a phylogeny can identify the sister-species of invasive *Aedes* of health concern. The

sister-species shares a common ancestor with the species of interest (is the closest related in the phylogenetic tree) and is very useful for correctly polarizing evolutionary novelties, such as new genes in phylogenomics and transcriptomics studies (Ramasamy et al. 2016; Crava, Brütting, and Baldwin 2016). Second, phylogenies may help to define taxonomy and classification. A recent classification (Reinert, Harbach, and Kitching 2004) has raised the number of genera from 10 to 79; the genera, however, have been later reduced to 10 (Wilkerson et al. 2015). Molecular investigations of Aedini relationships can help to clarify these taxonomical issues. Third, dated phylogenies help to characterize the paleoecological scenario in which mosquito radiations happen, thus providing evidence with clues about their pre-adaptations as it has been shown, for example, in *Drosophila* (Ometto et al. 2013; Rota-Stabelli et al. 2020). Molecular studies have addressed Aedini evolution by studying their phylogeny using both mitochondrial and nuclear markers. While relationships within the Aedini group has been studied in detail using a multimarker approach (Soghigian, Andreadis, and Livdahl 2017), the origin of the family and their reciprocal affinity with other Culicinae are not well studied or have not been addressed because datasets were centred only on Aedini (Soghigian, Andreadis, and Livdahl 2017). Furthermore, the stability of clades within the Aedini has never been addressed by comparing different statistical frameworks (e.g., maximum likelihood versus Bayesian), or by employing a different substitution model (e.g., homogeneous versus heterogeneous (Feuda et al. 2017)).

One key aspect so far neglected in Aedini phylogenetic studies is the direct comparison of the phylogenetic signal from the DNA of the two cellular compartments: nuclear (nDNA) and mitochondrial (mtDNA). It has been shown that the nDNA and mtDNA may carry different phylogenetic signals and produce conflicting phylogenies, in some cases, because of hybridization events affecting mtDNA (Hirano et al. 2019). MtDNA substitution rate is typically faster compared to the nuclear one; this can lead toward homoplasy caused by site saturation, which in turn may affect the topology and may underestimate the correct inference of substitution rates (Hirano et al. 2019).

Little in general is known about how mtDNA and nDNA conflict for what concerns estimation of divergence times. In some cases, chronological signal can be consistent between nuclear and mitochondrial genes, as in fish (Near et al.

2012) and amphibians (Zheng et al. 2011), with discrepancy just in the shallow time part of the tree. In *Drosophila*, the two types of markers recover similar divergences with mtDNA supporting slightly younger estimates than nDNA (Ometto et al. 2013). In butterflies, the chronological conflict between nDNA and mtDNA is more marked, although it seems to be restricted to a few species experiencing hybridization (Wahlberg et al. 2009). In the above cases, the confidence interval of the divergence estimates using the two types of markers largely overlap. Therefore, the conflict is not statistically significant. The molecular clocks of mtDNA and nDNA have never been systematically compared in Aedini.

A systematic comparison of chronological signal in mosquitoes has never been undertaken. An effort to date the Aedini mosquito using nDNA data in a Bayesian framework (Reidenbach et al. 2009) recovered the origin of Aedini at 157 (Credibility Intervals, CI: 187–124) millions of years ago (Ma) and a Culicidae radiation at 216 (229–192) Ma. A recent effort using a multigene (nDNA + mtDNA) strategy in a maximum likelihood framework (Soghigian, Andreadis, and Livdahl 2017) recovered an Aedini origin at circa 125 Ma. In the latter, the diversification of *A. albopictus* from its sister species *A. flavopictus* is circa 25 (14 – 43) Ma, while *A. albopictus* and *A. aegypti* common ancestor was set at approximately 55 Ma, a time compatible estimate, but quite distant from that based on whole genomes 71 (44–107) Ma (X.-G. Chen et al. 2015). A recent divergence estimate of Culicinae using complete mtDNA (A. F. da Silva et al. 2020) recovered the origin of Aedini at 130 (CI: 101–168) Ma and an *A. albopictus*-*A. aegypti* split at circa 67 (CI: 55–94) Ma. There are, therefore, certain discrepancies in the available literature for what concerns the timing of Aedini radiation.

This work aims at providing an updated picture of Aedini phylogeny and divergence times by disentangling the phylogenetic signal in available genetic markers. We used four nuclear and four mitochondrial genes in a Bayesian framework to study the evolutionary history of the Aedini, their relationship with other Culicinae, and the timing of their origin and diversification. Our results revealed previously under looked incongruences between nuclear and mitochondrial data, for what concerns both their rate of evolution and their posterior divergence estimates. This has an important implication for our

understanding of Aedini evolution and more generally for the long-lasting issue of incongruences between mitochondrial and nuclear data in inferring species phylogeny and divergences.

2.4 Materials and Methods

2.4.1 Genes and Taxa Selection

In our study, we employed four mitochondrial coded genes: Cytochrome c oxidase I (COI), Cytochrome c oxidase II (COII), NADH dehydrogenase subunit 4 (NAD4) and 16S, and four nuclear-coded genes: Enolase, Arginine Kinase, 18S, and 28S. We chose these genes after various rounds of literature and blast searches because they were the most evenly distributed through the Aedini tribe and the outgroup. Similarly to other recent *Aedes* phylogenetic studies (Soghigian, Andreadis, and Livdahl 2017), the current availability of genes in the database did not allow us to sample more genes. The number of annotated genes in Genebank for Aedes and other mosquitoes species is low, in general no more than 5 or 6 markers per species; many species are characterized by many variants of the same marker, for example, COI. Annotated genome and transcriptome data was present only for the model organisms *Aedes albopictus* and *Aedes aegypti*. We had to exclude from our gene list the genes encoding for white, hunchback, and Carbomoylphosphate synthase (CAD) because they are poorly sampled within the Culicidae family. We did not employ Internal Transcribed Spacer (ITS) because of poor and ambiguous alignment between Aedini and its outgroups. Since we were interested in studying the origin of Aedini, poor alignment of Aedini ITS sequence with those of the outgroups could have affected the correct inference of their phylogeny. We sampled genes from the same specimen whenever it was possible; in most cases, we concatenated genes from different specimens of the same species. This is the common procedure when concatenating genes for inter-specific phylogenetic studies (Soghigian, Andreadis, and Livdahl 2017; Feuda et al. 2017). We followed the nomenclature as in (W K Reisen 2016; Wilkerson et al. 2015). For more clarity, in some of our phylogenetic trees, we displayed in brackets the proposed subgenera. Each chosen gene for all the available Aedini taxa was downloaded from GenBank. To reduce missing data and promote a direct comparison between nuclear and

mitochondrial data, we selected a species only if at least two genes represented it in each of the two types of markers (nuclear and mitochondrial). Moreover, we excluded species that seemed to be ambiguously labelled. The final dataset was represented by 34 evenly phylogenetically distributed Aedini, plus 10 outgroups sampled from other Culicinae, Anopheline, and other Diptera samples (see Table S1 for the species list). The outgroup sequences were essential to root the Aedini phylogeny and to generate nodes for calibrating the molecular clock. For each of the eight markers, we filtered out ambiguous sequences. We used a fast bootstrap RaxML analysis (see details below) to preliminarily check if sequences were clustering within their expected group (e.g., sequences for Aedini to form a monophyletic Aedini group). Sequences clustering with a different group were considered unreliable and they were excluded from downstream analysis.

2.4.2 Alignments and Phylogenetic Analyses

We aligned each of the eight genes independently. We aligned protein-coding genes using MAFFT through TranslatorX (Abascal, Zardoya, and Telford 2010), and non-coding genes using MAFFT directly (Katoh, Rozewicki, and Yamada 2018). Finally, the genes were concatenated using FASconCAT (Kück and Meusemann 2010) and manually edited to detect a few misaligned sites. We generated three aligned datasets: nuclear, mitochondrial, and concatenated. The nuclear dataset (nDNA) is composed of the concatenation of Enolase, Arginine Kinase, 18S, and 28S; it is 3270 nucleotides (nt) in length. The mitochondrial dataset (mtDNA) is composed of the concatenation of COI, COII, NAD4, and 16S; it is 4224 nt in length. The third dataset (concatenated) is the concatenation of nDNA and mtDNA and is 7494 nt in length. To further study Aedini relationships, we generated a fifth dataset based on the original 6298 nt alignment of (Soghigian, Andreadis, and Livdahl 2017), increasing site occupancy by using Gblocks at default parameters. The final dataset (named Soghigian) was composed of 71 sequences and 3815 nt with 8% of missing data. Although there is some overlap of genes between concatenated and Soghigian datasets, they substantially differ because of the presence of 4 genes (COII, NAD4, and 16S present in concatenated, while ITS is absent from concatenated) and mostly because of very different taxon sampling. The concatenated dataset contains

various outgroups to the Aedini because the aim was to set the origin of Aedini. Phylogenetic analyses were performed using both maximum likelihood (ML) and Bayesian statistical frameworks, using, respectively, RAxML (Stamatakis 2014) and PhyloBayes (Lartillot, Lepage, and Blanquart 2009) or BEAST (see below). The RAxML analyses were performed on all datasets using the General Time Reversible (GTR) replacement model plus four discrete rate categories of gamma (G) and employing 100 bootstrap replicates. PhyloBayes analyses were performed using the same model and repeated using the heterogeneous CAT (plus G) replacement model.

2.4.3 Divergence Estimates

BEAST v2.5 was used to reconstruct phylogenies and to estimate divergence times (R. Bouckaert et al. 2019). We use BEAUti to set the analyses using the following prior information to calibrate the clock. We employed a root prior based on the fruit fly-mosquito split using a normal distribution with mean 260 Ma and a 95% prior distribution to be between 296 and 238 Ma, as indicated by (Benton and Donoghue 2007). We employed three minimum calibration points for the diversification of Anophelinae, Culicinae, and Culicidae, using, respectively, 34 Ma, 34 Ma, and 99 Ma, according to the three oldest fossils known for each of these groups (Misof et al. 2014; Borkent and Grimaldi 2004). Uniform distribution was assigned to each minimum bound with the maximum bounds set at the mean of origin of Diptera employed as root calibration 264 Ma. These calibrations were used for the mtDNA, nDNA, and concatenated datasets. We run BEAST using a different set of (model) priors and choose the most fitting combination of priors using the Harmonic mean, the Akaike Information Criterion (AICm), the stepping stone (SS), and the Path sampling (PS); for the latter, we used the Path sampler package and set the analysis at 50% burn-in with 40 steps of 500,000 chain length. The priors compared were GTR model versus HKY, the coalescent versus speciation (Birth and Death and Yule) tree priors and finally, we compare strict versus relaxed lognormal clock. All other chains were run for 100,000,000 generations until Beast log files indicated proper convergences of all posteriors and the likelihood using Tracer1.7 (R. Bouckaert et al. 2014).

2.5 Results

2.5.1 Conflicts between Nuclear and Mitochondrial Phylogenies

Bayesian inference of the eight concatenated genes dataset under a homogeneous replacement model (GTR, Figure 1A) reveals a generally well-supported tree with the *Aedes* genus divided into two distinct clades as in (Soghigian, Andreadis, and Livdahl 2017): Clade A (in pink, Posterior Probability (PP): 1.00) comprises various species including *A. albopictus* and *A. aegypti*; Clade B (in orange, PP: 1.00) comprises various species often regarded as *Ochlerotatus* plus others referred to as *Aedes* such as *A. koreicus*. Species of the *Psorophora* genus are the sister of Clade A + Clade B (PP: 1.00). Within Clade A we observed two groups (dark pink), one consisting of species attributed to *Stegomyia* + *Armigeres* (clade A1, PP: 1:00), the second containing four genera (*Aedimorphus*, *Catageiomyia*, *Diceromyia*, and *Scutomyia*; PP: 0.92). The mutual relationship of no-Aedini Culicinae is instead unresolved (PP: 0.48); four genera (*Sabethes*, *Wyeomyia*, *Malaya*, and *Toxorhynchites*) form however, a robustly supported (PP:1.00) group, which we have provisionally named Group C.

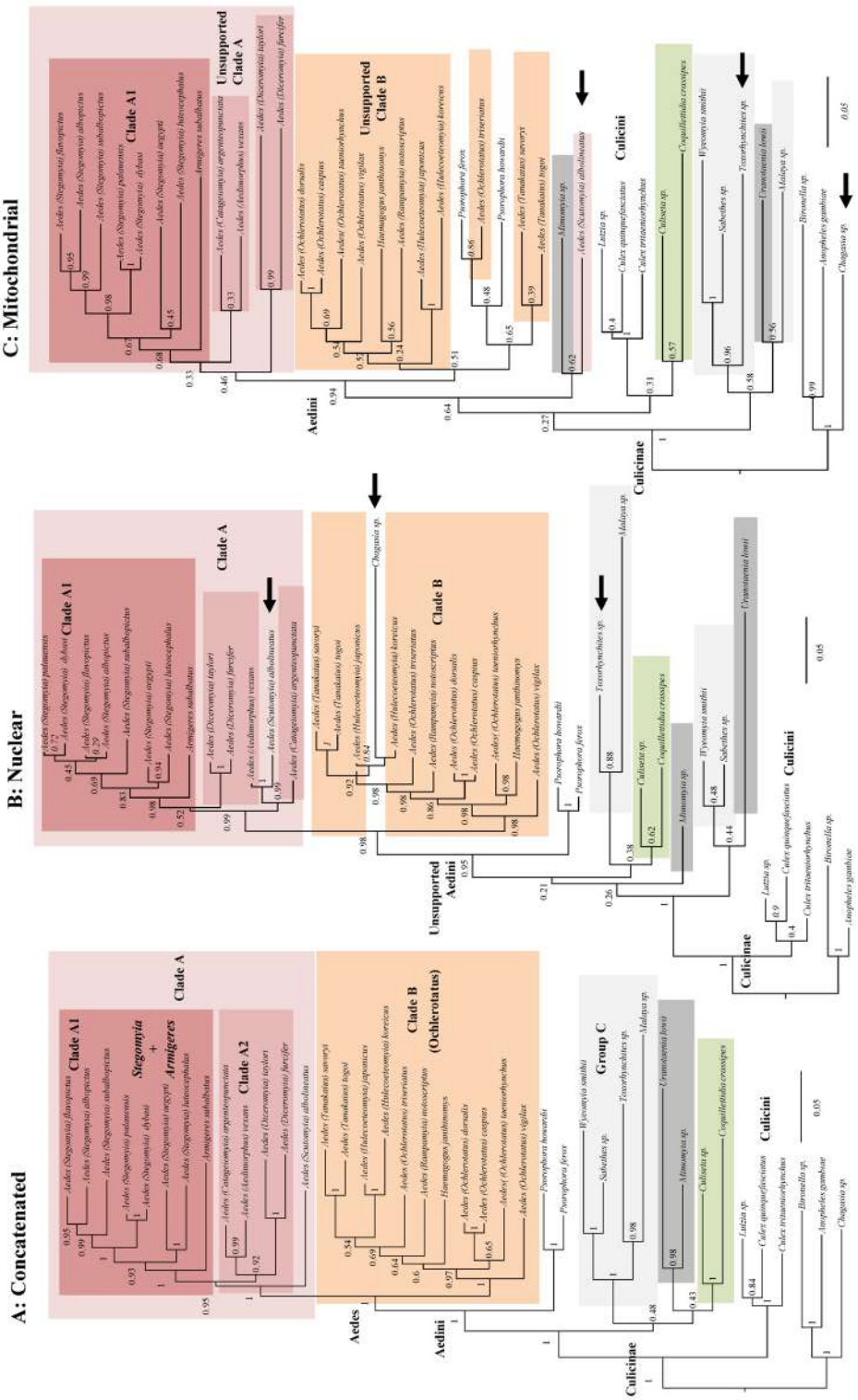


Figure 1. Topological incongruence between mitochondrial and nuclear data. (**A**): Bayesian consensus tree of concatenated dataset. (**B**): Bayesian consensus tree of nuclear (nDNA) dataset. (**C**): Bayesian consensus tree of mitochondrial (mtDNA) dataset. All analyses have been performed using a GTR+G model in PhyloBayes. Numbers at nodes are posterior probabilities (PP). Groups identified using the concatenated dataset have been coloured. Arrows indicate highly supported incongruences between the nDNA and mtDNA datasets. The corresponding Maximum Likelihood trees are in Supplementary Figure S1.

Our concatenated analysis of Figure 1A recovers, at least for most nodes, a robust topology. One of the aims of our study was, however, to disentangle the phylogenetic signal for the Aedini by exploring its consistency over different data types and methodological treatments. We, therefore, analyzed the nDNA and mtDNA datasets separately (respectively, Figure 1B, C), and revealed various instances of mitochondrial-nuclear incongruence. Overall, both trees are less resolved than the concatenated tree (for example, they both do not support Group C nor Group A2), pointing toward the utility of concatenating genes. The nuclear tree is, however, markedly more resolved (it has overall higher supports at nodes) than the mitochondrial one. It does support, for example, the monophyly of *Aedes* and both Groups A and B (all with PP > 0.9), while the mtDNA dataset does not support them. These differences may be explained by less phylogenetic signal in the mtDNA dataset. This is, however, not related to fewer nucleotide positions as the mtDNA alignment is larger than the nDNA one (4224 nt vs. 3270 nt). We identify some interesting cases of well-supported incongruences between the nDNA and the mtDNA trees involving *A. albolineatus*, *A. subalbopictus*, and *Toxorhynchites* sp. (depicted by arrows in Figure 1 B,C). There are various topological incongruences, for example, in the position of *A. subalbopictus*, the two *Psorophora* and *Uranotaenia lowii*, but their affinities did not receive high PP in at least one of the two trees, therefore the nodes that show incongruencies are not considered statistically significant.

2.5.2 A Conservative Picture of Aedini and Other Culicinae Phylogeny

To explore in more detail the phylogenetic signal behind our Bayesian trees of Figure 1, we further performed phylogenetic analyses employing different statistical frameworks, different models of replacement, and types of datasets (Figure 2). In panel A we depict the result of a Maximum Likelihood (ML) analysis of the concatenated dataset. In panel B is the same dataset analyzed in a Bayesian framework using an among-site heterogeneous CAT model more suitable for ancient radiations and saturated datasets (Rota-Stabelli, Daley, and Pisani 2013).

In panel C is the ML analysis of a dataset (named Soghigian) centred on Aedini and derived from (Soghigian, Andreadis, and Livdahl 2017). To provide a

conservative picture of Aedini phylogeny, we have collapsed a node if its bootstrap support (BS) from the Maximum Likelihood (ML) analysis was lower than 75% and if its posterior probability (PP) from Bayesian analysis was lower than 0.9. We found a consistent signal (compare Figure 1A with Figure 2A,B) for a group of *Sabethes*, *Wyeomyia*, and *Malaya* (Sabethini tribe), plus *Toxorhynchites* (Toxorhynchitini tribe) which we have provisionally named Group C. This group is monophyletic using both homogeneous and heterogeneous models of evolution, but its internal relationships, as well as its relative affinity with other Aedini, is inconsistent over different analyses and, in general, not strongly supported. This group is not consistent with a previous multigene phylogeny, which supports *Toxorhynchites* as closely related to *Mymoyia* than to the Sabethini (Reidenbach et al. 2009). Although highly supported in all our concatenated analyses, we advocate caution in considering the validity of Group C, as our analyses may have been biased by an unfortunate combination of reduced taxon and gene sampling; indeed, in most analyses, *Toxorhynchites* is the sister taxa of *Mymoyia*, therefore disrupting the monophyly of Sabethini. Our investigations are instead congruent with previous studies (Huber, Jansen, et al. 2014) in supporting Group A, and to a lesser extent Group A1 (*Stegomyia* + *Armigeres*). Group B is instead supported only by the homogeneous GTR model of evolution. Site heterogeneous models of replacement such as CAT have been repeatedly shown as being capable of reducing systematic errors (Drummond et al. 2006; Lartillot and Philippe 2004); we cannot, therefore, exclude that the signal responsible for Group B is artefactual and we advocate care in considering it as monophyletic. From a systematic point of view, our phylogenies support the classical 10 genera classification of Aedini (Huber, Jansen, et al. 2014; Pombi and Montarsi 2020). Overall, while some nodes are robustly supported in all analyses of Figure 2 (for example, Group A), other nodes are poorly supported or are supported only in one analysis.

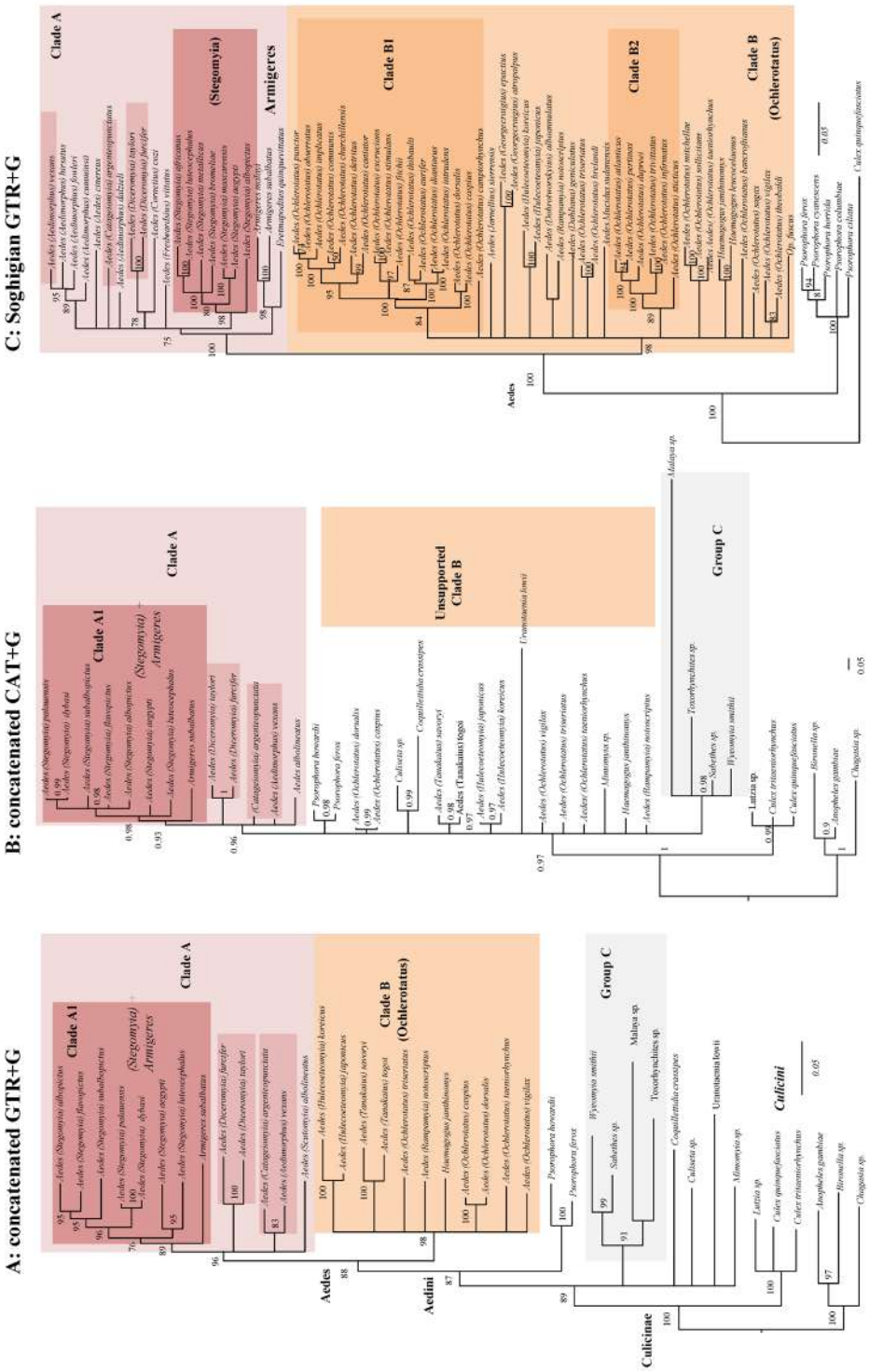


Figure 2. A conservative picture of Culicinae phylogeny using different models and datasets. To highlight lack of phylogenetic signal, all nodes below PP 0.90 and BS 75 have been collapsed. (**A**): Maximum likelihood tree of the concatenated dataset using the GTR+G model in RaXml. (**B**): Bayesian consensus tree of the concatenated dataset under the CAT+G model using PhyloBayes. (**C**): Maximum likelihood tree of a modified (Soghigian, Andreadis, and Livdahl 2017) dataset under the GTR+G model using RaXml. The number at nodes are bootstrap supports (BS) in panels A and C, and posterior probabilities (PP) in panel B. The backbone of the tree and the monophyly of Clade B (orange) are strongly supported using GTR, but not using CAT. Many relationships within Clade B are poorly supported in all analyses. Full trees with all nods and supports are in Supplementary Figures S2–S4.

2.5.3 Divergence estimates of the Aedini

To define which evolutionary models better describe the radiation of mosquitoes in our concatenated dataset, we contrasted the strict clock versus the log-normal relaxed clock models, the coalescent versus the speciation models (Yule and a Birth and Death), and the HKY versus the GTR replacement model (Table 1). The relaxed clock is favoured over the strict clock. Furthermore, under the relaxed clock, the coefficient of variation rate was approximately 0.4 for the mitochondrial data and roughly 1 for the nuclear, further indicating that an uncorrelated clock hypothesis suits better our datasets than a strict clock. This is because if a lognormal clock has a coefficient of variation close to 0, it could be considered clock-like, so comparable with a strict clock (Richards and Murali 2015). Demographic speciation models are more supported than coalescent model, but the stepping stone and path sampling could not discriminate between a Yule and a Birth-Death model; we chose the Yule model because it was favoured by the AICm, which penalizes based on the number of free parameters. The two models provided similar results (Table 1) nevertheless. Therefore, we used a combination of GTR+G, relaxed log-normal, and Yule models for our clock analyses.

Table 1. Model tested with divergence estimates for two nodes.

Clock Model	Substitution Model	Tree Prior	logLikelihood	AICm	Harmonic Mean	PS/SS	Culicidae	Aedini
Strict	GTR	Yule	54,908.6	109,892.3	−54,926.7	4	156 (114–204)	90 (75–104)
Relaxed (LogN)	HKY	Yule	54,624.8	109,467.9	−54,666.3	5	166 (119–215)	96 (68–125)
	GTR	Yule	54,362.7	109,108.1	−54,424.3	1	180 (137–228)	113 (83–143)
		Birth Death	54,363.6	109,115	−54,414.9	1	180 (135–227)	112 (82–142)
		Coalescent Constant	54,370.2	109,208.7	−54,415.7	3	173 (123–225)	100 (66–132)

Our analysis of the concatenated dataset using the best fitting models allows us to obtain a picture of Aedini evolutionary history, which we have contrasted with the appearance of some major vertebrate lineages and flowering plants in Figure 3. According to our posterior estimates, the mosquito family (Culicidae)

diversified in its two subfamilies—Culicinae and Anophelinae—approximately 180 Ma (95% High Posterior Densities, HPD 137–228 Ma) in the lower Jurassic. The earliest fossil of a Chaoboridae, the Culicidae sister group, is 187 Ma (R. Bouckaert et al. 2014). This would suggest a very rapid diversification of Culicomorpha. Our estimates tend to match the proposed origin of angiosperm (van der Kooi and Ollerton 2020). Culicinae diversified in two clades (Culicini and the clade leading to Aedini) between the end of the Jurassic and the early Cretaceous, at 146 (108–182) Ma, while the Aedini tribe diversified at 113 (83–143) Ma with the split of *Aedes* from *Psorophora* genus. Within Aedini, Clade A, and Clade B originated circa 106 (77–133) Ma. Within Clade A, the subgenus *Stegomyia* (which includes model organisms *A. albopictus* and *A. aegypti*) originated 84 (58–109) Ma, concomitantly with the diversification of Clade B (which include the subgenus *Ochlerotatus*) at 86 (61–111) Ma in the late Cretaceous. To test for the effect of outgroup on our dated phylogenies, we repeated the analysis of our concatenated alignment, excluding the Brachycera outgroup. This additional analysis shows that the calibration at the Diptera divergence, provided by Benton (Benton and Donoghue 2007), drives our divergence estimates toward the root. The median height is younger without outgroups, although the two analyses are compatible for what concerns their (overlapping) 95% HPD (Table 2). The rooted tree provided more precise estimates. The 95% HPD is smaller in the root-calibrated phylogeny than in the unrooted one. Overall, our date estimates tend to be slightly younger than the ones provided previously (Soghigian, Andreadis, and Livdahl 2017; Reidenbach et al. 2009; X.-G. Chen et al. 2015) for what concern the origin of Culicinae, but slightly older for what concern the origin of Aedini (see Table 2).

2.5.4 Chronological Incongruences between Nuclear and Mitochondrial Data

Clock analysis using separately nuclear and mitochondrial genes (Figure 4) revealed unexpected strong incongruences. The estimates for the origin of the main mosquito clades (deep nodes of the phylogeny) are similar using the two datasets and reinforce our findings using the concatenated data of Figure 3. For example, Culicinae originated in the early Jurassic and Aedini in the Cretaceous, both in the mtDNA and nDNA analyses. However, there is a strong discrepancy

for what concerns the diversifications within the Aedini lineages. For example, Group A diversified during the Cretaceous using mitochondrial (and also concatenated) data, but is much younger (Paleogene) using nuclear data (Table 2 for details). Even more discrepant is the origin of *Aedes* species: *A. aegypti* and *A. albopictus* split ranges from 81 (61–101) Ma using mitochondrial data, to 30 (15–45) Ma, using nuclear data; the split between *A. albopictus* and *A. flavopictus* is 32 (20–47) Ma using mitochondrial data and just 4 (0.5–11) Ma using nuclear data. Worryingly, those estimates do not overlap at their confidence interval. From a statistical point of view, this indicates that the two datasets reject each other. Overall, the estimates from the concatenated dataset are more similar to those of the mtDNA dataset than the nDNA dataset (Figure 4C–E).

Table 2. Divergence estimates of selected nodes from Figure 3 and other analyses. For each node, we provide the mean and the 95% high posterior density. On the right column, we provide estimates from previous studies.

Node	Taxonomic Level	Concatenated Dataset Figure 3	No Outgroup (Concatenated)	Nuclear Data Figure 4A	Mitochondrial Data Figure 4B	Others: Reidenbach09; Soghigian17 *; Da Silva 20 #; Chen 15 ^
a	Diptera	257 (223–294)		261 (225–296)	258 (224–293)	260 (239–295) ^
b	Culicidae split (Culicinae origin)	180 (137–228)	100 (50–185)	178 (113–245)	182 (143–223)	216 (229–192) 182 # 218 (181–260) ^
c	Culicinae split	146 (108–182)	92 (41–139)	139 (92–194)	150(118–184)	204 (226–172) 130 # 179 (148–217) ^
d		137 (103–173)	86 (38–127)	123 (79–171)	135 (104–164)	
e	Aedini split (Aedes origin)	113 (83–143)	64 (34–122)	92 (55–137)	111 (95–150)	123 (155–90) 125 * 102 #
f	*Aedes* split (Clades A-B split)	105 (77–133)	57 (28–110)	69 (42–103)	107 (85–133)	92 (123–61) 102 *
g	Clade A split	99 (72–126)	51 (24–100)	49 (29–76)	96 (73–118)	
h		83 (59–109)	50 (22–93)	36 (20–57)	92 (71–116)	
i	*Stegomyia* (*A. aegypti–A. albopictus*) split	73 (50–96)	36 (14–70)	27 (15–45)	81 (61–102)	55 * 67 # 71 (44–107) ^
j	*A. albopictus–A. flavopictus* split	28 (14–43)	36 (14–70)	3.7 (0.1–11.2)	33 (20–46)	25 *
l	*A. koreicus–A. japonicas* split	32 (15–51)	14 (3–31)	3.6 (0.2–10.9)	46 (24–71)	20 *

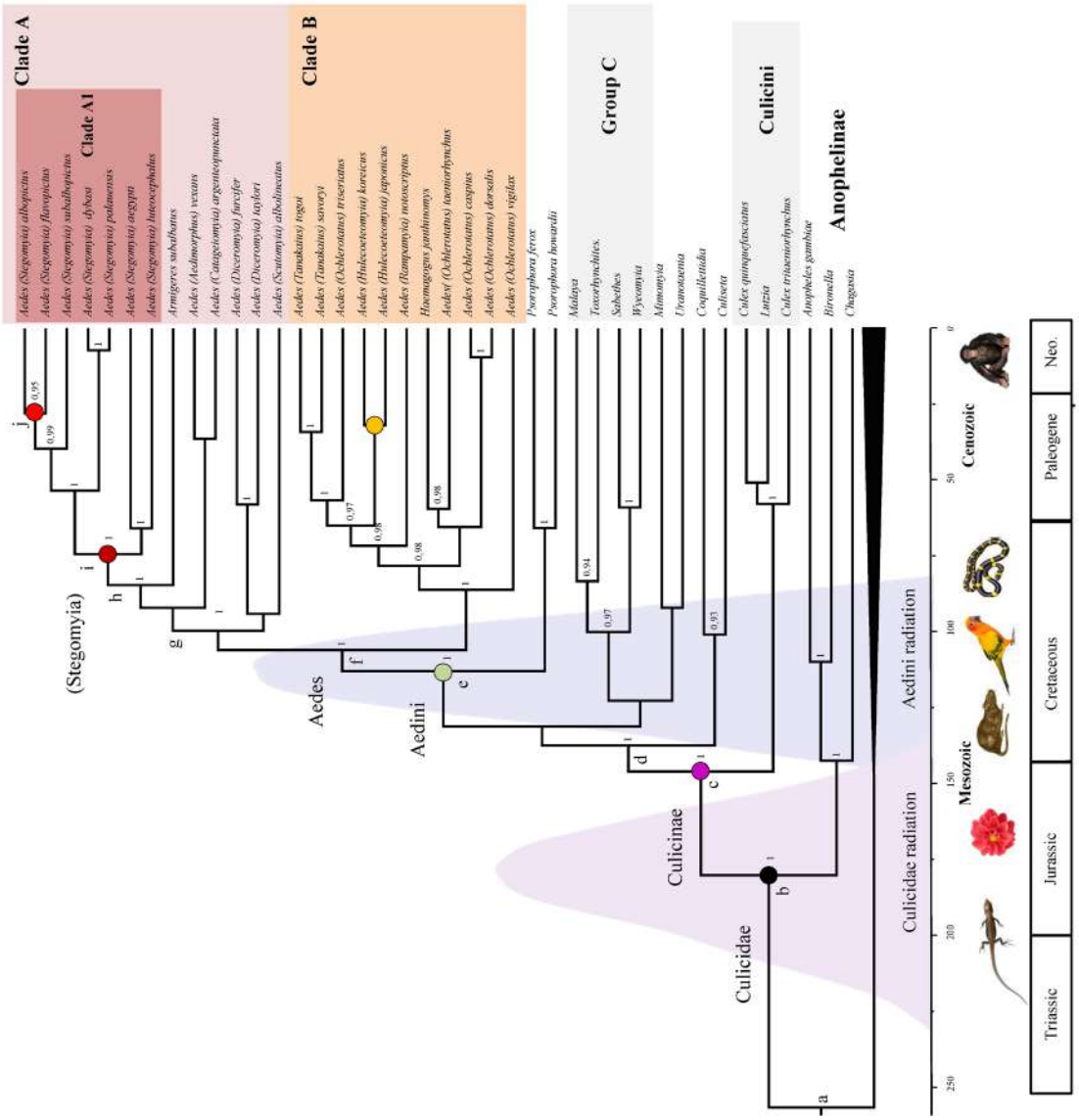


Figure 3. Bayesian estimates of Aedini divergence. Posterior consensus tree from the analysis of the concatenated dataset. The two shaded distributions highlight the distribution of the 95% HPD for the origin of the Culicinae (b node) and the split of the Aedini (e node): for precise estimates and the 95% HPD see Table 2. Supports at nodes are posterior probabilities higher than 0.95. Time is in millions of years before the present.

We further inspected the posterior rates of both the nDNA and the mtDNA trees (Figure 5A,B, respectively; tree topologies are similar to those of Figure 1B,C). Additionally, in the case of rates there are various discrepancies between the two datasets: *A. furcifer* and *A. taylori* are, for example, fast-evolving according to mtDNA, but slow evolving using nDNA. These high rates are likely responsible for the dubious position of these two species in the mtDNA tree of Figure 5B. Our clock analyses returned mean posterior evolutionary rate (calculated over the whole tree) of 1.01×10^{-3} (sd = 1.12×10^{-4}) mutation per site per millions of years (msm) for the mtDNA and of 9.93×10^{-4} msm (sd = 1.8×10^{-4}) for the nDNA.

2.5.5 Mitochondrial-Nuclear Chronological Incongruences Are Consistent over Different Analytical Conditions

We tested the robustness of the chronological incongruence observed between mitochondrial and nuclear data (Figure 4) by verifying if the results are biased by the taxon sampling and the number of gaps in our alignments. We first repeated the clock analyses using a reduced version of our dataset. We excluded three species (*Chagasia*, *Uranotenia*, *Aedes albolineatus*) which had an extremely different branching position in the phylogeny of the nDNA and mtDNA analysis. Results are very similar compared to when using the full dataset for what concerns both the divergence estimates (Figure S5) and the average mutation rate at branches (Figure S6). This indicates that the chronological discrepancy is not due to the presence of these inconsistent taxa (*Chagasia*, *Uranotenia*, *Aedes albolineatus*) in the dataset. We then tested if the pattern we observed is due to a particular taxon and site sampling by repeating the analyses using a different dataset. We inferred divergence estimates using separately the nuclear and mitochondrial partitions of the Soghigian alignment, derived from (Soghigian, Andreadis, and Livdahl 2017) and previously used for Figure 1C. This dataset is characterized by a different taxon representation compared to our dataset (it is centred on *Aedes* and contains few outgroups) and by a higher site representation (contains a lower amount of missing data, see methods for details). Results (Figure S7) provide a similar picture to when using our nDNA and mtDNA. Divergences closer to the root are similar, but those within *Aedes*, including the diversification of Clade A and Clade B are very different. This indicates that the

chronological discrepancy is not due to peculiar taxon or gene sampling nor is affected by the amount of missing data in the datasets.

Because of its higher mutation rate, MtDNA is, in general, more prone to saturation than the nuclear genome (Rota-Stabelli and Telford 2008; Bernt, Braband, et al. 2013). Accordingly, we would expect to underestimate the number of observed mutations in mtDNA dataset compared to the nDNA one, with the consequence that nodes using mtDNA dataset should appear younger than they are. We observe, however, exactly the opposite. Saturation and heterogeneity of the replacement pattern may have nevertheless played a certain role in overestimating the mitochondrial age in our mitochondrial phylogeny. We therefore tested our datasets for saturation by inferring divergences under the CAT model, a mixture model is known to be less sensitive to systematic error in the presence of site-specific saturations (Drummond et al. 2006). The CAT trees are indeed slightly different from those obtained using homogeneous models of replacement (Figure S8). The divergences become more similar between the two datasets, but the nDNA dataset consistently returns younger ages for recent nodes compared to the mtDNA dataset. We conclude that site heterogeneity is only partially responsible for the mitochondrial–nuclear chronological discrepancy.

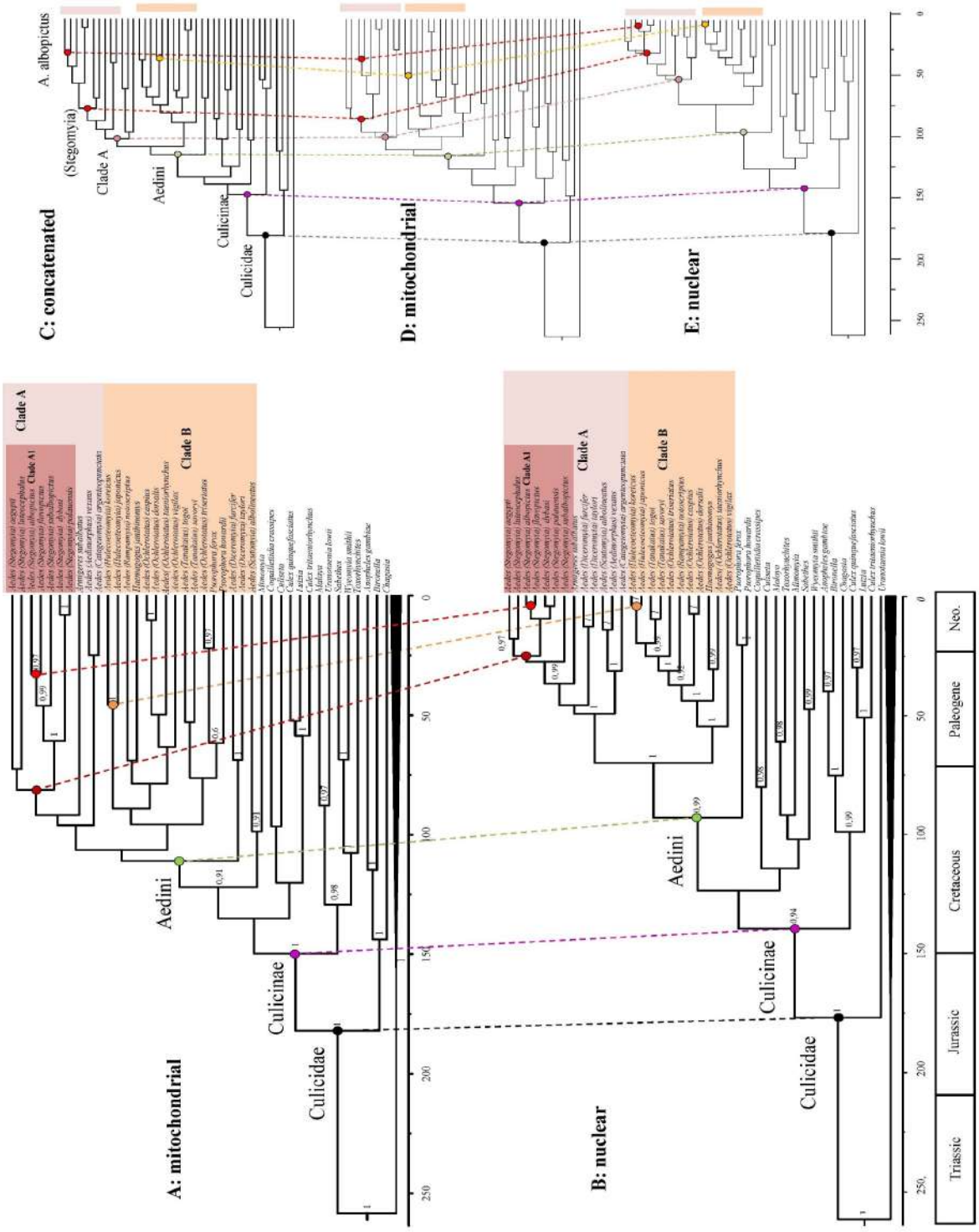


Figure 4. Chronological incongruence between mitochondrial and nuclear data. Note that while posterior estimates are similar for ancient nodes, there are strong incongruences for recent nodes. (**A**): posterior consensus tree from the analysis of the mtDNA dataset. (**B**): posterior consensus tree from the analysis of the nDNA dataset. (**C**–**E**): The concatenated, the mitochondrial, and the nuclear trees simplified for comparison. Supports at nodes are posterior probabilities higher than 0.95. Time is in millions of years before the present.

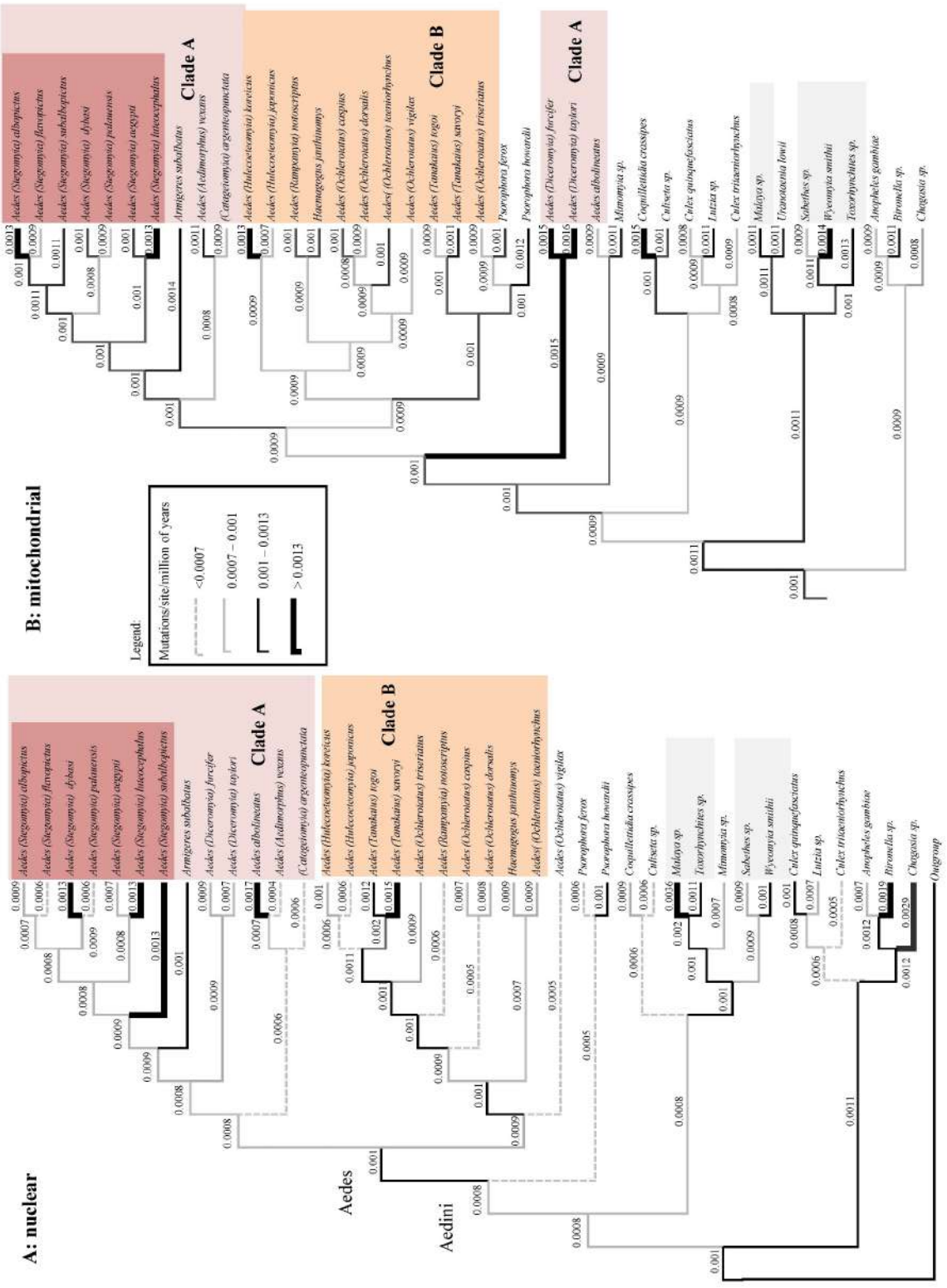


Figure 5. High degree of rate heterogeneity between nuclear and mitochondrial data. (**A**): the nDNA Bayesian phylogeny with mean posterior rates plotted on branches. (**B**): the Bayesian mtDNA phylogeny with mean posterior

rates plotted on branches. Note that there are local accelerations of rate (bold lines) in certain taxa only in one of the two data types.

2.6 Discussion

Our phylogeny of Aedini using a concatenated dataset of eight mtDNA and nDNA markers (Figure 1A) recovers, at least for most nodes, a robustly supported tree topology. Our comparison of mtDNA and nDNA datasets revealed however some unexpected highly supported phylogenetic discrepancies (Figure 1B, 1C). We suggest three explanations for these incongruences. The first is the wrong taxonomic assignment during field collection. Accordingly, one or more genes for some species may come from another (similar and mistaken for) species creating conflicting phylogenetic signals and wrong tree topology. Another, in our opinion less likely, explanation involves complex evolutionary events, such as past hybridization between species, which have resulted in different inheritance patterns for either the mtDNA or some regions of the nDNA. The final explanation is the stochasticity embedded in small (four genes) datasets, such as the ones we have used. The stochasticity in the mtDNA tree may have been exacerbated by systematic errors related to the fast-evolving nature of the mtDNA (Reeves et al. 2018; Bernt, Braband, et al. 2013) and evident high level of apomorphies, as revealed by the longer terminal branches in the mtDNA tree compared with the nDNA one. The different phylogenetic signal, however, does not seem to relate to the amount of missing data as the mitochondrial alignment is more complete than the nuclear one (38% of missing data in mtDNA vs. 45% in nDNA). Moreover, some inconsistent results between the two phylogenies (mDNA and nDNA) could be addressed to the incomplete linage sorting. Whatever the source of the topological discrepancies between datasets, our result points toward the limitation of a PCR-scaled approach for Aedini phylogeny and point toward future studies based on whole mtDNA and genome-scaled nDNA dataset. Indeed, undetected stochastic and systematic type of errors may also affect the concatenated dataset, as we have shown that the phylogenetic signal is unstable at many nodes when employing different replacement models and statistical frameworks (Figure 2B and Figure S4). In particular, the poor support using heterogeneous models may be due to the ability of this model to detect saturated or fast-evolving sites (Drummond et al. 2006). Under this scenario, the highly supported Clade B when using the homogeneous GTR model (Figures 1A and 2A) may be the result of a systematic error. The various phylogenetic

incongruences we observe using different replacement models (Figure 2A,B) reinforce what we have found when comparing nuclear and mitochondrial data. They alert us of possible systematic and stochastic errors. We advocate adopting a cautious, conservative way in interpreting our (but also other available (Soghigian, Andreadis, and Livdahl 2017; Reidenbach et al. 2009; A. F. da Silva et al. 2020)) trees of the Aedini based on few genetic makers, as seemingly high supports (as in our Figure 1A) are not consistent over data type (Figure 1B, C and Figure 2C), method of inference (Figure 2A) or replacement model (Figure 2B). In perspective, our data indicate that the phylogeny of Aedini should be resolved with confidence only using a genome-scaled nuclear and a complete mtDNA dataset as done in other dipteran studies (Ometto et al. 2013).

Our divergence estimates using both the concatenated and the mtDNA and nDNA datasets (Figures 3 and 4) are concordant in indicating that mosquitoes radiated from the mid-Jurassic on and that Aedini radiation started in the mid-Cretaceous, quite concomitant with the origin and the earliest diversifications of mammals first, and later birds during the Cretaceous. We cautiously speculate that there may have been a general history of co-radiation (the available data do not provide enough evidence to advocate co-evolution) between the Aedini and warm-blooded vertebrates. In support of this hypothesis, the Aedini group has a specific preference for mammals and birds (Rota-Stabelli et al. 2010; Rota-Stabelli and Telford 2008). The fact that a relaxed clock better fits our Aedini concatenated dataset is not surprising considering that a large variety of ecological traits characterizes mosquitoes and demographic habits (Soghigian, Andreadis, and Livdahl 2017), which can be responsible for different generation times and therefore different branch rates (Ometto et al. 2013). The mean posterior rate for the mtDNA dataset is 1.01×10^{-3} msm, higher than the 9.93×10^{-4} msm estimated for the nuclear genes. The higher mutation rate of mitochondrial genes is expected as the mtDNA is well known to evolve faster than the nuclear genome in animals (Papadopoulou, Anastasiou, and Vogler 2010). Our mean mtDNA rate estimates are, however, circa one order of magnitude smaller than the mitochondrial COI rate of coleopterans (1.17×10^{-2} msm) inferred by Papadopoulou et al. (Papadopoulou, Anastasiou, and Vogler 2010). This can be explained by different timespan between the latter and our dataset. Indeed, shallow phylogenetic studies consistently recover faster evolutionary rate than

deep phylogenies (S. Y. W. Ho and Lo 2013; S. Y. W. Ho et al. 2011). Our nuclear rate estimates are in line instead with those inferred over long phylogenetic distances using Ecdysozoa nuclear data (mean 1.01×10^{-3} msm), (Guidetti et al. 2017), but lower than those based on mitochondrial *Drosophila* data (7.9×10^{-3}) (Ometto et al. 2013); this indicates that mosquitoes may have been characterized during their radiation by an overall smaller number of generations per year compared to *Drosophila*. We found that the nDNA data of most lineages within Clade A evolves faster than in the lineages of Clade B; this pattern is less marked, but conserved in the mtDNA data (Figure 5). A possible explanation for this pattern is that species of Clade A have in general more generation per year than those of Clade B. The two important invasive *Aedes* species, *A. albopictus*, and *A. koreicus* are characterized by markedly higher replacement rate if compared with their respective sister species, *A. flavopictus* and *A. japonicus*; this pattern can be observed for both mitochondrial and nuclear data. Assuming that the instantaneous mutation rate is conserved within the genus, this result suggests that *A. albopictus* and *A. koreicus* are characterized by a higher number of generations per year compared to other closely related *Aedes*, a hypothesis which may at least partially explain their high invasive potential.

Our analyses revealed a consistent chronological incongruence between the phylogenetic signal of nuclear and mitochondrial genes. mtDNA provides divergence times within Aedini significantly older than nDNA. Previous clock studies in insects have shown poor (Ometto et al. 2013) to moderate (Wahlberg et al. 2009; Andújar, Serrano, and Gámez-Zurita 2012) incongruence between nuclear and mitochondrial data. In these analyses, mitochondrial and nuclear estimates, although different, were overlapping for what concerns their 95% HPD in older nodes. Whereas the 95% HPD of younger node do not overlap, indicating a statistically significant incongruence. We have shown that these incongruences do not depend on rough taxa (compare Figure 4 and Figure S5), nor on-site occupancy and gene sampling (compare with Figure S7), although there is a mitigation of the discrepancies when using a heterogeneous model of replacement (compare with Figure S8). We conclude that the mtDNA–nDNA chronological incongruence in Aedini data does not depend on analytical conditions, although the correct interpretation of saturation in both datasets, particularly the mtDNA one, may play a certain role. Based on our results, we

cannot exclude that there may have been a long history of multiple hybridization events within *Aedes* species, which have affected the mitochondrial genome differently than the nuclear one. Indeed, complex phylogenetic signal due to multiple hybridization events has been recently shown in the *Anopheles* mosquitos (Thawornwattana, Dalquen, and Yang 2018; P. G. Foster et al. 2017). The observed discrepancies prevent us from drawing a conclusion on the actual timing of diversification of model organisms such as *A. albopictus*, whose mean split from sister species *A. flavopictus* may dramatically range from 32 Ma using mitochondrial to just 4 Ma using nuclear data. In light of these results, we advocate that future research should concentrate on determining the biological (or methodological) reason for this discrepancy by comparing time-trees from whole mtDNA genomes with those from genome-scaled sampling of nuclear genes.

In conclusion, we have provided here a detailed analysis of the phylogenetic and chronological signal in currently available nuclear and mitochondrial genes of the Aedini. Overall, our data point toward the limitation of a multigene PCR-scaled approach for Aedini phylogeny and indicate that future research should be based on genome scaled data. Probably our most interesting finding is the strong chronological incongruence between the nuclear and the mitochondrial data. We could exclude various possible misleading factors such as taxa assignment, missing data, and saturation (Figures S5–S7), but could not ultimately test a stochastic effect related to using only eight genes. This is because at present there is not enough data in databases to build a taxon-rich genome-scaled dataset centred on Aedini. The incongruences we have identified do not currently allow defining the exact timing of evolution of important model organisms, such as *A. aegypti* and *A. albopictus* (Palatini et al. 2020). We advocate that these chronological incongruences should be investigated in future by comparing whole mitogenomes with genome-scaled nuclear data as we have done for example, in *Drosophila* (Wahlberg et al. 2009).

Supplementary Materials: The following are available online at https://www.mdpi.com/2075-1729/11/3/181/s1.

Author Contributions: O.R.-S. designed the study. N.Z. and O.R.-S. performed phylogenetic and molecular clock analyses, O.R.-S., N.Z. and A.R. interpreted the results, O.R.-S. and N.Z. wrote the article. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

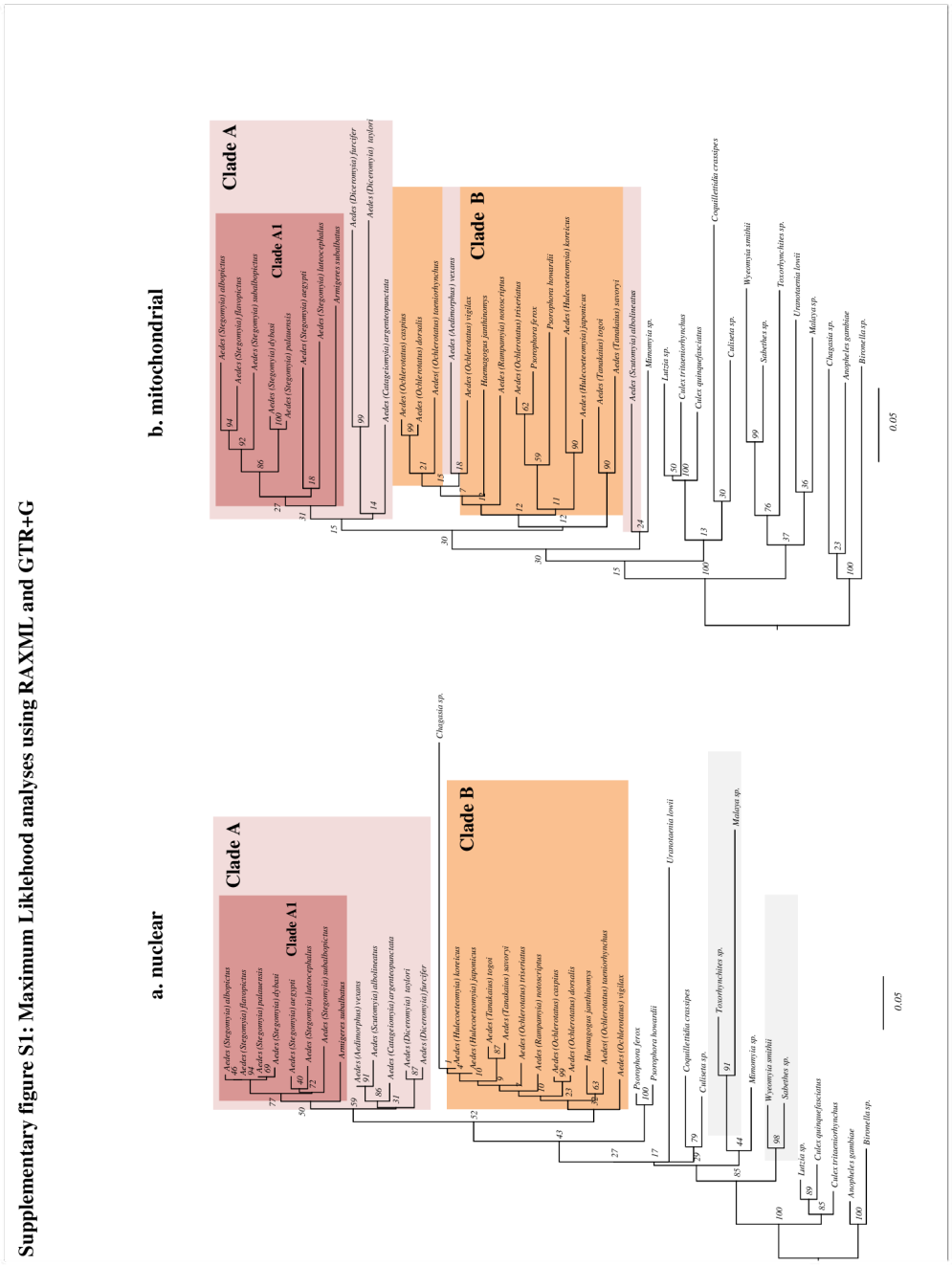
Informed Consent Statement: Not applicable.

Data Availability Statement: All data generated for this study are included in this article and its supplementary information file.

Conflicts of Interest: The authors declare no conflict of interest.

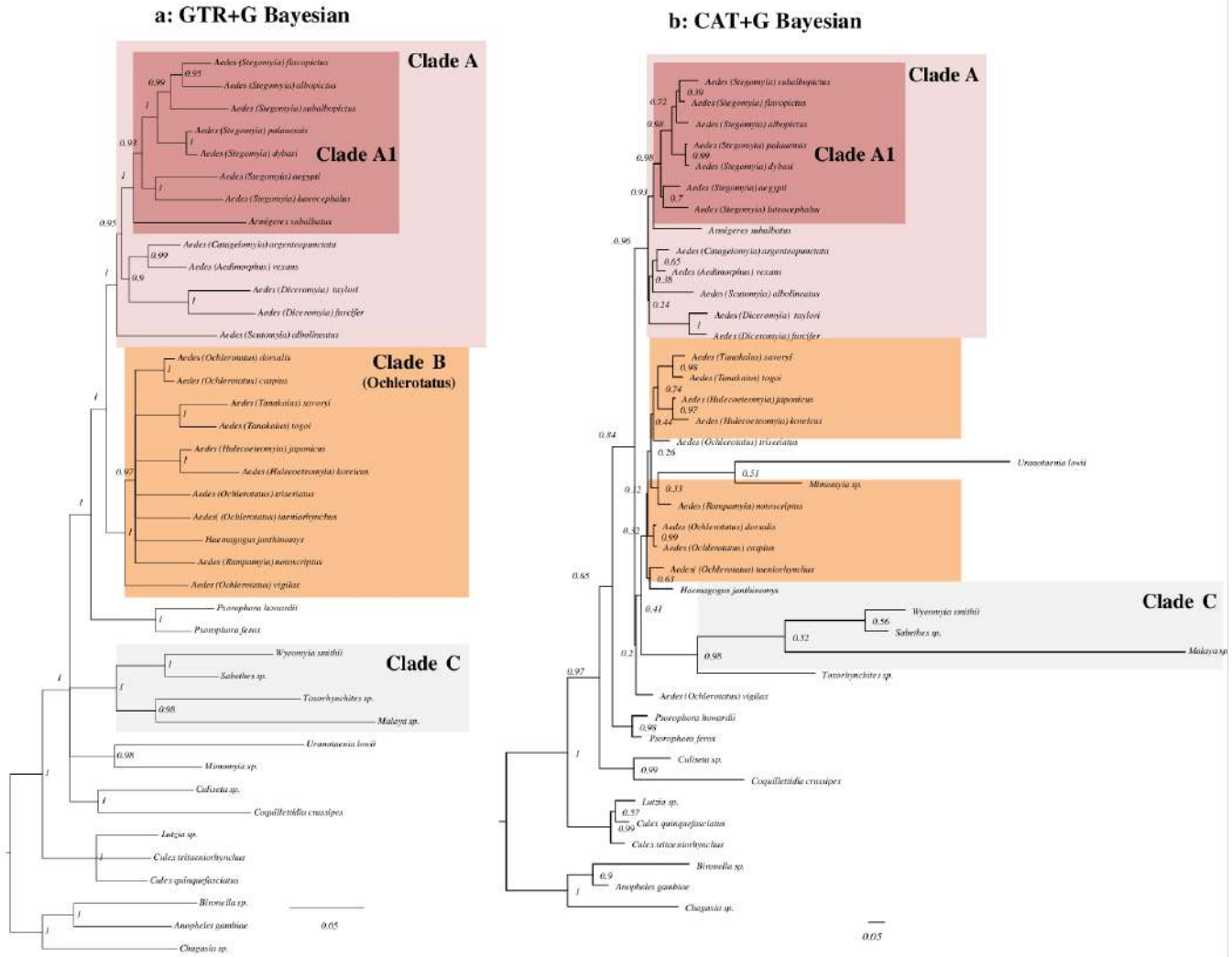
2.7 Supplementary material

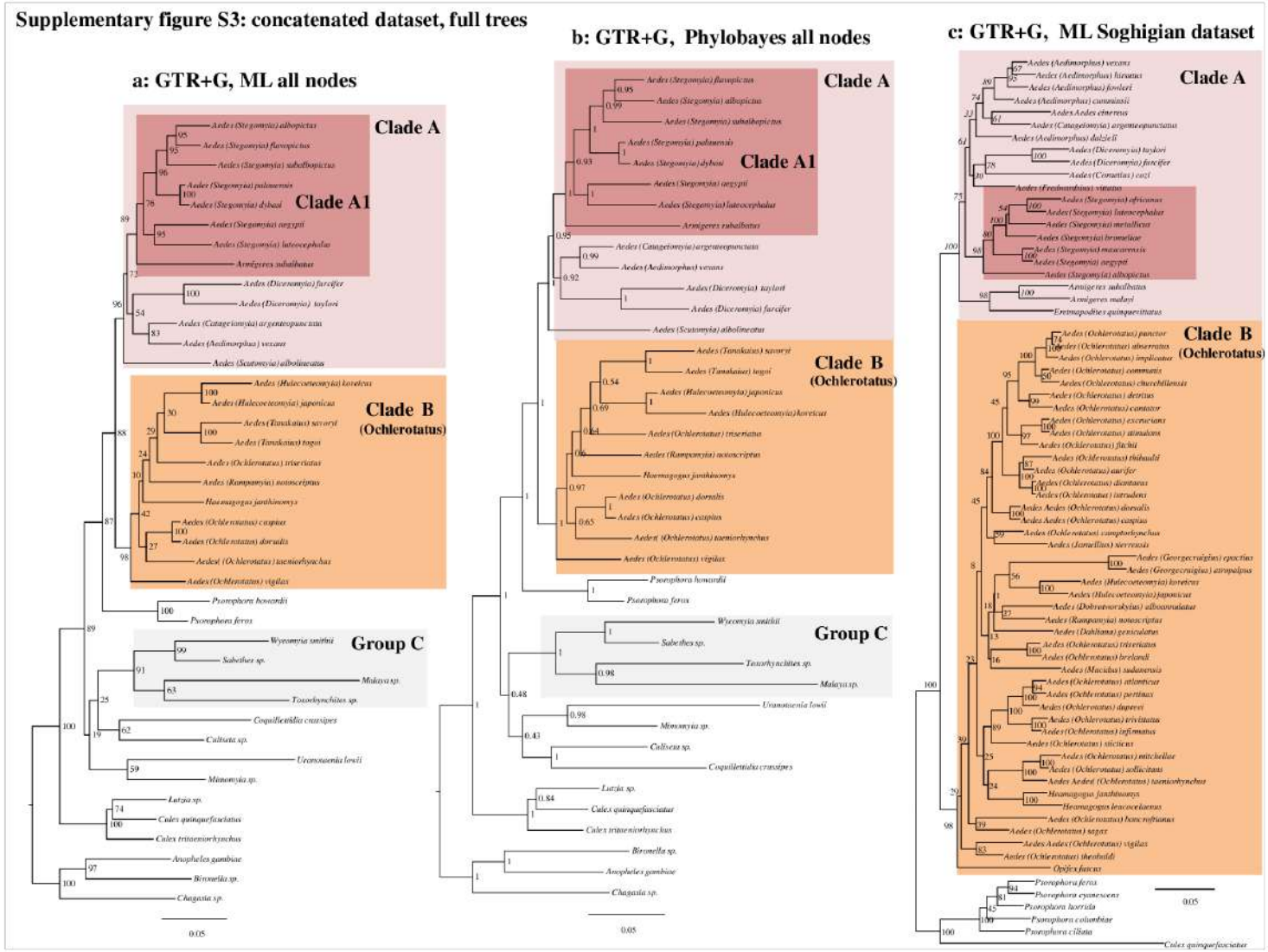
2.7.1 Figure



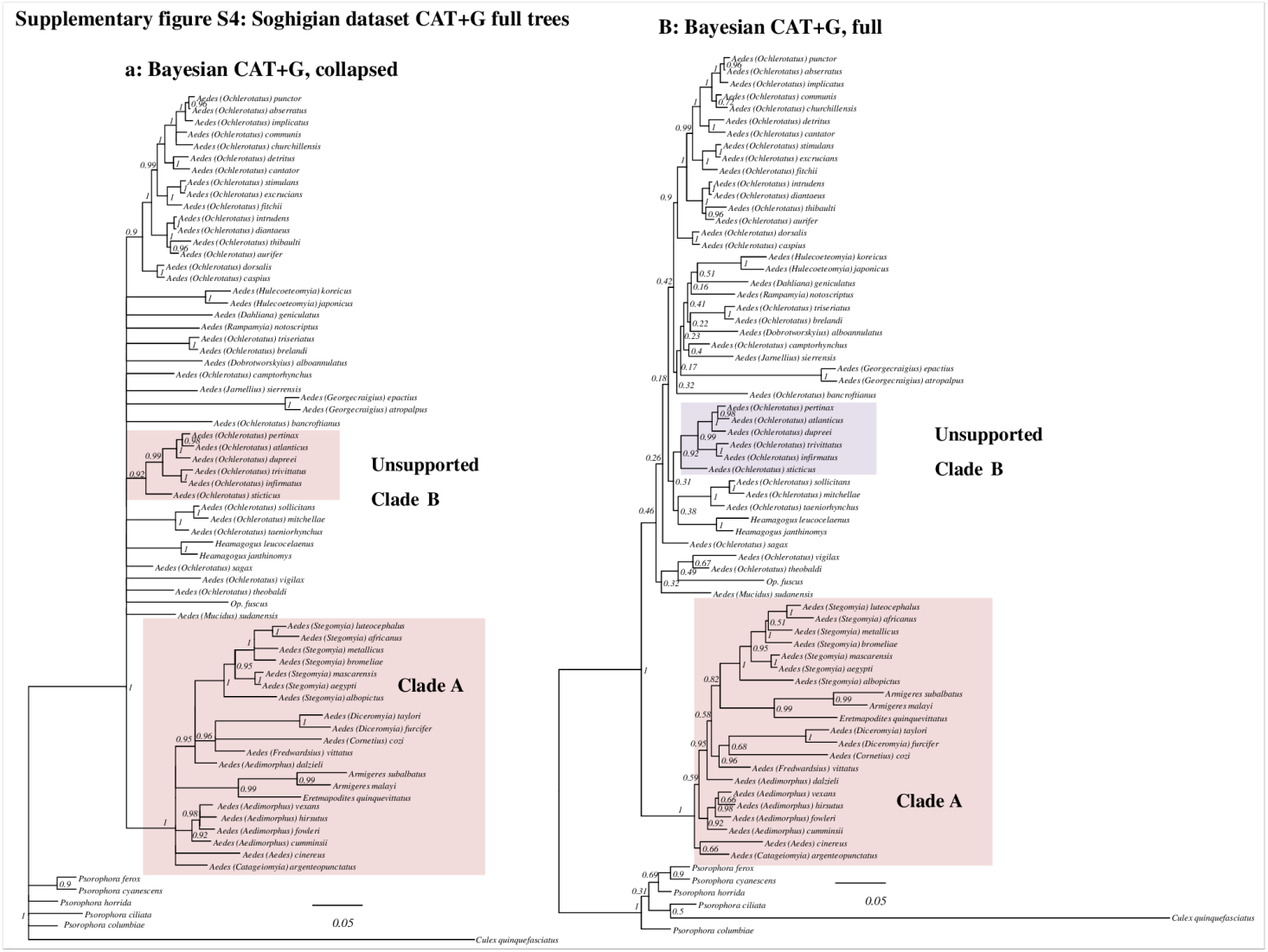
Supplementary figure S1: Maximum Likelihood analyses using RAXML and GTR+G

Supplementary figure S2: Bayesian trees of concatenated datasets

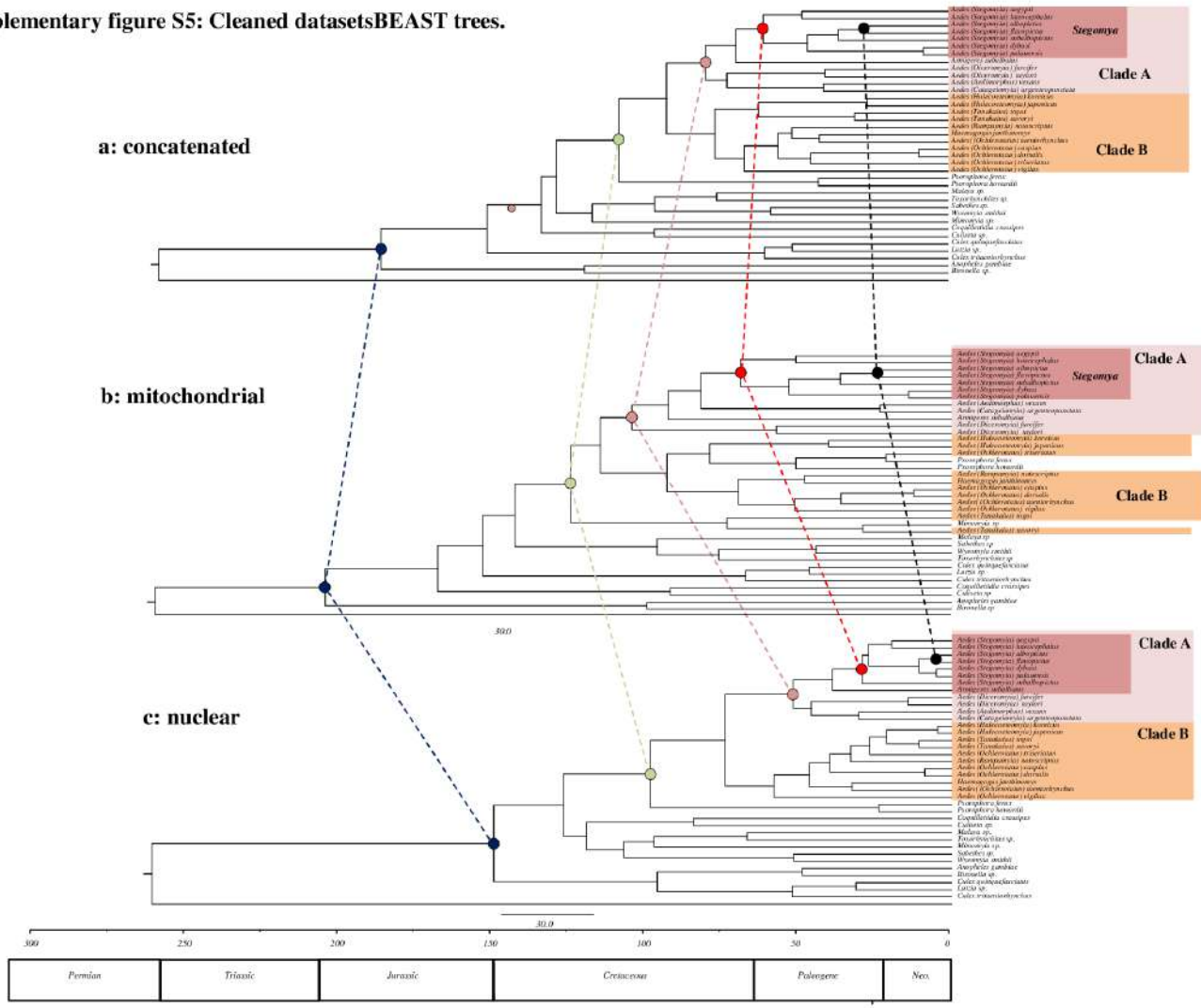



Supplementary figure S3: concatenated dataset, full trees
a: GTR+G, ML all nodes
b: GTR+G, Phylobayes all nodes
c: GTR+G, ML Soghigian dataset
Clade A
Clade A1
Clade B
(Ochlerotatus)
Group C
0.05

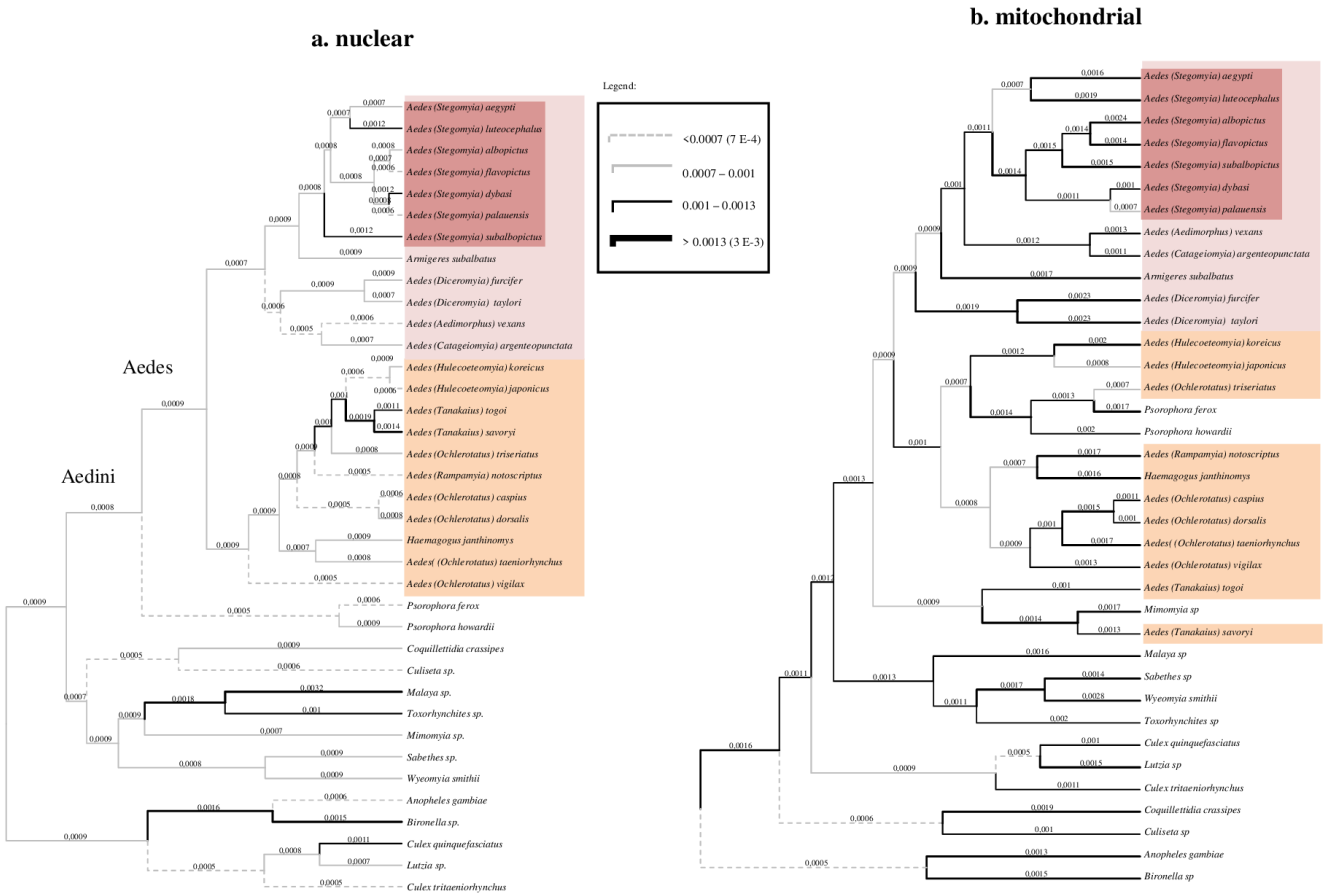
Supplementary figure S4: Soghigian dataset CAT+G full trees



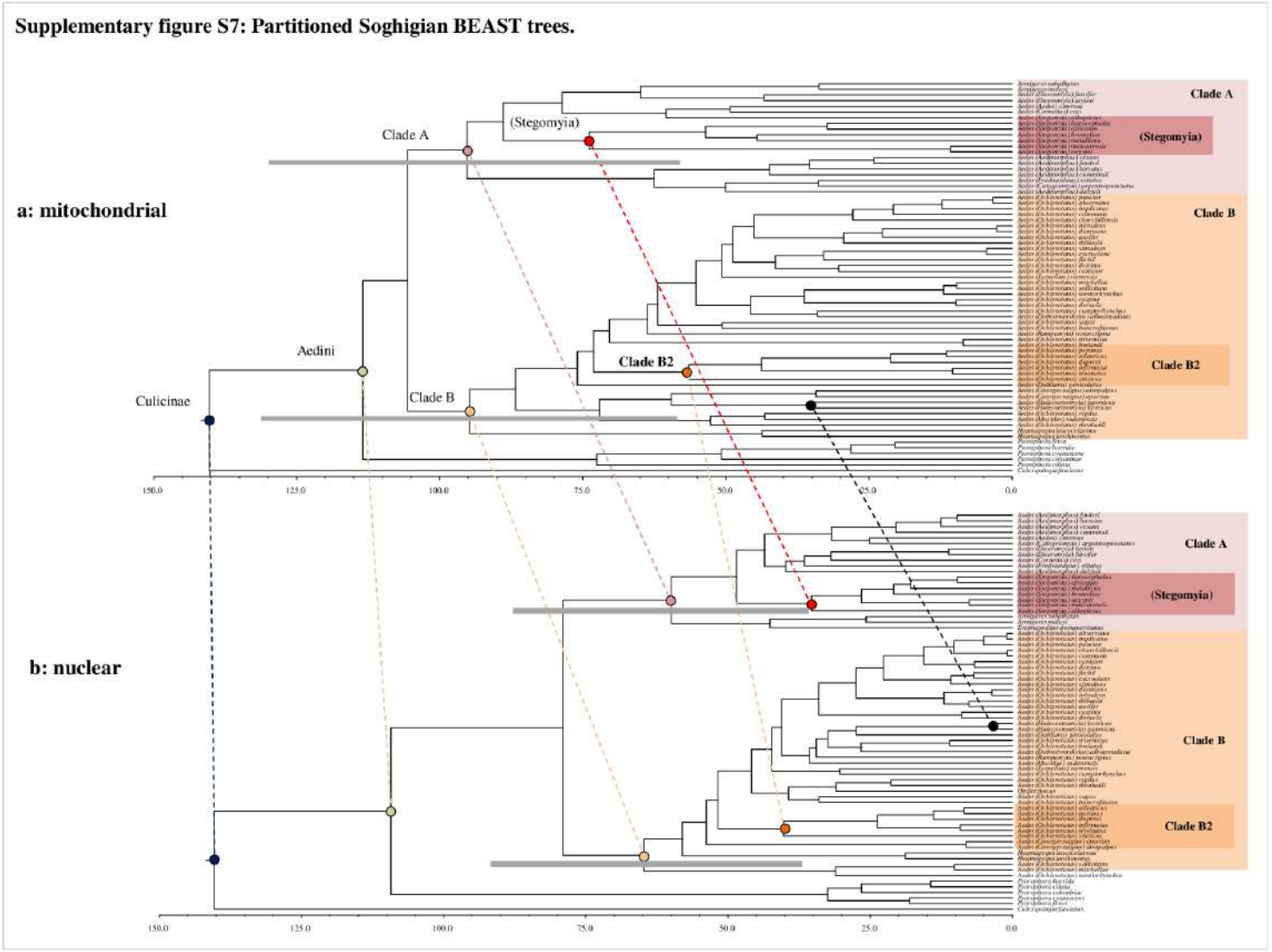
Supplementary figure S5: Cleaned datasetsBEAST trees.



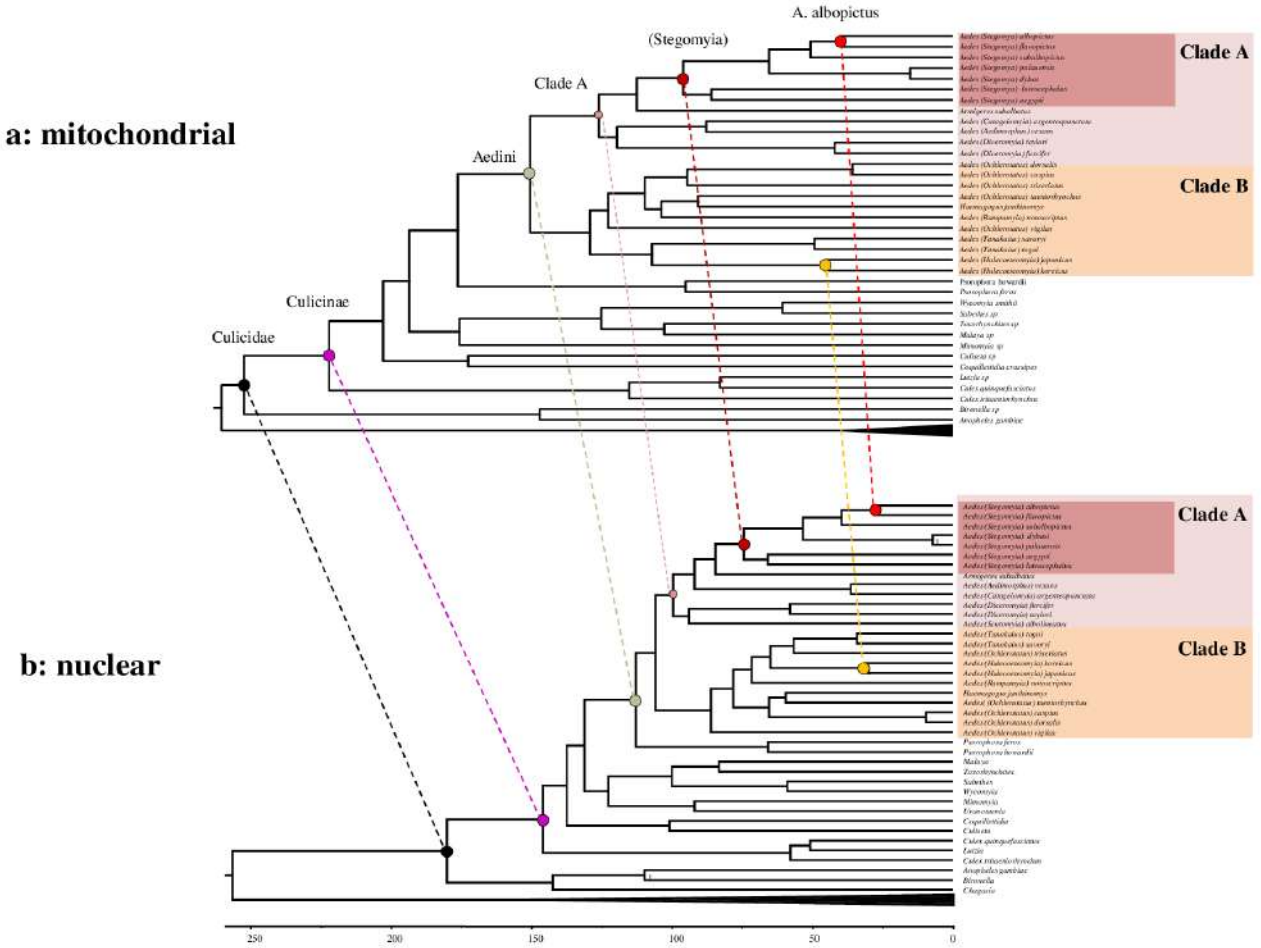
Supplementary figure S6: Cleaned datasets BEAST trees. Posterior mean rates at nodes



Supplementary figure S7: Partitioned Soghigian BEAST trees.



Supplementary figure S8: Phylobayes timetrees using the CAT+G model



2.7.2 Supplementary Table

Supplementary Table 1

Species name	Nuclear genes				Mitochondrial genes			
	28S	**18S**	**Enolase**	**Arg-kinase**	**16S**	**CoxI**	**CoxII**	**NAD4**
Aedes (Stegomyia) aegypti	MG242540.1	MG232525.1	MG232460.1	MG232457.1	EU352212.1			
Aedes (Scutomya) albolineatus	KF687500.1					KF564650.1	KC913592.1	
Aedes (Stegomyia) albopictus	MG242543.1	X57172.1	MG232463.1	MG232454.1	AY072044.1			
Aedes (Stegomyia) dybasi	DQ397943.1				DQ397925.1	DQ397913.1		
Aedes(Stegomyia) flavopictus	DQ397936.1				DQ397929.1	DQ397907.1		
Aedes (Diceromyia) furcifer	MG242572.1	MG232557.1	MG232487.1	MG232431.1		KU187185.1	AY645287.1	
Aedes (Huleocoetomya) koreicus	GU229913.1					KT962063.1	GU229897.1	KT945239
Aedes (Stegomyia) luteocephalus	MG242581.1	MG232566.1	MG232495.1	MG232424.1		AY645234.1	AY645300.1	JX427527.1
Aedes (Rampamyla)notoscriptus	MG242585.1	MG232570.1	MG232499.1	MG232420.1	KM676218.1			
Aedes (Stegomyia) palauensis	DQ397942.1				DQ397924.1	DQ397896.1		
Aedes (Stegomyia) subalbopictus	KF687510.1					EU259301.1	KC913583.1	
Aedes (Diceromyia) taylori	MG242596.1	MG232580.1	MG232508.1	MG232411.1		AY645228.1	AY645295.1	
Aedes (Tanakayus) togoi	GU229910.1		LC026054.1			LC025711.1	GU229895.1	GU229922.1
Aedes (Aedimorphus) vexans	MG242602.1	AM071382.1	MG232513.1	MG232406.1		AY645243.1	AY645304.1	
Anopheles gambiae	GCA_000005575.1 (genome)				MG753708.1			
Armigeres subalbatus	MG242607.1	MG232591.1	MG232516.1	MG232402.1	AY439688.1	KM497419.1	HQ398972.1	AY440746.1
Bironella_sp.		AF417796.1	GQ906881.1	GQ906808.1	MF381612.1			
Calliphora vomitoria	MF281681.1	MF992252.1			KT444440.1			

Aedes (Catageiomyia) argenteopunctata	MG242545.1	MG232530.1	MG232464.1	MG232453.1		MG242468.1		
Chagasia_sp.	AF417831.1	AF417797.1			MF381717.1			
Chrysomya megacephala	KY197787.1	FJ025483.1			KT272787.1			
Coquillettidia crassipes	KF687592.1	AY988454.1	GQ906883.1			KF564770.1	KF687365.1	
Culex quinquefasciatus	XM_001865926.1	XM_001849752.1	XM_001842566.1	XM_001849602.1	GU188856.2			
Culextritaeniorhynchus	KF687515.1	U48385.1			KT852976.1			
Culiseta sp.	KF687543.1	AY988451.1	GQ906882.1fla	GQ906809.1		AF425848.1		
Drosophila pseudoobscura	HQ110539.1	XR_053284.1	AF025805.1	XM_015187032.1	FJ899745.1			
Drosophila sechellia		XR_048770.1	XM_002041688.1		AF200832.1			
Drosophila simulans	HQ110541.1	AY037174.1	DQ864222.1	XM_016184539.1	AY518672.1			
Drosophila yakuba	HQ110545.1		DQ864189.1	XM_002093328.2	KF824886.1			
Haemagogus janthinomys	MG242608.1	MG232592.1	MG232517.1	MG232401.1	KT372555.1			
Lucilia caesar	MF694304.1	FJ025492.1			KM657111.1			
Lucilia cuprina	FR719299.1			JQ088101.1	JX913750.1			
Lucilia sericata	KY197804.1:	KR133393.1			KT272854.1			
Lutzia sp.	KF687516.1					HQ398896.1	HQ398958.1	
Malaya sp.	KF687541.1			GQ906814.1		EU259293.1	KF687373.1	
Mimomyia sp.	KF687538.1		GQ906889.1	GQ906816.1			KU380447.1	
Aedes (Ochlerotatus) caspius	MG242555.1	MG232540.1	MG232473.1	MG232444.1		KM452934.1		
Aedes (Ochlerotatus) dorsalis	MG242565.1		MG232480.1	MG232436.1	KU880657.1	KR691544.1	KC913569.1	
Aedes (Huleocoetomya) japonicus	MG242580.1	MG232565.1		MG232425.1	DQ397915.1	DQ397889.1 KC913579.1		DQ470164.1
Aedes (Ochlerotatus) taeniorhynchus	MG242595.1	MG232579.1	MG232507.1	MG232412.1		KT766538.1	HQ853680.1	

Aedes (Ochlerotatus) triseriatus	MG242600.1	MG232584.1	MG232511.1	MG232408.1		AF417730.1	AF417766.1	
Aedes (Ochlerotatus) vigilax	MG242603.1	MG232587.1	MG232514.1	MG232405.1	MK575484.1			
Psorophora ferox	MG242613.1	MG242613.1	MG232521.1	MG232396.1		MF172349.1		
Psorophora howardii	MG232599.1	MG232599.1	MG232523.1			MG242538.1		
Sabethes sp.				GQ906821.1	NC_037498.1			
Sarcophaga sp.	KU746469.1	AF322419.1			KM881633.1			
Aedes (Tanakaius) savoryi	LC025745.1		LC026055.1			LC025656.1		
Toxorhynchites sp.	KC177664.1	AY988455.1	GQ906895.1	GQ906823.1		AF425849.1		
Uranotaenia lowii	AF417833.1	AF417799.1		GQ906826.1		AF417728.1	AF417764.1	
Wyeomyia smithii			GQ906899.1	GQ906827.1		HM136816.1		

CHAPTER 3 - THE MITOCHONDRIAL GENOMES OF INVASIVE MOSQUITOES *AEDES KOREICUS* AND *AEDES JAPONICUS* FROM ITALY AND AN UPDATED CLOCK ANALYSIS OF MOSQUITO MITOGENOMES

3.1 Introduction to the Chapter

In this Chapter, I further investigated the clock signal in mitochondrial sequences of Aedini mosquitoes. I have used whole mitochondrial genomes to clarify the peculiar chronological incongruence revealed in **Chapter 2**. I employed the newly assembled mitogenomes of *A. koreicus* and *A. japonicus* exploiting the genome data presented in Chapter 4A. I investigated divergences in the light of new outgroups by adding *Corethrella condita, Dixella sp., Chaborus sp.* and Chironomidae samples to increase the phylogenetic resolution on the backbone of the Culicomorpha clade. Results revealed the importance of taxon sampling in molecular clock studies of Aedini. Indeed, increasing the taxon sampling within the Culicomorpha clade provides a more recent estimate for the diversification of Culicidae family. In addition, I investigate the recent diversification of *A. koreicus*, uncovering a high diversity between the strain we collected in Friuli and the strain from Korea. This difference is greater than expected, suggesting the presence of a possible species complex involving *A. koreicus* and *A. japonicus*. Further investigations of the inner diversity within this putative species complex may provide interesting insights into the evolution of these two related *Aedes* species that may be exploited for their management and control.

Contribution: For this study, I was in charge of all the steps from the DNA extraction to assemblies and phylogenetic analysis. Moreover, supervised by Omar Rota-Stabelli, I interpreted the results and drafted the article.

The mitochondrial genomes of invasive mosquitoes *Aedes koreicus* and *Aedes japonicus* from Italy and an updated clock analysis of mosquito mitogenomes.

Nicola Zadra[1,2], Daniele Arnoldi[1], Fabrizio Montarsi[3], Annapaola Rizzoli[2], Omar Rota Stabelli [1,2,4]

[1] CIBIO Department, University of Trento; [2] Fondazione Edmund Mach; [3] Laboratorio di Parassitologia, Istituto Zooprofilattico Sperimentale delle Venezie; [4] Center C3A, University of Trento.

3.2 Abstract

Aedes japonicus and *Aedes koreicus* are two mosquitoes that are quickly establishing in Europe. Here we present the mitochondrial genomes of these two species from individuals sampled in North Italy and provide an up to date estimate of mosquito divergence time based on calibrated relaxed clock analyses of whole mitochondrial DNA data.

The reconstructed mitogenomes possess the typical gene content and gene order of other Aedini mosquitoes. Genetic distances and divergence estimates between Korean and Italian samples of *A. koreicus* support the proposed hypothesis of a species complex involving *A. koreicus* and *A. japonicus*.

Compared to previous studies, we retrieved more recent estimates for the origin and radiation of Aedini and Culicini, whereas the estimates for shallower nodes are congruent with previous works. We hypothesize that the reason behind this discrepancy is that our mitogenomic analysis is the first to include more outgroups to the Culicidae: this may have adjusted previous estimates for deep nodes that were biased by a reduced outgroup sampling.

Our results recalibrate the origin of some mosquitoes clades by moving their diversification toward the present and provide new useful data to understand the diversity of *A. koreicus* and *A. japonicus* in newly invaded areas and in relation to the other mosquitoes.

3.3 Introduction

Aided by global trade, increased travelling, and global warming, various Aedini mosquitoes competent for arboviruses are becoming endemic in Europe, and North America. As a consequence, diseases that were previously only tropical are now arising in these invaded territories, posing public health concerns (Jolyon M Medlock et al. 2018; E. C. Cameron et al. 2010; F. Schaffner, Medlock, and Van Bortel 2013; Moin-Vaziri et al. 2019; Weger-Lucarelli et al. 2016). Because of their capability of spreading several arboviruses and adapting to urban environments, *Aedes aegypti* and *Aedes albopictus* quickly became model organisms within Aedini (Kraemer et al. 2015; N. R. Faria et al. 2017).

Two other Aedini are quickly establishing in temperate climates of Europe: *A. japonicus* is well established in central Europe (Francis Schaffner, Chouin, and Guilloteau 2003) and is now enlarging its areal to North Italy (Bernhard Seidel, Montarsi, et al. 2016) while *A. koreicus* is now well established in the Veneto region of Italy (Capelli et al. 2011) and enlarging his areal toward Trentino (F. Baldacchino et al. 2017; Montarsi et al. 2015) and Lombardia regions (Negri et al. 2021). Like most other *Aedes*, these invasive species show human blood-feeding behaviour and are competent for many arboviruses such as West Nile virus, Zika virus, and Yellow fever virus (Jolyon M Medlock et al. 2018; Huber, Jansen, et al. 2014; Capelli et al. 2011). *Aedes japonicus* has established in Europe in the mid-2000 (Versteirt et al. 2009; Francis Schaffner, Chouin, and Guilloteau 2003), and it has become stable in Italy only recently (Frédéric Baldacchino et al. 2015; Bernhard Seidel, Nowotny, et al. 2016). While species such as *A. albopictus* typically require warm temperatures, *A. koreicus* and *A. japonicus* seem particularly well adapted to more temperate environments (Montarsi et al. 2015; Bartlett-Healy et al. 2012). From an evolutionary point of view, *A. koreicus* and *A. japonicus*, are reciprocally sister species, with a mean divergence estimated at circa 32 my (Zadra, Rizzoli, and Rota-Stabelli 2021).

The phylogeny and the timing of Aedini and other mosquitoes divergence have been studied in detail using multimarker approaches (Soghigian, Andreadis, and Livdahl 2017). According to these studies, mosquitoes (Culicidae) originated in the lower Jurassic with mean estimates ranging from 182 Ma to 197 Ma, depending on the study (see right columns of Table 1). Some studies predate the

origin of Culicidae to the Upper lower Triassic with mean estimates of 216-218 Ma (Reidenbach et al. 2009; X.-G. Chen et al. 2015). All these estimates, except for one (Lorenz et al. 2021), were obtained employing only the Brachycera as outgroups. This procedure creates a long uninterrupted branch between the root and the Culicidae diversification. This long internal branch may be, however, broken by several other lineages of the Culicomorpha, the superfamily to which Culicidae belong. Phylogeny and divergences of Aedini and other Culicidae should be investigated by employing a richer set of outgroups to increase the node density along the branch that from the root leads to the Culicidae clade. Such an improvement in the rooting may fix possible issues related to inappropriate outgroup sampling.

Here we present the whole mitochondrial DNA of two individuals of *A. koreicus* and *A. japonicus* sampled in Nord Italy. Our general aim was to produce new data useful for molecular typing in the field for two species of medical relevance. We also present an updated time tree of mosquito evolution based on whole mitochondrial data, which we have used to clarify some previously reported methodological issues in Aedini phylogenetics.

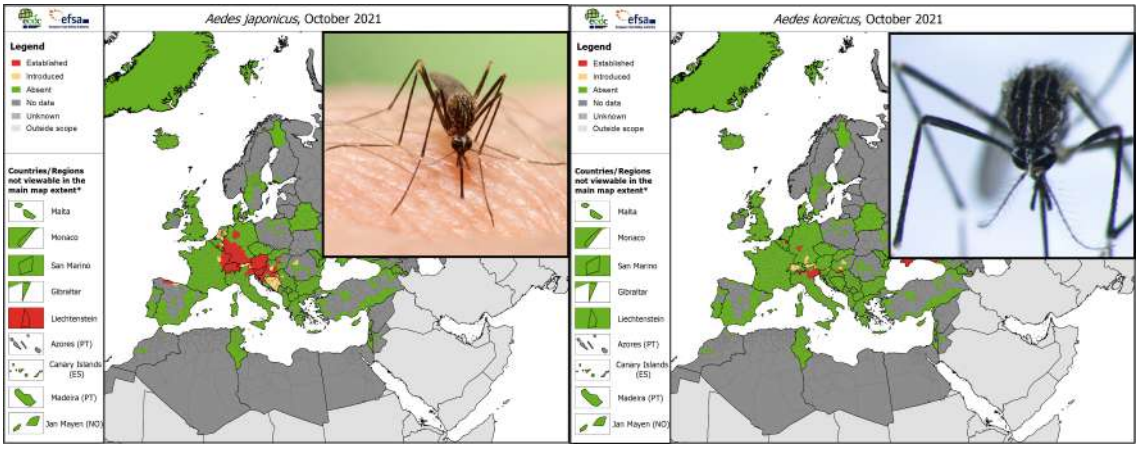


Figure 1. Distribution maps in Europe for *A. japonicus* and *A. koreicus* updated at October 2021, and the pictures of *A. japonicus* female (left panel) and *A. koreicus* female (right pane). Green: not present; red: present; yellow: uncertain. Photo credits: *A. japonicus* (https://www.ecdc.europa.eu/en/disease-vectors/facts/mosquito-factsheets/aedes-japonicus); *A. koreicus* Capelli (licensee BioMed Central Ltd. 2011). The maps and the relative mosquito distributions were realized by European Centre for Disease Prevention and Control (https://www.ecdc.europa.eu/en/disease-vectors).

3.4 Methods

3.4.1 Sampling and sequencing

Aedes koreicus and A. japonicus specimens came from lab-reared populations of respectively Fondazione Mach (FEM) of San Michele all' Adige and the Istituto Zooprofilattico Sperimentale delle Venezie (IZSV); founder individuals were collected in the wild respectively in Trentino province and Friuli regions (both in North-East Italy).

The DNA extraction was performed using the nucleon-spin tissue extraction kit Qiagen optimizing the protocol for insect DNA extraction. The extraction employed one pupa of *A. koreicus* and one pupa of *A. japonicus* in order to maximize the amount of total DNA, minimize the presence of environmental DNA, and reduce the heterozygosity of the DNA. After the extraction, library preparation and raw data sequencing was performed by the Next Generation Sequencing (NGS) facilities of the CIBIO department of Trento University using the NOVASEQ platform.

3.4.2 Mitogenomes assembly

The total raw reads were processed by MitoZ without filtering. MitoZ handles all the steps for the assembly of mitogenomes (Meng et al. 2019). It first filters the reads and trims adaptors; the clade flag is used to filter out all the reads that do not correspond with the specific clade. The other steps handle the assembly step and the annotation. After the annotation steps, we noticed that one gene was missing from the mitogenome of *A. koreicus*: this gene was successfully retrieved using the program MitoS (Bernt, Donath, et al. 2013).

3.4.3 Dataset preparation

We downloaded all the Aedini mitogenomes available on NCBI, plus various representatives of Culicomorpha, a monophyletic infraorder within (paraphyletic) Nematocera. We employed all the Aedini mitogenomes present on NCBI on 1/10/2021, and we further added the two newly assembled *A. koreicus* and *A. japonicus* mitogenomes for a total of 24 Aedini mitogenomes. To obtain a dataset comparable with previous phylogenetic studies that investigated

Culicidae evolution, we employed a taxon sampling that include 15 non-Aedini Culicinae and 11 Anophelinae as in (Zadra, Rizzoli, and Rota-Stabelli 2021; A. F. da Silva et al. 2020). We further added some Culicidae sister groups: Chaboridae, Corethrellidae, Dixidea; all these clades together are grouped in the Culicoidea clade. We further added two Chironomidae sequences to break a long internal branch between the *Drosophila*/Mosquitoes split and the diversification of the Culicoidea. From the Genbank file, we therefore, extracted t-RNAs, ribosomal, and protein-coding genes using custom scripts and aligned every gene independently using MAFFT (Katoh, Rozewicki, and Yamada 2018). We concatenated each gene with the FASconCATconcat script (Kück and Meusemann 2010), obtaining a multi-gene concatenated alignment of 15388 nucleotides.

3.4.4 Phylogenetic and Clock analyses.

We inferred a maximum likelihood tree with IQ-TREE version 1.6 using ultrafast bootstrapping with 1000 replicates under a GTR+G replacement model as defined by the program as the best fitting model (Minh, Nguyen, and Von Haeseler 2013). BEAST v2.6 was used to estimate divergence times (R. Bouckaert et al. 2014). We calibrated the clock using a combination of priors. We employed a root prior based on the fruit fly-mosquito split using a normal distribution with a mean set at 260 Ma and a 95% prior distribution to be between 296 and 238 Ma, as suggested by (Borkent and Grimaldi 2004; Benton and Donoghue 2007). We employed minimum calibration points for the diversification of Anophelinae and Culicinae, both at 34 Ma, according to the oldest fossils known for each group (Benton and Donoghue 2007; Misof et al. 2014). We add a minimum calibration of 99 Ma for the Culicidae, as the oldest known fossil, *Priscoculex burmanicus* (Poinar, Zavortink, and Brown 2020): there is, however, no indication if this fossil belongs to the crown or to the stem Anophelinae. Hence, we preferred to be as conservative as possible, placing this constraint on the origin of Anopheline/Culicinae split, the Culicidae crown. We apply a monophyletic constrain to the Culicomorpha, Culicidae, Aedini, and Anophelinae clade. The monophyly constraints on clades were based on previous morphological and molecular studies (Narayanan Kutty et al. 2018; Sæther 2000; Yeates et al. 2007;

Wiegmann et al. 2011). Moreover, we tested for putative outgroup sampling effect by running the analysis with and without the non-Culicidae Culicomorpha. We ran all MCMC chains for 200,000,000 generations twice. We checked the actual convergence with Tracer1.7 (Rambaut et al. 2018), discarded 20% of the trees as burn-in and summarised the Bayesian analyses using TreeAnnotator. We employed the Birth and Death model (BD) and a relaxed lognormal clock, as already shown to be the most suitable for mitogenome analysis in previous work (Zadra, Rizzoli, and Rota-Stabelli 2021; A. F. da Silva et al. 2020).

3. 5 Results

3.5.1 Mitogenome

We were able to fully reconstruct the mitogenomes of both the species sequenced using the MitoZ pipeline and starting from raw reads. We employed a set of 60 million reads for *A. koreicus* and 75 million reads for *A. japonicus*: this indicates a coverage of at least 4000X and high-quality mitochondrial genomes according to (Richter et al. 2015). The gene content does not show any difference from other members of the Culicidae clade, as shown in Figure 2. The gene order is also identical in *A. koreicus* and *A. japonicus*, the difference is only in the length of the sequence. This is not surprising as mosquito mitogenomes are rather conserved, all showing 37 genes comprising 13 protein-coding genes, 22 tRNAs and two rRNA genes (Lorenz et al. 2019; S. L. Cameron 2014). The gene order within the Diptera are almost identical, the only differences being the tRNA coding for Serine, which is on the + strand in *Drosophila,* whereas it is on the - strand in mosquitoes (S. L. Cameron 2014; S. L. Cameron et al. 2007). The overall genetic identity between *A. koreicus* and *A. japonicus* mitogenomes is *93 %*, whereas the identity between the *A. koreicus* from Italy sequenced in this work with the other available *A. koreicus* mitogenome from Korea (NC_046946) is 99.5 % caused by 80 SNP spread along the mitogenomes.

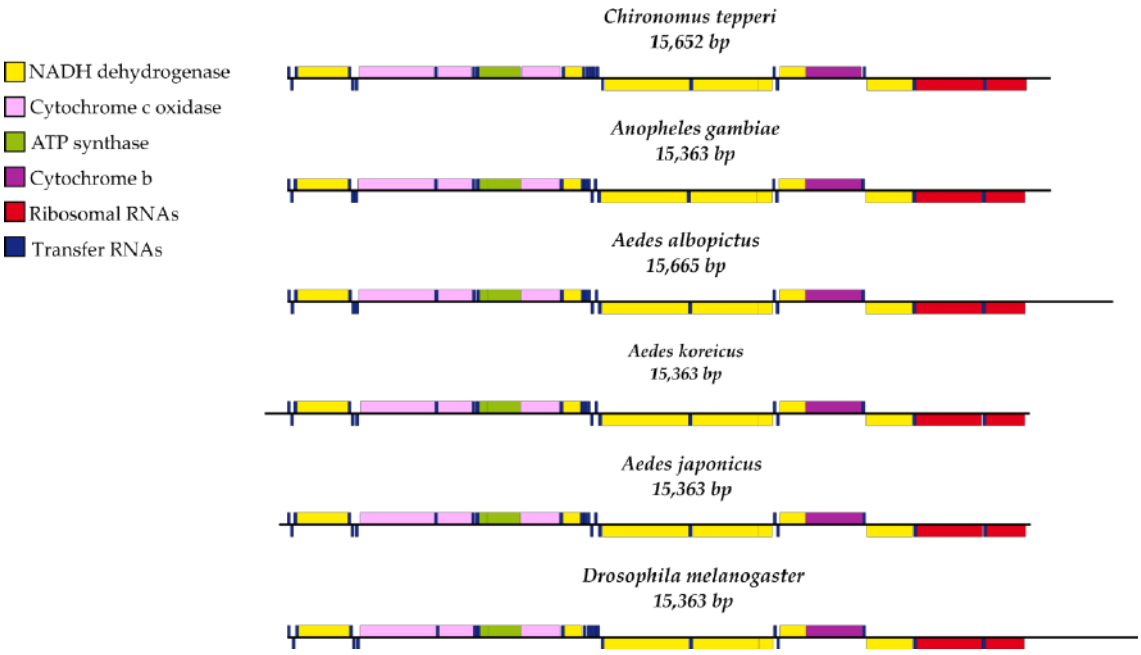


Figure 2. The mitogenome structure for six Diptera Species. *Aedes japonicus* and *A. koreicus* and the other Culicidae (*A. albopictus, Anopheles gambiae*) are characterized by an identical gene content and gene order (*A. albopictus, Anopheles gambiae*). Mosquitoes differ from other dipterans (*Chironomus tepperi,* and *Drosophila melanogaster* used here as examples*)* only for the strand in which the serine tRNA is located.

3.5.2 Divergence time based on mitogenomes

After preliminary alignments, we concluded that the promoter regions (d-loop) were unsuitable for phylogenetics and clock analysis because of their high divergence and the consequent difficulty in obtaining a meaningful alignment. Results of our time-calibrated mitogenomic analysis of mosquitoes are summarised in Table 1 and in Figure 3 (Supplementary Figure 1, provides the HPD 95% for all the estimates, with no collapsed nodes). In Figure 3 and Table 1, we compare our estimates for some key nodes with the estimates from three previous studies of mosquitoes' divergence times using mitochondrial data (Zadra, Rizzoli, and Rota-Stabelli 2021; A. F. da Silva et al. 2020; Lorenz et al. 2021). According to our analysis, the Culicomorpha (Node A) radiation started at a mean of 255 (HPD 95%: 226 -287) Ma at the turn of the Triassic period, close to the estimated Diptera origin 262 Ma. The Culicoidea (Node B, which comprises Culicidae, Dixidae, Corethrellidae, and Chaoboridae families) diversification took place through the Triassic period until the mid-Jurassic, with its first diversification at a mean of 226 (HPD 95%: 192 – 258) Ma. Corethrellidae and Chaboridae separated from the lineage that leads toward

Culicidae at a mean of 205 (HPD 95%: 172 – 238) Ma and 184 (HPD 95%: 151-214), respectively. The Culicidae (Node C) radiation occurred at a mean of 127 (HPD 95%: 101 – 150) Ma, this node representing the split between the Anophelinae and Culicinae clades. This is the most recent estimate retrieved for this node to date (Reidenbach et al. 2009; X.-G. Chen et al. 2015; Lorenz et al. 2021; Zadra, Rizzoli, and Rota-Stabelli 2021; Hao et al. 2017; A. F. da Silva et al. 2020). The previously mean estimated ages for the Culicidae node spanned between 145 and 190, much older than the estimates reported here in this work. Our results support a Culicidae diversification that falls entirely within the Early Cretaceous. The diversification of Anopheline (Node D) and Culicinae (Node E) took place almost simultaneously at a mean of 106 (HPD 95%: 84 - 132) and 106 (HPD 95%: 85 -128) Ma, respectively. Node F (split between Culicini and Aedini) occurred at 87 (HPD 95%: 70 -106) Ma. Previous estimates propose the divergence of the node F around 130 Ma (Table 1) (Zadra, Rizzoli, and Rota-Stabelli 2021; A. F. da Silva et al. 2020; Lorenz et al. 2021). Aedini tribe is the clade that gave rise to the species we have sequenced, *A. koreicus* and *A. japonicus,* as well as other species of medical relevance. We estimate the split of this clade (Node G) to have occurred at a mean of 75 (HPD 95%: 59 -90) Ma. This new proposed time-scale of mosquito evolution suggests the emergence of major mosquito lineages during the Cretaceous period. The *A. albopictus* split from its sister species *A. flavipictus* (Node H) took place at 23 (HPD 95%: 12 - 34) Ma, whereas the split between *A. koreicus* and *A. japonicus* is set at a mean of 27 (HPD 95%: 15 - 40) Ma, estimating a more recent time than Zadra et al. (2021) but in line with the estimate of Soghigian et al. (2017) (Soghigian, Andreadis, and Livdahl 2017). To explore better how the outgroups affect the tree topology and posterior, we ran the analysis without the non-Culicidae Culicomorpha species (namely, *Polypedilum vanderplanki, Chironomus tepperi, Dixella sp., Corethrella condita* and *Chaborus sp.*). In the absence of non-Culicidae outgroups, the mean and the 95% HPD estimates become older for the diversification of most clades, including Culicidae (node C), Culicinae (node E), and Aedini (node G), as indicated by the light bars in Figure 3.

From a methodological point of view, all our dated Bayesian phylogenies are well converged as the ESS estimates were well above 200: this indicates a good mixing of every parameter. Moreover, the coefficient of variation (σ_r, the parameter that

measures the clock-likeliness of the data) is 0.403. If the σ_r is close to zero, then the data have low rate variation, and they can be modelled as a strict clock); vice versa, higher values tend to favour the relaxed clock assumption that we have indeed implemented in all our analyses. A relaxed nature of the clock in the mitogenome of mosquitoes can be actually observed in Figure 4. The mean mutation rate calculated over the whole tree is estimated at $4.01 * 10^{-3}$ (mutation/site/MY from here on (msm)). The rates in the Aedini group are highly variable in some clades, ranging between $3 * 10^{-3}$ msm and $8 * 10^{-3}$ msm. *Aedes albopictus,* with a rate of $6.3 * 10^{-3}$ msm, has the higher rate among its close relatives besides the *Haemagogus* genus that shows high variability in the evolutionary rate, spanning from 7.7 to $3.2* 10^{-3}$. The Chironomidae clade shows a high rate of evolution, $8.2 * 10^{-3}$ msm, along with *Charoborus* $7.3 * 10^{-3}$ msm (Figure 4). These high rates reflect long branches as detected by the Maximum likelihood phylogeny (Supplementary Figure 2).

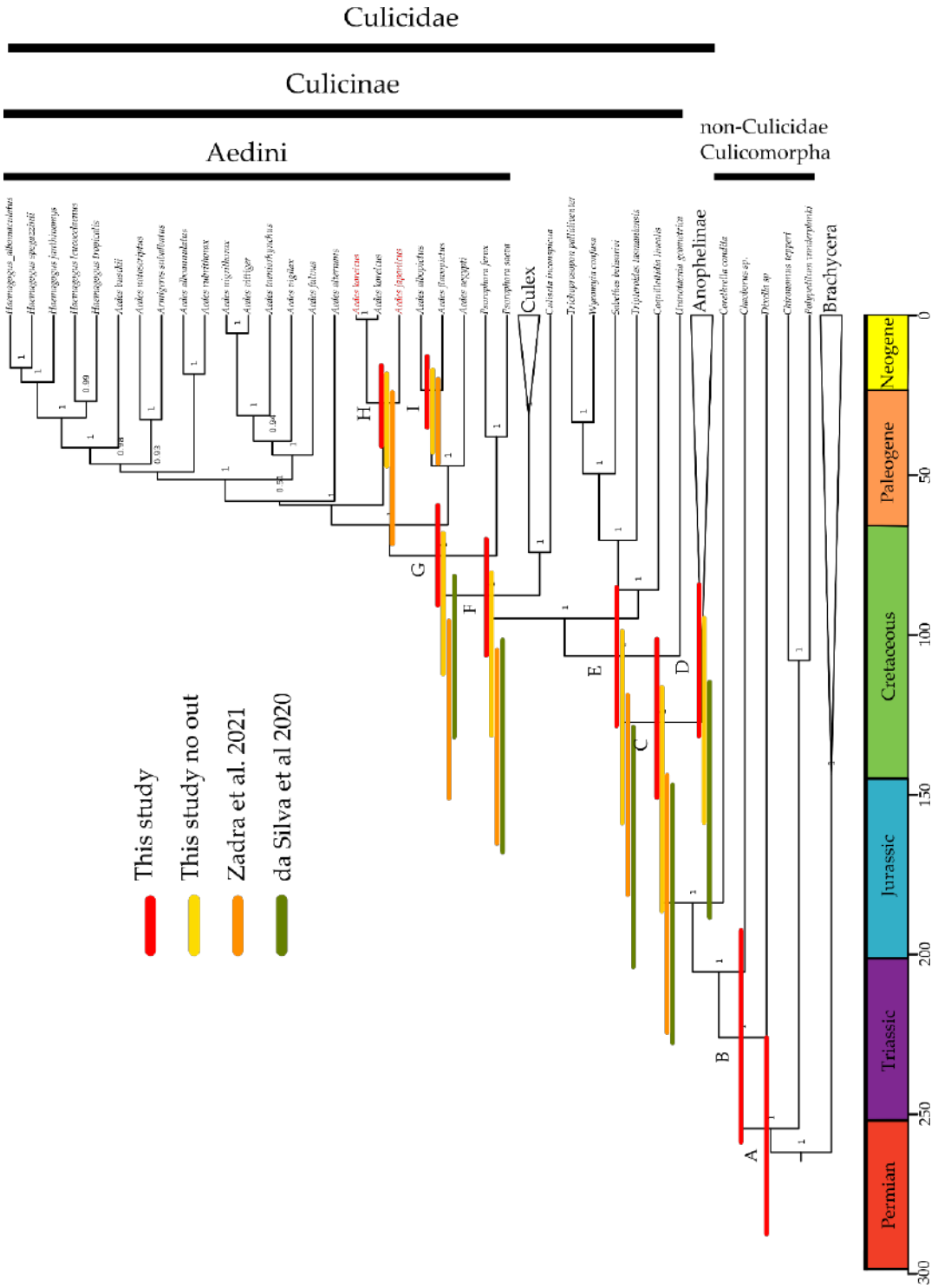


Figure 3. Divergence times of the Culicidae mosquitoes using non-Culicidae Culicomorpha outgroup. Red bars indicate the estimates from the present study, yellow bars from the same analysis carried out employing only the Culicidae taxa and no other Culicomorpha samples, orange bars indicate estimates of Zadra et al. (2021) (Zadra, Rizzoli, and Rota-Stabelli 2021), green bars indicate estimates from da Silva et al. (2020) (A. F. da Silva et al. 2020). The HPD 95% of the correspondent node are in Table 1).

Table 1 Divergence estimates of selected nodes from Figure 3 and other analyses. For each node, we provide the mean and the 95% high posterior density. We compare the result of our study with a previous study, Lorenz et al. (2021) did not provide HPD 95% estimates. All the estimates are provided in million years ago (Ma) (Lorenz et al. 2021). [1](Zadra, Rizzoli, and Rota-Stabelli 2021) [2](A. F. da Silva et al. 2020) [3](Lorenz et al. 2021).

Node	Taxonomic level	This study	This study only Culicidae	Zadra et al. (2021)[1]	da Silva et al. (2020)[2]	Lorenz et al. (2021)[3]
A	Culicomorpha	255 (226 -287)				
B	Culicoidea	226 (192 – 258)				220
C	Culicidae	127 (101 – 150)	151 (115 -186)	182 (143–223)	182 (146-233)	197
D	Anophelinae	106 (84 -132)	125 (94-159)		145 (114-187)	147
E	Culicinae	106 (85 -128)	128 (98-158)	150 (118–184)	160 (128-205)	153
F		87 (70 -106)	105 (79-131)	135 (104–164)	130 (101-168)	123
G	Aedini	75 (59 -90)	90 (67-113)	111 (95–150)	102 (81-132)	74
H	*A. albopictus – A. flavipictus*	23 (12 -34)	29 (16-43)	33 (20–46)		
I	*A. koreicus – A. japonicus*	27 (15 -40)	32 (18-48)	46 (24–71)		

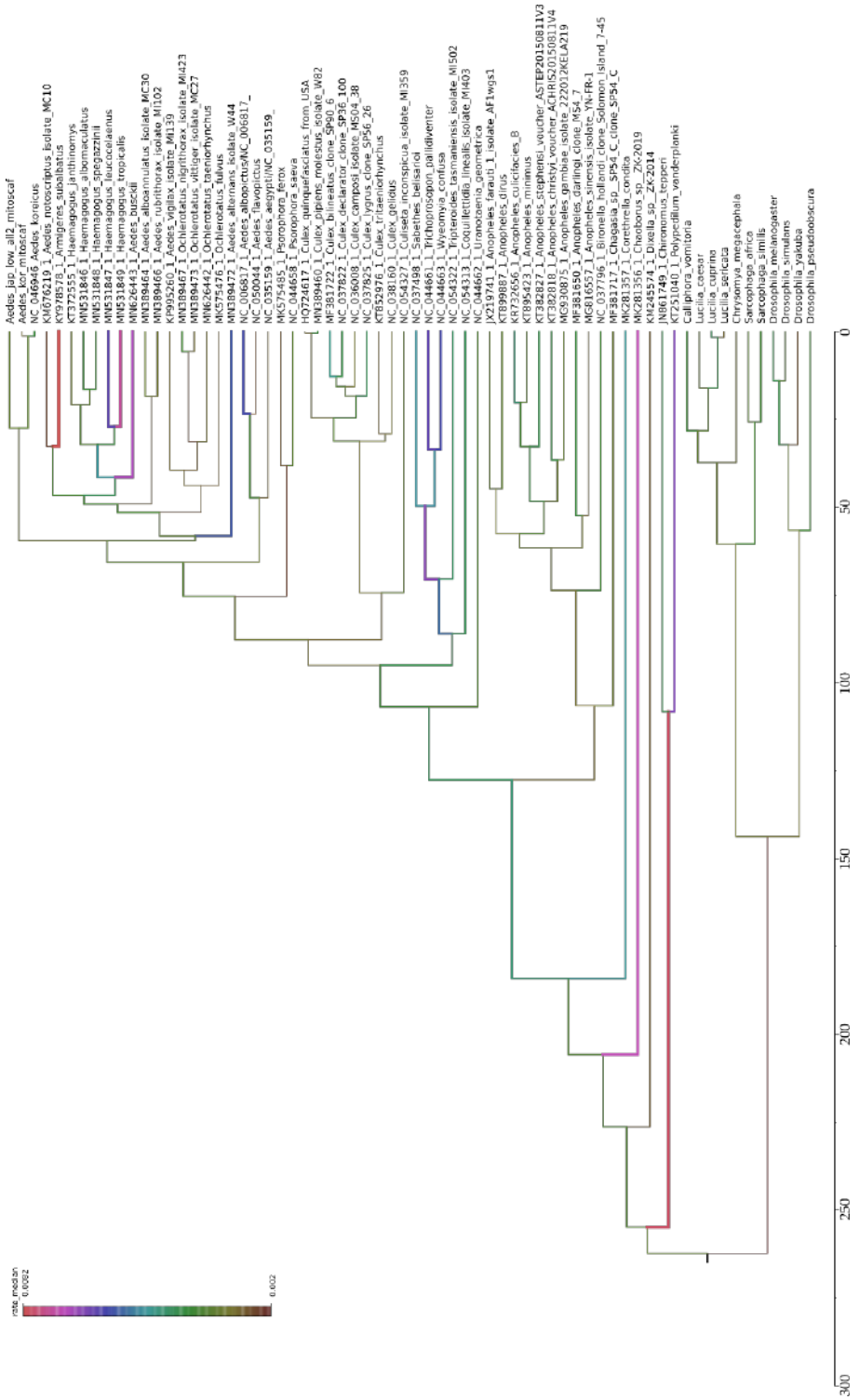


Figure 4. Rate heterogeneity within the Culicomorpha radiation. Specific rate branches are highlighted by the line colours. Some lineages, or species, are characterized by an increased mutation rate. Brown/green shades indicate slower evolutionary rates, whereas Blue/red shades indicate faster evolutionary rates. The lower rate (brown) is set at 2 10^{-3} msm the highest rate is set at 8.2 10^{-3} (red). We can appreciate a higher rate in the Culicidae outgroups and in some Aedini lineages.

3.6 Discussion

3.6.1 *Aedes koreicus – Aedes japonicus* divergence, are they a species complex?

Here we presented the mitochondrial genomes of *A. japonicus* and *A. koreicus,* two recently established invasive mosquitoes in Europe. Available data for *A. koreicus* is scarce, mainly consisting of few gene markers (Negri et al. 2021): this makes it challenging to quantify, for example, the genetic distance among samples and identify possible introduction routes. Mitochondrial genomes have a lower resolution power if compared with genome scaled nuclear markers or microsatellites; however, they are quick to sequence and reliable source of data for population and phylogenetic analysis. The two previously available mitogenomes for *A. koreicus* on NCBI were both collected in South Korea. To better understand the *A. koreicus* spread in Europe, we provided the mitogenome of an Italian sample of *A. koreicus*, the first to be sequenced outside its native range. The invasive *A. japonicus* has been more intensively studied. Microsatellites, some mitochondrial markers and related population genetics studies are already available for European and American populations (Fonseca et al. 2010; Widdel et al. 2005; Huber, Schuldt, et al. 2014; Smitz et al. 2021; Baharmand et al. 2020). A complete mitogenome sequence was not however available on the public database, and here we have provided the first *A. japonicus* mitogenome.

The mitogenomic identity between the *A. koreicus* from Italy that we have sequenced and the *A. koreicus* from Korea (NC_046946) is 99.5 %. This is in line with the similarity of 99.6 %, identified by Steinbrink et al. (2019) (Steinbrink et al. 2019) using the COI barcode between *A. koreicus* samples from Belgium and Germany. The genetic difference between the Italian and the Korean mitogenomes is reflected by a high estimated of their divergence time in Figure 3: mean 1.22 (HPD 95%: 0.62 - 2.06) Ma. If we compare the age of this *A. koreicus - A. koreicus* split with that of *Culex pipiens - Culex quinquefasciatus,* which occurred around 0.43 (HPD 95%: 0.17 – 0.75), we can notice that the divergence between two different *Culex* species is actually lower than the one estimated between the two *A. koreicus* mitogenomes. The two *Culex* species belong to the so-called species complex (Harbach 2012), but are still considered two different species. This high divergence between *A. koreicus* individuals suggests the possibility of

a species complex that include this species and *A. japonicus*. A previous phylogenetic investigation of the *A. japonicus* and *A. koreicus* using mitochondrial markers had indeed retrieved *A. koreicus* nested within a paraphyletic *A. japonicus* composed of likely subspecies *A. j. japonicus*, *A. j. yaeyamensis*, *A. j. amamiensis with* (E. C. Cameron et al. 2010): this further point toward a species-complex scenario. Although the *A. japonicus – A. koreicus* split is estimated at several millions of years ago in the early Neogene (node H in Figure 3), the morphological character employed for distinguishing the three *A. japonicus* subspecies and *A. koreicus* has a generous amount of overlap that is also suggestive of a species complex and that could lead to possible misidentification in adults. More data from Asian native range and newly introduced countries is needed to better understand the reciprocal affinities of *A. japonicus* and *A. koreicus.* If a species complex will be confirmed, it may imply chances of hybridisations in both native and invasive regions, making more challenging the identification and management of these species.

3.6.2 Mosquitoes radiation could be more recent than we previously thought

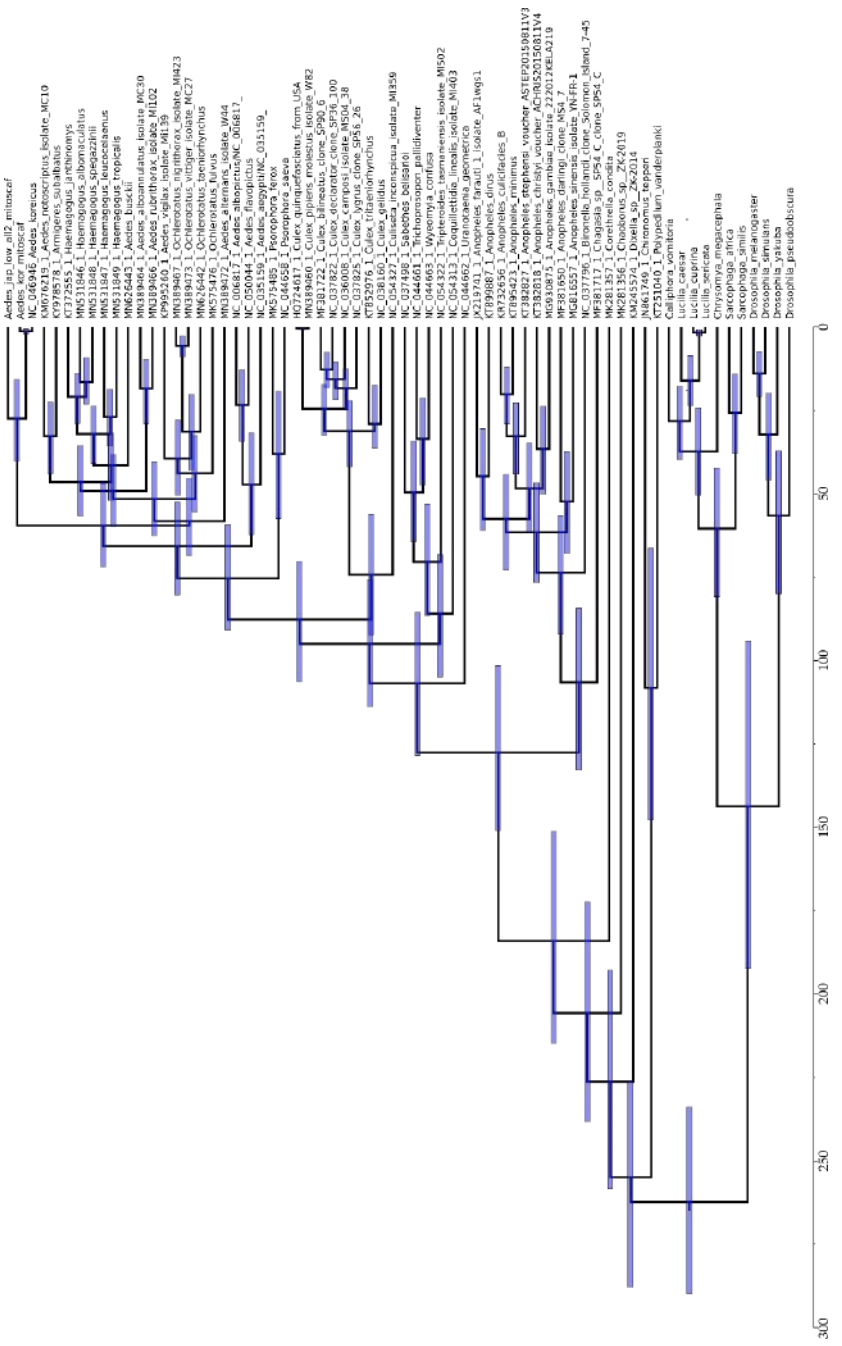
Our dated phylogeny is highly congruent in terms of topology with previous studies, but it retrieves interesting insight into the time-scale of mosquito evolution. As depicted in Figure 3, our analysis (in red) provided a more recent time-scale in contrast to other previous estimates (Table 1). The most important discordance between previous works and this study is for the Culicidae split: our 95% HPD estimates barely overlap with previous ones (Zadra, Rizzoli, and Rota-Stabelli 2021; A. F. da Silva et al. 2020; Reidenbach et al. 2009; Soghigian, Andreadis, and Livdahl 2017; X.-G. Chen et al. 2015).

The younger age of the Culicidae clade retrieved by this work can be explained by using a larger outgroup sampling which breaks the long branch that leads from the root to the Culicidae group. The ingroup and outgroup sampling have been shown to affect the divergence estimates in Bayesian frameworks (Soares and Schrago 2015; Spasojevic et al. 2021; Bromham et al. 2018). Adding the non-Culicidae Culicomorpha outgroup also increased the precision in the Culicidae estimates. The analysis carried out excluding the non-Culicidae Culicomorpha species (yellow bars in Figure 3) affected the HPD 95%, which becomes broader

in the analysis without the outgroup, whereas using the full dataset, the estimates range is narrow. The discrepancies between our and other studies can also be explained by the dataset employed: our analysis employed all the genes, whereas other used only the coding genes (A. F. da Silva et al. 2020) or only few mitochondrial markers (Zadra, Rizzoli, and Rota-Stabelli 2021). Another possible source of discrepancy is the taxon sampling within the Culicidae: Da Silva et al. (2021) employed, for example, more *Anopheles* and less Aedini than us, and this may have affected the distribution of rates in the tree and the age of ancestral nodes.

Regardless of the reason behind the observed discrepancies, our younger estimates allow us to suggest novel paleo-ecological scenario. In our study, the Culicidae diversification (node C of Figure 3) falls within the lower Cretaceous instead in the mid-Jurassic as in previous studies. The blood-feeding behaviour is a widespread trait within the Culicomorpha clade, and probably this behaviour is the ancestral trait of this group: according to our timetree, this character may have appeared at the Permian-Triassic boundary concomitantly with the Culicomorpha diversification (node A). This blood-sucking habit has later diversified to adapt to all the ecological niches provided by evolving tetrapods. An interesting novel paleo-ecological scenario is that our new estimate of Culicidae diversification (node C of Figure 3) tends to match the Angiosperm radiation and the origin of most of its extant diversity in the mid-Cretaceous around 120-130 Ma (van der Kooi and Ollerton 2020; H. T. Li et al. 2019). The origin of the angiosperm has occurred probably in the Triassic – Early Jurassic, but this estimate is still debated (Coiro, Doyle, and Hilton 2019). However, many studies agree to set the origin of the angiosperm somewhere in Triassic or Early-Jurassic period, especially studies carried out with many genera (Clarke, Warnock, and Donoghue 2011; Magallón and Castillo 2009; C. S. P. Foster et al. 2017; S. A. Smith, Beaulieu, and Donoghue 2010). The match between Culicidae and angiosperms is supported by our time trees but not by previous studies.

3.8 Supplementary Figure



Supplementary Figure 1. Here we show the complete phylogenetic tree used in Figure 3. All the nodes are coupled with its HPD 95%bar. No group is collapsed. Anophelinae and *Culex* clades are reported entirely (not collapsed).

Supplementary Figure 2. Maximum Likelihood tree shows the fast-evolving Chironomidae group and the *Chaoborus sp.*

CHAPTER 4 - SHORT-READS GENOME SEQUENCING OF PEST AND BIOCONTROL SPECIES

4.1 introduction to the Chapter

In this Chapter, I provide genome reports of the sequencing of three species: *A. koreicus*, *A. japonicus* and *Trissolcus japonicus*, the wasp involved in the biological control of the Brown Marmorated Stink Bug (*Halyomorpha halys*).

The Chapter is divided into two parts. In Part A, I present the methodology that I applied for every step of the experiment, from rearing to the assembly and the draft result of the genome skimming of *A. koreicus* and *A. japonicus*. The quality of the reads did not allow me to produce proper genome assemblies and the completeness assessment indicates that we can retrieve a good fraction of genes for *A. japonicus* and only a small amount for *A. koreicus*. This data provides, however, enough information for successful genome skimming of genes, mitochondrial DNA (used in Chapter 3), and associated microbiome.

In part B, I provided an overall picture of the quality of the genome assembly of two individuals of *Trissolcus japonicus*, one from the strain involved in the releasing for biological control and one from wild captured in Trentino. The genomes have a high BUSCO completeness and will provide a solid base for further genomic studies within this clade of miniaturised wasps.

Contribution: In this work, I reared *Aedes koreicus* and performed the DNA extraction, assembly, mitogenome assemblies, and orthology assessments for all three species. I received training from many of the co-authors listed. The metagenomic analyses were performed by colleagues. Supervised by my advisor Omar Rota-Stabelli, I designed most of the experiments and interpreted the results.

Part A. Genome skimming of the invasive mosquitoes *Aedes koreicus* and *Aedes japonicus*

Nicola Zadra[1,2], Andrea Silverj, Paula Escuer, Annapaola Rizzoli[1], Daniele Arnoldi[1], Gioia Capelli[3], Fabrizio Montarsi[3], Alejandro Gracia Sanchez[4] Omar Rota Stabelli [1,2,5]

[1] Fondazione Edmund Mach; [2] CIBIO department, University of Trento; [3] Laboratorio di Parassitologia, Istituto Zooprofilattico Sperimentale delle Venezie; [4] Barcelona University; [5] Center C3A, University of Trento.

Abstract

Despite their importance as a threat to human and animal health, knowledge of *Aedes* mosquitoes' evolutionary genomics is rather fragmentary, impairing a good understanding of their biology and effective control. In the Aedini lineage, only the genomes of *A. albopictus* and *A. aegypti* have been sequenced. Other species within this clade have been widely neglected, including *A. japonicus* and *A. koreicus,* two species that are established in temperate regions of Russia and Europe. Here, we present the genome skimming analysis of *A. japonicus* and *A. koreicus* individuals sampled in North Italy. The quality of the two assemblies of these two species is very different: *A. japonicus* assembly shows a fair level of completeness, whereas *A. koreicus* assembly is very fragmented. This data provides enough information for successful genome skimming: we could extract up to 90% of BUSCO genes from *A. japonicus* and examine the microbiome and virome associated with the reads. These data need to be integrated by long read and by more short read sequencing; it provides, however, the first framework to investigate the genome evolution of these two species and an initial repository to extract genes for downstream analyses.

4.2 Introduction

Anopheles mosquitoes have been studied in deep, and a large panel of genomes from 25 *Anopheles* species have helped characterize better their biology (Isaacs et al. 2018; Anselmetti et al. 2018; Wei et al. 2017; Jiang et al. 2014). Conversely, within the Aedini mosquitoes, only two species have been sequenced so far: *A. albopictus* and *A. aegypti*. Both species are quickly establishing in urban environments outside their original Asian areal, show human blood-feeding behaviour and are competent for many arboviruses such as West Nile virus, Zika virus, and Yellow fever virus (Jolyon M Medlock et al. 2018; L. H. Chen and Wilson 2020; Capelli et al. 2011; Huber, Jansen, et al. 2014). These characteristics have made *A. albopictus* and *A. aegypti* worldwide health concerns and important model organisms. While *A. albopictus* presence is now reported from every continent and most of the Mediterranean basin (Figure 1), *A. aegypti* seems not capable of settling stable populations in continental Europe with the exception of southern Russia and Georgia (N. R. Faria et al. 2017; Kraemer et al. 2015). Both *A. aegypti* and *A. albopictus* belong to the subgenus *Stegomyia* and can be currently compared from a genomic point of view only against each other. For example, it would be impossible to discriminate between genomic characters specific to *Stegomyia* (black dots of Figure 1) from those shared with all other *Aedes* mosquitoes. In a comparative genomic framework, the availability of genome data from other *Aedes* species may help increase the correct interpretations of *A. aegypti* and *A. albopictus* genomes.

Other *Aedes* species are competent for arboviruses, but they are less studied for at least three reasons: they are confined to certain tropical regions, they are not/less efficient vector of arboviruses, and they have been introduced in new territories only recently. Among them, there are two *Aedes* that are currently establishing in temperate climates of Europe. *A. japonicus* is quickly establishing in central Europe (Francis Schaffner, Chouin, and Guilloteau 2003; Francis Schaffner et al. 2009) and its presence is confirmed in Italy since 2015 (Bernhard Seidel, Nowotny, et al. 2016). *A. koreicus* is enlarging its new areal in North Italy and the Alpine region (Capelli et al. 2011). While *A. albopictus* typically requires warm temperatures, *A. koreicus* and *A. japonicus* seem particularly well adapted to temperate climates. *A. koreicus* prefers urban habitats, laying eggs in man-

made containers, whereas *A. japonicus* prefers sylvan and rural habitats, but it could tolerate suburban and urban habitats (Montarsi et al. 2013).

Aedes koreicus and *A. japonicus* are reciprocally sister species (a sister species is the most closely related species on the phylogenetic tree), with a mean divergence estimated at circa 32 Ma (Zadra, Rizzoli, and Rota-Stabelli 2021). Because they are sister species, their genome sequencing would reciprocally ease the interpretation of comparative genomic studies by, for example, assigning species-specific genes to certain species (orange and red dots of Figure 1). The genome sequencing of other *Aedes* species is anticipated to be not an easy task as these species are often characterized by large, highly repetitive genomes. The assembly of a European (therefore invasive strain) *A. albopictus* genome (Dritsou et al. 2015) has proven extremely difficult if compared to a native Asian strain (X.-G. Chen et al. 2015). The issues encountered during genome assembly in Dritsou et al. (2015) (Dritsou et al. 2015) can be explained by at least three factors: the large genome size, the high frequency of repetitive regions, and the short reads approach employed for the sequencing. Subsequent resequencing using long-read technology has strongly bypassed assemblies issues (Palatini et al. 2020). *Aedes japonicus* and *A. koreicus* are invasive of Europe, and it may be possible that they show a similar pattern of transposon enrichment, which would further unease their genome assembly.

Here we report the results of genome sequencing for *A. japonicus* and *A. koreicus* using short (Illumina) reads. We obtained a genome assembly with good completeness for *A. japonicus* and a poor assembly with low level of completeness for *A. koreicus*. These data can provide genetic material for a genome-skimming approach.

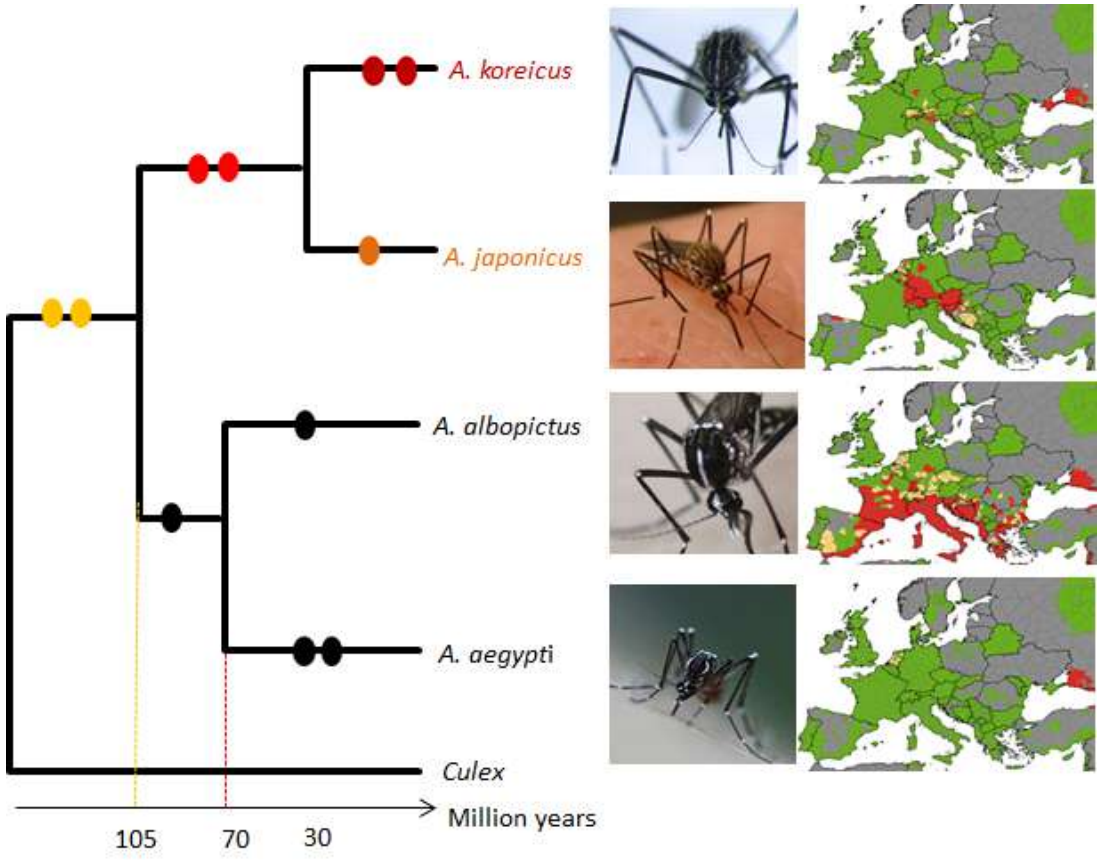


Figure 1. Phylogeny and European distribution of four *Aedes* species. In the left panel, we show a simplified tree of the four *Aedes* mosquitoes. The dots along the branches highlight hypothetical genomic changes that occurred in mosquitoes' lineage, showing that increasing the number of genomes available can increase the resolution of the phylogenomic analysis. Adding a temporal scale helps to understand when a certain change occurred, providing insight into the paleoecology of the species and in which climatic scenario the genomic changes took place. In the right panel, we show the distribution of the four *Aedes* species in Europe. Maps are taken from the European Centre for Disease Prevention and Control, updated in April 2021. Green: not present; red: present; yellow: uncertain. Photo credits: *A. japonicus*: cydno (CC BY-NC); *A. koreicus* and *A. albopictus*: Capelli (licensee BioMed Central Ltd. 2011); *A. aegypti*: Monica Ridlehoover (CC BY 2.0).

4.3 Methods

4.3.1 Sampling and rearing

A. japonicus specimens were reared in captivity from samples collected in the wild in Friuli and supplied by the Istituto Zooprofilattico Sperimentale delle Venezie (IZSV). We performed DNA extraction on an *A. japonicus* pupae provided by IZSV. The *A. koreicus* specimens were reared in captivity at Fondazione Edmund Mach from samples collected in the Trentino area. To lower *A. koreicus* heterozygosis, we rear an inbreeding line. To avoid unwanted multiple breeding, we isolated a couple (a male and a female) from the rest of the

siblings, placing them in a separated cage as soon as they hatched. In the cage, we place cotton soaked with sugar water and a black glass fill with water with a blotting paper for aiding the oviposition. A blood meal was proposed to the mosquito female every two days until oviposition. The procedure was repeated for two generations: in the second generation, high mortality was present in the larval stage, whereas the third generation produced non fecundated eggs. Therefore we performed the DNA extraction on an *A. koreicus* pupa of the second generation. The rearing conditions were set as follows: temperature 23/26 C°, stable humidity to 70 % and the day-night cycle was kept 16 h of light and 8 h of dark, following the protocol of Marini et al. (2019) (Marini et al. 2019).

4.3.2 DNA extraction, libraries and sequencing

The DNA extraction was performed using the nucleon-spin tissue extraction kit Qiagen optimizing the protocol for insect DNA extraction. The extraction was performed on single pupae of *Aedes koreicus* and *Aedes japonicus*. The DNA yielded was assessed with Qubit. We extracted DNA from one single individual to reduce heterozygosity of the genome data and increase the quality of the assembly. After the extraction, library preparation and raw data sequencing was performed by Next Generation Sequencing (NGS) facilities of Trento that used the NOVASEQ platform to obtain pair-end reads of 150 bp length each, with an average insert size of 650 bp. Raw reads from the two *Aedes* species were quality checked using FastQC.

4.3.3 Data processing, assemblies, quality assessment

Raw reads were processed using MaSuRCA3.1.2 assembler (Zimin et al. 2013). The Kmer were estimated during the procedure. As suggested by the MaSuRCA developer, we performed the assembly without any further trimming steps. We employed the default parameter except for the JF_SIZE parameter adjusted for the expected size of or genome. The *De novo* assembly quality was assessed using BUSCO v5.0 against the Insect datasets (Simão et al. 2015; Waterhouse et al. 2019).

4.3.4 Microbial and viral profiling

The trimming step was performed using Trimmomatic, using the option "ILLUMINACLIP:/Trimmomatic-0.39/adapters/NexteraPE PE.fa:2:30:10:1:TRUE LEADING:3 TRAILING:3 SLIDINGWINDOW:4:5 MINLEN:25". A further run of FastQC on trimmed data confirmed the effectiveness of the trimming phase. Cleaned reads (paired and unpaired) were used as input of MetaPhlAn-3.0, which was run using the mpa_v30_CHOCOPhlAn_201901 database with and without the option "--add_viruses". Relative abundance plots were obtained separately for bacteria and viruses using hclust2.

4.4 Results and discussion

4.4.1 Genome data for genome skimming

We obtained a total of 220 Million reads for *A. japonicus* and 150 Million read for *A. koreicus*. The average quality (Phred score) for the *A. japonicus* reads was 35 for the forwards and 29 for the reverse, whereas the quality was slightly lower for the *A. koreicus* reads, 35 and 27, respectively (Table 1). To generate our assemblies, we used all the reads without any further filtering or trimming as MaSuRCA implements its own automatic cleaning procedure (Zimin et al. 2013). However, the low quality of the reverse reads files (in particular for *A. koreicus*) makes the program discard many reads, lowering the amount of total reads available for the assembly. The *A. japonicus* assembly resulted in an N50 of 2986, with an average coverage of 7.4X, whereas for *A. koreicus*, the assembly provided a dramatically low N50 of 765 bp, with an average coverage of only 4.5X. The unsuccessful assembly of *A. koreicus* is mirrored by the BUSCO assessment of gene completeness (Figure 2). While for *A. japonicus* we could retrieve 91% of genes, for *A. koreicus* we could find only 29% of genes, most of which were fragmented.

	Reads	*N° sequence*	*Duplicates (%)*	*Quality score*
A. koreicus	R1	79.4	11.0	35.5
	R2	79.4	33.9	27.3
A. japonicus	R1	110.7	14.1	35.2
	R2	110.7	26.5	29

Table 1. The reads quality check obtained with Fastqc, shows high duplicated rates in both, especially in *A. koreicus*, moreover, the amount of reads in *A. koreicus* is substantially lower than in *A. japonicus*. The quality scores show that probably there was a problem during the sequencing of the reverse strand of the pair-end reads.

	N of reads	*Coverage*	*N50 (bp)*	*GC content (%)*	*Genome size*
A. koreicus	10148019	4.51	751	40.51	1005618802
A. japonicus	42966735	7.37	2986	39.86	1146538077

Table 2 Here, we display some key statistics for the assembled genome. The table shows the difference in assembly quality obtained. The *A. japonicus* has a higher number of reads and, the coverage and the N50.

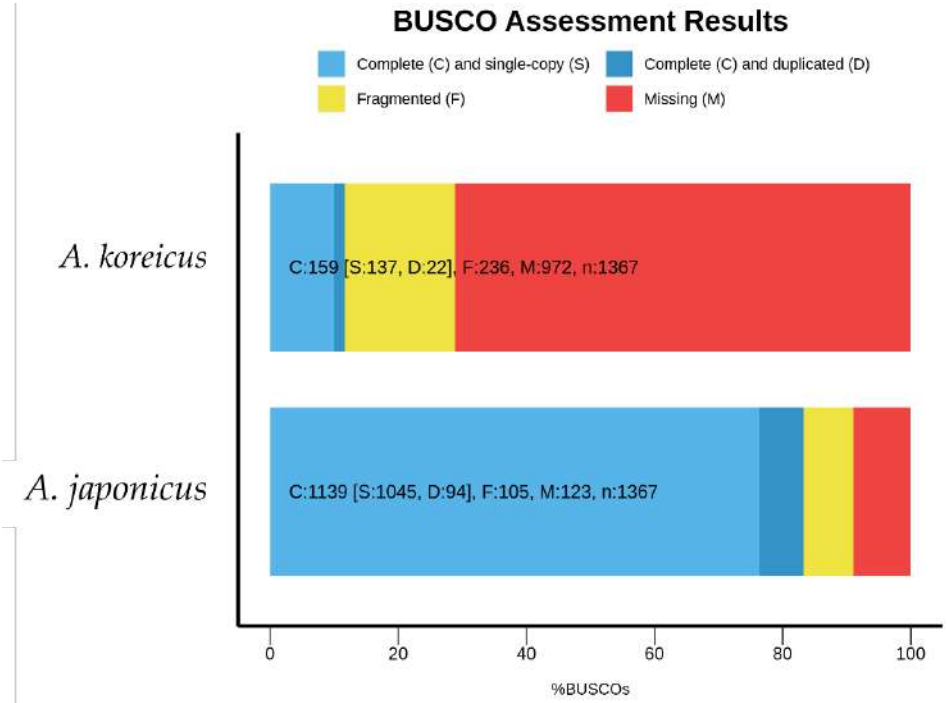


Figure 2. BUSCO assessment of the genome assemblies. The analysis was carried out employing the Insect single-copy orthologous dataset (insecta_odb10). The light blue bar indicates the percentage of complete genes found by BUSCO in single copy; the dark blue bar represents the duplicated complete genes found; the yellow bar shows the percentage of fragmented genes; eventually, the red bar indicates the missing genes, namely the genes that are supposed to be present in the species in question but not found in the genome assembly provided to BUSCO.

Overall, the sequencing of *A. koreicus* has encountered sequencing issues (low quality reads and low output) which allowed us to retrieve only a small fraction of genes and a highly fragmented assembly. On the other hand, *A. japonicus* assembly provided a reasonable number of expected genes: this indicates that this genome data can be analysed for the presence of genes belonging to gene families of ecological and management interest such as opsin genes, odorant-receptor (OR) and odorant-binding proteins (OBP). Phylogenomic studies of these gene families have been already performed in insect genomes of similar quality (Ometto et al. 2013). Due to the large genomes that characterized the *Aedes* mosquitoes, the assembly within this clade has never been straightforward. Indeed, for *A. albopictus,* many attempts (X.-G. Chen et al. 2015; Dritsou et al. 2015) and a mixture of the short and long reads had to be employed before reaching high-quality genome assembly and annotation (Palatini et al. 2020).

4.4.2 Different microbial and viral profiles and absence of *Wolbachia* and *Asaia*

As in other whole body DNA extraction sequencing projects, some of the reads do not come from the fragmented mosquito genome but from other biological entities such as bacteria and viruses. We characterize the overall microbial and viral signal associated with the two mosquito species by profiling the filtered genome reads (see Methods) using MetaPhlAn-3.0. The microbial content associated with the reads was very different in the two species (Figure 3a). The microbial profiles seem to be similar to those obtained by other authors who analysed mosquito microbiome, but we found no evidence of *Wolbachia* and *Asaia*: these two genera are, however, in high prevalence only in adults, not in pupae (Coon et al. 2014; X. Wang et al. 2018; Alfano et al. 2019; Möhlmann et al. 2020). We rerun the analysis, including the low-quality reads, as they could contain some extra-signal, but we were not able to detect *Wolbachia* or *Asaia,* and the overall scenario did not change significantly. We detected significantly more viruses in *A. japonicus* than in *A. koreicus* (Figure 3b). These differences could be guided by the sequencing depth of the samples, which is higher for *A. japonicus*.

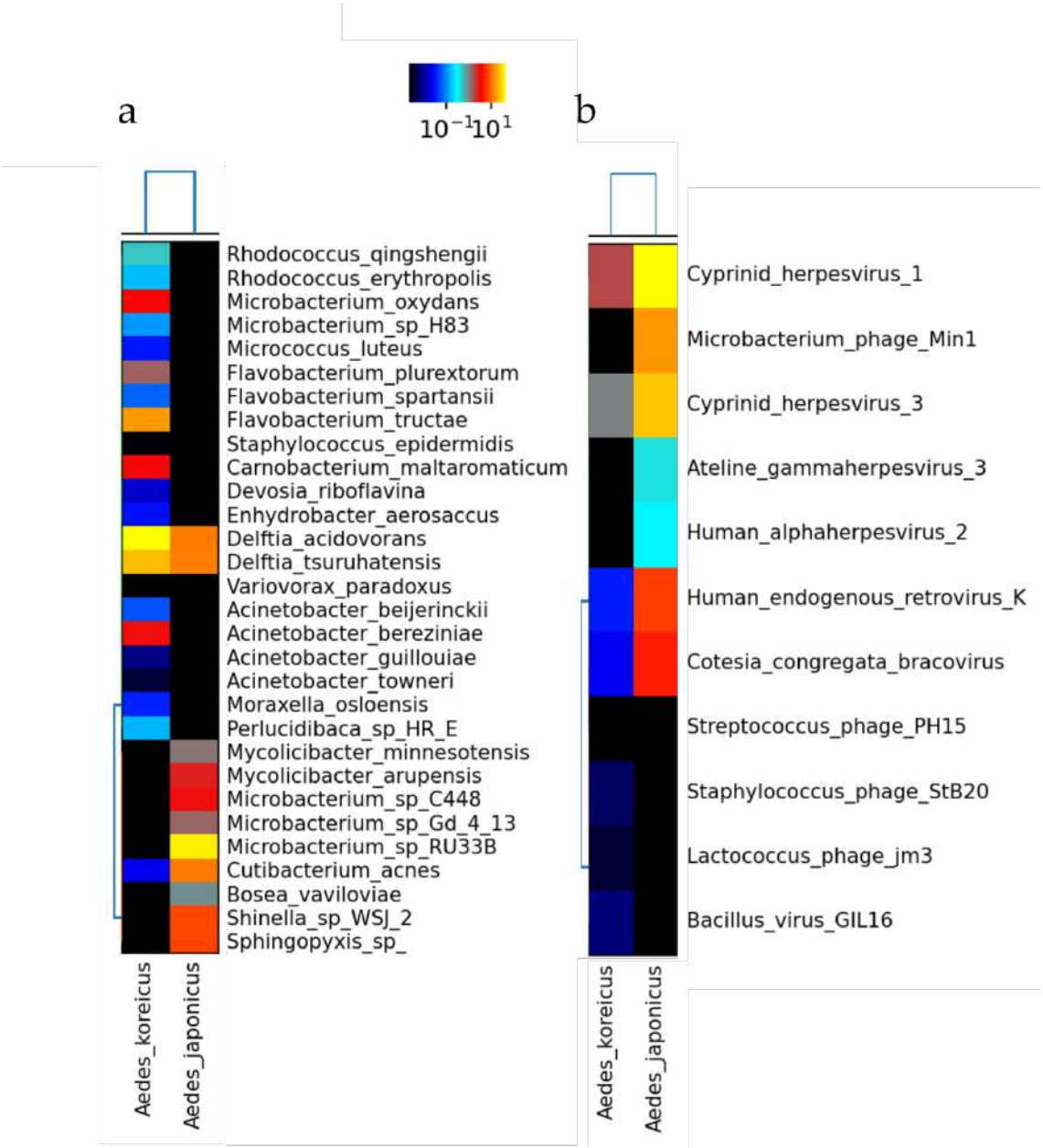


Figure 3. **Microbes and viruses associated with *A. koreicus* and *A. japonicus***. The colours indicate the relative abundance of the species detected in each of the two samples. We can observe a great difference between the microbial content in the two sequenced species panel *a*. This can be addressed to the different rearing conditions of the *A. japonicus* and *A. koreicus*. Panel *b* shows the relative abundance of target viruses associated with *A. koreicus* and *A. japonicus*.

4.4.3 Limitations of the single individual approach

Our sequencing approach was based on the sequencing of one single individual. The benefit of this approach is that it reduces the amount of heterozygosity: only two copies of the same chromosome are sequenced in a diploid organism rather than many as in a pooled experiment. It is, however, challenging because of the low amount of DNA yield that can be recovered from small insects like

mosquitoes. According to our experimental design we were expecting 250 million reads per genome, enough to provide a coverage exceeding 30X for a genome of circa 1Gb. After trimming and quality check, we could use for assembly less than 20% of the expected reads: this resulted in low coverage and poor assemblies. It is possible that the poor quality of the reads was caused by our decision of sequencing from one single individual. For doing this, we had to build libraries from slightly less than 1µg of DNA : this low amount of DNA was likely responsible for a reduced quality of the libraries and a subsequent low quality of reads. The risk of sequencing from a single individual is more relevant for long reads approaches which require a higher amount of input DNA compared to the short reads approach (Richards 2019; Kingan et al. 2019; F. Li et al. 2019). We tried to use long reads for *A. koreicus* employing Nanopore MinION, but we repeatedly failed to obtain reads: this again was likely due to the low amount of starting DNA (1µg) coupled with the many passages for nanopore library preparation that has damaged the integrity of DNA molecules. Based on our results, we advocate extreme care in sequencing from a single individual that can provide less than 1 µg of DNA.

Conclusions

While the quality and the low number of reads did not allow us to perform a proper genome assembly, our data is suitable for a genome skimming approach. We could, for example, extract high-quality whole metagenomes (Zadra et al. in prep) and we could retrieve, at least for *A. japonicus*, a reasonable amount of genes for downstream genetic and phylogenomic application (Benjamin Linard et al. 2015; Malé et al. 2014; Denver et al. 2016; B. Linard et al. 2016).

Part B. The short-read genome of two populations of the parasitoid wasp *Trissolcus japonicus* from Italy: male haploidy, metagenomic screening, and identical mitogenomes

Zadra Nicola[1,2], Andrea Silverj[2], Battiada Matteo[1], Pindo Massimo[1], Anfora Gianfranco[3], Mazzoni Valerio[1], Rota-Stabelli Omar[3].

[1]Fondazione Edmund Mach; [2]CIBIO department, University of Trento; [3] Center C3A, University of Trento.

Abstract

Trissolcus japonicus is an egg parasitoid wasp widely recognized as the best candidate for the classical biological control of the invasive bug *Halyomorpha halys*. Practical use of this wasp in the field relies on a good understanding of its biology and on an intra-specific typing system to differentiate strains for assessing releases. Here we used short-read sequencing to generate genome data from a single female from a lab-reared population used for biocontrol studies (CREA) and a male from a recently established population (TN) from Trentino, North Italy. We assembled two distinct draft genomes with an N50 of up to 150 Kb, characterised their associated metagenome, and extracted full-length mitogenomes. The total DNA yield after extraction was circa double in the female with respect to the male, suggesting that males are haploid in this species. The mitogenomes of the two individuals were identical to each other and identical to the previously published CREA mitogenome: this suggests that the TN and the CREA individuals cannot be discriminated against using mitochondrial markers. Ongoing studies are directed towards characterizing and validating differences in the nuclear genome that may be used for typing in release and re-catch experiments. The assembled genome is a source of data for a variety of evolutionary and applied studies.

4.5 Introduction

4.5.1 *Trissolcus japonicus*: the *Halyomoprha halys* control agent

Trissolcus japonicus (Ashmead, 1904) is a tiny hymenopteran of the Scelionidae family; originally from Asia, it is now present as an adventive organism in various countries, including USA and Italy (Talamas, Masner, and Johnson 2013; Stahl et al. 2019). *Trissolcus japonicus* has become popular as a natural parasitoid of *H. halys*, an invasive pest from Asia which is now posing a threat to agriculture in many European and American countries (Lee et al. 2013). Different strategies have been investigated to contrast *H. halys* invasion, but biological control seems the most promising technique. Although *H. halys* eggs can be attacked by several parasitoids of the genera *Trissolcus*, *Telenomus*, *Oencyrtus* and *Anastatus* (Lee et al. 2013; Arakawa, Namura, and others 2002; Z.-Q. Yang et al. 2009), various studies identified *T. japonicus* as the candidate with the highest potential for classical biological control programmes (Z.-Q. Yang et al. 2009; J. Zhang et al. 2017). *Trissolcus japonicus* is considered the natural enemy of *H. halys* because it has an egg parasitisation rate of 80% in its native range (Z.-Q. Yang et al. 2009). For this reason, various countries are developing programmes for the control of the bug through the introduction of this wasp. In the Trentino region, *T. japonicus* was released in 2020 for biological control, and the sample investigated in this paper come from a Trentino wild population and released individual (CREA).

4.5.2 Typing *Trissolcus japonicus* for release assessment

The ineffectiveness of nets and chemical methods to fight *H. halys* in Italy has led the National Plant Health Committee (Comitato Fitosanitario Nazionale, CFN) to promote classical biological control programmes through a complex network of local institutions coordinated by CREA-DC. One of the main steps of this program is the release of *T. japonicus* in 712 sites distributed in six northern Italian regions during the 2020-2023 period. The CFN has identified a particular *T. japonicus* population of American origin for the release (identified as CREA population, from here on, this population is referred to as CREA-*T. japonicus*), which is now reared in the various involved institutions. An important phase of the release programme is the one following the actual CREA-*T. japonicus* release: the post-release environmental monitoring. This delicate phase aims at

evaluating the impact and the diffusion of *T. japonicus* in the agricultural contexts affected by *H. halys*. One type of control consists of collecting *H. halys* eggs from the wild and, at the time of their hatching, estimating if the level of *T. japonicus* parasitisation is actually increasing. It is, however, important to assess how much of a hypothetical increase in *H. halys* parasitisation is due to the release of the CREA population rather than to the already present adventive *T. japonicus* population. This can be done by performing a recapturing experiment and by typing the captured specimens using genetic markers.

Previous analyses indicated that the barcode COI is identical in the reared CREA population and in the wild type captured in Trentino before the first release of CREA. This impairs the actual typing of the different populations using the current canonical barcode COI (Cytochrome c oxidase subunit I) marker and suggests that more data is necessary for strain discrimination, possibly by employing mtDNA specific markers designed on more variable regions of the mitogenome.

To help with the typing of *T. japonicus* releasing in Italy and to obtain genome data to make genetic studies easier, here we present the draft genome of two individuals: one member of a Trentino population and one from the reared CREA population.

4.6 Methods

4.6.1 Sampling

The CREA *T. japonicus* samples were reared at the Agricultural Entomology (EA) unit of Fondazione Edmund Mach from samples distributed by CREA and initially imported from the USA (USDA-ARS, Beneficial Insects Introduction Research Unit, Delaware). The Trento *T. japonicus* samples come from a lab-reared population established from individual hutched from *H. halys* eggs collected during a 2019 *H. halys* monitoring in Trentino. The rearing has been maintained isolated (no other introduction of individuals from outside) at Fondazione Edmund Mach. CREA and the Trento strains were reared in vials containing 20/30 individuals, fed with honey drop and regularly supplied with *H. halys* eggs.

4.6.2 Extraction and sequencing

We used a Nucleospin Tissue extraction kit to obtain DNA from two different samples. The samples consist of one female from the Trentino population and a female from the CREA population. We performed the extraction on adults. The DNA extraction was performed on two separate days to minimise the chances of accidental sample swapping and cross-contamination. Cell lysis was performed by incubating the sample with Proteinase K and Sodium Dodecyl Sulfate (SDS) solution. After the cell lysis, the DNA was isolated using silica membrane in combination with salt and ethanol. At the end of the procedure, DNA was eluted in 50 µl volume, and quality/quantity was assessed using Qbit fluorometers.

The total DNA of the two samples was ligated into libraries and sequenced in NOVASEQ SP as 200 PE. The library insert size with its max and min are 550 (1000 - 300) bp. The reads output yield was roughly 500 million reads per sample for all the samples of the fastq.gz file. The reads quality was assessed using fastQC tool (https://github.com/s-andrews/FastQC).

4.6.3 Assembly and BUSCO assessment

We employed the MaSuRCA 3.8 assembler for the assembly using the short Illumina PE reads with a reads length of 100 nt (Zimin et al. 2013). The assembly statistic shows a high coverage of 100x. GenomeQC was employed for retrieving genome statistics from the two assemblies (Manchanda et al. 2020). In light of the genome size of the only Scelionidae that has been sequenced (*Telenomus remus*, 130 Mb (H. Xu et al. 2021)) and the notoriously small genomes of parasitoid wasps, we can speculate that the genome of *T. japonicus* probably has a similar size and infer a good coverage of our assembly (>100x). In BUSCO v5.2 (Simão et al. 2015; Waterhouse et al. 2019), we assessed the completeness of the two genomes obtained in the assembly. We ran two analyses for each genome, one using the orthologous Insect dataset and one using the Hymenoptera orthologous dataset. Testing different orthologous datasets allowed us to evaluate if the genome completeness is consistent through different gene sets.

4.6.4 Mitochondrial extraction

To assemble the mitogenomes of the two samples, we employed MitoZ 2.3 (Meng et al. 2019) starting from the raw data. We employed the entire MitoZ pipeline from reads filtering through assembly annotation. MitoS software was employed to recover missing genes (Bernt, Donath, et al. 2013).

4.6.5 Microbial and viral profiling

Raw reads were quality checked using FastQC and trimmed with Trimmomatic. Cleaned reads (paired and unpaired) were used as input of MetaPhlAn-3.0, which was run using the mpa_v30_CHOCOPhlAn_201901 database with and without the option "--add_viruses". Relative abundance plots were obtained separately for bacteria and viruses using hclust2.

4.7 Results and Discussion

4.7.1 Aploid males?

We extracted the DNA from two different populations, the wild introduced population of Trentino and the *T. japonicus* from CREA, that are being released for biological control in Italy. The DNA yield after extraction was circa twice as high in the females than in the corresponding males: CREA female 4.67 ng/µl, TN female 4.83 ng/µl, CREA male 1.73 ng/µl, TN male 2.00 ng/µl (we sequenced only two individuals, but extracted DNA from four). This suggests that females are diploids and males are haploids in *T. japonicus*.

4.7.2 Good quality genomes using only short reads

The CREA and the Trentino samples have a high average quality score (PHRED average score of 36), no drop of quality or any other issues were detected. The genome size for every assembly is around 160 Mb (see table 1 for detailed information). The coverage obtained in the four runs is very similar to that expected from the amount of the initial reads (~100x). The genome was successfully assembled by the MaSuRCA assembler, and the assembly statistics obtained using GenomeQC show a high-quality assembly (Table 1). BUSCO

assessment shows a high-quality assembly, retrieving 98.4 % (1342/1367) of gene completeness in the Trento sample and 96 % (1312/1367) in the CREA sample employing the insect gene orthologous dataset (insecta_odb10).

This is the second Scelionidae species to have a genome sequenced, whereas other Hymenoptera clades have received more attention from genome studies. This lack could be explained by the size of the species belonging to the Platygastroidea clade. Platygastroidea clade is known as miniaturised wasps, which have undergone a dramatic reduction in body size throughout their evolution (body sizes span between 0.5 mm and 10 mm). Small body size has been unsuitable for sequencing from a single animal because of the insufficient DNA yield that can be obtained from small insects (Richards and Murali 2015). As shown by our sequencing, this limitation is being surpassed by the amelioration of short-read chemistry. The likely high homozygosity of this parasitoid wasp (see next paragraph) likely helped the effective assembly in the absence of long reads.

	CREA Male (DMSO_3)	***Trento Female (DMSO_4)***
Total scaffold length	156 Mp	168 Mp
N50	24640 bp	150093 bp
L50[1]	1473	290
Longest scaffold	298943 bp	12225666 bp
Estimated genome size	189157107 bp	187864075 bp
Number of scaffolds	58707	18593

Table 1. The table shows the assembly statistics for the two genomes. As we can see, the Trento sample resulted in a better assembly than the other. However, both the genomes show good quality statistics that make the assembly suitable for further genomic studies.

[1] L50 is defined as the count of the smallest number of contigs whose length reaches 50% of the genome size.

4.7.3 Small genome and a likely gene family loss

The genome size of *T. japonicus* has been evaluated to be ~160 Mb, which is smaller than those of other sequenced hymenopteran species (80% of sequenced species hymenopteran genomes range between 180 to 340 Mb) (Branstetter et al. 2018). The genome length is, however, similar to the newly assembled genome of *Telenomus remus* (129 Mb), the only other Scelionidae genome sequenced so far (H. Xu et al. 2021).

When we employ the Hymenoptera dataset (hymenoptera_odb10) in BUSCO, we notice a drop in the percentage of genes retrieved. Trento sample shows 85.6% (5133/5991), whereas CREA sample showed 81.7% (4895/5991) of completeness (Figure 1). . This increased amount of missing genes is likely due to the fact that the Hymenoptera dataset is biased towards bees and wasp clades. Until a few months ago, there was no Platygastroidea genome record, so the BUSCO orthologous dataset is clearly biased toward other clades. This points toward a genuine lack of some orthologous genes from the parasitoid wasp genome: this hypothesis should be tested in future by performing a thorough gene modelling (proteome) of *T. japonicus* genome and by conducting comparative genomic studies. It would be, for example, interesting to understand the genomic bases behind the driving force toward small body that has evolved independently in many lineages of Hymenoptera.

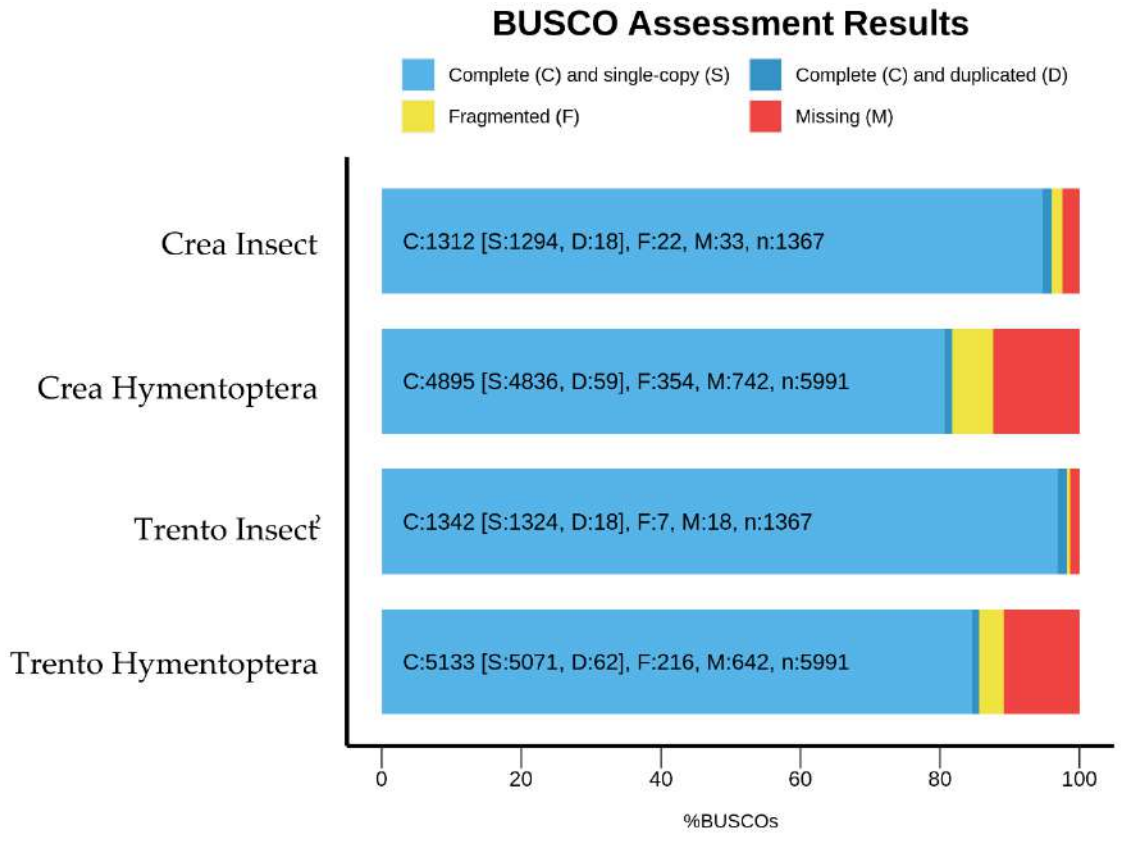


Figure 1. Results of BUSCO analyses. The analysis was carried out using two different datasets Hymenoptera and insect. The light blue bar indicates the percentage of complete genes in a single copy; the dark blue bar represents the duplicated complete genes found; the yellow bar shows the percentage of fragmented genes; the red bar indicates the missing genes, namely the genes that are supposed to be present in the species in question but not found in the genome assembly provided to BUSCO. The percentage of missing genes using the Hymenoptera dataset (hymenoptera_odb10) is higher than the missing genes found employing the Insect dataset (insecta_odb10).

4.7.4 Identical mitogenomes

The mitogenomes were successfully assembled for the CREA and Trentino samples, with sizes of 16142 bp and 16141 bp, respectively. MitoZ failed to retrieve five tRNAs: tRNA-Arg, tRNA-Asp, tRNA-Glu, tRNA-Gly; tRNA-Ser. Reannotation using MitoS resolved this issue. Unfortunately, the two mitogenomes show an identity of 100%, which means that the comparison cannot provide an SNP to develop a rapid molecular identification toolkit to distinguish the two populations for re-capturing experiments. Small close populations can fix the same haplotype due to stochastic effects, or the haplotype can come from the same population where the haplotype was already fixed. Interestingly, in this species' mating behaviour, the offspring tend to reproduce with siblings because they are already sexually mature at the moment of hatching. This may prevent genotype mixing even among sympatric populations and would promote genetic isolation and homozygosis. Despite the result of our analysis, an SNP was detected in the mitochondrial marker (COI) among local and native range *T. japonicus* individuals (Stahl et al. 2019).

4.7.5 Metagenomic screening indicates a difference between males and females

To detect possible contaminants and characterise the microbes associated with the two sequenced *Trissolcus* individuals, we ran MetaPhlAn-3.0 (see Methods). The female individual presented a higher diversity and abundance of microbes, with some species detected only in the male. Interestingly, we detected *Wolbachia* only in the female. As the genome of *Wolbachia* is particularly short, we decided to control for its presence by directly mapping raw reads against a large panel of *Wolbachia* reference genomes (Figure 2).

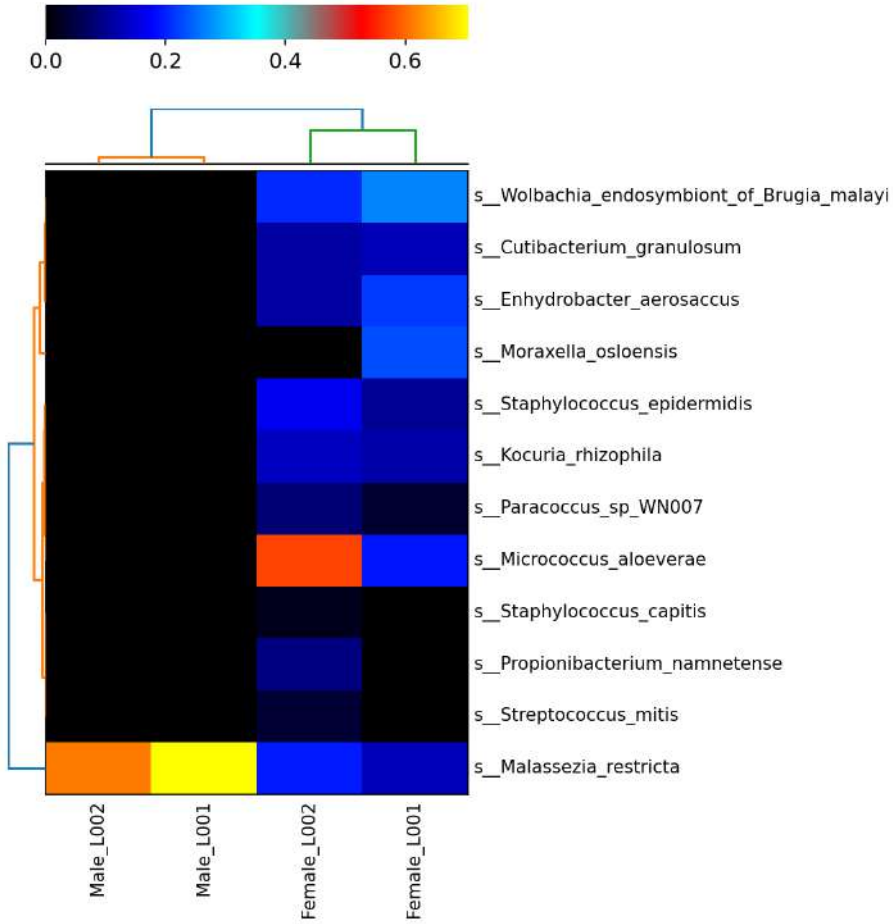


Figure 2. Metagenomic profile of the *T. japonicus* sequencing lanes from this study. The female individual is characterised by a higher richness of bacterial species compared to the male. Results are coherent among different sequencing lanes, indicating that there is no specific bias at this level.

4.8 Conclusion

In this study, we presented the genome *T. japonicus* genus and provided a solid reference for developing rapid genetic-driven measures for the management and control of native and released *T. japonicus* populations. The Trentino and the released populations have an identical mitogenome. Nuclear heterogeneous markers may, however, provide a useful tool to rapid discriminate between the two populations. Molecular characterisation and building a vast set of markers could highlight probable breeding between CREA (the released population for biological control purposes) and the Trentino *T. japonicus* populations. Studying the *H. halys* parasitoid's population dynamics would help evaluate the impact of the biological control measure taken in 2020. If one population is taking over the other population, and if there will be some degree of genetic admixture. Eventually, this practice would help assess the effectiveness of releasing CREA

T. japonicus population and control the population dynamics of this species in the non-native environment.

From a methodological point of view, this work has shown that high-quality genomes can be retrieved even for small insects using the "basic" short reads sequencing.

CHAPTER 5 - EXTENSIVE PHYLOGENOMIC ANALYSIS OF ZIKA VIRUS PROVIDES AN UPDATED SCENARIO OF ITS ORIGIN AND EVOLUTION

5.1 Introduction to the Chapter

In Chapter 5, I present a study that extended the molecular clock study on Zika virus (ZIKV) using a different approach than other phylogenetic analyses presented in this thesis. Employing different datasets, I aim to analyze some neglected aspects of ZIKV evolution and dynamics, such as dating deep nodes and investigating Thailand's role as a source of ZIKV infection in Southeast Asia. In this study, I reconstruct the ZIKV phylogeny using the BEAST2 software, testing several tree priors and clock models through model selection analysis.

I first analyzed the ZIKV recombination and the early Asian-African divergence to date the origin of the two main lineages circulating today. I put effort into reconstructing the ancestral state location in ZIKV evolution, dating the re-introduction of ZIKV in Africa from the Americas. Secondly, I investigated the divergence time between the ZIKV lineage and the Spondweni virus (SPOV), its closest known sister species, for dating ZIKV origin. Indeed, the recent ZIKV dynamics have been extensively explored, while the investigation of deep nodes has been neglected. Furthermore, the study highlighted the error borne by the sequences obtained from samples with a long cell passage history. The results provide an updated picture of ZIKV evolution and propose the timing of many nodes that have not been investigated so far. We suggest a pivotal role of Thailand as a source of infection based on the node ancestral state reconstruction and a ZIKV diversification dated around 1850. This Chapter extends the molecular clock search to an arbovirus vectored by *Aedes* mosquitoes.

Contribution: The study was started at the beginning of the PhD along with the project on Aedini phylogenies (Chapter 2). I performed the experiment design, the evolutionary analysis, and the writing, guided by my advisor Omar Rota-Stabelli. In this study, I got my first introduction to bioinformatics, and I compiled python scripts for formatting the input data and for selecting suitable sequences based on .gb metadata. This Chapter is in preparation for Plos Neglected Disease.

Extensive phylogenomic analysis of Zika virus provides an updated scenario of its origin and evolution

Nicola Zadra[1,2], Annapaola Rizzoli[1] and Omar Rota-Stabelli[1]

[1]Research and Innovation Centre, Fondazione Edmund Mach, San Michele all'Adige, Italy

[2]CIBIO University of Trento, Italy

[3]C3A University of Trento, Italy

5.2 Abstract

While the patterns of recent Zika virus (ZIKV) outbreaks have been studied in detail, the origin and the early timing of its evolution are less clear. To address these issues, we collected a large set of nearly 500 ZIKV genomes, performed thorough phylogenetic studies, checked for recombination events and applied a robust model selection. Our phylogenies identify recent recombination between Singaporean and African lineages, suggesting that multiple distant lineages are co-circulating during outbreaks. We confirm Thailand's pivotal role in the origin and spread of the virus in Asia, and find that the phylogenetic signal to define early African divergences is too weak to establish definite conclusions. Overall, our analysis points toward an eleventh-century origin of the ZIKV lineage, followed by a recent first diversification in the nineteenth century. We finally analyze the timing of the re-introduction to the African continent of the Asian lineage and address the delay between probable introduction and outbreak onset. Our study provides novel insight into the origin and early dynamics of ZIKV, which will improve our knowledge of its biology and help forecast future outbreaks.

5.3 Introduction

Zika virus (ZIKV) is a positive-stranded ssRNA mosquito-borne Arbovirus, belonging to the Flaviviridae family and Flavivirus genus. ZIKV was first reported during a Yellow fever survey in Uganda in 1947 (Kirya, Mukwaya, and Sempala 1977; Kirya and Okia 1977; Dick G 1952), while the first human case was reported in 1954 in Nigeria (MacNamara 1954). Since its discovery and proven pathogenicity, this virus has been considered a neglected tropical disease involved in a few local outbreaks, mainly in the Pacific Islands, such as Yap Island 2007 (M. R. Duffy et al. 2009), French Polynesia 2013(Cao-Lormeau 2014) and New Caledonia (Hayes 2009; Ioos et al. 2014; Lanciotti et al. 2008). However, after the emergence of ZIKV in the Americas, the interest in this virus rose rapidly. In March 2015, ZIKV was first detected in Brazil (Zanluca et al. 2015), when ZIKV had the chance to spread all over the Americas, from South America to Florida (Q. Zhang et al. 2017; N. R. Faria et al. 2017; Metsky et al. 2017; Massad et al. 2017; Nuno Rodrigues Faria et al. 2016). ZIKV infection usually causes mild effects that last less than seven days. However, it can cause microcephaly and birth defects in the fetuses of infected pregnant women (Mlakar et al. 2016; Rasmussen et al. 2016; Ventura et al. 2016).

The epidemiology of late Zika evolution has been well established, and many studies have addressed the dynamics of Zika virus circulation since the onset in Brazil, which began in 2013 (Metsky et al. 2017; Nuno Rodrigues Faria et al. 2016; Musso 2015). These studies outlined a single introduction from the Pacific Islands into the Americas. At the same time, other work describes ZIKV epidemiology in South American countries and the timing of the multiple introductions in Florida, Mexico, and Caribbean islands (Grubaugh et al. 2017; Costa et al. 2021). Despite all the information gathered in the last few years regarding recent Zika dynamics (N. R. Faria et al. 2017; Nuno Rodrigues Faria et al. 2016; Hu et al. 2019; L. Wang et al. 2016), ZIKV evolutionary history before American outbreaks is still lacking and unclear. Some recent studies suggest a long term circulation of this virus in Thailand (Ruchusatsawat et al. 2019; Phumee et al. 2019), India (Yadav et al. 2019) and other countries in Southeast Asia (Z. Y. Liu, Shi, and Qin 2019); moreover, a ZIKV seroprevalence was observed in many countries in Southeast Asia (Musso 2015). However, the phylogenetic relationship between the strain involved in subsequent Zika onset in Southeast Asia remains unclear (Z. Y. Liu, Shi, and Qin

2019; Duong, Dussart, and Buchy 2017). Furthermore, the origin and the early diversification of ZIKA evolution have not yet been investigated thoroughly.

In particular, the origin of ZIKV and the early divergence of its Asian and African strains are still unclear. In recent years, new ZIKV infections caused by American strains were reported in many African countries, e.g. Angola (Hill et al. 2019) and Cape Verde (Oumar Faye et al. 2020), underlying the ZIKV spreading potential. ZIKV re-emerging in Africa is a concern because of the putative increased pathogenicity it has acquired (Z. Y. Liu, Shi, and Qin 2019; Pettersson, Bohlin, et al. 2018). In addition, the American ZIKV strain may come in contact with many mosquito species in Africa and, once in sympatry, new potential mosquito vectors could be involved in the infection routes (Kauffman and Kramer 2017; Weger-Lucarelli et al. 2016; Epelboin et al. 2017).

The paper presents comprehensive results that shed light on Southeast Asia's – Thailand's in particular - importance as an infection reservoir for the three outbreaks that have occurred so far in Asia (Singapore, Yap Islands, French Polynesia). Furthermore, the paper investigates the origin of ZIKV and the ZIKV African – Asian split in more detail. Zika phylogeny was extensively investigated in a shallow time window and on a local scale, with particular attention to its emergence in Brazil. The authors want to address the need and the usefulness of a more comprehensive look into Zika evolution, investigating the origin of ZIKV and the pre-pandemic dynamics on a global scale. In conclusion, we address the pivotal role of Thailand in nourishing the ZIKV spread in East Asia and the lack of detection of ZIKV circulation in countries that faced outbreaks.

5.4 Methods

5.4.1 Datasets preparation and alignment

ZIKV sequences were downloaded from GenBank on November 1, 2019, with a total of 1733 hits (Sayers et al. 2019). Sequences were filtered by the following criteria: GenBank files must include collection date, sampling location and only sequences longer than 600 bp were considered. Around 550 sequences were gathered from which duplicates and ambiguous sequences were removed manually, leading to a final set of 479 sequences. The history of the sample was

analyzed for some sequences (full list in Supplementary Table 1), as many old African samples had gone through many cell line passages, while others had an unknown passage history. Hence, removing the unreliable sequences from the dataset prevents the risk of introducing sequences that display a biased phylogenetic signal due to the mutations accumulated during the cell passage. These sequences are not consistent with the tip dating approach and they could lead to undesired phylogenetic artefacts. Applying the same criteria, we downloaded five Spondweni viruses (SPOV), only one of which satisfied the criteria. Misaligned sites were corrected manually and flanking regions (5' and 3' UTR) were removed from the alignment due to the high variability and high amount of missing data. We employed three different datasets to investigate ZIKV evolution, the first involving all 474 sequences. The second dataset was prepared using a subset of the previous dataset using only complete CDS sequences and filtering the sequences in over-represented clades (mainly American sequences and the Singapore sequence) in order to reduce the taxon sampling bias and obtain a more suitable dataset to investigate the deeper phylogeny of ZIKV. The third dataset was obtained by adding the Spondweni sequence to estimate the Zika virus origin. All the datasets were aligned using MAFFT (Katoh, Rozewicki, and Yamada 2018).

5.4.2 Recombination

The dataset was analyzed with RDP v4.4.8 (Martin et al. 2015) to prevent possible phylogenetic biases due to recombination events. This software allows analyzing the same datasets with different tools at the same time and includes different methods, such as GENECONV (Padidam, Sawyer, and Fauquet 1999), Chimaera (Posada and Crandall 2002), MaxChi, Bootscan (Salminen et al. 1995) and 3Seq (Boni, Posada, and Feldman 2007). A recombination event was considered significant if detected by at least four methods out of five, with a p-value <= 0.01, applying the Bonferroni correction to avoid false positives. IQ-TREE 1.6.12 was employed to assess the differences in evolutionary history between the recombinant sequence and the rest of the genome (for the phylogenetic method, see section 1.3, Methods). The sequences that show recombination patterns were removed from the alignment for further analysis.

5.4.3 Phylogenetic analysis

IQ-TREE 1.6.12 (Nguyen et al. 2015; Trifinopoulos et al. 2016) was used for Maximum-likelihood (ML) analysis. The ML phylogenies were obtained for all the datasets tested, using ultrafast bootstrapping (Minh, Nguyen, and Von Haeseler 2013) with 1000 bootstrap alignments, 1000 maximum iterations, approximate Bayes test and the SH-like approximate likelihood ratio test.

BEAST v2.5 (R. Bouckaert et al. 2014; Baele et al. 2013) was employed to explore the timing of Zika evolution from the early divergence time to its origin. We generated a xml files with multiple combinations of clock models and tree priors to assess which prior model performed better on the dataset. Two runs per each xml file were executed. In total, we use seven prior combinations, relaxed clock combined with three coalescent tree priors Coalescent Constant, Coalescent Exponential and Coalescent Bayesian skyline. The same tree priors were used in combination with the strict clock. In addition, we employed the Birth-Death Serial (BDS) tree prior in combination with the relaxed clock to compare the speciation tree prior (BDS) against the coalescent priors. All the clock priors were set with a minimum of 10^{-5} and a maximum of 10^{-2} using a uniform distribution. This value range is observed by Duffy et al. (2010) for the RNA viruses, and it comprises the clock rate observed in the previous ZIKV study (Metsky et al. 2017; S. Duffy, Shackelton, and Holmes 2008; Pettersson, Eldholm, et al. 2018). The substitution model employed in every run was GTR + γ. All the chains were run for 200,000,000 generations until they reached convergence, which was assessed using Tracer1.7 (Rambaut et al. 2018). All BEAST runs were calibrated with tip dates, where the most recent sample is set as zero time in the tree (10/10/2018). Moreover, we used the collection location as a discrete trait to infer the node's state in the phylogeny. To investigate ZIKV evolutionary dynamics, we employed the three datasets described in Paragraph 5.4.1. We run BEAST analysis on each dataset employing the same priors and parameters suggested by the model selection (see next Paragraph 5.4.4). The reduced dataset containing 118 ZIKV sequences uses the collection location as a discrete trait to infer the node's ancestral state (R. Bouckaert and Xie 2016).

5.4.4 Model selection

BEAST2 analyses were run using different sets of priors. Stepping Stone (SS) method was employed to compare all the evolutionary models applied in this study. SS allows for the comparison of different analyses using the marginal likelihood and the Bayesian factor. This type of model has several advantages over other alternatives, such as AICm and harmonic mean (Baele et al. 2013). The MODEL SELECTION package was used to perform Log marginal likelihood estimates for a different combination of the molecular clock and coalescent tree prior. The best-fitting model combination is ranked 1, while the worst is ranked 7 (see Table 1). Model selection was performed on every xml compiled for this analysis described in the previous paragraph (5.4.3). The evolutionary models were tested along with both strict clock and uncorrelated relaxed clock. In addition, we run the model selection on the BDS to compare a speciation model against the coalescent that are supposed to be the most fitting models for describing ZIKV evolutionary dynamics. The Bayesian factor was calculated as described in the BEAST tutorial (Barido-Sottani et al. 2018).

5.5. Results and discussion

5.5.1 Dataset employed in the analysis

After filtering the sequence following the criteria listed in paragraph 5.4.1, we obtained three different datasets. The larger dataset, the first employed in the analysis, contains 479 sequences of 10811 nt length, with 28 sequences collected in Africa, 123 Asian, 24 from Pacific Islands and 305 sampled in the Americas. Moreover, we carefully include in our datasets the French Polynesian sequence KX447518 which was found the closer relative to the American outbreak by Pettersson (Pettersson, Eldholm, et al. 2018) to obtain a comparable node for the origin of the American outbreak. The second dataset employed is a subsample of the first, including 117 samples that comprise 8 African samples, 35 Asian samples, and 74 American samples. The final analyzed dataset includes 118 sequences. It was obtained by adding the only suitable SPOV sequence to the second dataset described here.

5.5.2 Recombination

RDP4 analysis detected two recombination events, suggesting the Singapore strain as a major contributor and the Uganda strains (1947; Accession: HQ234498, 1962; Accession: KY288905) as minor contributors. A breakpoint was detected in the E genomic region, while another similar breakpoint was detected in the African sequences (O. Faye et al. 2014). RDP4 detected this recombination event with all the methods employed, given a p-value threshold of 0.01. The phylogenetic analysis highlighted the recombination event (see Figure 1). The samples KY241717 and KY241717 collected in Singapore appear in different positions in the two trees, one using the recombinant region and one the rest of the genome. In the phylogeny of the recombinant region the two Singapore samples cluster with the African clade with strong support (bootstrap support: 99; SH-aLRT: 98.8). The phylogenetic tree obtained by the remaining genome sequences places the two recombinant samples with the remaining Singapore samples. These sequences were excluded from further analysis to avoid systematic error (Posada and Crandall 2002). The recombination event suggests a cocirculation of an undetected African strain in Singapore, with a secondary re-introduction of ZIKV in Asia. The recombination event in Flavivirus seems not to be as common as in other groups of positive-stranded ssRNA(+) viruses (Taucher, Berger, and Mandl 2010); however, recombination occurrence was observed especially in the Dengue virus group with which ZIKV is related to (Holmes, Worobey, and Rambaut 1999; Simon-Loriere and Holmes 2011; Tolou et al. 2001).

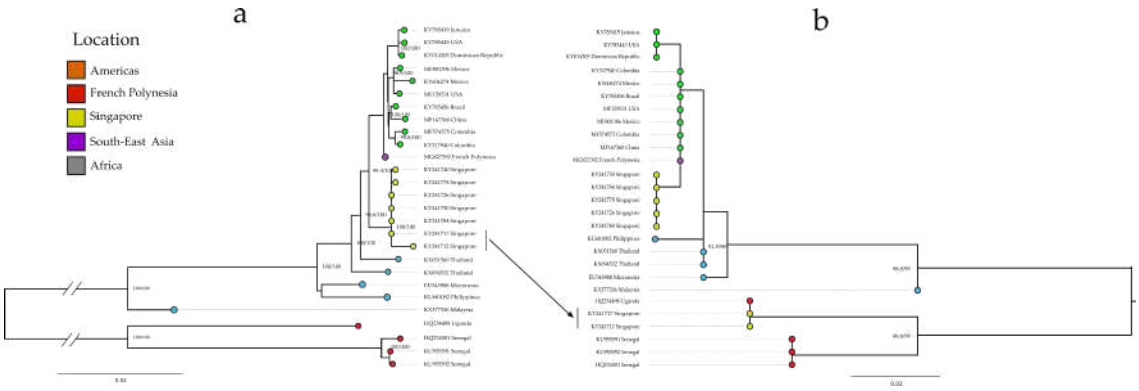


Figure 1. Phylogeny of recombinant and non-recombinant regions in a Singapore sample. a) ML tree of the non-recombinant region using IQ-tree using a subset of sequences: all the Singapore sequences cluster together. b) ML tree of the recombinant region. Singapore samples (KY241712 and KY241717) cluster with the African sequence. Tips are coloured in accordance with the sampled location; this criterion is consistent in all the figures in this paper.

5.5.3 Model selection

Marginal likelihood provided support for the relaxed molecular clock, with the best-fitting tree prior being a nonparametric Bayesian Skyline. Table 1 shows six combinations of parameters, and it ranks them via Bayes factor (BF) (Kass and Raftery 1995). Following Kass and Raftery, the evidence strength of a hypothesis against H_0 was rated as BF<3=no evidence; 3–20 = positive support; 20-150 = strong support, and > 150 = overwhelming support (Kass and Raftery 1995).

Clock prior	Tree Prior	Marginal Likelihood	BF*	BF[1]	Rank
Relaxed Clock	Coalescent Bayesian Skyline	-37482.14905	108.05095	15.95496897	**1**
	Coalescent Constant	-37489.91496	100.2850412	8.189060159	2
	Coalescent Exponential	-37534.1891	56.01090227	-36.08507881	4
Strict clock	Coalescent Bayesian Skyline	-37498.10402	92.09598107	-	3
	Coalescent Constant	-37555.80642	34.39358112	-57.70239995	5
	Coalescent Exponential	-37561.11131	29.08869762	-63.00728345	6
Relaxed Clock	BDS	-37590.2321	-	-92.12808	7

Table 1. The table shows the model selection using the Stepping Stone (SS) analysis. Comparing the marginal likelihood value using the Bayes factor (BF) and ranking it in a readable format. The last column ranks the models based on the marginal likelihood analysis.
*Bayesian factor calculate using the Birth and Death Serial (BDS) as a comparison for other models
[1] Bayesian factor calculated using the Calescent Bayesian Skyline strict clock as a comparison for other models

In Table 1, we can see that the lognormal uncorrelated clock using a coalescent tree prior is favoured over the other strict clock. The analysis slightly rejected the coalescent constant model (see Table 1), but the model provides similar tree topology and posterior estimates, as we can see in Supplementary Figure 1 and in Table 2.

Clock prior	Tree Prior	Rank	Clock rate* (10^{-4})	Root age (95% HPD)	American crown (95% HPD)	σ_r
Relaxed Clock	**CBS**	**1**	**7.01**	**135.2 (85.67 - 197.24)**	**5.96 (5.3 - 6.86) 2012/10/17**	**0.539**
	CC	2	7.25	146.7 (86.07 - 227-62)	6.23 (5.49 - 7.2) 2012/07/16	0.624
	CE	4	7.22	107.8 (62.93 - 177.07)	6.25 (6.48 - 7.17) 2012/07/22	0.649
Strict clock	CBS	3	5.59	176.4 (159.8 - 193-03)	6.33 (5.9 - 6.79) 2012/06/19	-
	CC	5	5.63	178.5 (157.64 - 200.49)	6.48 (5.97 - 7.06) 2012/05/05	-
	CE	6	5.61	177.4 (158.48 - 199.69)	6.48 (5.98 - 7.06) 2012/05/05	-
Relaxed Clock	BDS	7	9.87	66.3 (55.42 - 79.84)	6.04 (5.33 - 6.92) 2012/9/10	0.783

Table 2. Here we report key results for the different models compared in the analysis: the mean clock rate, the root age and the American crown age are shown for all the combinations. For the model where we employ the relaxed clock, we provided the coefficient of variation (σ_r = the clock standard deviation divided by its mean), this parameter measures the clock-likeliness of the rate. BDS (Birth-Death Serial) model was chosen to represent the outlier. the coalescent models are preferred over a speciation model as BDS.
*(mutation/site/year)

BF does not definitely discriminate between all the models (meaning there is only positive support and not strong support) and provides a similar estimate for the root and the American crown group. Even though the Bayesian Skyline tree prior is slightly favoured over the Coalescent Constant, the posterior estimates are constant. Moreover, we observed that models providing older node estimates are rejected along with the models that provide younger node estimates, showing that the data are better explained by a model that places the mean age of the tree at the end of the 19th century (Supp. Figure 1). The molecular clock does not vary much across the models employed in this study (Table 2), even though the selected clock (relaxed clock) shows a higher overall molecular rate than the strict clock.

In addition, the coefficient of variation (σ_r; the standard deviation divided by the mean in Table 2) suggests that the relaxed clock assumption is theoretically sound since this parameter is estimated to be 0.53. σ_r measures the clock-likeliness of the data; if σ_r is close to zero (lower than 0.2), then the data have low rate variation and they can be modelled as a strict clock; vice versa, values that range between 0.2 and 1 show the relaxed clock assumption over the strict clock (Barido-Sottani et al. 2018; Drummond et al. 2006). The selected model was then employed in the analysis with SPOV as well.

5.5.4 Sample history can affect the mutation rate

The phylogenetic analysis provides results for nodes that have not been studied yet in other works. Investigating the deep nodes for ZIKV is challenging and needs specific care in the data curation. In particular, several old samples from Africa were retained in the dataset in many studies (O. Faye et al. 2014; Pettersson, Bohlin, et al. 2018; Freire et al. 2015), which can represent a possible source of error for the calibration analysis and become an issue for estimating the topology of the basal node, which is essential to establish ZIKV's origin. Therefore, all samples showing a long or unspecified passage cell history were removed from the final dataset (Supplementary Table 1). An analysis was performed using a complete dataset to check the effect of cell passage history on the calibration and rates using BEAST2. The dataset employed 474 sequences containing old samples that went through cell passage history (Supplementary

Table 1). The rate behaves erratically in the tree obtained using a complete dataset without filtering out the sequences with many or ambiguous cell passages. The branch that leads to the Senegal samples, comprising sequences KF383085 and KF383034, has a mutation rate estimated at 0.0041 mutation/site/year (m/s/y), roughly five times the rate estimated here for the ZIKV phylogeny. Moreover, in the African lineage, the rate varies extensively across branches, including the maximum and the minimum mutation rate of the whole phylogeny. These differences can affect the estimated mutation rate, especially in the case of the coefficient of variation. In such a scenario, the mutation rate is likely overestimated because of two factors: i) evolution did not stop at the moment of the reported collection date. Hence, employing the tip dating approach on these sequences can bias the posterior estimates; ii) the cell culture does not apply the same evolutionary pressure on viruses as the natural environment. The selective pressure in some sites is relaxed and is no longer under purifying selection; hence the passage history can affect the analysis (Haddow et al. 2012; Bush et al. 2000). Secondly, the natural transmission bottleneck during Flavivirus infections lowers the overall virus diversity, which does not happen in cell culture. Diversity reduction through transmission bottlenecks is well documented for mosquito-borne Flaviviruses, as is the effect of purifying selection played by the host species on the virus (Forrester, Coffey, and Weaver 2014; Lequime et al. 2016; Grubaugh et al. 2016; Ciota et al. 2012). The transmission bottlenecks are missing in cell culture, providing an unrealistic high diversity in viruses reared in cell culture. Eventually, cell culture passages promote an artificial enhancement of the mutation rate, which is deleterious for evolutionary analysis.

5.5.5 Zika phylogeny

Previous studies have shown that ZIKV diverged into two lineages at the beginning of its diversification: the American and the African (Pettersson, Eldholm, et al. 2018; Gong, Xu, and Han 2017). The phylogenetic tree presented in Figure 2 provides little evidence for the monophyly of the African clade and the Asian clade because we excluded many African samples due to their unreliability for a dated phylogenetic analysis. For this reason, we can not claim the monophyly of the African clade. In addition, our African sampling (as

mentioned in paragraph 2) is incomplete and can only provide a narrow view of the diversity of the African clade, limiting our knowledge of the broad scenario of ZIKV evolution in Africa. Hence, more ZIKV sequences are needed from the African continent to better address this issue and try to answer Gong's question, 'Zika virus two or three lineages?' (Gong, Gao, and Han 2016; Gong, Xu, and Han 2017). Faye et al. [32] contributed to this topic with a deep analysis of Africa's ZIKV circulation, though this question remains unanswered.

5.5.6 Node age and trait estimates, comparison with previous work

The paragraph above shows that the relaxed clock paired with a Bayesian Skyline coalescent tree prior is the best fitting model describing our data. The tree presented in Figure 2 displays the phylogeny of ZIKV with the American and the Singapore outbreak in a collapsed format (full tree provided in supplementary material). In addition, it includes a histogram displaying the posterior probability (PP) for the ancestral reconstruction analysis. Table 3 provides the posterior estimates for the node age, the 95% HPD and the comparison with Patterson's estimates (Patterson, Sammon, and Garg 2016).

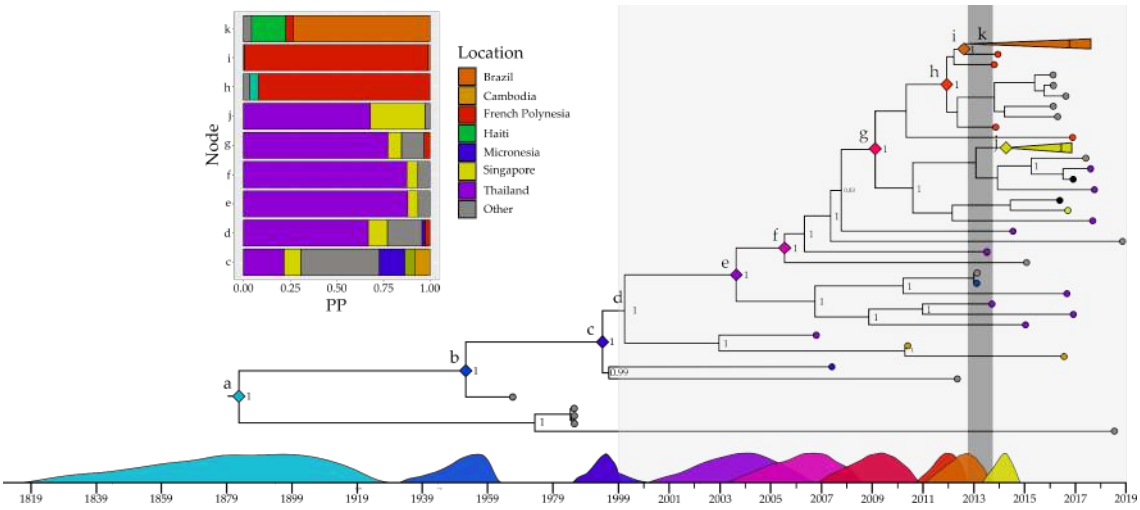


Figure 2. ZIKV phylogeny of a subset of 117 sequences. The figure has two different scales, the light grey box define a zoomed view in the recent phylogeny, the dark grey bar represents the interval of the node *i* determined by Metsky et al. (Metsky et al. 2017). The tips are colored by their correspondent collection location in the legend. On the *x*-axis, the 95% HPD are plotted; the colour corresponds with the colour of the node above. The histogram provides the posterior probability of the node location. The grey shade indicates a change of scale in the time plotted on the *x*-axis.

The African and Asian-American clades (node *a*) diverged around 135 years before the most recent sample, in other words, in 1883 (HPD 95% 1933/03 - 1821/08). The origin of the African and the Asian lineage estimates partially overlap the estimate in Pettersson, 1834-08 (1814-11 – 1852-08) (see Table 3). Pettersson's estimates employed a strict clock prior; this can explain the differences between our findings and Patterson's estimates (Brown and Yang 2011). Node *b* represents the Asian radiation that occurred in 1952 (HPD 95% 1962/07 - 1933/11), meaning that ZIKV was already circulating in South-East Asia at that time. This estimate supports the finding of Faria, Pettersson that date the emergence of ZIKV in Asia between the forties and fifties (Pettersson, Bohlin, et al. 2018; Nuno Rodrigues Faria et al. 2016; Pettersson, Eldholm, et al. 2018). The ancestral reconstruction did not provide a reasonable result for nodes *a* and *b* because of the signal paucity contributing to the inference of the ancestral state. The root is distant from its proximal nodes meaning that the node's state information is not available for a wide time span, making the inference of the nodes challenging (*a* and *b* PP are not shown in Figure 2). The following node *c* (1994/02 HPD 95% 2000/07 - 1984/11) shows the coalescent event of all the lineages involved in recent epidemics (Yap Island, Singapore and the Americas). The role of these lineages in local outbreaks makes sense in the light of Pettersson's findings, in which a mutation in the E protein is detected. This particular mutation has been shown to be closely associated with the enhanced spreading potential in Flaviviruses (Fritz et al. 2011). The histogram in Figure 2 provides the ancestral state reconstruction for this node. However, the posterior probability (PP) does not favour one location over another.

Node*	**Age**[1]	**Age** (yy/mm)	**Age** (95% HP age)	**95 HPD age** (yy/mm)	**Patterson 2018**[3] (yy/mm)
a	135.2	1883/08	85.67 - 197.24	1933/03 - 1821/08	1834-08 (1814-11 – 1852-08)
b	66.4	1952/06	56.34 - 85	1962/07 - 1933/11	1948-03 (1953-10 – 1942-04)
c	24.8	1994/02	18.34- 33.98	2000/07 - 1984/11	1994-03 (1996-11 – 1991-08)
d	19.4	1999/06	15.14 -24.66	2003/09 - 1994/03	1999-02 (2001-04 – 1997-12)
e	15.03	2003/10	11.9 - 18.66	2006/10 - 2000/03	2003-10 (2005-06 – 2001-11)
f	13.12	2005/9	10.15 - 16.2	2008/09 - 2002/08	/
g	9.59	2009/4	7.58 -11.71	2011/04 - 2007/02	/
h	8.36	2012/2	5.87 - 7.99	2012/10 - 2010/11	2012-04 (2012-17 – 2011-08)
i	6.18	2012/9	5.46 - 7.12	2013/05 - 2011/09	2012-10 (2013-02 – 2012-03)
k	5.97	2012/10	5.3-6.81	2013/07 - 2012/02	2012-11 (2013-04 – 2012-05)[2]
j	4.56	2013/11	3.8 - 5-4	2015/02 - 2013/06	/
Angola radiation	3.76	2015/2	2.58 - 4.72	2016/04 - 2014/03	/

Cape - America split	4.6	2014/4	4.14 - 5.19	2014/08 - 2013/08	/
Cape radiation	3.65	2015/3	3.19 - 4.27	2015/08 - 2014/08	/

Table 3. Estimated node age for key events in the ZIKV evolution. We provide the time expressed in years before the 0 time of the tree and in the date format (yyyy/mm), the estimations are compared with the date provided by Patterson et al. (2018). Different taxon sampling does not allow to provide all the comparisons. *(pp>0.95)

[1]Age is provided in years before the present.

[2]This node defines the American ZIKV radiation, but the topology provided by Pettersson and by us is different due to low node support within the American outbreak.

[3](Pettersson, Eldholm, et al. 2018)

In the upper nodes, the analysis shows Thailand playing a pivotal role in the source of the epidemic in Asia (see the histogram in Figure 2). Node *d* has a posterior probability (PP) of 0.67 of being of Thailand origin, which is estimated to be in 1999/06 (HPD 95% 2003/09 - 1994/03). Ancestral inference places the nodes *d*, *e*, *f*, *g* and *j* in Thailand with high PP (d:0.67; e:0.88; f:0.88; g:0.77; j:0.68). These data are supported in observational studies that address a long-circulation of ZIKV in Thailand (Buathong et al. 2015; Ruchusatsawat et al. 2019). The earliest Thailand sequence was collected in 2006/10/28 (MG645981), and there is serological evidence that Zika was circulating in Thailand before 1954 [71,72]. These facts are indicative of the presence of ZIKV in South-East Asia long before its first detection, suggesting a long adaptation of ZIKV to the local environment occurred between 1954 and 2006.

In the last few decades, Thailand seems to be the source of infection in many Southeast Asian countries, like Singapore, Indonesia, India, and Cambodia. The evidence suggested that ZIKV probably was already circulated in the country 20 years ago, in 1999/06 (HPD 95% 2003/09 - 1994/03), node *d*. Furthermore, other cases of introduction (Japan, Europe, China) were reported to be connected with tourism but never rose into local outbreaks in the country where the virus was imported.

Node *j* highlights the split between a Thailand sample and the Singapore lineage; the split is estimated to have occurred around 2014/4 (HPD 95% 2015/02 - 2013/06), whereas the Singapore lineage started to diversify at 2014/8 (HPD 95% 2015/5 – 2014/1). Even though the ZIKV seems to start circulating in Singapore

earlier than 2015/5, the infection cluster was detected only in August 2016 (Maurer-Stroh et al. 2016; Z. J. M. Ho et al. 2017). In addition, two different lineages were detected during the Singapore outbreak: the main lineage (yellow triangle in Figure 2) and another sample related to the Thailand sequences; this evidence reveals two independent ZIKV introductions in Singapore contributed to the outbreak.

Patterson observed the same phenomenon between phylogeny employing the African outgroup or omitting it. The MRCA estimates for *k* and *i* nodes provided here are 2012/10 (HPD 95% 2013/07 - 2012/02) and 2012/9 (HPD 95% 2015/02 - 2013/06), respectively. These posteriors estimations presented here represent older estimates than others previously proposed dates suggested by Faria and Metsky, that date the introduction of ZIKV in the Americas between October 2013–April 2014. The differences among estimates are probably due to different taxon sampling, indeed, the analyses of Faria and Metsky analyses are focused on the American outbreak (Metsky et al. 2017; Nuno Rodrigues Faria et al. 2016). Our dataset can provide a better result for analyzing the event that led to the American and Singaporean outbreak since it is less skewed toward America's epidemic. Our dataset contains all the suitable sequences of ZIKV in East Asia, and we select a subsample of American sequences. Our dataset contains samples that break the long branches between node *b* and node *i* we include ten nodes in our phylogeny, whereas Pettersson et al. (2018) include only five nodes (Pettersson, Eldholm, et al. 2018). This approach increases the accuracy of branch-length estimates by reducing the node-density effect and the variance of the estimates (Bromham et al. 2018).

It is interesting to note that the outbreaks in Singapore and South America were reported long after the estimated introduction of ZIKV. This suggests a long circulation of ZIKV before the outbreak onset, supporting an ecological scenario in which the virus takes time to start circulating consistently in a new population. The time of an emerging epidemic is proportional to the virus's basic reproductive number (R0). Therefore, this delay between molecular estimation and detecting potential emerging outbreaks in time is extremely important for preventing further circulation of the virus, making prevention measures efficient before the infection is out of control, leading to an epidemic. Waiting until the number of infected is so high that the silent spreading of an infectious disease

becomes clear is too late. Prevention, vector control, and monitoring are the only way to face future epidemics and manage the emerging disease. Constantly testing the mosquito vector for target virus presence could help identify a new outbreak's potential onset. For instance, Dengue virus (DENV) surveillance in vectors is Brazil, the Philippines and other countries where DENV is endemic (De Figueiredo et al. 2010; I. C. Dos Reis et al. 2019; Balingit et al. 2020; Lau et al. 2015); however, the active DENV surveillance in vectors is carried out also in Spain where DENV is not endemic (Aranda et al. 2018). However, the frequency of infected mosquitoes could be really low and difficult to detect; hence, this practice has to be coupled with the other kind of intervention already suggested by WHO (World Health Organization). The trees commented in this section are provided in supplementary material as .tree.

5.5.7 Re-introduction of ZIKV in Africa

Re-introductions of ZIKV in Africa were detected in Angola and Cape Verde. In Angola, ZIKV was confirmed in 2016/12, but evidence of the previous circulation set a probable first case at 2016/9 (Hill et al. 2019). Our analysis points to 2015/2 (HPD 95% 2016/4 -2014/3) the divergence of all Angolan sequences (see Table 3), meaning that the ZIKV was circulating before this date. It is not possible to provide with precision the date of divergence of the Angolan samples from the most closely related American sample due to the low support at the nodes, but it appears to have occurred in mid-2014. The timing of the introduction of ZIKV in Cape Verde should fall in the timespan between the American-Cape Verde split and Cape Verde radiation. These two events are estimated to have occurred in 2014/4 (HPD 95% 2014/8-2013/08) and 2015/3 (HPC 95% 2015/8-2014/8), whereas the first reported case of the epidemic is dated in October 2015 (Oumar Faye et al. 2020). The results provided here shows that the Asian strain of ZIKV was re-introduced in Africa (Cape Verde and Angola) before the epidemic in Brazil was confirmed in May 2015. This suggests that the spread of ZIKV out of South America had started probably before the ZIKV epidemic was detected in Brazil. These results put emphasis on the delay in detecting wide-reaching ZIKV circulation. The delay in detecting ZIKV onset is a constant characteristic in all the outbreaks in recent decades.

5.5.8 Spondweni divergence

The origin and the divergence time of the deep nodes in Zika evolution have not been sufficiently investigated yet. This section tries to answer some still-open questions about how old the ZIKV circulating nowadays really is. SPOV, the sister group of ZIKV, can infect humans and its consequences are usually mild and less dangerous than Zika's.

Our results date the origin of ZIKV to ~800 (HPD 95% 294 B.C. – 1516 A.D.), pointing to the Middle Ages as the probable origin of ZIKV in Africa. The estimate for the Asian-African split of ZIKV (node *a*) is 1852 (HPD 95% 1916-1776). The ZIKV diversification date estimate considerably overlaps the posterior date of our previous analysis, meaning that adding the SPOV outgroup does not bias the analysis. The only estimates for the age of the node in question are reported by Pettersson et al. (2014) (Pettersson and Fiz-Palacios 2014); however, it is much older than the one reported here. This is probably due to the taxon sampling and the calibration employed in the analysis; indeed, the paper aimed to date the Flavivirus evolution, not the ZIKV origin.

The clock rate estimated in SPOV tree is 6.42 10^{-4} (m/s/y), whose value is in line with the clock rate of Flavivirus and +ssRNA viruses that are supposed to range between 10^{-3} and 10^{-4} (m/s/y) (S. Duffy, Shackelton, and Holmes 2008). Moreover, the clock estimated here is slower than the values calculated using only the outbreak sequences, roughly 10^{-3} (Fajardo et al. 2016; N. R. Faria et al. 2017) as expected because the clock depends on the timespan of the phylogeny (S. Y. W. Ho et al. 2011).

The divergence between ZIKV and SPOV occurred a long time before the ZIKV diversification. The ZIKV lineage and the SPOV lineage probably went through many evolutionary novelties in their lineage, and it can not be excluded that other viruses related to these two may still remain undetected but circulating via mosquitoes vector infection route. This partially explains the high ZIKV seroprevalence in countries where ZIKV is endemic (Musso 2015).

In conclusion, we want to stress a limitation in inferring the date of ZIKV-SPOV split. The taxon sampling in this phylogenetic analysis is greatly biased toward ZIKV, we only employed one SPOV sequence, the only one suitable for our analysis. This could affect the inference of the ZIKV-SPOV estimate by increasing

the variance of the clock estimates, indeed, the HPD 95% of this node is wide ~800 (HPD 95% 294 B.C. – 1516 A.D.). We cautiously suggest this timescale for the ZIKV origin, more sequences that could increase the node density throughout the branches leading to SPOV and ZIKV lineage.

5.6 Concluding remarks

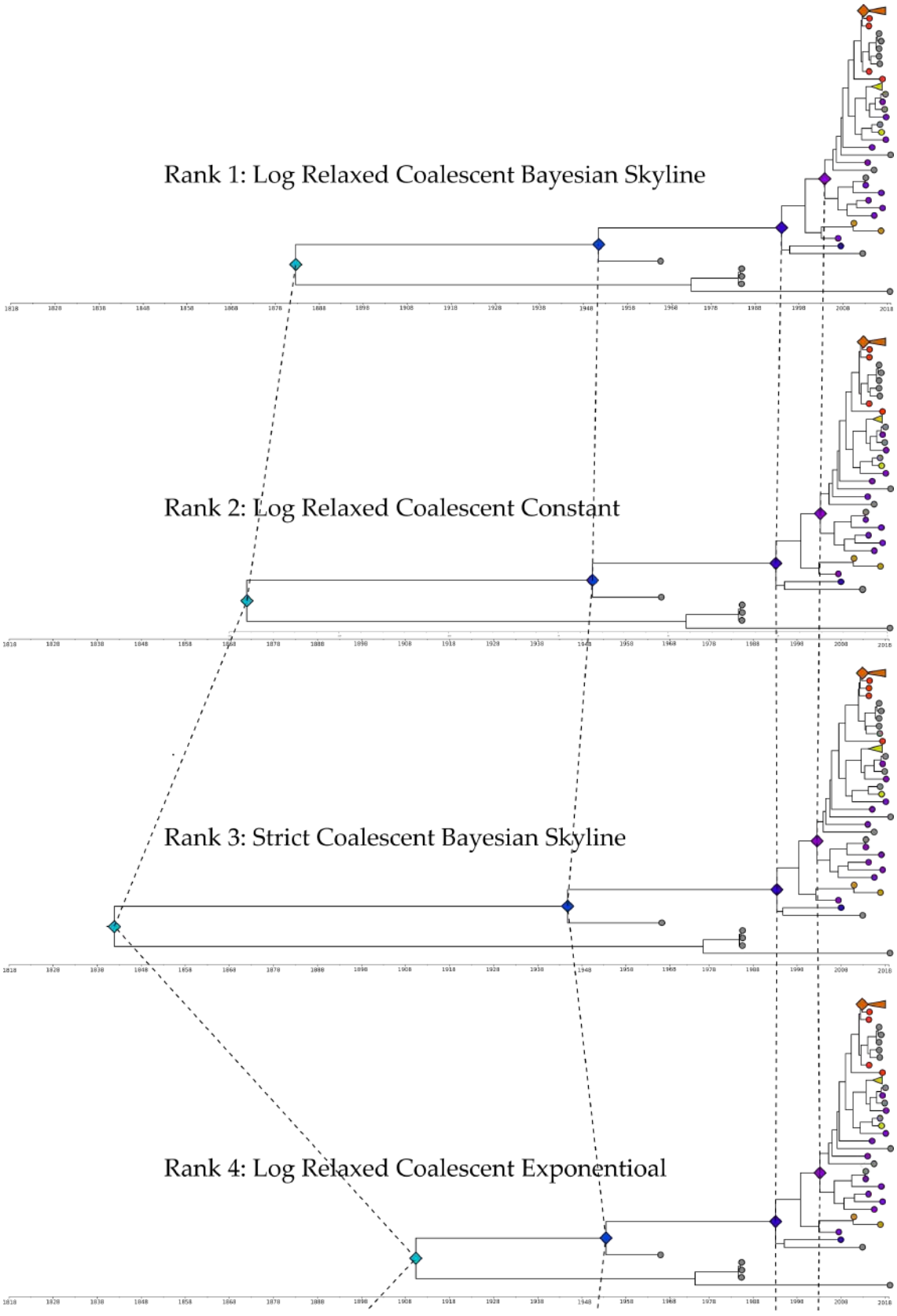
Interest in ZIKV has come and gone, leaving behind unanswered questions that we partially tried to address here. Our phylogenies provide useful information about the origin and the early diversification of ZIKV. The most important topic discussed in this work is the evident delay in detecting ZIKV before the outbreak onset: ZIKV was circulating within communities for at least one year before outbreaks become evident in Brazil, Angola and Cape Verde. This unmanaged viral spread has led to sequential outbreaks (e.g., the spread of ZIKV in Micronesia, French Polynesia, the Americas, Angola and Cape Verde). The ancestral state reconstruction analysis clarifies the role of Thailand in sustaining the ZIKV circulation in South-East Asia and its sequential outbreaks in Asia and the Americas.

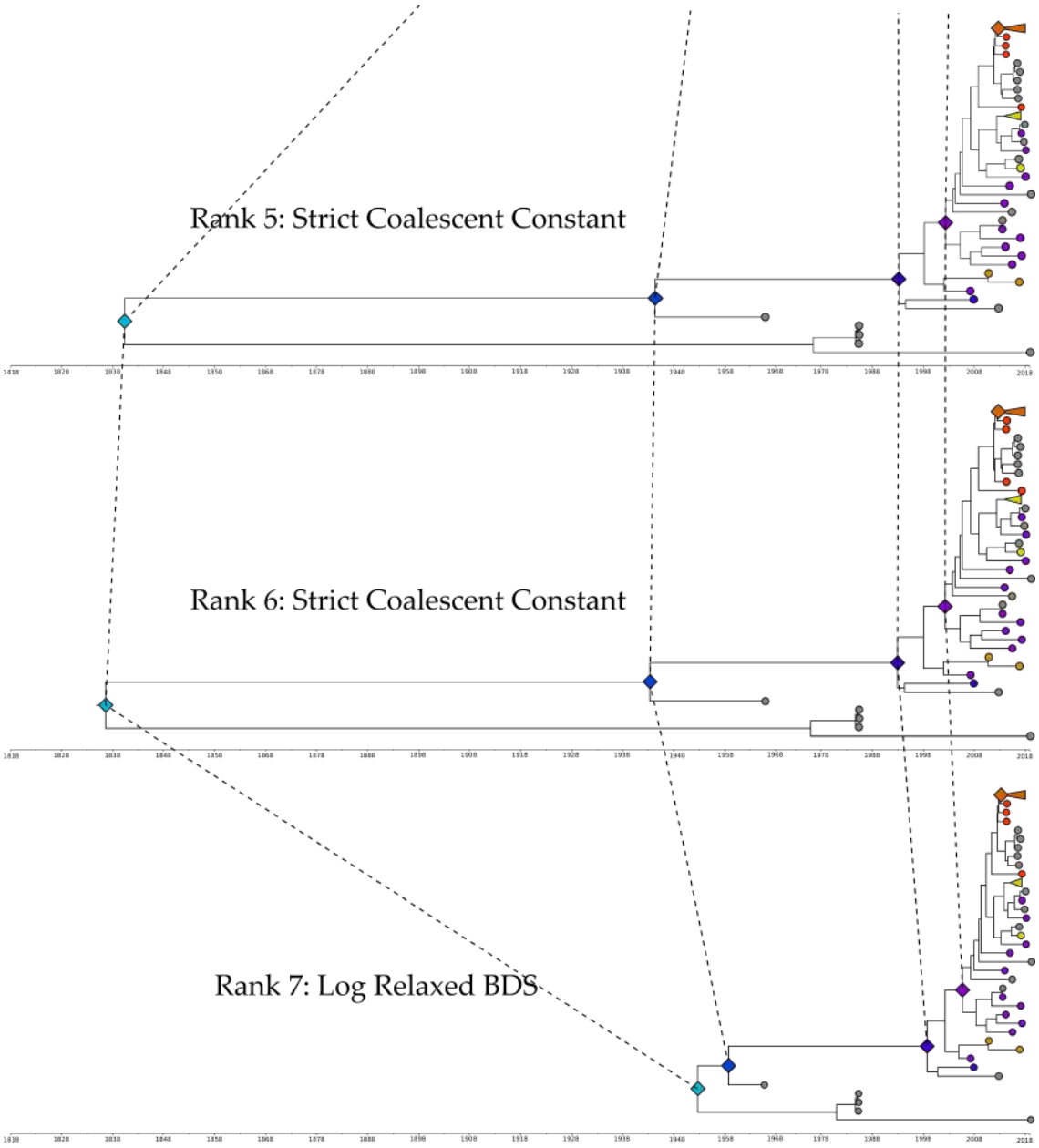
Studying ZIKV evolution has to be paired with its vector spread and distribution. ZIKV is transmitted by *Aedes aegypti* and, less efficiently by *Aedes albopictus;* these two major vectors are spreading all over the globe, bearing their infection potential with them (Kraemer et al. 2015; 2019). Climate change and rising temperature in temperate regions can favour emerging diseases. The efficiency of virus transmission for these two species has been shown to correlate with the environmental temperature (Reinhold, Lazzari, and Lahondère 2018; Chouin-Carneiro et al. 2020). Circulation of ZIKV can lead to further outbreaks if not controlled and monitored constantly, even in areas where it is not endemic yet. In recent years, Europe has faced several local infections of DENV and Chikungunya viruses (Lazzarini et al. 2020; Grandadam et al. 2011). These events were possible due to the presence of vectors, increased connectivity and globalization, and they may have been favoured by climate change. DENV distribution is, for example, not only correlated with the vector presence but also with climate factors as rainfall, temperature and humidity (Brady et al. 2014; Francis Schaffner and Mathis 2014). Current evidence suggests that climate

change have partially guided recent outbreaks of several arboviruses (Jolyon M Medlock and Leach 2015; E. A. Gould and Higgs 2009; Whitehorn and Yacoub 2019). Reducing the spread of invasive vectors worldwide is essential to prevent new viral threats. Our results on the early evolution of Zika reinforce the idea that increasing anthropization and natural niche disruption due to human activities, together with a globalized society favouring mobility of people among different countries, are favouring the emergence of novel arbovirus threats (E. Gould et al. 2017; Kraemer et al. 2015).

5.7 Supplementary material

5.7.1 Supplementary Figure





Supplementary Figure 1. Here we show graphically how the time tree changes across the different models, as in Table 1 the differences are more stressed in the deep nodes than in the upper. As detected in the model selection, the differences found between the tree rank 1 and the tree rank 2 are negligible; indeed, the trees look very similar in all the topology and in the posterior distribution of all the values.

5.7.2 Supplementary Table

Accession	Collection	Geo	Length	Publications	Release	Usable
HQ234498	1947	Uganda	10269	22389730	19/03/2012	no, too many passages
HQ234500	1968	Nigeria	10251	22389730	20/03/2012	no, too many passages
HQ234501	1984	Senegal	10269	22389730	21/03/2012	no, too many passages
KF268948	1976	Central African Republic	10788	25514122	22/03/2012	no, too many passages; not clear
KF270886	2007	Gabon	841	24516683	23/03/2012	yes
KF270887	2007	Gabon	772	24516683	24/03/2012	yes
KF383015	2001	Senegal	753	24421913	25/03/2012	no, too short
KF383016	2001	Senegal	753	24421913	26/03/2012	no, too short
KF383017	2001	Senegal	753	24421913	27/03/2012	no, too short
KF383018	2000	Senegal	753	24421913	28/03/2012	no, too short
KF383019	1998	Senegal	753	24421913	29/03/2012	no, too short
KF383020	1980	Cote d'Ivoire	753	24421913	30/03/2012	no, too short
KF383021	1998	Senegal	753	24421913	31/03/2012	no, too short
KF383022	1997	Senegal	753	24421913	01/04/2012	no, too short
KF383023	1997	Senegal	753	24421913	02/04/2012	no, too short
KF383024	1997	Senegal	753	24421913	03/04/2012	no, too short
KF383025	1997	Senegal	753	24421913	04/04/2012	no, too short
KF383026	1997	Senegal	753	24421913	05/04/2012	no, too short
KF383027	1997	Senegal	753	24421913	06/04/2012	no, too short
KF383028	2002	Senegal	753	24421913	07/04/2012	no, too short
KF383029	2002	Senegal	753	24421913	08/04/2012	no, too short
KF383030	1981	Burkina Faso	753	24421913	09/04/2012	no, too short
KF383031	1969	Senegal	753	24421913	10/04/2012	no, too short
KF383032	1979	Senegal	753	24421913	11/04/2012	no, too short
KF383033	1979	Senegal	753	24421913	12/04/2012	no, too short
KF383034	1979	Senegal	753	24421913	13/04/2012	no, too short
KF383035	1963	Uganda	735	24421913	14/04/2012	no, too short
KF383036	1999	Cote d'Ivoire	753	24421913	15/04/2012	no, too short
KF383037	1996	Cote d'Ivoire	753	24421913	16/04/2012	no, too short
KF383038	1999	Cote d'Ivoire	753	24421913	17/04/2012	no, too short
KF383039	1991	Senegal	753	24421913	18/04/2012	no, too short
KF383040	1990	Cote d'Ivoire	753	24421913	19/04/2012	no, too short
KF383041	1990	Cote d'Ivoire	753	24421913	20/04/2012	no, too short
KF383042	1990	Cote d'Ivoire	753	24421913	21/04/2012	no, too short
KF383043	1990	Cote d'Ivoire	753	24421913	22/04/2012	no, too short
KF383044	1990	Cote d'Ivoire	753	24421913	23/04/2012	no, too short
KF383045	1990	Cote d'Ivoire	753	24421913	24/04/2012	no, too short
KF383046	1999	Cote d'Ivoire	753	24421913	25/04/2012	no, too short
KF383084	1991	Senegal	708	24421913	26/04/2012	no, too short
KF383085	1969	Senegal	708	24421913	27/04/2012	no, too short
KF383086	1999	Cote d'Ivoire	708	24421913	28/04/2012	no, too short
KF383087	1979	Senegal	708	24421913	29/04/2012	no, too short
KF383088	1979	Senegal	708	24421913	30/04/2012	no, too short

KF383089	2002	Senegal	708	24421913	01/05/2012	no, too short
KF383090	2002	Senegal	708	24421913	02/05/2012	no, too short
KF383091	2001	Senegal	708	24421913	03/05/2012	no, too short
KF383092	2001	Senegal	708	24421913	04/05/2012	no, too short
KF383093	2001	Senegal	708	24421913	05/05/2012	no, too short
KF383094	2000	Senegal	708	24421913	06/05/2012	no, too short
KF383095	1998	Senegal	708	24421913	07/05/2012	no, too short
KF383096	1998	Senegal	708	24421913	08/05/2012	no, too short
KF383097	1997	Senegal	708	24421913	09/05/2012	no, too short
KF383098	1997	Senegal	708	24421913	10/05/2012	no, too short
KF383099	1997	Senegal	708	24421913	11/05/2012	no, too short
KF383100	1997	Senegal	708	24421913	12/05/2012	no, too short
KF383101	1997	Senegal	708	24421913	13/05/2012	no, too short
KF383102	1997	Senegal	708	24421913	14/05/2012	no, too short
KF383103	1999	Cote d'Ivoire	708	24421913	15/05/2012	no, too short
KF383104	1999	Cote d'Ivoire	708	24421913	16/05/2012	no, too short
KF383105	1996	Cote d'Ivoire	708	24421913	17/05/2012	no, too short
KF383106	1990	Cote d'Ivoire	708	24421913	18/05/2012	no, too short
KF383107	1990	Cote d'Ivoire	708	24421913	19/05/2012	no, too short
KF383108	1990	Cote d'Ivoire	708	24421913	20/05/2012	no, too short
KF383109	1981	Burkina Faso	708	24421913	21/05/2012	no, too short
KF383110	1990	Cote d'Ivoire	708	24421913	22/05/2012	no, too short
KF383111	1990	Cote d'Ivoire	708	24421913	23/05/2012	no, too short
KF383112	1990	Cote d'Ivoire	708	24421913	24/05/2012	no, too short
KF383113	1980	Cote d'Ivoire	708	24421913	25/05/2012	no, too short
KF383114	1979	Senegal	708	24421913	26/05/2012	no, too short
KU720415	1947	Uganda	10766		27/05/2012	no, too many passages
KU955591	1984-11-20	Senegal	10806	27174284	28/05/2012	yes,3 passages,
KU955592	1984-12-06	Senegal	10806	27174284	29/05/2012	yes,3 passages,
KU955594	1947-04	Uganda	10795	27174284	30/05/2012	no, too many passages
KU955595	1984-12-14	Senegal	10806	27174284	31/05/2012	yes,3 passage,
KX377335	1947-04	Uganda	10807		01/06/2012	no, too many passages
KX421193	1947	Uganda	10269	27443522	02/06/2012	no, too many passages
KX601166	1984-11-17	Senegal: Kedougou	10771		03/06/2012	Do not know
KX601169	1947-04-20	Uganda: Entebbe	10648		04/06/2012	no, too many passages
KY288905	1962-11	Uganda	10752		05/06/2012	unknow history, NCPV:1308258v
KY576904	1989	Central African Republic: Bangui	1358		06/06/2012	no, unknown cell passages
KY989511	1947	Uganda	10807		07/06/2012	no, too many passages
MF510857	1984-06-12	Senegal	10802		08/06/2012	yes
MF629796	2011-04-05	Nigeria	1482	28398562	09/06/2012	yes
MF629797	2013-07-05	Nigeria	1427	28398562	10/06/2012	yes
MF629798	2013-07-10	Senegal	1512	28398562	11/06/2012	yes
MF629799	2000-11-11	Senegal	1390	28398562	12/06/2012	yes
MF926508	2016-10-13	Nigeria	841	29885620	13/06/2012	yes
MK028860	1984	Senegal: Kedougou	10771		14/06/2012	no

MK241415	2015-12-03	Cape Verde: Santiago	10617		15/06/2012	yes
MK241416	2015-11-27	Cape Verde: Fogo	10617		16/06/2012	yes
MK241417	2016-06-04	Cape Verde: Fogo	10617		17/06/2012	yes
MK829152	2017-05-25	Angola: Bengo, Caxito	10164	31559967	18/06/2012	yes
MK829153	2017-06	Angola: Luanda, Bairro Prenda	10007	31559967	19/06/2012	yes
MN025403	2018-08	Guinea: Faranah	10710		20/06/2012	yes

CHAPTER 6 - MOLECULAR CLOCK APPLICATION AND OTHER COLLABORATIONS

6.1 Molecular clock application and other collaborations

In this Chapter, I present other applications of the molecular clock on genetic and genomic data. Moreover, I describe my participation in a phylogenomic study on Diptera Opsin genes. These collaborations were carried out during my doctoral training. In the first collaboration (**Paragraph 6.1**), I analyzed the opsin evolution of the Diptera clade, focusing on the mosquito clade. This involved manual curation of data and alignment obtained from an automated pipeline that extracts opsin genes from whole-genome assembly. I checked for misaligned genes, verified the opsin dataset and curated the phylogenetic reconstruction of opsin evolution, targeting homologous genes and their clade-specific duplication. In **Paragraph 6.2,** the second co-authored paper analyzes the evolution of *Arundo* species and their place in the Graminacae phylogeny. In this work, I calculated the divergence time within the *Arundo* genus using a transcriptomic dataset. I investigated the BOP-PACMAD (**B**ambusoideae **O**ryzoideaeclade **P**ooideae – **P**anicoideae **A**ristidoideae **C**hloridoideaetime **M**icrairoideae **A**rundinoideae **D**anthonioideae) divergence, focusing on the *Arundo* genus employing a transcriptomic dataset, a plastidial dataset and fossil data. In the article in **Paragraph 6.3**, I collaborate in reconstructing the phylogeny of *Bactrocera* species. I used a phylogenomic pipeline to investigate the speciation of this genus and the incongruence among different tree topologies.

Every paper presented in this Chapter is contextualized in the light of the PhD training and is consistent with the outline discussed in the introduction. Each paper is briefly described, and I outline my contribution. Here, I present only the abstract and the part of methods and results where my work was involved.

6.2 Phylogenomics of opsin genes in Diptera reveals lineage-specific events and contrasting evolutionary dynamics in *Anopheles* and *Drosophila*

Diptera is an insect order including flies, mosquitoes, and various other species of economic importance. Their vision is mediated by the opsin genes, which have been studied in a few key model species. However, a comprehensive comparative genomic analysis does not exist, impairing our understanding of the evolutionary history of these genes in this order. In this work, we perform the first genome-scale analysis of opsin gene evolution in Diptera. We investigate their pattern of duplication, selection, and expression in more than 60 species that belong to 10 different families. Our results clarify the evolution of the opsin genes in dipterans, in particular in fruit flies and mosquitoes, and represent the foundation for functional studies on their visual system. Mosquitoes' RH6 opsin gene has undergone many duplications; since this gene is involved in colour vision, this event probably has relevance in mosquito ecology, for example, in the feeding behaviour or crepuscular activity of many species of this clade, but this needs further investigation. Manual curation of a dataset, if possible, removes systematic errors encountered in automated pipeline dataset reconstruction and alignment. Fragmented genomes are sometimes unreliable, and they require particular care in limiting systematic errors.

Contribution: In this study, I mainly contributed to the manual curation of the opsin datasets, identifying the misaligned sequence and retrieving the missing genes using blast. The gene prediction pipeline can fail to retrieve genes in low-quality assemblies; on the other hand, it can overestimate the opsin number. Manually curating the datasets allowed us to reduce the systematic error which affects the automated pipeline in fragmented genomes. I eventually curated the figure that provides information on the duplication pattern in the Diptera phylogeny.

The original article abstract and the section related to my contribution to this article are reported below. Full article link:

https://academic.oup.com/gbe/article/13/8/evab170/6322995?login=true

Phylogenomics of Opsin Genes in Diptera Reveals Lineage-Specific Events and Contrasting Evolutionary Dynamics in *Anopheles* and *Drosophila*

Roberto Feuda[1,2], Matthew Goulty[1], Nicola Zadra[3,4], Tiziana Gasparetti, Ezio Rosato[1], Davide Pisani[5], Annapaola Rizzoli[3], Nicola Segata[4], Lino Ometto[6], and Omar Rota Stabelli[3,7]

[1]Department of Genetics and Genome Biology, University of Leicester, UK, [2]Department of Biology and Evolution of Marine Organisms, Stazione Zoologica Anton Dohrn, Naples, Italy, [3] Research and Innovation Centre, Fondazione Edmund Mach (FEM), San Michele all'Adige, Italy, [4]Department CIBIO, University of Trento, Italy; [5] School of Earth Sciences, University of Bristol, UK, [6]Department of Biology and Biotechnology, University of Pavia, Italy, [7] Center Agriculture Food Environment (C3A), University of Trento, Italy

6.2.1 Abstract

Diptera is one of the biggest insect orders and displays a large diversity of visual adaptations. Similarly to other animals, the dipteran visual process is mediated by opsin genes. Although the diversity and function of these genes are well-studied in key model species, a comprehensive comparative genomic study across the dipteran phylogeny is missing. Here we mine the genomes of 61 dipteran species, reconstructed the evolutionary affinities of 528 opsin genes, and determined the selective pressure acting in different species. We found that opsins underwent several lineage-specific events, including an independent expansion of Long Wave Sensitive opsins in flies and mosquitoes and numerous family-specific duplications and losses. Both the Drosophila and the *Anopheles* complement are derived in comparison with the ancestral dipteran state. Molecular evolutionary studies suggest that gene turnover rate, overall mutation rate, and site-specific selective pressure are higher in *Anopheles* than in *Drosophila*. Overall, our findings indicate an extremely variable pattern of opsin evolution in dipterans, showcasing how two similarly aged radiations, *Anopheles* and Drosophila, are characterized by contrasting dynamics in the evolution of this gene family. These results provide a foundation for future studies on the dipteran visual system.

Key words: Diptera, evolution, opsin, flies, mosquitoes

6.2.2 Manual curation

The dataset obtained was eventually manually curated. For example, we first checked for missing data. We selected sequences that lacked part of the opsin protein, and, where possible, we retrieved the missing data using BLAST (tblastn) on the assembled genomes. Second, we looked for putative false duplications in the tree, and in the case where we found a species-specific duplication in our subsequent analyses, we removed the incomplete sequence. Moreover, we looked for unexpected opsin losses to assess whether the missing genes were true losses or artefacts (false negatives). In some cases, we found the missing gene in the genome of interest and the sequence was added manually to the alignment.

For some mosquito species, we lacked well-assembled genomes and, therefore, accurate gene models, which may have caused misrepresentation in the exact number of Rh6 copies in each *Anopheles* lineage and blurred the fine-scale duplications/losses pattern. We, therefore, carefully and manually validated the Rh6 genes in the *Anopheles* species. Using such an approach allowed us to increase the length of many orthologs, most importantly, allowing us to detect instances of false positives: cases where putative duplicated contigs or allele variants from heterozygote genomes could be mistaken for species-specific duplications.

We further manually inspected for possible pseudogenes. For the Drosophila and *Anopheles* species, we manually curated all the alignments in order to perform dN/dS studies (see below) to exclude pseudogenes because we could not find signature of pseudogenes (dN/dS = 1), nor could we detect internal stop codons. For all other species, we inspected the alignment by eye when the gene was characterized by extremely long branches.

6.2.3 Dipteran Opsins Have Undergone Lineage-Specific Diversification

To better understand the opsin distribution and evolutionary dynamics in the various dipteran groups, we mapped their presence/absence on the Diptera phylogeny (fig. 1A) and performed a manual as well as a statistical gene-tree/species-tree reconciliation (fig. 1B and 1C). The results indicate that the opsin repertoire underwent significant rearrangements on the dipteran phylogeny in a

lineage-specific manner. In Brachycera (the clade comprising Drosophila), the opsin complement is derived in comparison to the ancestral dipteran condition. We confirm previous findings that c-opsin and RGR/Go have been lost in all Brachycera (Feuda et al. 2016) and provide evidence that four paralogs—*Rh1*, *Rh2*, *Rh3*, and *Rh4*—are present only in this group. The observation that at least one duplication from the ancestral *Rh1/2/6* and *Rh3/4* genes is shared between *Drosophila*, tephritid fruit flies, Muscidae house flies, and *Glossina* tze-tze flies indicates that these duplications happened early in Brachycera evolution (fig. 1B). We further observe various lineage-specific events, such as the loss of *Rh4* in the common ancestor of Glossinidae, Muscidae, and Calliphoridae, duplications of *Rh1* in Muscidae, the loss of *Rh2* in the tze-tze fly *Glossina morsitans* (Attardo et al. 2019), and the loss of all opsins except for *Rh2* and *Rh6* in Diopside stalk-eyed flies. Interestingly, when we map introns' presence/absence in the different opsins, the results indicate that *Rh3* genes in all *Drosophila* species are intronless, suggesting their possible origin as retrotransposons (Booth and Holland 2004; P. Xu et al. 2016).

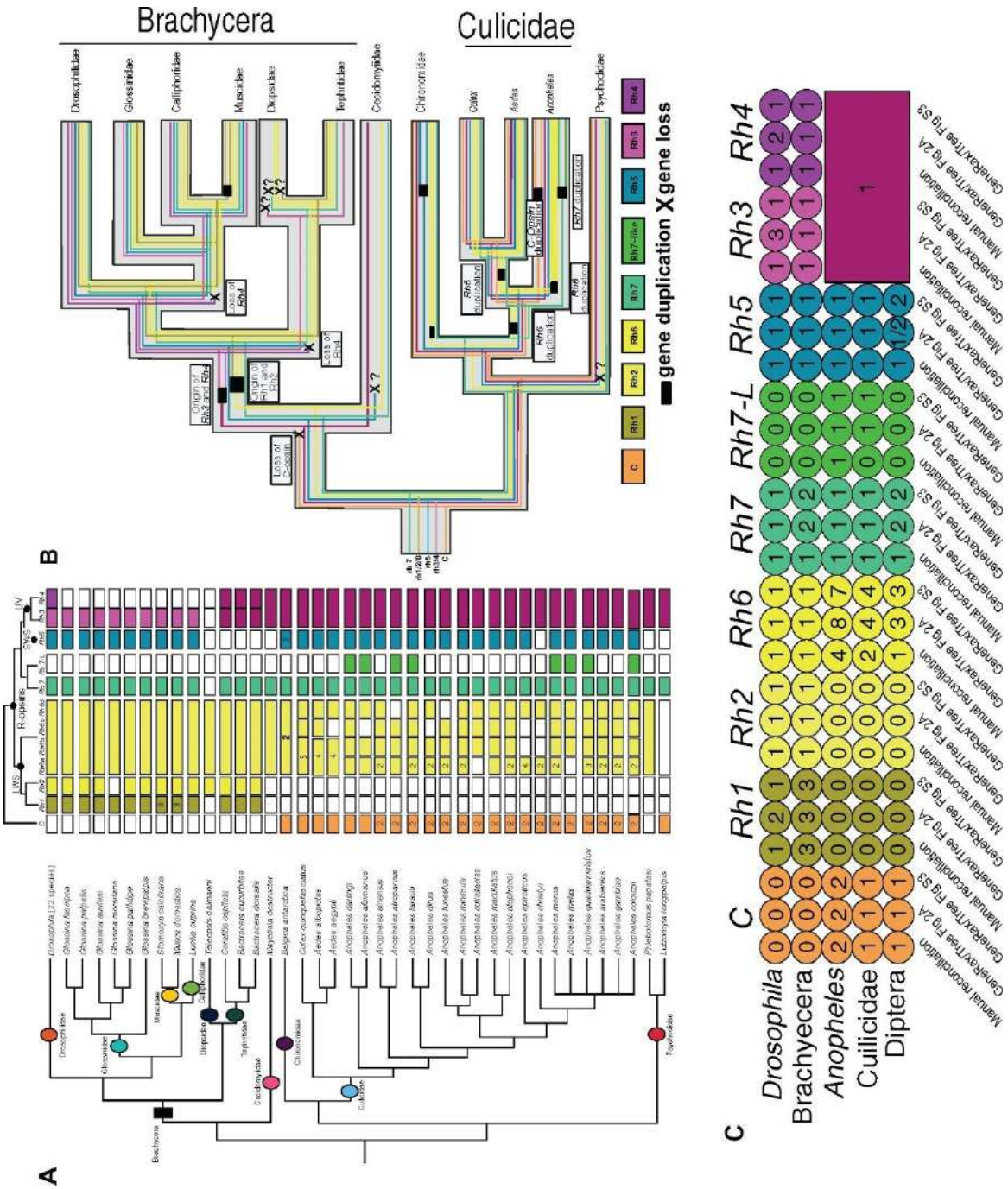


Figure 1 Opsins evolution in Diptera. (*A*) Opsin gene complements in Diptera. The phylogenetic tree was obtained from (Wiegmann et al. 2011). Gene nomenclature has been obtained from *Drosophila melanogaster*. The numbers in the boxes indicate the copies of opsin genes identified; white boxes indicate that genes have not been found. (*B*) Synopsis of the patterns of opsin duplications and losses in Diptera subgroups. Lineage-specific events are marked with a question mark if they were inferred from one single representative genome. (*C*) Estimated number of ancestral Rh across five nodes. For each opsin paralog, we report the estimate using three different analytical procedures (manual reconciliation, GeneRax on tree of fig. 1A.

6.3 Phylogenomic proof of Recurrent Demipolyploidization and Evolutionary Stalling of the "Triploid Bridge" in *Arundo* (Poaceae)

I co-authored this article that has now been published in the International Journal of Molecular Sciences (MDPI). The article explores the polyploidization event within the *Arundo* genus. This genus comprises eight putative species and is known for its large perennial species with ornamental and economic value. A better understanding of the evolution of the genus *Arundo* is thus relevant to understanding how the very high productivity of some of the *Arundo* species originated. The analysis is based on a robust phylogenomic and statistical framework that allows reconstructing the evolutionary history of a key trait of these reeds. The paper contains many interesting findings, such as hybridization events in the origin of *Arundo micrantha* and the dating of polyploydization events. Moreover, we re-visited the phylogeny, identifying *Arundo formosana* as basal in this clade. Moreover, we determine a more recent *Arundo* divergence than previous work based on a mitogenomic approach. Understanding the evolution of this important genus of grasses lays the foundations for future comparative genomics studies.

Contribution: Here, I applied my expertise on molecular clock dating methods to provide the evolutionary timespan for the *Arundo* genus using a genomic/transcriptomic dataset and the BOP-PACMAD clade employing a plastidial dataset. To calibrate the *Arundo* genus, we employed a rate on the four-fold site, whereas fossil calibration was used to calibrate the analysis on the plastidial dataset. Afterwards, I curated the model selection on the nuclear dataset, relying on the Stepping Stone (SS) method. The clock result indicates a probable time window for the demipolyploidization event in the *Arundo* evolution original article. In addition, I curated all the phylogenetic figures and the data interpretation for discussing the key event of the *Arundo* evolution. In conclusion, we proposed a rate for the genetic dataset calculated with a weighted mean among all the partitions.

Full article link:

https://www.mdpi.com/1422-0067/21/15/5247/htm

Phylogenomic proof of Recurrent Demipolyploidization and Evolutionary Stalling of the "Triploid Bridge" in *Arundo* (Poaceae)

Wuhe Jike[1,2], Mingai Li [1], Nicola Zadra[3], Enrico Barbaro[1], Gaurav Sablok[4], Giorgio Bertorelle[2], Omar Rota-Stabelli[3] and Claudio Varotto[1]

[1]Department of Biodiversity and Molecular Ecology, Research and Innovation Centre, Fondazione Edmund Mach, 38010 San Michele all'Adige (TN), Italy, [2]Dipartimento di Scienze della Vita e Biotecnologie, Università degli Studi di Ferrara, 44121 Ferrara, Italy, [3]Department of Sustainable Ecosystems & Bioresources, Research and Innovation Centre, Fondazione Edmund Mach, 38010 San Michele all'Adige (TN), Italy, [4]Department of Biosciences, University of Helsinki, 00014 Helsinki, Finland

6.3.1 Abstract

Polyploidization is a frequent phenomenon in plants, which entails the increase from one generation to the next by multiples of the haploid number of chromosomes. While tetraploidization is arguably the most common and stable outcome of polyploidization, over evolutionary time triploids often constitute only a transient phase, or a "triploid bridge", between diploid and tetraploid levels. In this study, we reconstructed, in a robust phylogenomic and statistical framework, the evolutionary history of polyploidization in *Arundo*, a small genus from the Poaceae family with promising biomass, bioenergy and phytoremediation species. Through the obtainment of 10 novel leaf transcriptomes for *Arundo* and outgroup species, our results prove that recurrent demiduplication has likely been a major driver of evolution in this species-poor genus. Molecular dating further demonstrates that the species originating by demiduplication stalled in the "triploid bridge" for evolutionary times in the order of millions of years without undergoing tetratploidization. Nevertheless, we found signatures of molecular evolution highlighting some of the processes that accompanied the genus radiation. Our results clarify the complex nature of *Arundo* evolution and are valuable for future gene functional validation as well as reverse and comparative genomics efforts in the *Arundo* genus and other Arundinoideae.

6.3.2 Dating methods

Divergence times were estimated using both the nuclear and the plastidial datasets described above using BEAST2 (R. Bouckaert et al. 2019). According to model selection (see below), for both datasets, an uncorrelated lognormal relaxed clock (not considering autocorrelations between adjacent branches (Drummond et al. 2006)) coupled with a Yule demographic prior was used. Substitutions were modeled using the GTR replacement model with four discrete categories of the gamma distribution. For all datasets and analyses, the MCMC was run for 200,000,000 generations, and checked for convergence using Tracer 1.7 (Rambaut et al. 2018) to ensure that the effective sample size (ESS) values were greater than 200 for every posterior and for the likelihoods. Consensus trees were obtained by TreeAnnotator 2.5.1 as maximum clade credibility (MCC) trees of all BEAST trees with a burnin of 20% and median heights of nodes. The nuclear tree was calibrated with the substitution rate previously estimated for Poaceae (Christin et al. 2014) using a normal distribution to cover between 6 × 10−3 and 7 × 10−3 (mean 6.5 × 10−3 ; SD 5 × 10−4) substitutions/site/million year. The root (*Arundo* stem) was further calibrated with a maximum of 40 MY (million years) that corresponds to the minimum for the origin of PACMAD clade (Christin et al. 2014). Because we used a mutation rate previously inferred from four-fold degenerate sites (Christin et al. 2014), in our analysis, we split our nucleotide alignment into two different partitions: one containing the four-fold degenerate sites, which were calibrated using the rate prior; the other containing all other nucleotides (Christin et al. 2014) which were left free to be estimated. To determine which clock and demographic prior fit better the nuclear dataset, the BEAST2 model comparison package (R. R. Bouckaert and Drummond 2017) was used employing different statistics: AICm (Akaike Information Criterion model), hMean (harmonic mean) and SS (Stepping Stone). The latter was needed because AICm, hMean can be affected by systematic error (Xie et al. 2011; Baele et al. 2012). Three calibration priors were used for the plastid dataset. Previous molecular dating studies have estimated the origin of Poales to be between 120–175 Ma (Christin et al. 2014): there is no fossil evidence that place the monocots in this age range (Eguchi and Tamura 2016), but we used 125 MY as a lower bound (maximum) for the BOP-PACMAD stem as in Burke (Saarela et al. 2018; Burke et al. 2016). The phytolites (fossil evidence) attributed to the Oryzoideae

subfamily (Prasad et al. 2011) were used to calibrate the Oryzoidae stem at 66 MY (upper bound or minimum). As a general prior for all branches the plastidial rate inferred by Christin and colleagues was used (Christin et al. 2014): 1× 10−3 with a lognormal distribution that ranges between 4.5 × 10−3 and 1.2 × 10−4 with a median of 6.0 × 10−4 substitutions/site/million year.

6.3.3 Dating Results

We estimated the divergence of *Arundo* species using both the nuclear and the plastid datasets. Model selection indicated that the nuclear dataset is best fitted by a combination of Yule demographic prior and a lognormally distributed relaxed clock (Table 1). The summary of the clock rates estimated from the nuclear dataset with different models and data partitions are provided in Table 2. The Bayesian consensus tree for the most fitting model in the nuclear dataset is reported in Figure 2, while in Table 3 we report the mean age for each of the nodes, together with the corresponding 95% High Posterior Densities (HPD, a type of confidence interval). According to Figure 3, all the posterior estimates for the age of nodes are characterized by a normal distribution which is indicative of good convergence of the analysis. The origin of the *Arundo* genus is set at 17.8 MY (mean estimate) with HPD ranging from circa 14 to 22 MY. The diversification of the extant *Arundo* started circa 7.9 MY (6.2–9.8 HPD), which corresponds to the split of *A. formosana*. The nuclear estimate for the diversification of extant *Arundo* is compatible with divergence estimates based on plastidial data (mean 9.6 MY, HPD 7.6–11.5 MY, Figure 2) particularly because there is a large overlap of posterior densities (Figure 3; Table 4). The lineage leading to *A. donax* diversified circa 5.7 Ma, and the split of the two varieties of *A. donax* sampled in this study is placed at 3.7 Ma, thus providing a conservative estimate of the time required for the generation of morphological differentiation within this clonal species. Divergence of *A. micrantha* is estimated at 4.7 Ma. Finally, the divergence of the *A. plinii* s.l. clade took place around 3.4 Ma, followed by the split of *A. donaciformis* from *A. plinii* around 2.5 Ma.

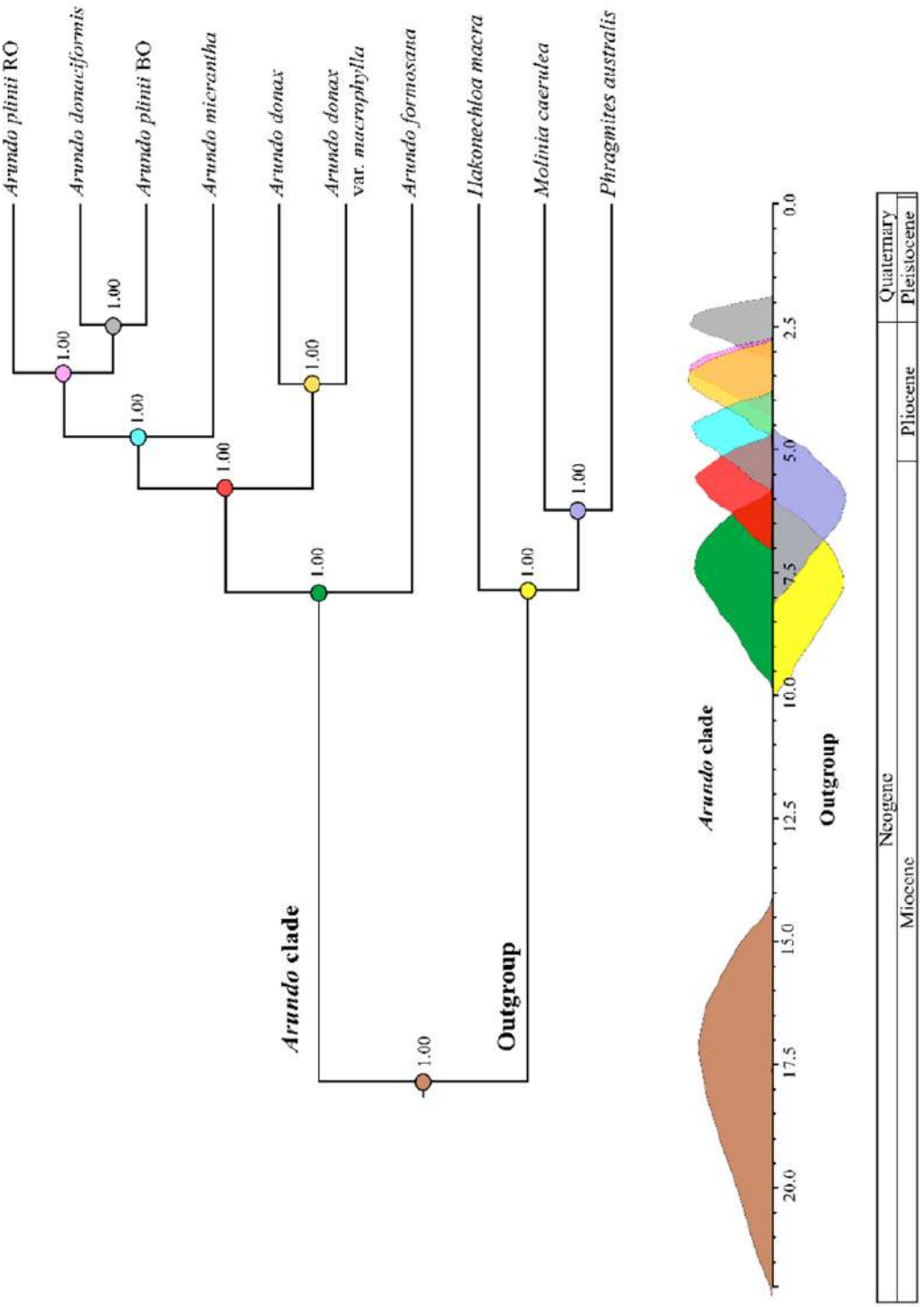


Figure 2. Chronogram of the *Arundo* genus based on the set of 144 nuclear genes. Numbers close to nodes indicate posterior support. The scale under the tree is in million years and the 95% CIs are drawn in the colour of the node they correspond to in the chronogram. The 95% CI for the *Arundo* ingroup is above the scale and under the scale for outgroups

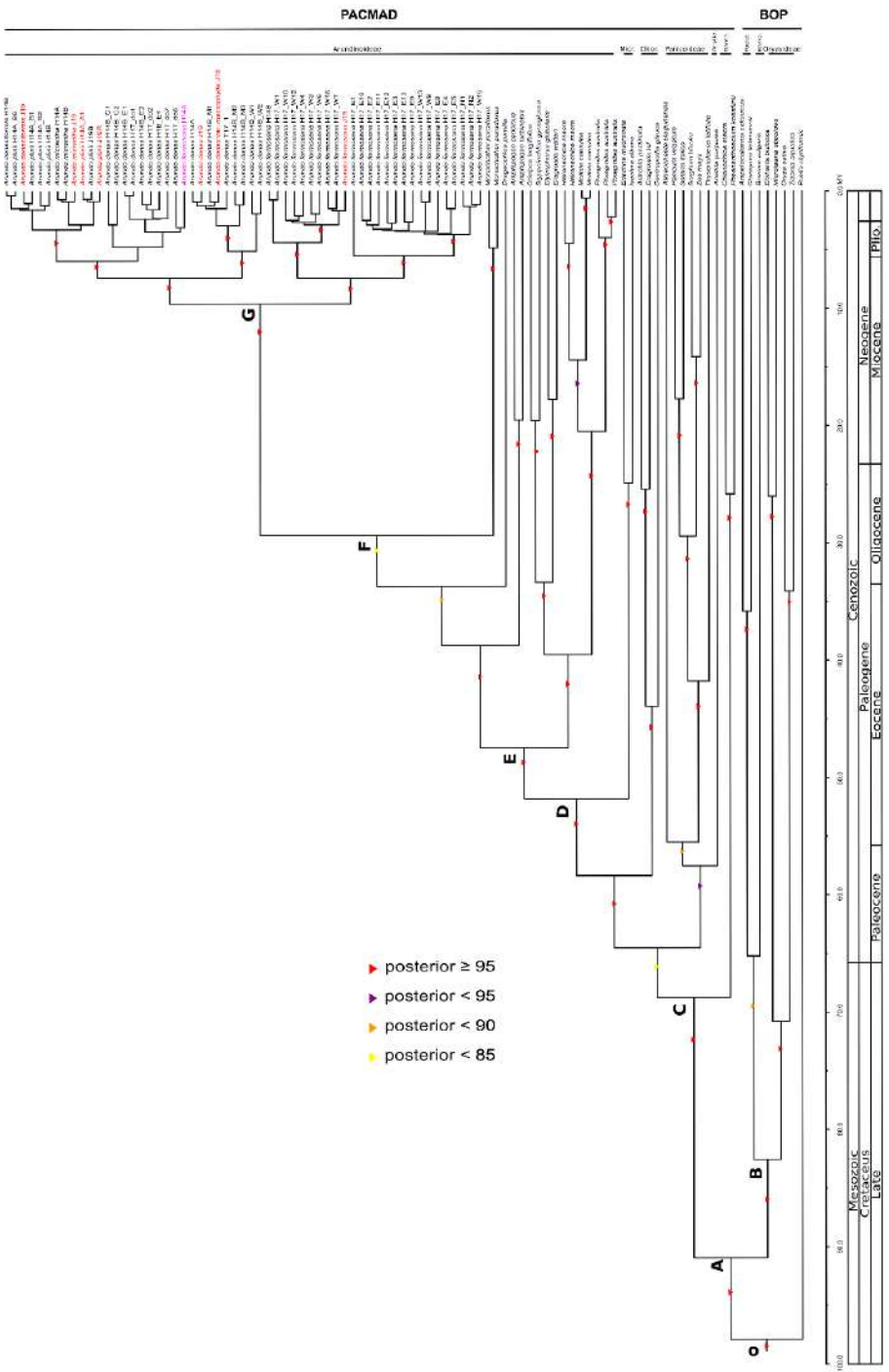


Figure 3. This figure was added in the Supplementary material in the original article. Here, we show the plastidial Phylogeny of the BOP-PACMAD clade.

Table 1. Summary of clock models and priors tested for molecular dating of the *Arundo* genus and selected outgroups divergence based on the nuclear dataset. Chain length 200,000,000. Numbers in bold indicate optimal (minimal) values. The best model is highlighted in red.

ID	Clock	Rate distribution	Demographic prior	Likelihood	Posterior	AICm	H mean	Stepping Stone
1	Relaxed	LogNormal	Yule	**-403604**	**-403651**	807262	-403623	-403820
2	Relaxed	Normal	Yule	**-403604**	**-403651**	**807261**	**-403621**	**-403778**
3	Relaxed	Normal	Birth and death	**-403604**	-403658	**807261**	-403623	-404302
4	Relaxed	Normal	Coal constant	-403636	-403714	807342	-403658	-403964
5	Strict	Normal	Yule	-403825	-403871	807693	-403839	-403958

Table 2. Summary of clock rates estimated from the nuclear dataset with different models and data partitions.

	Partition	Mean Rate[a] 10^{-3}	Rate[a] 95% 10^{-3}	Overall clock[a, b]	Overall variance 10^{-7}	Root age (95% CI)	*Arundo* split (95% CI)
1	Not 4 Fold	1.758	1.288 -2.227	2.453	2.7487	17.5 (13.3 - 22.5)	7.8 (5.8 – 9.8)
	4 Fold	6.621	5.401 -7.989				
2	Not 4 Fold	1.74	1.318 - 2.161	2.329	0.3926	17.8 (13.9 – 22.2)	7.9 (6.2 – 9.8)
	4 Fold	6.575	5.581 - 7.453				
3	Not 4 Fold	2.01	9.933 - 3.052	2.807	0.582	16.4 (8.9 – 26.4)	7.3 (3.9 – 11.7)
	4 Fold	7.601	3.913 - 11.2				
4	Not 4 Fold	220.9	158.6 - 287.5	19.053	799.827	39.4 (38.3 – 40)	8.5 (7.8 – 9.3)
	4 Fold	8.539	7.866 - 9.227				
5	Not 4 Fold	1.561	1.235 - 1.882	2.278	0.2498	18 (14.3 – 21.8)	7.8 (6.1 – 9.4)
	4 Fold	6.58	5.573 - 7.523				

[a] the clock rate is calculated in substitutions/site/million year.

[b] the overall clock is calculated using the weighted mean of the two partitions as follows: weighted clock $C = \sum_{i=1}^{n} \omega_i' c_i$, where c_i are the estimated mean rates for the 4-fold degenerate and non-degenerate partitions and $\omega_i = \frac{\text{No.of partition site}}{\text{Total No.of site}}$ are the respective normalized weights.

[c] the overall variance is calculated using the weighted variance of the two partitions as follows: weighted variance $\sigma_C^2 = \sum_{i=1}^{n} \omega_i'^2 \sigma_{c_i}^2$, where $\sigma_{c_i}^2$ are the estimated mean rates for the 4-fold degenerate and non-degenerate partitions and $\omega_i = \frac{\text{No.of partition codons}}{\text{Total No.of codons}}$ are the respective normalized weights.

Table 3. Summary of ages for *Arundo* clades. Mean age and 95% confidence intervals were estimated from the nuclear dataset with the best fitting model. Node color-coding refers to Fig. 4 in the main text.

Node	**Node name**	**Mean age (Million of years)**	**95 % C.I (Million of years)**
	Arundo origin	17.8	13.9 -22.2
	Arundo split (*A. formosana* split)	7.9	6.2 – 9.8
	A. donax split	5.7	4.5 _ 7.1
	A. micrantha split	4.7	3.8 – 5.9
	A. donax varieties split	3.6	2.6 -4.7
	A. plinii s. l. crown	3.4	2.6 -4.3
	A. donaciformis split	2.4	1.8 -3.1

Table 4: Summary of ages for BOP-PACMAD and Arundinoideae clades. Mean age and 95% confidence intervals were estimated from the chloroplast dataset with the best fitting model. Node IDs (letters) refer to Suppl. Fig. 1.

Node	**Node name**	**Mean age (Million of years)**	**95 % C.I (Million of years)**
0	BOP –PACMAD origin	97.9	76 -124
A	BOP-PACMAD split	90.4	73 - 113
B	BOP	82.5	69 -100
C	PACMAD	68.7	47 -90
D	Arundinoidea origin	51.8	34 -70
E	Arundinoidea split	47.4	30 -64
F	*Arundo* origin	29.3	14-43
G	*Arundo* split	9.7	7.5 - 11

6.4 The impact of fast radiation on the phylogeny of *Bactrocera* fruit flies

In this paper, we try to outline the evolution of a species-rich *Bactrocera* genus. The lack of prior information led us to make some strong assumptions to calibrate the tree, as we constrained the mutation rate of the four-fold degenerate sites, where we applied the instantaneous mutation rates calculated for *Drosophila*. However, the mutation rate between these two related genera is supposed to be similar on the neutral evolving site. Using StarBEAST we detected the species tree that helped us identify the suitable genes for the analysis, or rather we identified the genes that supported the species tree. This process of selecting suitable genes provides an unbiased dataset that supports a univocal topology. The results show a more recent timescale than the previous estimate, obtained using the mitogenomic dataset. An interesting result is a correlation between the speciation of Southeast Asian species and the Australian *Bactrocera* species that occurred during the rise in sea level in the Pleistocene. The sea level rising caused the creation of islands in the Sunda peninsula and a wider distancing between Australia and Asia, which could have pushed the allopatric speciation and fast radiation of the *Bactrocera* species.

Contribution: In this article, I contributed mainly to guiding and suggesting the best practice for analyzing the genomic dataset. The limited source of priors forces us to use a rate as a prior instead of fossils or other more reliable calibrations. I was involved in the experiment design and in part of the analysis, model selection and dating, as well as in the interpretation of the results. The paper is in preparation for BMC Ecology and Evolution.

Preprint article is available here:

https://www.biorxiv.org/content/10.1101/2021.09.07.459237v1.abstract

The impact of fast radiation on the phylogeny of *Bactrocera* fruit flies

Federica Valerio, Nicola Zadra, Omar Rota Stabelli, Lino Ometto

6.4.1 Abstract

True fruit flies (Tephritidae) include several species that cause extensive damage to agriculture worldwide. Among them, species of the genus *Bactrocera* are widely studied to understand the traits associated to their invasiveness and ecology. Comparative approaches based on a reliable phylogenetic framework are particularly effective, but, to date, molecular phylogenies of *Bactrocera* are still controversial. Here, we employed a comprehensive genomic dataset to infer a robust backbone phylogeny of eleven representative *Bactrocera* species and two outgroups. We further provide the first genome scaled inference of their divergence using calibrated relaxed clock. The results of our analyses support a closer relationship of *B. dorsalis* to *B. latifrons* than to *B. tryoni*, in contrast to all mitochondrial-based phylogenies. By comparing different evolutionary models, we show that this incongruence likely derives from the fast and recent radiation of these species that occurred around 2 million years ago, which may be associated with incomplete lineage sorting and possibly (ongoing) hybridization. These results can serve as basis for future comparative analyses and highlight the utility of using large datasets and efficient phylogenetic approaches to study the evolutionary history of species of economic importance.

6.4.2 Dating analysis methods

Because of the numerous incongruences between the species tree obtained by the multi-locus analyses and the single gene trees (see Results section), which could bias the dating analysis, we produced a conservative dataset by: i) limiting the species sample to 10 representative species (*C. capitata, Z. cucurbitae , B. bryoniae, B. dorsalis, B. jarvisi, B. latifrons, B. minax, B. musae, B. oleae, B. tryoni*) and ii) considering only those 37 genes that produced a ML tree supporting the consensus species tree with minimum ML bootstrap values of 50 at each node. Divergence times were then estimated by Beast v. 2.5.1 using the 4-fold

degenerate sites of the concatenated dataset (11,768 nt). This dataset allowed us to use an instantaneous (neutral) mutation rate as prior. Since the mutation rate in Tephritidae is not known yet, we assumed it to be similar to that of *Drosophila* (another Diptera) and used the estimate of 0.0346 (SD = 0.0028) substitutions per base pair per million years provided by (Obbard et al. 2012). Because in *Bactrocera* we assumed eight generations per year (in nature, they range from 3-5 of *B. oleae* and sub-tropical *B. dorsalis* populations, to >12 for the tropical species (X. Z. Li et al. 2019; Stephens, Kriticos, and Leriche 2007; Theron, Manrakhan, and Weldon 2017; Vargas et al. 1997)) and to account for uncertainty, we finally set as a prior a normally distributed mean of 0.028 (SD = 0.03). In a second approach, we set a mutation rate lognormal distributed with 'mean in real space' M = 0.028 and S = 0.82 (to produce the same 95% quantile – 0.077 – as the normal distribution). For both approaches, we performed a model selection to choose the most fitting clock and demographic prior based on the marginal likelihood values with the nested sampling approach implemented in the NS package (Russel et al. 2019). We tested the strict and the LOGN relaxed clock and the Yule and the Birth-Death models, for a total of eight different combinations. Following the recommendations provided by the dedicated Taming the Beast tutorial (Barido-Sottani et al. 2018), sub-chain length was set at 50,000, which corresponds to the length of the MCMC run (i.e., 5x107) divided by the smallest ESS value observed across the eight model runs (i.e., ~1,000), and the number of particles was set at 10. A model was considered favoured over another model if the difference between the two marginal likelihoods (i.e., the Bayes Factor (BF) in log space) was more than twice the sum of the corresponding standard deviations (SD). We ran eight different analyses that used different combinations of priors and model settings and performed a model selection to identify the most appropriate for our data. In particular, the nested sampling approach allowed us to estimate the marginal likelihoods of the different models and make pairwise comparisons using the associated Bayes Factor. All models had marginal likelihoods with a standard deviation ranging from 2.5 to 2.9, which was small enough to assess whether a model was favoured over another one. In all cases, we employed a GTR+G replacement model and a root prior uniformly distributed between 6 and 65 million years ago (Ma), which correspond to the age of a *Ceratitis* fly fossil (Norrbom 1994) and of the Schizophora radiation (Junqueira et al. 2016;

Wiegmann et al. 2011). Because the Bayesian phylogenetic analysis on the concatenated 4-fold degenerate sites resulted in a topology incongruent to the one supported by all other ML and Bayesian analyses (see Results), the species tree was fixed according to the latter consensus topology. All analyses were run twice, with chains run for 5x107 generations, sampling trees and parameters every 1,000 generations and inspecting convergence and likelihood plateauing in Tracer. Both chains resulted well mixed, with average effective sample size (ESS) values across posterior values being well above 200. The consensus trees (Maximum Clade Credibility trees) were generated after discarding the first 20% of generations as burn-in.

6.4.3 Dating analysis suggests fast and recent radiation in Bactrocera

Our dating analysis is based on the best combination of priors according to a model selection (Table 4) that indicated as favoured model the one where we set the mutation prior with a log-normal distribution, a strict clock, and a Birth-Death model. The Bayes Factor values, even after correcting for uncertainty by subtracting the corresponding standard deviations, are well above two, which provides overwhelming support for that model (Kass and Raftery 1995). The fact that a strict clock is favoured over a relaxed clock is consistent with the low mean value of the coefficient of variation parameter (i.e., the standard deviation of branch rates divided by the mean rate), which equals 0.24. Therefore, we will report the results obtained by this analysis. Incomplete lineage sorting is expected for rapid radiations e.g., (Pollard et al. 2006), which is exactly what it is revealed by our molecular clock analyses (Figure 4). Consistent with a rapid radiation of the (*B. dorsalis, B. latifrons, B. tryoni*) clade, the results of the clock analysis place its origin in the mid-Pliocene, at ~2.08 Ma, with a subsequent very close cladogenesis, at ~1.87 Ma, separating *B. dorsalis* and *B. latifrons* (Fig. 2). Interestingly, during this period sea rose at peak levels (Zhong et al. 2004) and thus increased distances between islands and island groups, possibly facilitating allopatric speciation (the three species have native ranges in south-east Asia and Australia). The proximity of the two cladogenetic events and the large overlap of their 95% confidence intervals agrees with a rapid radiation, which could have resulted in frequent incomplete lineage sorting. This would also explain the

discordant results between the nuclear and the mitochondrial phylogenies, a finding that is commonly reported in many organisms, including insects (Beltrán et al. 2002; DeSalle and Giddings 1986; Toews and Brelsford 2012). Moreover, we cannot exclude the possibility that these species experienced, or even still experience, hybridization events, which could then result in widespread introgression events. Indeed, hybrids have been reported for several closely related *Bactrocera* species (Augustinos et al. 2014; Bo et al. 2014; Cruickshank, Jessup, and Cruickshank 2001; Yeap et al. 2020; Pike, Wang, and Meats 2003), and although none of the published studies involved a pair of species analyzed in our analyses, possible introgression can occur via direct hybridization or via intermediate hybridization events involving other closely related species.

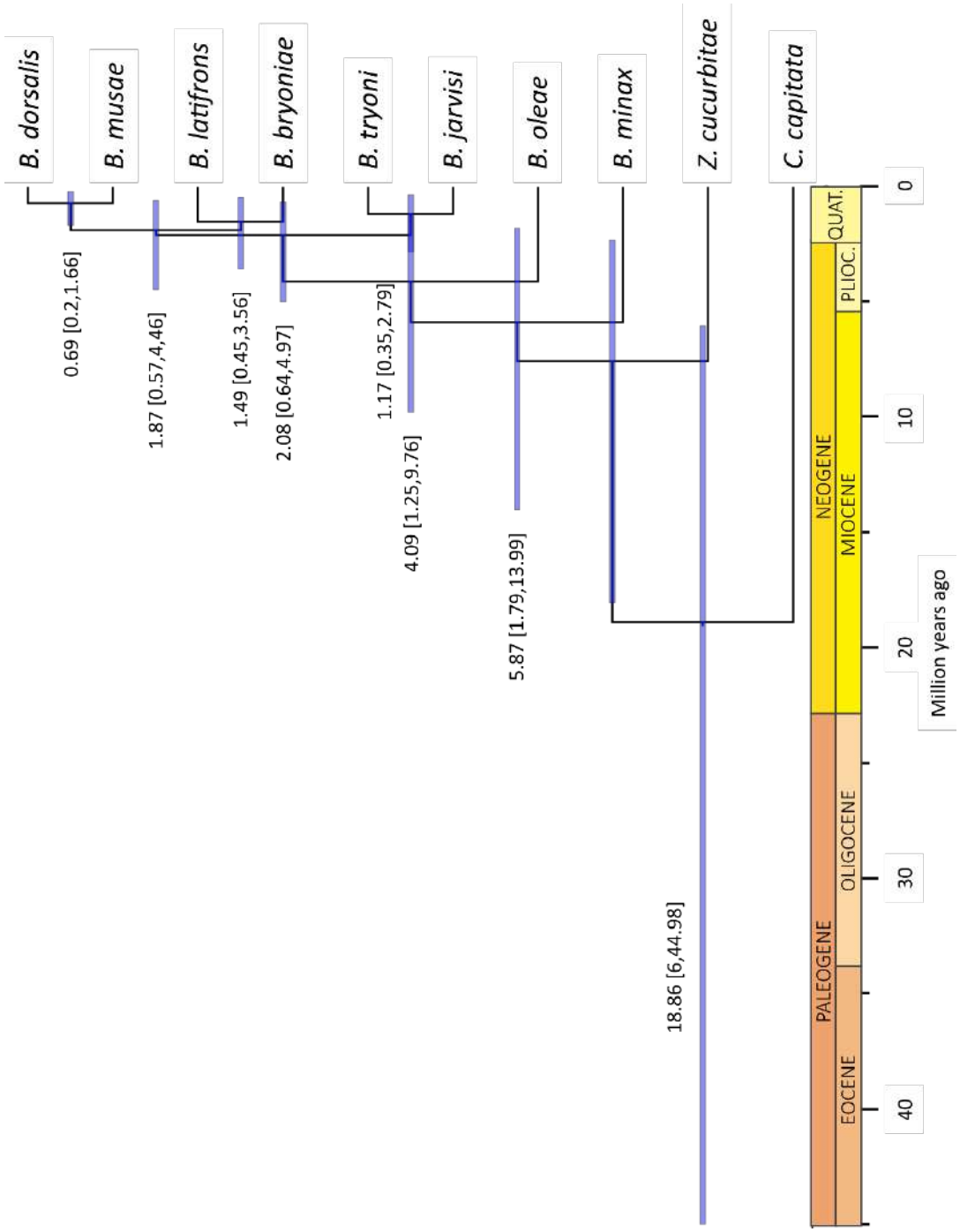


Fig 4. Molecular time tree of *Bactrocera* (using *C. capitata* as outgroup). *Bactrocera* (plus *Zeugodacus*) originated during the Miocene optimum (around 19 million years ago, Ma) and experienced recent fast cladogenetic events around 2 Ma. The analysis was done setting the mutation rate employing a log-normally distributed prior, a strict clock and a Birth-Death model. Mean and 95% highest posterior density of the inferred age (blue bars) are reported for each node.

Model 1		Model 2								
		Mut-normal				Mut-lognormal				
		yule_strict	*yule_rlxln*	*bd_strict*	*bd_rlxln*	*yule_strict*	*yule_rlxln*	*bd_strict*	*bd_rlxln*
Mut-normal	*yule_strict*	-	62.2	*	54.3	9.2	59.6	#	57.2
	yule_rlxln	#	-	#	#	#	#	#	#
	bd_strict	#	58.6	-	50.7	-	55.9	#	53.5
	bd_rlxln	#	7.9	#	-	#	5.2	#	*
Mut-lognormal	*yule_strict*	#	53.1	#	45.2	-	50.4	#	48
	yule_rlxln	#	*	#	#	#	-	#	#
	bd_strict	**6.1**	**68.3**	**9.8**	**60.4**	**15.3**	**65.7**	-	**63.3**
	bd_rlxln	#	*	#	#	#	*	#	-

Table 5. Results of the model selection of the different BEAST analyses used to estimate divergence times. Bayes Factors (BF) were estimated comparing eight BEAST models that combined a mutation prior with either a normal (*Mut-normal*) or lognormal (*Mut-lognormal*) distribution, either a Yule (*yule*) or a Birth-Death (*bd*) model, and either a *strict* or a LOGN relaxed (*rlxln*) clock. Models were compared in a pairwise fashion, by first estimating their Marginal Likelihoods (mL) and corresponding Standard Deviation (SD) and then calculating the Bayes Factors as $BF = mL_1\text{-}mL_2$, where model 1 and model 2 are those given in the respective row and column. We only report BF values that satisfied the conditions $BF > 0$ and $BF\text{-}(SD_1+SD_2) > 0$, where SD_1 and SD_2 are the SD values estimated for model 1 and 2. Highlighted on bold are the values for the model (*Mut-lognormal + bd + strict*) favored over all other seven models. # = BF < 0; * = $BF\text{-}(SD_1+SD_2) < 0$.

Conclusion

Reconstructing and dating the divergence time of organisms has become extremely important in the last decades, and the number of tools and methods developed is increasing exponentially. Molecular clock studies have been proven effective in unveiling many aspects of the evolution of life on global and local scale and have been applied to a plethora of organisms, from viruses to vertebrates, from insects to bacteria. In the cases studied in this thesis, I proposed new time-scaled phylogenies for a variety of invasive non-model species using different methodologies and datasets, providing interesting insights into the paleoecology of the species studied. The limited available genetic data made studying these species challenging; indeed, one of the major limitations of exploring the evolution of invasive species is, in some cases the paucity of molecular data. I investigated some unexplored fields of invasive species evolution and produced new genomic data by sequencing three species of interest for a large community of researchers (from zoologists to pathologists): *Aedes japonicus, Aedes koreicus,* and *Trissolcus japonicus.* In this thesis, I proposed an answer to several evolutionary issues, and the results could be of great interest for future comparative genomic studies. These results would be of great interest for field studies aimed at ameliorating the management of pests.

In **Chapter 2,** I investigated some aspects of Aedini mosquito evolution, examining the topology, the dated phylogeny, and the rate variance across branches and between nuclear and mitochondrial datasets. This work underlines a probable discordance between the mitochondrial and the nuclear phylogenies, which needs to be considered for future investigation to understand the past dynamics of the Aedini clade and to calibrate genomics studies in light of the right evolutionary timespan. We confirm that *A. flavopictus* is the closest relative presented in our dataset. Despite the attempts to provide molecular clock studies and time-calibrated trees of the Aedini clade, a satisfactory answer has not been provided yet. With this study, we underline the importance of properly applied molecular clock studies for understanding and driving future analysis in this field.

Unveiling the incongruence issues between nuclear and mitochondrial datasets suggested taking with caution the current and the previous evidence of dated phylogeny based only on mitochondrial data.

In **Chapter 3,** I updated the results of Chapter 2, employing the newly assembled mitogenomes from *A. koreicus* and *A. japonicus.* Here, I underlined the role of taxon sampling in reconstructing the phylogeny and the extremely high evolutionary rate in the Chironomidae clade and in many taxa within the Aedini clade. Moreover, the results uncover an unexpectedly high divergence between the two *A. koreicus* genomes and between *A. koreicus* and *A. japonicus.* This work constitutes the most comprehensive investigation into Aedini evolution using full mitochondrial sequences, which may be essential to scale future evolutionary analysis within this tribe. Here, I underlined the problem of the taxon sampling/outgroup sampling in inferring reliable evolutionary timescale in deep phylogeny. The knowledge acquired in the Chapters 2 and 3 would be of great interest to scale phylogenomic studies and understand when and in what ecological scenario the mosquitoes related novelties emerged. Macroevolutionary insights into the biology of mosquitoes and especially into the four invasive *Aedes* mosquitoes now can carry out in the light of molecular clock studies.

In **Chapters 2 and 3**, we proposed that the Aedini radiation can match the mammal and bird radiation. In the dated phylogeny proposed in Chapter 3, the Aedini crown date matches the K-Pg (66 Ma) boundary. This boundary designed the turning point for the evolution of many orders of placental mammals and birds (Leary et al. 2013; Prum et al. 2015). Aedini mosquitoes could have faced fast radiation in relation to new exploitable niches given by a rising number of mammals and birds that appeared after the K-Pg boundary. In addition, I tentatively proposed that the Angiosperm split is congruent with the initial Culicomorpha diversification. However, these speculations have to take with caution, and more specific studies have to be performed to define a tight correlation between the mosquitoes' evolution and other clades' evolutionary histories.

In **Chapter 4,** I presented the preliminary analysis of genome skimming for *A. japonicus* and *A. koreicus,* and a successful genome assembly for *Trissolcus*

japonicus, three species of growing interest for agriculture and human health**.** The two Aedini genomes are of extreme interest for future analysis of mosquito evolution. Indeed, these data increase the phylogenetic resolution of *Aedes* early divergence. The *T. japonicus* genome is of utmost relevance for integrated studies on the biological control of *Halyomorpha halys.* Metagenomic screening revealed the presence of the intracellular parasite *Wolbachia* in *T. japonicus*: this may be relevant for field applications because this bacterium is known to manipulate the reproductive behaviour of its host.

In **Chapter 5,** I presented an investigation of the deep-time evolution of ZIKV, updating knowledge on this virus. I expand the understanding of its origin, the role of South-East Asia as a source of the infection, and the delay that may be present between actual introduction and detection. This knowledge is useful to ameliorate our understanding of pre-epidemic dynamics of ZIKV and other viruses. This study provides a global picture of ZIKV evolution that was missing in the record, in particular a timescale of the virus origin (SPOV-ZIKV split) which has to be further investigated with more SPOV samples.

In **Chapter 6,** I further show how the molecular clock can be successfully applied to other organisms and adapted to answer different evolutionary questions. I show how the molecular clock can be applied to pests of agriculture by investigating the recent diversification of *Arundo* and *Bactrocera,* providing new insights into their evolution. Our results for *Bactrocera* revealed a very recent radiation for this genus and the possibility that different species of agricultural relevance have a history of reciprocal hybridization: this is relevant for their management because we cannot exclude that hybridization is currently ongoing both in their native and in the newly invaded regions. Conversely, our clock studies indicate that *Arundo* species are well separated, having originated more than 2.5 Ma: this is also interesting for the management of this pest because it excludes hybridization events. The opsin phylogenomics unveiled some macroevolutionary changes in the opsin genes composition in Brachycera and mosquitoes clade. Some of these findings suggest a correlation between the ecological adaptation in dipteran pests and genetic novelties in their Repertoire of opsins: these novelties may be exploited to define, for example, more efficient colors of traps, therefore, ameliorating their management.

Overall, the work presented in this dissertation shows the benefit of apply molecular clock methods to a variety of datasets: using a few genes, full mitochondrial genomes, and genomic or transcriptomic datasets.. I successfully applied molecular clock methods to uncover evolutionary scenarios, showing how to deal with problematic datasets and providing time-calibrated trees. Time-dated phylogenies represent the backbone of comparative genomic studies and provide a solid base for investigating genome evolution in light of sound hypotheses. Paleoecological scenarios, speciation timescale, rate of gene family contraction and expansion are extremely helpful information for understanding the more intimate biology of pest species and ameliorating their management.

Future perspectives

The studies carried out during my PhD provided interesting results on the timing of the evolution of invasive species and Zika virus. The evolutionary history of many groups of invasive species has not been examined in-depth: they are frequently studied only for direct management and eradication. In the work presented here, the molecular clock has been shown to be a powerful approach to understanding the intimate dynamics of invasive species. More can be done to increase knowledge of mosquito evolution and of their associated viruses. I would propose two main research tracks to follow up in further investigation of the evolution of invasive mosquitoes and of invasive species in general.

Analysing genomic data and providing a nuclear-only molecular clock of Aedini mosquitoes. Mitochondrial and nuclear data can provide contrasting results. This is not only due to systematic or stochastic errors that could affect the analysis, but also to the fact that nuclear and mitochondrial DNA may tell genuinely different stories because the DNA material follows different inheritance patterns. A time-scaled phylogeny of nuclear data could provide a clearer scenario of mosquito evolution, revealing for example, events of hybridization that involves only nuclear DNA.

Investigating opsin evolution in the Aedini tribe in light of the two genomic data collected in this work (Chapter 4). Increasing the genomic data in the

Aedini clade opened the opportunity to better identify the evolutionary novelties that characterize this group. First, using the *Aedes japonicus* genome will make it possible to verify whether the evolutionary pattern in the Rh6 opsin is maintained throughout Aedini evolution (Chapter 6.1). Moreover, this approach may be extended to the analysis of other gene families to understand how different mosquitoes adapt to their ecological niches and what genes are involved.

Macroevolution, taxon sampling and phylogenomics. In Chapter 6, I provided some evidence of how phylogenomics can provide interesting insights into macroevolution. This application can be extended to invasive mosquitoes studies. One of the first things to investigate is if there are some evolutionary patterns in some gene families that are in common among invasive mosquitoes. However, we need to increase the taxon sampling within the Aedini clade, sequencing the non-invasive sister species of the mosquitoes in question. We can track better what genes are more involved in the anthropophilic behaviour of some species.

In addition, I want to address the limitations and challenges I faced during my work. Aedini is a highly diverse clade that comprises many species, but the genetic sequences for this group are very limited. The paucity of available data, especially nuclear data, makes it difficult to build a proper dataset. Beyond this technical limitation, studying the clock requires calibration. Unfortunately, the fossilization rate is extremely low in insects, especially in small insects like mosquitoes; this reduces the calibration points exploitable for inferring evolutionary time. The third limitation in studying Aedini mosquitoes is the low number of taxa represented by genetic data; only 24 full mitochondrial genomes are now available for more than 1260 species described. This limitation could be overcome only by increasing taxon sampling, which would make it possible to employ more fossils for tree calibration and to enhance the phylogenetic resolution, empowering comparative genomic studies.

In conclusion, in light of increasing genomic data, the molecular clock methodology can be exploited to reconstruct a more accurate and complete evolutionary history, becoming an essential tool within the integrated approaches for managing and controlling invasive mosquitoes and other

invasive species. Insect evolution is full of ecological and behavioural novelties worth further investigation.

References

Abascal, Federico, Rafael Zardoya, and Maximilian J. Telford. 2010. "TranslatorX: Multiple Alignment of Nucleotide Sequences Guided by Amino Acid Translations." *Nucleic Acids Research* 38 (SUPPL. 2): 1–31. https://doi.org/10.1093/nar/gkq291.

Adachi, Jun, and Masami Hasegawa. 1996. "Model of Amino Acid Substitution in Proteins Encoded by Mitochondrial DNA." *Journal of Molecular Evolution* 42 (4): 459–68. https://doi.org/10.1007/BF02498640.

Alfano, Niccolò, Valentina Tagliapietra, Fausta Rosso, Mattia Manica, Daniele Arnoldi, Massimo Pindo, and Annapaola Rizzoli. 2019. "Changes in Microbiota across Developmental Stages of Aedes Koreicus, an Invasive Mosquito Vector in Europe: Indications for Microbiota-Based Control Strategies." *Frontiers in Microbiology* 10: 2832.

Andújar, Carmelo, José Serrano, and Jesús Gámez-Zurita. 2012. "Winding up the Molecular Clock in the Genus Carabus (Coleoptera: Carabidae): Assessment of Methodological Decisions on Rate and Node Age Estimation." *BMC Evolutionary Biology* 12 (1). https://doi.org/10.1186/1471-2148-12-40.

Anselmetti, Yoann, Wandrille Duchemin, Eric Tannier, Cedric Chauve, and Sèverine Bérard. 2018. "Phylogenetic Signal from Rearrangements in 18 Anopheles Species by Joint Scaffolding Extant and Ancestral Genomes." *BMC Genomics* 19 (2): 131–45.

Arakawa, Ryo, Yuki Namura, and others. 2002. "Effects of Temperature on Development of Three Trissolcus Spp.(Hymenoptera: Scelionidae), Egg Parasitoids of the Brown Marmorated Stink Bug, Halyomorpha Halys (Hemiptera: Pentatomidae)." *Entomological Science* 5 (2): 215–18.

Aranda, Carles, Miguel J Mart\'\inez, Tomas Montalvo, Roger Eritja, Jessica Navero-Castillejos, Eva Herreros, Eduard Marqués, et al. 2018. "Arbovirus Surveillance: First Dengue Virus Detection in Local Aedes Albopictus Mosquitoes in Europe, Catalonia, Spain, 2015." *Eurosurveillance* 23 (47): 1700837.

Aris-Brosou, Stéphane, and Ziheng Yang. 2002. "Effects of Models of Rate Evolution on Estimation of Divergence Dates with Special Reference to the Metazoan 18S Ribosomal RNA Phylogeny." *Systematic Biology* 51 (5): 703–14.

Aris-Brosou, Stéphane, and Ziheng Yang. 2003. "Bayesian Models of Episodic Evolution Support a Late Precambrian Explosive Diversification of the Metazoa." *Molecular Biology and Evolution* 20 (12): 1947–54.

Asplen, Mark K, Gianfranco Anfora, Antonio Biondi, Deuk-Soo Choi, Dong Chu, Kent M Daane, Patricia Gibert, et al. 2015. "Invasion Biology of Spotted Wing Drosophila (Drosophila Suzukii): A Global Perspective and Future Priorities." *Journal of Pest Science* 88 (3): 469–94.

Attardo, Geoffrey M., Adly M.M. Abd-Alla, Alvaro Acosta-Serrano, James E. Allen, Rosemary Bateta, Joshua B. Benoit, Kostas Bourtzis, et al. 2019. "Comparative Genomic Analysis of Six Glossina Genomes, Vectors of African Trypanosomes." *Genome Biology* 20 (1): 1–31. https://doi.org/10.1186/s13059-019-1768-2.

Augustinos, Antonios A, Elena Drosopoulou, Aggeliki Gariou-Papalexiou, Kostas Bourtzis, Penelope Mavragani-Tsipidou, and Antigone Zacharopoulou. 2014. "The Bactrocera Dorsalis Species Complex: Comparative Cytogenetic Analysis in Support of Sterile Insect Technique Applications." *BMC Genetics* 15 (2): 1–10.

Badii, M H, J Landeros, E Cerna, and J L Abreu. 2007. "Ecologia e Historia Del Dengue En Las Américas." *Daena: Internacional J Good Consciente* 2.

Baele, Guy, Philippe Lemey, Trevor Bedford, Andrew Rambaut, Marc A. Suchard, and Alexander V. Alekseyenko. 2012. "Improving the Accuracy of Demographic and Molecular Clock Model Comparison While Accommodating Phylogenetic Uncertainty." *Molecular Biology and Evolution* 29 (9): 2157–67. https://doi.org/10.1093/molbev/mss084.

Baele, Guy, Philippe Lemey, and Marc A Suchard. 2016. "Genealogical Working Distributions for Bayesian Model Testing with Phylogenetic Uncertainty." *Systematic Biology* 65 (2): 250–64.

Baele, Guy, Wai Lok Sibon Li, Alexei J. Drummond, Marc A. Suchard, and Philippe Lemey. 2013. "Accurate Model Selection of Relaxed Molecular Clocks in Bayesian Phylogenetics." *Molecular Biology and Evolution* 30 (2): 239–43. https://doi.org/10.1093/molbev/mss243.

Baharmand, Iman, Heather Coatsworth, Daniel A.H. Peach, Peter Belton, and Carl Lowenberger. 2020. "Molecular Relationships of Introduced Aedes Japonicus (Diptera: Culicidae) Populations in British Columbia, Canada Using Mitochondrial DNA." *Journal of Vector Ecology* 45 (2): 285–96. https://doi.org/10.1111/jvec.12399.

Baldacchino, F., F. Montarsi, D. Arnoldi, C. Barategui, N. Ferro Milone, G. Da Rold, G. Capelli, A. Rizzoli, and Phillip Kaufman. 2017. "A 2-Yr Mosquito Survey Focusing on Aedes Koreicus (Diptera: Culicidae) in Northern Italy and Implications for Adult Trapping." *Journal of Medical Entomology* 54 (3): 622–30. https://doi.org/10.1093/jme/tjw216.

Baldacchino, Frédéric, Beniamino Caputo, Fabrice Chandre, Andrea Drago,

Alessandra della Torre, Fabrizio Montarsi, and Annapaola Rizzoli. 2015. "Control Methods against Invasive Aedes Mosquitoes in Europe: A Review." *Pest Management Science* 71 (11): 1471–85. https://doi.org/10.1002/ps.4044.

Balingit, Jean Claude, Thaddeus M Carvajal, Mariko Saito-Obata, Maribet Gamboa, Amalea Dulcene Nicolasora, Ava Kristy Sy, Hitoshi Oshitani, and Kozo Watanabe. 2020. "Surveillance of Dengue Virus in Individual Aedes Aegypti Mosquitoes Collected Concurrently with Suspected Human Cases in Tarlac City, Philippines." *Parasites & Vectors* 13 (1): 1–13.

Barido-Sottani, Joëlle, Veronika Bošková, Louis Du Plessis, Denise Kühnert, Carsten Magnus, Venelin Mitov, Nicola F. Müller, et al. 2018. "Taming the BEAST - A Community Teaching Material Resource for BEAST 2." *Systematic Biology* 67 (1): 170–74. https://doi.org/10.1093/sysbio/syx060.

Barrera-Redondo, Josué, Santiago Ramírez-Barahona, and Luis E. Eguiarte. 2018. "Rates of Molecular Evolution in Tree Ferns Are Associated with Body Size, Environmental Temperature, and Biological Productivity." *Evolution* 72 (5): 1050–62. https://doi.org/10.1111/evo.13475.

Bartlett-Healy, Kristen, Isik Unlu, Peter Obenauer, Tony Hughes, Sean Healy, Taryn Crepeau, Ary Farajollahi, et al. 2012. "Larval Mosquito Habitat Utilization and Community Dynamics of Aedes Albopictus and Aedes Japonicus (Diptera: Culicidae)." *Journal of Medical Entomology* 49 (4): 813–24.

Bass, Chris, Alin M Puinean, Christoph T Zimmer, Ian Denholm, Linda M Field, Stephen P Foster, Oliver Gutbrod, Ralf Nauen, Russell Slater, and Martin S Williamson. 2014. "The Evolution of Insecticide Resistance in the Peach Potato Aphid, Myzus Persicae." *Insect Biochemistry and Molecular Biology* 51: 41–51.

Battistuzzi, Fabia U., Paul Billing-Ross, Oscar Murillo, Alan Filipski, and Sudhir Kumar. 2015. "A Protocol for Diagnosing the Effect of Calibration Priors on Posterior Time Estimates: A Case Study for the Cambrian Explosion of Animal Phyla." *Molecular Biology and Evolution* 32 (7): 1907–12. https://doi.org/10.1093/molbev/msv075.

Beaumier, Coreen, Melissa N Garcia, and Kristy O Murray. 2014. "The History of Dengue in the United States and Its Recent Emergence." *Current Tropical Medicine Reports* 1 (1): 32–35.

Beaver, Jacob T., Nadia Lelutiu, Rumi Habib, and Ioanna Skountzou. 2018. "Evolution of Two Major Zika Virus Lineages: Implications for Pathology, Immune Response, and Vaccine Development." *Frontiers in Immunology* 9 (JUL). https://doi.org/10.3389/fimmu.2018.01640.

Belshaw, Robert, Andy Gardner, Andrew Rambaut, and Oliver G. Pybus. 2008.

"Pacing a Small Cage: Mutation and RNA Viruses." *Trends in Ecology and Evolution* 23 (4): 188–93. https://doi.org/10.1016/j.tree.2007.11.010.

Beltrán, Margarita, Chris D Jiggins, Vanessa Bull, Mauricio Linares, James Mallet, W Owen McMillan, and Eldredge Bermingham. 2002. "Phylogenetic Discordance at the Species Boundary: Comparative Gene Genealogies among Rapidly Radiating Heliconius Butterflies." *Molecular Biology and Evolution* 19 (12): 2176–90.

Benedict, Mark Q, Rebecca S Levine, William A Hawley, and L Philip Lounibos. 2007. "Spread of the Tiger: Global Risk of Invasion by the Mosquito Aedes Albopictus." *Vector-Borne and Zoonotic Diseases* 7 (1): 76–85.

Benton, Michael J., and Philip C.J. Donoghue. 2007. "Paleontological Evidence to Date the Tree of Life." *Molecular Biology and Evolution* 24 (1): 26–53. https://doi.org/10.1093/molbev/msl150.

Bergsten, Johannes, Anders N. Nilsson, and Fredrik Ronquist. 2013. "Bayesian Tests of Topology Hypotheses with an Example from Diving Beetles." *Systematic Biology* 62 (5): 660–73. https://doi.org/10.1093/sysbio/syt029.

Bernt, Matthias, Anke Braband, Martin Middendorf, Bernhard Misof, Omar Rota-Stabelli, and Peter F. Stadler. 2013. "Bioinformatics Methods for the Comparative Analysis of Metazoan Mitochondrial Genome Sequences." *Molecular Phylogenetics and Evolution* 69 (2): 320–27. https://doi.org/10.1016/j.ympev.2012.09.019.

Bernt, Matthias, Alexander Donath, Frank Jühling, Fabian Externbrink, Catherine Florentz, Guido Fritzsch, Joern Pütz, Martin Middendorf, and Peter F. Stadler. 2013. "MITOS: Improved de Novo Metazoan Mitochondrial Genome Annotation." *Molecular Phylogenetics and Evolution* 69 (2): 313–19. https://doi.org/10.1016/j.ympev.2012.08.023.

Bertrand, Yann, Mats Töpel, Annelie Elväng, Wessam Melik, and Magnus Johansson. 2012. "First Dating of a Recombination Event in Mammalian Tick-Borne Flaviviruses." *PLoS One* 7 (2): e31981.

Bhatt, Samir, Peter W Gething, Oliver J Brady, Jane P Messina, Andrew W Farlow, Catherine L Moyes, John M Drake, et al. 2013. "The Global Distribution and Burden of Dengue." *Nature* 496 (7446): 504–7.

Biffin, Ed, Robert S. Hill, and Andrew J. Lowe. 2010. "Did Kauri (Agathis: Araucariaceae) Really Survive the Oligocene Drowning of New Zealand?" *Systematic Biology* 59 (5): 594–602. https://doi.org/10.1093/sysbio/syq030.

Bo, W, S Ahmad, T Dammalage, U Sto Tomas, Viwat Wornoayporn, I Ul Haq, C Cáceres, M J B Vreysen, J Hendrichs, and M K Schutze. 2014. "Mating Compatibility between Bactrocera Invadens and Bactrocera Dorsalis (Diptera: Tephritidae)." *Journal of Economic Entomology* 107 (2): 623–29.

Boni, Maciej F., David Posada, and Marcus W. Feldman. 2007. "An Exact Nonparametric Method for Inferring Mosaic Structure in Sequence Triplets." *Genetics* 176 (2): 1035–47. https://doi.org/10.1534/genetics.106.068874.

Booth, H. Anne F., and Peter W.H. Holland. 2004. "Eleven Daughters of NANOG." *Genomics* 84 (2): 229–38. https://doi.org/10.1016/j.ygeno.2004.02.014.

Borkent, Art, and David A. Grimaldi. 2004. "The Earliest Fossil Mosquito (Diptera: Culicidae), in Mid-Cretaceous Burmese Amber." *Annals of the Entomological Society of America* 97 (5): 882–88. https://doi.org/10.1603/0013-8746(2004)097[0882:TEFMDC]2.0.CO;2.

Bos, Kirsten I., Verena J. Schuenemann, G. Brian Golding, Hernãn A. Burbano, Nicholas Waglechner, Brian K. Coombes, Joseph B. McPhee, et al. 2011. "A Draft Genome of Yersinia Pestis from Victims of the Black Death." *Nature* 478 (7370): 506–10. https://doi.org/10.1038/nature10549.

Bouckaert, Remco, Joseph Heled, Denise Kühnert, Tim Vaughan, Chieh Hsi Wu, Dong Xie, Marc A Suchard, Andrew Rambaut, and Alexei J Drummond. 2014. "BEAST 2: A Software Platform for Bayesian Evolutionary Analysis." *PLoS Computational Biology* 10 (4): 1–6. https://doi.org/10.1371/journal.pcbi.1003537.

Bouckaert, Remco R, and Alexei J Drummond. 2017. "BModelTest: Bayesian Phylogenetic Site Model Averaging and Model Comparison." *BMC Evolutionary Biology* 17 (1): 1–11.

Bouckaert, Remco, Timothy G Vaughan, Joëlle Barido-Sottani, Sebastián Duchêne, Mathieu Fourment, Alexandra Gavryushkina, Joseph Heled, et al. 2019. "BEAST 2.5: An Advanced Software Platform for Bayesian Evolutionary Analysis." *PLoS Computational Biology* 15 (4): 1–28. https://doi.org/10.1371/journal.pcbi.1006650.

Bouckaert, Remco, and Walter Xie. 2016. "Ancestral Reconstruction/Discrete Phylogeography with BEAST 2.3.X," 1–21. http://tree.bio.ed.ac.uk/software/figtree/.

Brady, Oliver J, Nick Golding, David M Pigott, Moritz U G Kraemer, Jane P Messina, Robert C Reiner Jr, Thomas W Scott, David L Smith, Peter W Gething, and Simon I Hay. 2014. "Global Temperature Constraints on Aedes Aegypti and Ae. Albopictus Persistence and Competence for Dengue Virus Transmission." *Parasites & Vectors* 7 (1): 1–17.

Brady, Oliver J, Michael A Johansson, Carlos A Guerra, Samir Bhatt, Nick Golding, David M Pigott, Hélène Delatte, et al. 2013. "Modelling Adult Aedes Aegypti and Aedes Albopictus Survival at Different Temperatures in

Laboratory and Field Settings." *Parasites & Vectors* 6 (1): 1–12.

Branstetter, Michael G., Anna K. Childers, Diana Cox-Foster, Keith R. Hopper, Karen M. Kapheim, Amy L. Toth, and Kim C. Worley. 2018. "Genomes of the Hymenoptera." *Current Opinion in Insect Science* 25 (November 2017): 65–75. https://doi.org/10.1016/j.cois.2017.11.008.

Broeck, Frederik Van den, Nicholas J. Savill, Hideo Imamura, Mandy Sanders, Ilse Maes, Sinclair Cooper, David Mateus, et al. 2020. "Ecological Divergence and Hybridization of Neotropical Leishmania Parasites." *Proceedings of the National Academy of Sciences of the United States of America* 117 (40): 25159–68. https://doi.org/10.1073/pnas.1920136117.

Bromham, Lindell, Sebastián Duchêne, Xia Hua, Andrew M. Ritchie, David A. Duchêne, and Simon Y.W. Ho. 2018. "Bayesian Molecular Dating: Opening up the Black Box." *Biological Reviews* 93 (2): 1165–91. https://doi.org/10.1111/brv.12390.

Bromham, Lindell, and Remko Leys. 2005. "Sociality and the Rate of Molecular Evolution." *Molecular Biology and Evolution* 22 (6): 1393–1402. https://doi.org/10.1093/molbev/msi133.

Brown, Richard P, and Ziheng Yang. 2011. "Rate Variation and Estimation of Divergence Times Using Strict and Relaxed Clocks." *BMC Evolutionary Biology* 11 (1): 1–12.

Bryant, Juliet E., Edward C. Holmes, and Alan D.T. Barrett. 2007. "Out of Africa: A Molecular Perspective on the Introduction of Yellow Fever Virus into the Americas." *PLoS Pathogens* 3 (5): 0668–73. https://doi.org/10.1371/journal.ppat.0030075.

Buathong, Rome, Laura Hermann, Butsaya Thaisomboonsuk, Wiriya Rutvisuttinunt, Chonticha Klungthong, Piyawan Chinnawirotpisan, Wudtichai Manasatienkij, et al. 2015. "Detection of Zika Virus Infection in Thailand, 2012-2014." *American Journal of Tropical Medicine and Hygiene* 93 (2): 380–83. https://doi.org/10.4269/ajtmh.15-0022.

Burke, Sean V, Choun-Sea Lin, William P Wysocki, Lynn G Clark, and Melvin R Duvall. 2016. "Phylogenomics and Plastome Evolution of Tropical Forest Grasses (Leptaspis, Streptochaeta: Poaceae)." *Frontiers in Plant Science* 7: 1993.

Bush, Robin M., Catherine B. Smith, Nancy J. Cox, and Walter M. Fitch. 2000. "Effects of Passage History and Sampling Bias on Phylogenetic Reconstruction of Human Influenza A Evolution." *Proceedings of the National Academy of Sciences of the United States of America* 97 (13): 6974–80. https://doi.org/10.1073/pnas.97.13.6974.

Cameron, Emilie C, Richard C Wilkerson, Motoyoshi Mogi, Ichiro Miyagi,

Takako Toma, Heung Chul Kim, and Dina M Fonseca. 2010. "Molecular Phylogenetics of Aedes Japonicus, a Disease Vector That Recently Invaded Western Europe, North America, and the Hawaiian Islands." *Journal of Medical Entomology* 47 (4): 527–35. https://doi.org/10.1603/ME09259.

Cameron, Stephen L. 2014. "Insect Mitochondrial Genomics: Implications for Evolution and Phylogeny." *Annual Review of Entomology* 59: 95–117. https://doi.org/10.1146/annurev-ento-011613-162007.

Cameron, Stephen L., Christine L. Lambkin, Stephen C. Barker, and Michael F. Whiting. 2007. "A Mitochondrial Genome Phylogeny of Diptera: Whole Genome Sequence Data Accurately Resolve Relationships over Broad Timescales with High Precision." *Systematic Entomology* 32 (1): 40–59. https://doi.org/10.1111/j.1365-3113.2006.00355.x.

Campos, Túlio De Lima, Ricardo Durães-Carvalho, Antonio Mauro Rezende, Otávio Valério de Carvalho, Alain Kohl, Gabriel Luz Wallau, and Lindomar José Pena. 2018. "Revisiting Key Entry Routes of Human Epidemic Arboviruses into the Mainland Americas through Large-Scale Phylogenomics." *International Journal of Genomics* 2018.

Cao-Lormeau, Van Mai. 2014. "RE: Zika Virus, French Polynesia, South Pacific, 2013." *Emerging Infectious Diseases* 20 (11): 1960. https://doi.org/10.3201/eid2011.141380.

Capelli, Gioia, Andrea Drago, Simone Martini, Fabrizio Montarsi, Mauro Soppelsa, Nicola Delai, Silvia Ravagnan, et al. 2011. "First Report in Italy of the Exotic Mosquito Species Aedes (Finlaya) Koreicus, a Potential Vector of Arboviruses and Filariae." *Parasites and Vectors* 4 (1): 188. https://doi.org/10.1186/1756-3305-4-188.

Cavalli-Sforza, Luigi L, and Anthony W F Edwards. 1967. "Phylogenetic Analysis. Models and Estimation Procedures." *American Journal of Human Genetics* 19 (3 Pt 1): 233.

Charlesworth, Brian. 2009. "Fundamental Concepts in Genetics: Effective Population Size and Patterns of Molecular Evolution and Variation." *Nature Reviews Genetics* 10 (3): 195–205. https://doi.org/10.1038/nrg2526.

Chen, Lin H., and Mary E. Wilson. 2020. "Yellow Fever Control: Current Epidemiology and Vaccination Strategies." *Tropical Diseases, Travel Medicine and Vaccines* 6 (1): 1–10. https://doi.org/10.1186/s40794-020-0101-0.

Chen, Xiao-Guang, Xuanting Jiang, Jinbao Gu, Meng Xu, Yang Wu, Yuhua Deng, Chi Zhang, et al. 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.

https://doi.org/10.1073/pnas.1516410112.

Cheng, Gong, Yang Liu, Penghua Wang, and Xiaoping Xiao. 2016. "Mosquito Defense Strategies against Viral Infection." *Trends in Parasitology* 32 (3): 177–86.

Chibueze, Ezinne C., Veronika Tirado, Katharina da Silva Lopes, Olukunmi O. Balogun, Yo Takemoto, Toshiyuki Swa, Amarjargal Dagvadorj, et al. 2017. "Zika Virus Infection in Pregnancy: A Systematic Review of Disease Course and Complications." *Reproductive Health* 14 (1): 28. https://doi.org/10.1186/s12978-017-0285-6.

Choo, Amanda, Thu N M Nguyen, Christopher M Ward, Isabel Y Chen, John Sved, Deborah Shearman, Anthony S Gilchrist, Peter Crisp, and Simon W Baxter. 2019. "Identification of Y-Chromosome Scaffolds of the Queensland Fruit Fly Reveals a Duplicated Gyf Gene Paralogue Common to Many Bactrocera Pest Species." *Insect Molecular Biology* 28 (6): 873–86.

Chouin-Carneiro, Thais, Mariana Rocha David, Fernanda de Bruycker Nogueira, Flavia Barreto Dos Santos, and Ricardo Lourenço-De-oliveira. 2020. "Zika Virus Transmission by Brazilian Aedes Aegypti and Aedes Albopictus Is Virus Dose and Temperature-Dependent." *PLoS Neglected Tropical Diseases* 14 (9): 1–14. https://doi.org/10.1371/journal.pntd.0008527.

Christin, Pascal-Antoine, Elizabeth Spriggs, Colin P Osborne, Caroline A E Strömberg, Nicolas Salamin, and Erika J Edwards. 2014. "Molecular Dating, Evolutionary Rates, and the Age of the Grasses." *Systematic Biology* 63 (2): 153–65.

Ciota, Alexander T., Dylan J. Ehrbar, Greta A. Van Slyke, Anne F. Payne, Graham G. Willsey, Rachael E. Viscio, and Laura D. Kramer. 2012. "Quantification of Intrahost Bottlenecks of West Nile Virus in Culex Pipiens Mosquitoes Using an Artificial Mutant Swarm." *Infection, Genetics and Evolution* 12 (3): 557–64. https://doi.org/10.1016/j.meegid.2012.01.022.

Clarke, John T, Rachel C M Warnock, and Philip C J Donoghue. 2011. "Establishing a Time-Scale for Plant Evolution." *New Phytologist* 192 (1): 266–301.

Coiro, Mario, James A. Doyle, and Jason Hilton. 2019. "How Deep Is the Conflict between Molecular and Fossil Evidence on the Age of Angiosperms?" *New Phytologist* 223 (1): 83–99. https://doi.org/10.1111/nph.15708.

Coon, Kerri L, Kevin J Vogel, Mark R Brown, and Michael R Strand. 2014. "Mosquitoes Rely on Their Gut Microbiota for Development." *Molecular Ecology* 23 (11): 2727–39.

Costa, Larissa Catharina, Rafael Valente Veiga, Juliane Fonseca Oliveira,

Moreno S. Rodrigues, Roberto F.S. Andrade, Enny S. Paixão, Maria Glória Teixeira, et al. 2021. "New Insights on the Zika Virus Arrival in the Americas and Spatiotemporal Reconstruction of the Epidemic Dynamics in Brazil." *Viruses* 13 (1). https://doi.org/10.3390/v13010012.

Couper, Lisa I., Johannah E. Farner, Jamie M. Caldwell, Marissa L. Childs, Mallory J. Harris, Devin G. Kirk, Nicole Nova, et al. 2021. "How Will Mosquitoes Adapt to Climate Warming?" *ELife* 10: 1–36. https://doi.org/10.7554/eLife.69630.

Crava, Cristina M, Christoph Brütting, and Ian T Baldwin. 2016. "Transcriptome Profiling Reveals Differential Gene Expression of Detoxification Enzymes in a Hemimetabolous Tobacco Pest after Feeding on Jasmonate-Silenced Nicotiana Attenuata Plants." *BMC Genomics* 17 (1): 1–15. https://doi.org/10.1186/s12864-016-3348-0.

Crava, Cristina M, Sukanya Ramasamy, Lino Ometto, Gianfranco Anfora, and Omar Rota-Stabelli. 2016. "Evolutionary Insights into Taste Perception of the Invasive Pest Drosophila Suzukii." *G3: Genes, Genomes, Genetics* 6 (12): 4185–96.

Cruickshank, Leanne, Andrew J Jessup, and David J Cruickshank. 2001. "Interspecific Crosses of Bactrocera Tryoni (Froggatt) and Bactrocera Jarvisi (Tryon)(Diptera: Tephritidae) in the Laboratory." *Australian Journal of Entomology* 40 (3): 278–80.

Damiens, David, Audrey Ayrinhac, Wim Van Bortel, Veerle Versteirt, Wouter Dekoninck, and Thierry Hance. 2014. "Invasive Process and Repeated Cross-Sectional Surveys of the Mosquito Aedes Japonicus Japonicus Establishment in Belgium." *PLoS One* 9 (4): e89358.

Dayhoff, M, R Schwartz, and B Orcutt. 1978. "A Model of Evolutionary Change in Proteins." *Atlas of Protein Sequence and Structure* 5: 345–52.

Denver, Dee R., Amanda M.V. Brown, Dana K. Howe, Amy B. Peetz, and Inga A. Zasada. 2016. "Genome Skimming: A Rapid Approach to Gaining Diverse Biological Insights into Multicellular Pathogens." *PLoS Pathogens* 12 (8): 1–7. https://doi.org/10.1371/journal.ppat.1005713.

DeSalle, Rob, and L Val Giddings. 1986. "Discordance of Nuclear and Mitochondrial DNA Phylogenies in Hawaiian Drosophila." *Proceedings of the National Academy of Sciences* 83 (18): 6902–6.

Dick G. 1952. "Zika Isolation and Serological Specificity." *Trans Royal Soc Trop Med Hyg* 46 (5): 509–20.

Doolittle, Russell F., and Birger Blombäck. 1964. "Amino-Acid Sequence Investigations of Fibrinopeptides from Various Mammals: Evolutionary Implications." *Nature* 202 (4928): 147–52. https://doi.org/10.1038/202147a0.

Dritsou, Vicky, Pantelis Topalis, Nikolai Windbichler, Alekos Simoni, Ann Hall, Daniel Lawson, Malcolm Hinsley, et al. 2015. "A Draft Genome Sequence of an Invasive Mosquito: An Italian *Aedes Albopictus*." *Pathogens and Global Health* 109 (5): 207–20. https://doi.org/10.1179/2047773215Y.0000000031.

Drummond, Alexei J., Simon Y.W. Ho, Matthew J. Phillips, and Andrew Rambaut. 2006. "Relaxed Phylogenetics and Dating with Confidence." *PLoS Biology* 4 (5): 699–710. https://doi.org/10.1371/journal.pbio.0040088.

Drummond, Alexei J, Geoff K Nicholls, Allen G Rodrigo, and Wiremu Solomon. 2002. "Estimating Mutation Parameters, Population History and Genealogy Simultaneously from Temporally Spaced Sequence Data." *Genetics* 161 (3): 1307–20.

Drummond, Alexei J, and Andrew Rambaut. 2007. "BEAST: Bayesian Evolutionary Analysis by Sampling Trees." *BMC Evolutionary Biology* 7 (1): 1–8.

Drummond, Alexei J, and Marc A Suchard. 2010. "Bayesian Random Local Clocks, or One Rate to Rule Them All." *BMC Biology* 8 (1): 1–12.

Duchêne, Sebastián, and Simon Y W Ho. 2014. "Using Multiple Relaxed-Clock Models to Estimate Evolutionary Timescales from DNA Sequence Data." *Molecular Phylogenetics and Evolution* 77: 65–70.

Duchêne, Sebastián, Martyna Molak, and Simon Y W Ho. 2014. "ClockstaR: Choosing the Number of Relaxed-Clock Models in Molecular Phylogenetic Analysis." *Bioinformatics* 30 (7): 1017–19.

Duffy, Mark R., Tai-Ho Chen, W. Thane Hancock, Ann M. Powers, Jacob L. Kool, Robert S. Lanciotti, Moses Pretrick, et al. 2009. "Zika Virus Outbreak on Yap Island, Federated States of Micronesia." *New England Journal of Medicine* 360 (24): 2536–43. https://doi.org/10.1056/nejmoa0805715.

Duffy, Siobain, Laura A. Shackelton, and Edward C. Holmes. 2008. "Rates of Evolutionary Change in Viruses: Patterns and Determinants." *Nature Reviews Genetics* 9 (4): 267–76. https://doi.org/10.1038/nrg2323.

Duong, Veasna, Philippe Dussart, and Philippe Buchy. 2017. "Zika Virus in Asia." *International Journal of Infectious Diseases* 54: 121–28. https://doi.org/10.1016/j.ijid.2016.11.420.

Eguchi, Satoshi, and Minoru N Tamura. 2016. "Evolutionary Timescale of Monocots Determined by the Fossilized Birth-Death Model Using a Large Number of Fossil Records." *Evolution* 70 (5): 1136–44.

Epelboin, Yanouk, Stanislas Talaga, Loïc Epelboin, and Isabelle Dusfour. 2017. "Zika Virus: An Updated Review of Competent or Naturally Infected Mosquitoes." *PLoS Neglected Tropical Diseases* 11 (11): 1–22. https://doi.org/10.1371/journal.pntd.0005933.

Eritja, Roger, Ignacio Ruiz-Arrondo, Sarah Delacour-Estrella, Francis Schaffner, Jorge Álvarez-Chachero, Mikel Bengoa, Mar\'\ia-Ángeles Puig, Rosario Melero-Alc\'\ibar, Aitana Oltra, and Frederic Bartumeus. 2019. "First Detection of Aedes Japonicus in Spain: An Unexpected Finding Triggered by Citizen Science." *Parasites & Vectors* 12 (1): 1–9.

Erlanger, Tobias E, Svenja Weiss, Jennifer Keiser, Jürg Utzinger, and Karin Wiedenmayer. 2009. "Past, Present, and Future of Japanese Encephalitis." *Emerging Infectious Diseases* 15 (1): 1.

Fajardo, Alvaro, Martín Soñora, Pilar Moreno, Gonzalo Moratorio, and Juan Cristina. 2016. "Bayesian Coalescent Inference Reveals High Evolutionary Rates and Diversification of Zika Virus Populations." *Journal of Medical Virology* 88 (10): 1672–76. https://doi.org/10.1002/jmv.24596.

Faria, N. R., J. Quick, I. M. Claro, J. Thézé, J. G. De Jesus, M. Giovanetti, M. U.G. Kraemer, et al. 2017. "Establishment and Cryptic Transmission of Zika Virus in Brazil and the Americas." *Nature* 546 (7658): 406–10. https://doi.org/10.1038/nature22401.

Faria, Nuno Rodrigues, Raimunda Do Socorro Da Silva Azevedo, Moritz U.G. Kraemer, Renato Souza, Mariana Sequetin Cunha, Sarah C. Hill, Julien Thézé, et al. 2016. "Zika Virus in the Americas: Early Epidemiological and Genetic Findings." *Science* 352 (6283): 345–49. https://doi.org/10.1126/science.aaf5036.

Faye, O., C.C.D.M. Freire, O. Faye, J.V. de Oliveira, P.M.D.A. Zanotto, M. Diallo, and A.A. Sall. 2014. "Molecular Evolution of Zika Virus during Its Emergence in the 20th Century." *International Journal of Infectious Diseases* 21 (1): 2–3. https://doi.org/10.1016/j.ijid.2014.03.411.

Faye, Oumar, Maria de Lourdes Monteiro, Bram Vrancken, Matthieu Prot, Sebastian Lequime, Maryam Diarra, Oumar Ndiaye, et al. 2020. "Genomic Epidemiology of 2015-2016 Zika Virus Outbreak in Cape Verde." *Emerging Infectious Diseases* 26 (6): 1084–90. https://doi.org/10.3201/EID2606.190928.

Felsenstein, Joseph. 1981. "Evolutionary Trees from DNA Sequences: A Maximum Likelihood Approach." *Joseph Felsenstein* 17 (6): 368–76. https://doi.org/10.13110/humanbiology.89.3.05.

Feuda, Roberto, Martin Dohrmann, Walker Pett, Hervé Philippe, Omar Rota-Stabelli, Nicolas Lartillot, Gert Wörheide, and Davide Pisani. 2017. "Improved Modeling of Compositional Heterogeneity Supports Sponges as Sister to All Other Animals." *Current Biology* 27 (24): 3864-3870.e4. https://doi.org/10.1016/j.cub.2017.11.008.

Feuda, Roberto, Ferdinand Marletaz, Michael A Bentley, and Peter W H Holland. 2016. "Conservation, Duplication, and Divergence of Five Opsin

Genes in Insect Evolution." *Genome Biology and Evolution* 8 (3): 579–87.

Figueiredo, Mario Lg De, Almério de C Gomes, Alberto A Amarilla, André de S Leandro, Agnaldo de S Orrico, Renato F De Araujo, Jesu\'\ina do SM Castro, Edison L Durigon, Victor H Aquino, and Luiz T M Figueiredo. 2010. "Mosquitoes Infected with Dengue Viruses in Brazil." *Virology Journal* 7 (1): 1–5.

Fischer, Dominik, Stephanie Margarete Thomas, Franziska Niemitz, Björn Reineking, and Carl Beierkuhnlein. 2011. "Projection of Climatic Suitability for Aedes Albopictus Skuse (Culicidae) in Europe under Climate Change Conditions." *Global and Planetary Change* 78 (1–2): 54–64. https://doi.org/10.1016/j.gloplacha.2011.05.008.

Fisher, Ronald Aylmer. 1930. *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.

Fonseca, D. M., A. K. Widdel, M. Hutchinson, S. E. Spichiger, and L. D. Kramer. 2010. "Fine-Scale Spatial and Temporal Population Genetics of Aedes Japonicus, a New US Mosquito, Reveal Multiple Introductions." *Molecular Ecology* 19 (8): 1559–72. https://doi.org/10.1111/j.1365-294X.2010.04576.x.

Forrester, Naomi L., Lark L. Coffey, and Scott C. Weaver. 2014. "Arboviral Bottlenecks and Challenges to Maintaining Diversity and Fitness during Mosquito Transmission." *Viruses* 6 (10): 3991–4004. https://doi.org/10.3390/v6103991.

Foster, Charles S P, and Simon Y W Ho. 2017. "Strategies for Partitioning Clock Models in Phylogenomic Dating: Application to the Angiosperm Evolutionary Timescale." *Genome Biology and Evolution* 9 (10): 2752–63.

Foster, Charles S P, Hervé Sauquet, Marlien der Merwe, Hannah McPherson, Maurizio Rossetto, and Simon Y W Ho. 2017. "Evaluating the Impact of Genomic Data and Priors on Bayesian Estimates of the Angiosperm Evolutionary Timescale." *Systematic Biology* 66 (3): 338–51.

Foster, Peter G., Tatiane Marques Porangaba de Oliveira, Eduardo S. Bergo, Jan E. Conn, Denise Cristina Sant'Ana, Sandra Sayuri Nagaki, Silvio Nihei, et al. 2017. "Phylogeny of Anophelinae Using Mitochondrial Protein Coding Genes." *Royal Society Open Science* 4 (11): 170758. https://doi.org/10.1098/rsos.170758.

Freire, C. C. M., A. Iamarino, D. F. L. Neto, A. A. Sall, P. Marinho, and A. Zanotto. 2015. "Spread of the Pandemic Zika Virus Lineage Is Associated with NS1 Codon Usage Adaptation in Humans." *BioRxiv* 2015 (August): 1–8. https://doi.org/10.1101/032839.

Fritz, R., J. Blazevic, C. Taucher, K. Pangerl, F. X. Heinz, and K. Stiasny. 2011. "The Unique Transmembrane Hairpin of Flavivirus Fusion Protein E Is

Essential for Membrane Fusion." *Journal of Virology* 85 (9): 4377–85. https://doi.org/10.1128/jvi.02458-10.

Fuehrer, Hans-Peter, Ellen Schoener, Stefanie Weiler, Bita Shahi Barogh, Carina Zittra, and Gernot Walder. 2020. "Monitoring of Alien Mosquitoes in Western Austria (Tyrol, Austria, 2018)." *PLoS Neglected Tropical Diseases* 14 (6): e0008433.

Gillespie, J. H. 1989. "Lineage Effects and the Index of Dispersion of Molecular Evolution." *Molecular Biology and Evolution* 6 (6): 636–47. https://doi.org/10.1093/oxfordjournals.molbev.a040576.

Gillespie, J H. 1991. "The Causes of Molecular Evolution." *Oxford University Press*.

Gillooly, James F., Andrew P. Allen, Geoffrey B. West, and James H. Brown. 2005. "The Rate of DNA Evolution: Effects of Body Size and Temperature on the Molecular Clock." *Proceedings of the National Academy of Sciences of the United States of America* 102 (1): 140–45. https://doi.org/10.1073/pnas.0407735101.

Gloria-Soria, Andrea, Diego Ayala, Ambicadutt Bheecarry, Olger Calderon-Arguedas, Dave D Chadee, Marina Chiappero, Maureen Coetzee, et al. 2016. "Global Genetic Diversity of Aedes Aegypti." *Molecular Ecology* 25 (21): 5377–95.

Goldman, Nick, and Ziheng Yang. 1994. "A Codon-Based Model of Nucleotide Substitution for Protein-Coding DNA Sequences." *Molecular Biology and Evolution* 11 (5): 725–36. https://doi.org/10.1093/oxfordjournals.molbev.a040153.

Gong, Zhen, Yuxia Gao, and Guan Zhu Han. 2016. "Zika Virus: Two or Three Lineages?" *Trends in Microbiology* 24 (7): 521–22. https://doi.org/10.1016/j.tim.2016.05.002.

Gong, Zhen, Xiaoyu Xu, and Guan Zhu Han. 2017. "The Diversification of Zika Virus: Are There Two Distinct Lineages?" *Genome Biology and Evolution* 9 (11): 2940–45. https://doi.org/10.1093/gbe/evx223.

Gould, Ernest A, and Stephen Higgs. 2009. "Impact of Climate Change and Other Factors on Emerging Arbovirus Diseases." *Transactions of the Royal Society of Tropical Medicine and Hygiene* 103 (2): 109–21.

Gould, Ernest, John Pettersson, Stephen Higgs, Remi Charrel, and Xavier de Lamballerie. 2017. "Emerging Arboviruses: Why Today?" *One Health* 4 (April): 1–13. https://doi.org/10.1016/j.onehlt.2017.06.001.

Grandadam, Marc, Valérie Caro, Sébastien Plumet, Jean Michel Thiberge, Yvan Souarès, Anna Bella Failloux, Hugues J. Tolou, et al. 2011. "Chikungunya Virus, Southeastern France." *Emerging Infectious Diseases* 17 (5): 910–13.

https://doi.org/10.3201/eid1705.101873.

Grard, Gilda, Grégory Moureau, Rémi N. Charrel, Edward C Holmes, Ernest A Gould, and Xavier de Lamballerie. 2010. "Genomics and Evolution of Aedes-Borne Flaviviruses." *Journal of General Virology* 91 (1): 87–94. https://doi.org/10.1099/vir.0.014506-0.

Grubaugh, Nathan D., Jason T. Ladner, Moritz U.G. Kraemer, Gytis Dudas, Amanda L. Tan, Karthik Gangavarapu, Michael R. Wiley, et al. 2017. "Genomic Epidemiology Reveals Multiple Introductions of Zika Virus into the United States." *Nature* 546 (7658): 401–5. https://doi.org/10.1038/nature22400.

Grubaugh, Nathan D., Claudia Ruckert, Philip M. Armstrong, Angela Bransfield, John F. Anderson, Gregory D. Ebel, and Doug E. Brackney. 2016. "Transmission Bottlenecks and RNAi Collectively Influence Tick-Borne Flavivirus Evolution." *Virus Evolution* 2 (2): 1–14. https://doi.org/10.1093/ve/vew033.

Gryseels, Sophie, Thomas D. Watts, Jean Marie Kabongo Mpolesha, Brendan B. Larsen, Philippe Lemey, Jean Jacques Muyembe-Tamfum, Dirk E. Teuwen, and Michael Worobey. 2020. "A near Full-Length HIV-1 Genome from 1966 Recovered from Formalin-Fixed Paraffin-Embedded Tissue." *Proceedings of the National Academy of Sciences of the United States of America* 117 (22). https://doi.org/10.1073/pnas.1913682117.

Guidetti, Roberto, Sandra J. McInnes, Michele Cesari, Lorena Rebecchi, and Omar Rota-Stabelli. 2017. "Evolutionary Scenarios for the Origin of an Antarctic Tardigrade Species Based on Molecular Clock Analyses and Biogeographic Data." *Contributions to Zoology* 86 (2): 97–110. https://doi.org/10.1163/18759866-08602001.

Haddow, Andrew D., Amy J. Schuh, Chadwick Y. Yasuda, Matthew R. Kasper, Vireak Heang, Rekol Huy, Hilda Guzman, Robert B. Tesh, and Scott C. Weaver. 2012. "Genetic Characterization of Zika Virus Strains: Geographic Expansion of the Asian Lineage." *PLoS Neglected Tropical Diseases* 6 (2). https://doi.org/10.1371/journal.pntd.0001477.

Halasa, Yara A, Donald S Shepard, Dina M Fonseca, Ary Farajollahi, Sean Healy, Randy Gaugler, Kristen Bartlett-Healy, Daniel A Strickman, and Gary G Clark. 2014. "Quantifying the Impact of Mosquitoes on Quality of Life and Enjoyment of Yard and Porch Activities in New Jersey." *PloS One* 9 (3): e89221.

Hao, You-Jin, Yi-Lin Zou, Yi-Ran Ding, Wen-Yue Xu, Zhen-Tian Yan, Xu-Dong Li, Wen-Bo Fu, Ting-Jing Li, and Bin Chen. 2017. "Complete Mitochondrial Genomes of Anopheles Stephensi and An. Dirus and Comparative Evolutionary Mitochondriomics of 50 Mosquitoes." *Scientific Reports* 7 (1):

1–13.

Harapan, Harapan, Alice Michie, R. Tajo Sasmono, and Allison Imrie. 2020. "Dengue: A Minireview." *Viruses* 12 (8): 1–35. https://doi.org/10.3390/v12080829.

Harbach, Ralph E. 2012. "Culex Pipiens: Species versus Species Complex--Taxonomic History and Perspective." *Journal of the American Mosquito Control Association* 28 (4s): 10–23.

Hardion, Laurent, Régine Verlaque, Kristin Saltonstall, Agathe Leriche, and Bruno Vila. 2014. "Origin of the Invasive Arundo Donax (Poaceae): A Trans-Asian Expedition in Herbaria." *Annals of Botany* 114 (3): 455–62.

Hasegawa, Masami, Hirohisa Kishino, and Taka aki Yano. 1985. "Dating of the Human-Ape Splitting by a Molecular Clock of Mitochondrial DNA." *Journal of Molecular Evolution* 22 (2): 160–74. https://doi.org/10.1007/BF02101694.

Hautmann, Michael. 2020. "What Is Macroevolution?" *Palaeontology* 63 (1): 1–11.

Hayes, Edward B. 2009. "Zika Virus Outside Africa." *Emerging Infectious Diseases* 15 (9): 1347–50. https://doi.org/10.3201/eid1509.090442.

Heath, Tracy A, Mark T Holder, and John P Huelsenbeck. 2012. "A Dirichlet Process Prior for Estimating Lineage-Specific Substitution Rates." *Molecular Biology and Evolution* 29 (3): 939–55.

Heath, Tracy A, John P Huelsenbeck, and Tanja Stadler. 2014. "The Fossilized Birth--Death Process for Coherent Calibration of Divergence-Time Estimates." *Proceedings of the National Academy of Sciences* 111 (29): E2957--E2966.

Heled, Joseph, and Alexei J Drummond. 2015. "Calibrated Birth--Death Phylogenetic Time-Tree Priors for Bayesian Inference." *Systematic Biology* 64 (3): 369–83.

Higuera, Adriana, and Juan David Ramírez. 2019. "Molecular Epidemiology of Dengue, Yellow Fever, Zika and Chikungunya Arboviruses: An Update." *Acta Tropica* 190 (November 2018): 99–111. https://doi.org/10.1016/j.actatropica.2018.11.010.

Hill, Sarah C, Jocelyne Vasconcelos, Zoraima Neto, Domingos Jandondo, Líbia Zé-zé, Renato Santana Aguiar, Joilson Xavier, and Julien Thézé. 2019. "Emergence of the Asian Lineage of Zika Virus in Angola : An Outbreak Investigation." *The Lancet Infectious Diseases* 19 (10): 1138–47. https://doi.org/10.1016/S1473-3099(19)30293-2.

Hirano, Takahiro, Takumi Saito, Yoshihiro Tsunamoto, Joichiro Koseki, Bin Ye, Van Tu Do, Osamu Miura, Yoshihisa Suyama, and Satoshi Chiba. 2019.

"Enigmatic Incongruence between MtDNA and NDNA Revealed by Multi-Locus Phylogenomic Analyses in Freshwater Snails." *Scientific Reports* 9 (1): 1–13. https://doi.org/10.1038/s41598-019-42682-0.

Ho, Simon Y.W., Robert Lanfear, Lindell Bromham, Matthew J. Phillips, Julien Soubrier, Allen G. Rodrigo, and Alan Cooper. 2011. "Time-Dependent Rates of Molecular Evolution." *Molecular Ecology* 20 (15): 3087–3101. https://doi.org/10.1111/j.1365-294X.2011.05178.x.

Ho, Simon Y.W., and Nathan Lo. 2013. "The Insect Molecular Clock." *Australian Journal of Entomology* 52 (2): 101–5. https://doi.org/10.1111/aen.12018.

Ho, Simon Y.W., and Matthew J. Phillips. 2009. "Accounting for Calibration Uncertainty in Phylogenetic Estimation of Evolutionary Divergence Times." *Systematic Biology* 58 (3): 367–80. https://doi.org/10.1093/sysbio/syp035.

Ho, Simon Y W, Matthew J Phillips, Alexei J Drummond, and Alan Cooper. 2005. "Accuracy of Rate Estimation Using Relaxed-Clock Models with a Critical Focus on the Early Metazoan Radiation." *Molecular Biology and Evolution* 22 (5): 1355–63.

Ho, Zheng Jie Marc, Hapuarachchige Chanditha Hapuarachchi, Timothy Barkham, Angela Chow, Lee Ching Ng, Jian Ming Vernon Lee, Yee Sin Leo, et al. 2017. "Outbreak of Zika Virus Infection in Singapore: An Epidemiological, Entomological, Virological, and Clinical Analysis." *The Lancet Infectious Diseases* 17 (8): 813–21. https://doi.org/10.1016/S1473-3099(17)30249-9.

Hoebeke, E Richard. 2003. "Halyomorpha Halys (Stal)(Heteroptera: Pentatomidae): A Polyphagous Plant Pest from Asia Newly Detected in North America." *Proc. Entomol. Soc. Wash.* 105: 225–37.

Höhna, Sebastian, Michael J Landis, Tracy A Heath, Bastien Boussau, Nicolas Lartillot, Brian R Moore, John P Huelsenbeck, and Fredrik Ronquist. 2016. "RevBayes: Bayesian Phylogenetic Inference Using Graphical Models and an Interactive Model-Specification Language." *Systematic Biology* 65 (4): 726–36.

Holmes, Edward C., Michael Worobey, and Andrew Rambaut. 1999. "Phylogenetic Evidence for Recombination in Dengue Virus." *Molecular Biology and Evolution* 16 (3): 405–9. https://doi.org/10.1093/oxfordjournals.molbev.a026121.

Hu, Tao, Juan Li, Michael J. Carr, Sebastián Duchêne, and Weifeng Shi. 2019. "The Asian Lineage of Zika Virus: Transmission and Evolution in Asia and the Americas." *Virologica Sinica* 34 (1): 1–8. https://doi.org/10.1007/s12250-018-0078-2.

Huber, Katrin, Stephanie Jansen, Mayke Leggewie, Marlis Badusche, Jonas Schmidt-Chanasit, Norbert Becker, Egbert Tannich, and Stefanie C Becker. 2014. "Aedes Japonicus Japonicus (Diptera: Culicidae) from Germany Have Vector Competence for Japan Encephalitis Virus but Are Refractory to Infection with West Nile Virus." *Parasitology Research* 113 (9): 3195–99. https://doi.org/10.1007/s00436-014-3983-9.

Huber, Katrin, Kathrin Schuldt, Martin Rudolf, Marco Marklewitz, Dina M. Fonseca, Christian Kaufmann, Yoshio Tsuda, et al. 2014. "Distribution and Genetic Structure of Aedes Japonicus Japonicus Populations (Diptera: Culicidae) in Germany." *Parasitology Research* 113 (9): 3201–10. https://doi.org/10.1007/s00436-014-4000-z.

Huelsenbeck, John P, Bret Larget, and David Swofford. 2000. "A Compound Poisson Process for Relaxing the Molecular Clock." *Genetics* 154 (4): 1879–92.

Huelsenbeck, John P, and Fredrik Ronquist. 2001. "MRBAYES: Bayesian Inference of Phylogenetic Trees." *Bioinformatics* 17 (8): 754–55.

Hulme, Philip E. 2009. "Trade, Transport and Trouble: Managing Invasive Species Pathways in an Era of Globalization." *Journal of Applied Ecology* 46 (1): 10–18.

Ioos, S., H. P. Mallet, I. Leparc Goffart, V. Gauthier, T. Cardoso, and M. Herida. 2014. "Current Zika Virus Epidemiology and Recent Epidemics." *Medecine et Maladies Infectieuses* 44 (7): 302–7. https://doi.org/10.1016/j.medmal.2014.04.008.

Isaacs, Alison T, Henry D Mawejje, Sean Tomlinson, Daniel J Rigden, and Martin J Donnelly. 2018. "Genome-Wide Transcriptional Analyses in Anopheles Mosquitoes Reveal an Unexpected Association between Salivary Gland Gene Expression and Insecticide Resistance." *BMC Genomics* 19 (1): 1–12.

Jiang, Xiaofang, Ashley Peery, A Brantley Hall, Atashi Sharma, Xiao-Guang Chen, Robert M Waterhouse, Aleksey Komissarov, et al. 2014. "Genome Analysis of a Major Urban Malaria Vector Mosquito, Anopheles Stephensi." *Genome Biology* 15 (9): 1–18.

Jukes, Thomas H, Charles R Cantor, and others. 1969. "Evolution of Protein Molecules." *Mammalian Protein Metabolism* 3: 21–132.

Junqueira, Ana Carolina M, Ana Maria L Azeredo-Espin, Daniel F Paulo, Marco Antonio T Marinho, Lynn P Tomsho, Daniela I Drautz-Moses, Rikky W Purbojati, Aakrosh Ratan, and Stephan C Schuster. 2016. "Large-Scale Mitogenomics Enables Insights into Schizophora (Diptera) Radiation and Population Diversity." *Scientific Reports* 6 (1): 1–13.

Kass, Robert E, and Adrian E Raftery. 1995. “Bayes Factors.” *Journal of the American Statistical Association* 90 (430): 773–95. https://doi.org/10.1080/01621459.1995.10476572.

Katoh, Kazutaka, John Rozewicki, and Kazunori D Yamada. 2018. “MAFFT Online Service: Multiple Sequence Alignment, Interactive Sequence Choice and Visualization.” *Briefings in Bioinformatics* 20 (4): 1160–66. https://doi.org/10.1093/bib/bbx108.

Kauffman, Elizabeth B., and Laura D. Kramer. 2017. “Zika Virus Mosquito Vectors: Competence, Biology, and Vector Control.” *Journal of Infectious Diseases* 216 (Suppl 10): S976–90. https://doi.org/10.1093/infdis/jix405.

Kaufman, Michael G, and Dina M Fonseca. 2014. “Invasion Biology of Aedes Japonicus Japonicus (Diptera: Culicidae).” *Annual Review of Entomology* 59: 31–49.

Keightley, Peter D., Ana Pinharanda, Rob W. Ness, Fraser Simpson, Kanchon K. Dasmahapatra, James Mallet, John W. Davey, and Chris D. Jiggins. 2015. “Estimation of the Spontaneous Mutation Rate in Heliconius Melpomene.” *Molecular Biology and Evolution* 32 (1): 239–43. https://doi.org/10.1093/molbev/msu302.

Keightley, Peter D., Urmi Trivedi, Marian Thomson, Fiona Oliver, Sujai Kumar, and Mark L. Blaxter. 2009. “Analysis of the Genome Sequences of Three Drosophila Melanogaster Spontaneous Mutation Accumulation Lines.” *Genome Research* 19 (7): 1195–1201. https://doi.org/10.1101/gr.091231.109.

Kendall, David G. 1948. “On the Generalized" Birth-and-Death" Process.” *The Annals of Mathematical Statistics* 19 (1): 1–15.

Kikuno, Reiko, Hidenori Hayashida, and Takashi Miyata. 1985. “Rapid Rate of Rodent Evolution.” *Proceedings of the Japan Academy* 61 (4): 153–56.

Kilpatrick, A Marm, and Sarah E Randolph. 2012. “Drivers, Dynamics, and Control of Emerging Vector-Borne Zoonotic Diseases.” *The Lancet* 380 (9857): 1946–55. https://doi.org/10.1016/S0140-6736(12)61151-9.

Kimura, Mooto. 1987. “Molecular Evolutionary Clock and the Neutral Theory.” *Journal of Molecular Evolution* 26 (1): 24–33. https://doi.org/10.1007/s00239-009-9284-8.

Kimura, Motoo. 1968. “Evolutionary Rate at the Molecular Level.” *Nature*, 624–626. https://doi.org/10.1006/rwgn.2001.0432.

Kimura, Motoo. 1969. “The Rate of Molecular Evolution Considered from the Standpoint of Population Genetics.” *Proceedings of the National Academy of Sciences* 63 (4): 1181 LP – 1188. https://doi.org/10.1073/pnas.63.4.1181.

Kimura, Motoo, and Tomoko Ohta. 1971. “On the Rate of Molecular Evolution.”

Journal of Molecular Evolution 1 (1): 1–17. https://doi.org/10.1007/BF01659390.

Kingan, Sarah B., Haynes Heaton, Juliana Cudini, Christine C. Lambert, Primo Baybayan, Brendan D. Galvin, Richard Durbin, Jonas Korlach, and Mara K.N. Lawniczak. 2019. "A High-Quality de Novo Genome Assembly from a Single Mosquito Using Pacbio Sequencing." *Genes* 10 (1). https://doi.org/10.3390/genes10010062.

Kirya, B. G., L. G. Mukwaya, and S. D.K. Sempala. 1977. "A Yellow Fever Epizootic in Zika Forest, Uganda, during 1972: Part 1: Virus Isolation and Sentinel Monkeys." *Transactions of the Royal Society of Tropical Medicine and Hygiene* 71 (3): 254–60. https://doi.org/10.1016/0035-9203(77)90020-7.

Kirya, B. G., and N. O. Okia. 1977. "A Yellow Fever Epizootic in Zika Forest, Uganda, during 1972: Part 2: Monkey Serology." *Transactions of the Royal Society of Tropical Medicine and Hygiene* 71 (4): 300–303. https://doi.org/10.1016/0035-9203(77)90104-3.

Kishino, Hirohisa, Jeffrey L Thorne, and William J Bruno. 2001. "Performance of a Divergence Time Estimation Method under a Probabilistic Model of Rate Evolution." *Molecular Biology and Evolution* 18 (3): 352–61.

Klobučar, A, I Lipovac, N Žagar, S Mitrović-Hamzić, V Tešić, T Vilibić-Čavlek, and E Merdić. 2019. "First Record and Spreading of the Invasive Mosquito Aedes Japonicus Japonicus (Theobald, 1901) in Croatia." *Medical and Veterinary Entomology* 33 (1): 171–76.

Kohne, E. David. 1970. "Evolution of Higher-Organism DNA." *Quarterly Reviews of Biophysics* 3 (3): 327–75. https://www.cambridge.org/core/journals/quarterly-reviews-of-biophysics/article/evolution-of-higherorganism-dna/619FCA38AF4B9BC51AF76429F344CD26.

Kooi, Casper J. van der, and Jeff Ollerton. 2020. "The Origins of Flowering Plants and Pollinators." *Science*. American Association for the Advancement of Science. https://doi.org/10.1126/science.aay3662.

Kraemer, Moritz U.G., Robert C. Reiner, Oliver J. Brady, Jane P. Messina, Marius Gilbert, David M. Pigott, Dingdong Yi, et al. 2019. "Past and Future Spread of the Arbovirus Vectors Aedes Aegypti and Aedes Albopictus." *Nature Microbiology* 4 (5): 854–63. https://doi.org/10.1038/s41564-019-0376-y.

Kraemer, Moritz U.G., Marianne E Sinka, Kirsten A Duda, Adrian Q.N. Mylne, Freya M Shearer, Christopher M Barker, Chester G Moore, et al. 2015. "The Global Distribution of the Arbovirus Vectors Aedes Aegypti and Ae. Albopictus." *ELife* 4 (JUNE2015): 1–18. https://doi.org/10.7554/eLife.08347.

Kück, Patrick, and Karen Meusemann. 2010. "FASconCAT: Convenient Handling of Data Matrices." *Molecular Phylogenetics and Evolution* 56 (3):

1115–18. https://doi.org/10.1016/j.ympev.2010.04.024.

Kunkel, T.A. 2009. "Evolving Views of DNA Replication (In) Fidelity OVERLAPPING FUNCTIONS." *Cold Spring Harbor Laboratory Press* LXXIV: 1–11.

Kuno, Goro. 2020. "The Absence of Yellow Fever in Asia: History, Hypotheses, Vector Dispersal, Possibility of Yf in Asia, and Other Enigmas." *Viruses* 12 (12). https://doi.org/10.3390/v12121349.

Lanciotti, Robert S., Olga L. Kosoy, Janeen J. Laven, Jason O. Velez, Amy J. Lambert, Alison J. Johnson, Stephanie M. Stanfield, and Mark R. Duffy. 2008. "Genetic and Serologic Properties of Zika Virus Associated with an Epidemic, Yap State, Micronesia, 2007." *Emerging Infectious Diseases* 14 (8): 1232–39. https://doi.org/10.3201/eid1408.080287.

Lanfear, Robert, Simon Y.W. Ho, T. Jonathan Davies, Angela T. Moles, Lonnie Aarssen, Nathan G. Swenson, Laura Warman, Amy E. Zanne, and Andrew P. Allen. 2013. "Taller Plants Have Lower Rates of Molecular Evolution." *Nature Communications* 4 (May): 1–7. https://doi.org/10.1038/ncomms2836.

Lanfear, Robert, Hanna Kokko, and Adam Eyre-Walker. 2014. "Population Size and the Rate of Evolution." *Trends in Ecology and Evolution* 29 (1): 33–41. https://doi.org/10.1016/j.tree.2013.09.009.

Langley, Charles H., and Walter M. Fitch. 1974. "An Examination of the Constancy of the Rate of Molecular Evolution." *Journal of Molecular Evolution* 3 (3): 161–77. https://doi.org/10.1007/BF01797451.

Larget, Bret, and Donald L. Simon. 1999. "Markov Chain Monte Carlo Algorithms for the Bayesian Analysis of Phylogenetic Trees." *Molecular Biology and Evolution* 16 (6): 750–59. https://doi.org/10.1093/oxfordjournals.molbev.a026160.

Lartillot, Nicolas, Thomas Lepage, and Samuel Blanquart. 2009. "PhyloBayes 3: A Bayesian Software Package for Phylogenetic Reconstruction and Molecular Dating." *Bioinformatics* 25 (17): 2286–88. https://doi.org/10.1093/bioinformatics/btp368.

Lartillot, Nicolas, and Hervé Philippe. 2004. "A Bayesian Mixture Model for Across-Site Heterogeneities in the Amino-Acid Replacement Process." *Molecular Biology and Evolution* 21 (6): 1095–1109. https://doi.org/10.1093/molbev/msh112.

Lartillot, Nicolas, and Hervé Philippe. 2006. "Computing Bayes Factors Using Thermodynamic Integration." *Systematic Biology* 55 (2): 195–207. https://doi.org/10.1080/10635150500433722.

Lau, Sai-Ming, Indra Vythilingam, Jonathan Inbaraj Doss, Shamala Devi Sekaran, Tock H Chua, Wan Yusof Wan Sulaiman, Karuthan Chinna,

Yvonne Ai-Lian Lim, and Balan Venugopalan. 2015. "Surveillance of Adult Aedes Mosquitoes in Selangor, Malaysia." *Tropical Medicine & International Health* 20 (10): 1271–80.

Lazzarini, Luca, Luisa Barzon, Felice Foglia, Vinicio Manfrin, Monia Pacenti, Giacomina Pavan, Mario Rassu, et al. 2020. "First Autochthonous Dengue Outbreak in Italy, August 2020." *Eurosurveillance* 25 (36): 2001606.

Leary, Maureen a O, Jonathan I Bloch, John J Flynn, Timothy J Gaudin, Andres Giallombardo, Norberto P Giannini, Suzann L Goldberg, et al. 2013. "The Placental Mammal Ancestor and the Post–K-Pg Radiation of Placentals." *Science* 339 (February): 662–67.

Lee, Doo-Hyung, Brent D Short, Shimat V Joseph, J Christopher Bergh, and Tracy C Leskey. 2013. "Review of the Biology, Ecology, and Management of Halyomorpha Halys (Hemiptera: Pentatomidae) in China, Japan, and the Republic of Korea." *Environmental Entomology* 42 (4): 627–41.

Leftwich, Philip T, Michael Bolton, and Tracey Chapman. 2016. "Evolutionary Biology and Genetic Techniques for Insect Control." *Evolutionary Applications* 9 (1): 212–30.

Leparc-Goffart, Isabelle, Antoine Nougairede, Sylvie Cassadou, Christine Prat, and Xavier De Lamballerie. 2014. "Chikungunya in the Americas." *The Lancet* 383 (9916): 514. https://doi.org/10.1016/S0140-6736(14)60185-9.

Lequime, Sebastian, Albin Fontaine, Meriadeg Ar Gouilh, Isabelle Moltini-Conclois, and Louis Lambrechts. 2016. "Genetic Drift, Purifying Selection and Vector Genotype Shape Dengue Virus Intra-Host Genetic Diversity in Mosquitoes." *PLoS Genetics* 12 (6). https://doi.org/10.1371/JOURNAL.PGEN.1006111.

Lewald, Kyle M, Antoine Abrieux, Derek A Wilson, Yoosook Lee, William R Conner, Felipe Andreazza, Elizabeth H Beers, et al. 2021. "Population Genomics of Drosophila Suzukii Reveal Longitudinal Population Structure and Signals of Migrations in and out of the Continental United States." *G3* 11 (12): jkab343.

Li, F., X. Zhao, M. Li, K. He, C. Huang, Y. Zhou, Z. Li, and J. R. Walters. 2019. "Insect Genomes: Progress and Challenges." *Insect Molecular Biology* 28 (6): 739–58. https://doi.org/10.1111/imb.12599.

Li, Hong Tao, Ting Shuang Yi, Lian Ming Gao, Peng Fei Ma, Ting Zhang, Jun Bo Yang, Matthew A. Gitzendanner, et al. 2019. "Origin of Angiosperms and the Puzzle of the Jurassic Gap." *Nature Plants* 5 (5): 461–70. https://doi.org/10.1038/s41477-019-0421-0.

Li, X Z, H Y Yang, Tao Wang, J G Wang, and H Y Wei. 2019. "Life History and Adult Dynamics of Bactrocera Dorsalis in the Citrus Orchard of Nanchang,

a Subtropical Area from China: Implications for a Control Timeline." *Scienceasia* 45: 212–20.

Linard, B., P. Arribas, C. Andújar, A. Crampton-Platt, and A. P. Vogler. 2016. "Lessons from Genome Skimming of Arthropod-Preserving Ethanol." *Molecular Ecology Resources* 16 (6): 1365–77. https://doi.org/10.1111/1755-0998.12539.

Linard, Benjamin, Alex Crampton-Platt, Conrad P.D.T. Gillett, Martijn J.T.N. Timmermans, and Alfried P. Vogler. 2015. "Metagenome Skimming of Insect Specimen Pools: Potential for Comparative Genomics." *Genome Biology and Evolution* 7 (6): 1474–89. https://doi.org/10.1093/gbe/evv086.

Liu-Helmersson, Jing, Joacim Rocklöv, Macquin Sewe, and Åke Brännström. 2019. "Climate Change May Enable Aedes Aegypti Infestation in Major European Cities by 2100." *Environmental Research* 172 (February): 693–99. https://doi.org/10.1016/j.envres.2019.02.026.

Liu, Haoxuan, and Jianzhi Zhang. 2019. "Yeast Spontaneous Mutation Rate and Spectrum Vary with Environment." *Current Biology* 29 (10): 1584-1591.e3. https://doi.org/10.1016/j.cub.2019.03.054.

Liu, Zhong Yu, Wei Feng Shi, and Cheng Feng Qin. 2019. "The Evolution of Zika Virus from Asia to the Americas." *Nature Reviews Microbiology* 17 (3): 131–39. https://doi.org/10.1038/s41579-018-0134-9.

Lorenz, Camila, João M.P. Alves, Peter G. Foster, Lincoln Suesdek, and Maria Anice M. Sallum. 2021. "Phylogeny and Temporal Diversification of Mosquitoes (Diptera: Culicidae) with an Emphasis on the Neotropical Fauna." *Systematic Entomology* 46 (4): 798–811. https://doi.org/10.1111/syen.12489.

Lorenz, Camila, João Marcelo Pereira Alves, Peter Gordon Foster, Maria Anice Mureb Sallum, and Lincoln Suesdek. 2019. "First Record of Translocation in Culicidae (Diptera) Mitogenomes: Evidence from the Tribe Sabethini." *BMC Genomics* 20 (1): 1–8.

Lumsden, William H R. 1955. "An Epidemic of Virus Disease in Southern Province, Tanganyika Territory, in 1952--1953 II. General Description and Epidemiology." *Transactions of the Royal Society of Tropical Medicine and Hygiene* 49 (1): 33–57.

Lynch, Michael, Matthew S. Ackerman, Jean Francois Gout, Hongan Long, Way Sung, W. Kelley Thomas, and Patricia L. Foster. 2016. "Genetic Drift, Selection and the Evolution of the Mutation Rate." *Nature Reviews Genetics* 17 (11): 704–14. https://doi.org/10.1038/nrg.2016.104.

MacNamara, F. N. 1954. "Zika Virus: A Report on Three Cases of Human Infection during an Epidemic of Jaundice in Nigeria." *Transactions of the*

Royal Society of Tropical Medicine and Hygiene 48 (2): 139–45. https://doi.org/10.1016/0035-9203(54)90006-1.

Magallón, Susana, and Amanda Castillo. 2009. "Angiosperm Diversification through Time." *American Journal of Botany* 96 (1): 349–65.

Maistrello, Lara, Elena Costi, Stefano Caruso, Giacomo Vaccari, Paolo Bortolotti, Roberta Nannini, Luca Casoli, Anselmo Montermini, Massimo Bariselli, and Roberto Guidetti. 2016. "Halyomorpha Halys in Italy: First Results of Field Monitoring in Fruit Orchards."

Malé, Pierre Jean G., Léa Bardon, Guillaume Besnard, Eric Coissac, Frédéric Delsuc, Julien Engel, Emeline Lhuillier, Caroline Scotti-Saintagne, Alexandra Tinaut, and Jérôme Chave. 2014. "Genome Skimming by Shotgun Sequencing Helps Resolve the Phylogeny of a Pantropical Tree Family." *Molecular Ecology Resources* 14 (5): 966–75. https://doi.org/10.1111/1755-0998.12246.

Manchanda, Nancy, John L. Portwood, Margaret R. Woodhouse, Arun S. Seetharam, Carolyn J. Lawrence-Dill, Carson M. Andorf, and Matthew B. Hufford. 2020. "GenomeQC: A Quality Assessment Tool for Genome Assemblies and Gene Structure Annotations." *BMC Genomics* 21 (1): 1–9. https://doi.org/10.1186/s12864-020-6568-2.

Manni, Mosè, Carmela R. Guglielmino, Francesca Scolari, Anubis Vega-Rúa, Anna Bella Failloux, Pradya Somboon, Antonella Lisa, et al. 2017. "Genetic Evidence for a Worldwide Chaotic Dispersion Pattern of the Arbovirus Vector, Aedes Albopictus." *PLoS Neglected Tropical Diseases* 11 (1). https://doi.org/10.1371/journal.pntd.0005332.

Margoliash, Emanuel. 1963. "Primary Structure and Evolution of Cytochrome C." *Proceedings of the National Academy of Sciences of the United States Of* 50 (1961): 672–79. https://doi.org/10.1073/pnas.50.4.672.

Marini, Giovanni, Daniele Arnoldi, Frederic Baldacchino, Gioia Capelli, Giorgio Guzzetta, Stefano Merler, Fabrizio Montarsi, Annapaola Rizzoli, and Roberto Rosà. 2019. "First Report of the Influence of Temperature on the Bionomics and Population Dynamics of Aedes Koreicus, a New Invasive Alien Species in Europe." *Parasites and Vectors* 12 (1): 1–12. https://doi.org/10.1186/s13071-019-3772-5.

Martin, Darren P., Ben Murrell, Michael Golden, Arjun Khoosal, and Brejnev Muhire. 2015. "RDP4: Detection and Analysis of Recombination Patterns in Virus Genomes." *Virus Evolution* 1 (1): 1–5. https://doi.org/10.1093/ve/vev003.

Massad, E., M. Nascimento Burattini, K. Khan, C. J. Struchiner, F. A.B. Coutinho, and A. Wilder-Smith. 2017. "On the Origin and Timing of Zika

Virus Introduction in Brazil." *Epidemiology and Infection* 145 (11): 2303–12. https://doi.org/10.1017/S0950268817001200.

Matthews, Benjamin J., Olga Dudchenko, Sarah B. Kingan, Sergey Koren, Igor Antoshechkin, Jacob E. Crawford, William J. Glassford, et al. 2018. "Improved Reference Genome of Aedes Aegypti Informs Arbovirus Vector Control." *Nature* 563 (7732): 501–7. https://doi.org/10.1038/s41586-018-0692-z.

Maurer-Stroh, Sebastian, T. Mak, Y. Ng, S. Phuah, R. G. Huber, J. K. Marzinek, D. A. Holdbrook, R. T. Lee, L. Cui, and R. T. Lin. 2016. "South-East Asian Zika Virus Strain Linked to Cluster of Cases in Singapore, August 2016." *Eurosurveillance* 21 (38): 2–5. https://doi.org/10.2807/1560-7917.ES.2016.21.38.30347.

Medlock, J. M., K. M. Hansford, V. Versteirt, B. Cull, H. Kampen, D. Fontenille, G. Hendrickx, H. Zeller, W. Van Bortel, and F. Schaffner. 2015. "An Entomological Review of Invasive Mosquitoes in Europe." *Bulletin of Entomological Research* 105 (6): 637–63. https://doi.org/10.1017/S0007485315000103.

Medlock, Jolyon M, Kayleigh M Hansford, Francis Schaffner, Veerle Versteirt, Guy Hendrickx, Herve Zeller, and Wim Van Bortel. 2012. "A Review of the Invasive Mosquitoes in Europe: Ecology, Public Health Risks, and Control Options." *Vector-Borne and Zoonotic Diseases* 12 (6): 435–47. https://doi.org/10.1089/vbz.2011.0814.

Medlock, Jolyon M, Kayleigh M Hansford, Alexander G C Vaux, Ben Cull, Emma Gillingham, and Steve Leach. 2018. "Assessment of the Public Health Threats Posed by Vector-Borne Disease in the United Kingdom (UK)." *International Journal of Environmental Research and Public Health* 15 (10): 2145.

Medlock, Jolyon M, and Steve A Leach. 2015. "Effect of Climate Change on Vector-Borne Disease Risk in the UK." *The Lancet Infectious Diseases* 15 (6): 721–30.

Medlock, Jolyon M, Alexander G C Vaux, Benjamin Cull, Francis Schaffner, Emma Gillingham, Valentin Pfluger, and Steve Leach. 2017. "Detection of the Invasive Mosquito Species Aedes Albopictus in Southern England." *The Lancet Infectious Diseases* 17 (2): 140.

Menardo, Fabrizio, Sebastian Duchêne, Daniela Brites, and Sebastien Gagneux. 2019. "The Molecular Clock of Mycobacterium Tuberculosis." *PLoS Pathogens* 15 (9): 1–24. https://doi.org/10.1371/journal.ppat.1008067.

Meng, Guanliang, Yiyuan Li, Chentao Yang, and Shanlin Liu. 2019. "MitoZ: A Toolkit for Animal Mitochondrial Genome Assembly, Annotation and

Visualization." *Nucleic Acids Research* 47 (11). https://doi.org/10.1093/nar/gkz173.

Metsky, Hayden C., Christian B. Matranga, Shirlee Wohl, Stephen F. Schaffner, Catherine A. Freije, Sarah M. Winnicki, Kendra West, et al. 2017. "Zika Virus Evolution and Spread in the Americas." *Nature* 546 (7658): 411–15. https://doi.org/10.1038/nature22402.

Minh, Bui Quang, Minh Anh Thi Nguyen, and Arndt Von Haeseler. 2013. "Ultrafast Approximation for Phylogenetic Bootstrap." *Molecular Biology and Evolution* 30 (5): 1188–95. https://doi.org/10.1093/molbev/mst024.

Misof, Bernhard, Shanlin Liu, Karen Meusemann, Ralph S. Peters, Alexander Donath, Christoph Mayer, Paul B. Frandsen, et al. 2014. "Phylogenomics Resolves the Timing and Pattern of Insect Evolution." *Science* 346 (6210): 763–67. https://doi.org/10.1126/science.1257570.

Mlakar, Jernej, Misa Korva, Nataša Tul, Mara Popović, Mateja Poljšak-Prijatelj, Jerica Mraz, Marko Kolenc, et al. 2016. "Zika Virus Associated with Microcephaly." *New England Journal of Medicine* 374 (10): 951–58. https://doi.org/10.1056/nejmoa1600651.

Möhlmann, Tim W R, Chantal B F Vogels, Giel P Göertz, Gorben P Pijlman, Cajo J F Ter Braak, Dennis E Te Beest, Marc Hendriks, et al. 2020. "Impact of Gut Bacteria on the Infection and Transmission of Pathogenic Arboviruses by Biting Midges and Mosquitoes." *Microbial Ecology* 80: 703–17.

Moin-Vaziri, Vahideh, Remi N. Charrel, Mehdi Badakhshan, Xavier De Lamballerie, Nourina Rahbarian, Mulood Mohammadi Bavani, and Shahyad Azari-Hamidian. 2019. "A Molecular Screening of Mosquitoes (Diptera: Culicidae) for Flaviviruses in a Focus of West Nile Virus Infection in Northern Iran." *Journal of Arthropod-Borne Diseases* 13 (4): 391–98. https://doi.org/10.18502/jad.v13i4.2235.

Molaei, Goudarz, Ary Farajollahi, Jamesina J Scott, Randy Gaugler, and Theodore G Andreadis. 2009. "Human Bloodfeeding by the Recently Introduced Mosquito, Aedes Japonicus Japonicus, and Public Health Implications." *Journal of the American Mosquito Control Association* 25 (2): 210–14.

Montarsi, Fabrizio, Andrea Drago, Simone Martini, Mattia Calzolari, Francesco De Filippo, Alessandro Bianchi, Matteo Mazzucato, et al. 2015. "Current Distribution of the Invasive Mosquito Species, Aedes Koreicus [Hulecoeteomyia Koreica] in Northern Italy." *Parasites and Vectors* 8 (1): 1–5. https://doi.org/10.1186/s13071-015-1208-4.

Montarsi, Fabrizio, Simone Martini, Marco Dal Pont, Nicola Delai, Nicola Ferro

Milone, Matteo Mazzucato, Fabio Soppelsa, et al. 2013. “Distribution and Habitat Characterization of the Recently Introduced Invasive Mosquito Aedes Koreicus [Hulecoeteomyia Koreica], a New Potential Vector and Pest in North-Eastern Italy.” *Parasites & Vectors* 6 (1): 1–10.

Montarsi, Fabrizio, Simone Martini, Alice Michelutti, Graziana Da Rold, Matteo Mazzucato, Davide Qualizza, Domenico Di Gennaro, et al. 2019. “The Invasive Mosquito Aedes Japonicus Japonicus Is Spreading in Northeastern Italy.” *Parasites and Vectors* 12 (1): 1–8. https://doi.org/10.1186/s13071-019-3387-x.

Mooers, Arne, and Paul H. Harvey. 1994. “Metabolic Rate, Generation Time, and the Rate of Molecular Evolution in Birds.” *Molecular Phylogenetics and Evolution*. https://doi.org/10.1006/mpev.1994.1040.

Mooney, Harold A, and Elsa E Cleland. 2001. “The Evolutionary Impact of Invasive Species.” *Proceedings of the National Academy of Sciences* 98 (10): 5446–51.

Moran, Emily V., and Jake M. Alexander. 2014. “Evolutionary Responses to Global Change: Lessons from Invasive Species.” *Ecology Letters* 17 (5): 637–49. https://doi.org/10.1111/ele.12262.

Morgan, Gregory J. 1998. “Emile Zuckerkandl, Linus Pauling, and the Molecular Evolutionary Clock, 1959-1965.” *Journal of the History of Biology*, 155–78.

Muñoz, Joaqu\'\in, Santiago Ruiz, Ramón Soriguer, Miguel Alcaide, Duarte S Viana, David Roiz, Ana Vazquez, and Jordi Figuerola. 2012. “Feeding Patterns of Potential West Nile Virus Vectors in South-West Spain.” *PloS One* 7 (6): e39549.

Musso, Didier. 2015. “Zika Virus Transmission from French Polynesia to Brazil.” *Emerging Infectious Diseases* 21 (10): 1887–89. https://doi.org/10.3201/eid2110.151125.

Narayanan Kutty, Sujatha, Wing Hing Wong, Karen Meusemann, Rudolf Meier, and Peter S. Cranston. 2018. “A Phylogenomic Analysis of Culicomorpha (Diptera) Resolves the Relationships among the Eight Constituent Families.” *Systematic Entomology* 43 (3): 434–46. https://doi.org/10.1111/syen.12285.

Neafsey, Daniel E, Robert M Waterhouse, Mohammad R Abai, Sergey S Aganezov, Max A Alekseyev, James E Allen, James Amon, et al. 2015. “Highly Evolvable Malaria Vectors: The Genomes of 16 Anopheles Mosquitoes.” *Science* 347 (6217): 1258522.

Near, Thomas J., Ron I. Eytan, Alex Dornburg, Kristen L. Kuhn, Jon A. Moore, Matthew P. Davis, Peter C. Wainwright, Matt Friedman, and W. Leo Smith.

2012. “Resolution of Ray-Finned Fish Phylogeny and Timing of Diversification.” *Proceedings of the National Academy of Sciences of the United States of America* 109 (34): 13698–703. https://doi.org/10.1073/pnas.1206625109.

Negri, Agata, Irene Arnoldi, Matteo Brilli, Claudio Bandi, Paolo Gabrieli, and Sara Epis. 2021. “Evidence for the Spread of the Alien Species Aedes Koreicus in the Lombardy Region, Italy.” *Parasites and Vectors* 14 (1): 4–9. https://doi.org/10.1186/s13071-021-05031-7.

Newton, Michael A, and Adrian E Raftery. 1994. “Approximate Bayesian Inference with the Weighted Likelihood Bootstrap.” *Journal of the Royal Statistical Society: Series B (Methodological)* 56 (1): 3–26.

Nguyen, Lam Tung, Heiko A. Schmidt, Arndt Von Haeseler, and Bui Quang Minh. 2015. “IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies.” *Molecular Biology and Evolution* 32 (1): 268–74. https://doi.org/10.1093/molbev/msu300.

Norrbom, Allen L. 1994. “New Genera of Tephritidae (Diptera) from Brazil and Dominican Amber, with Phylogenetic Analysis of the Tribe Ortalotrypetini.” *Insecta Mundi*, 289.

Nunney, Leonard. 1999. “Lineage Selection and the Evolution of Multistage Carcinogenesis.” *Proceedings of the Royal Society B: Biological Sciences* 266 (1418): 493–98. https://doi.org/10.1098/rspb.1999.0664.

Nunney, Leonard. 2018. “Size Matters: Height, Cell Number and a Person’s Risk of Cancer.” *Proceedings of the Royal Society B: Biological Sciences* 285 (1889). https://doi.org/10.1098/rspb.2018.1743.

Obbard, Darren J., John MacLennan, Kang Wook Kim, Andrew Rambaut, Patrick M. O’Grady, and Francis M. Jiggins. 2012. “Estimating Divergence Dates and Substitution Rates in the Drosophila Phylogeny.” *Molecular Biology and Evolution* 29 (11): 3459–73. https://doi.org/10.1093/molbev/mss150.

Ohta, Tomoko, and John H Gillespie. 1996. “Development of Neutral and Nearly Neutral Theories.” *Theoretical Population Biology* 49 (0007): 128–42. http://ac.els-cdn.com/S0040580996900076/1-s2.0-S0040580996900076-main.pdf?_tid=db129526-08e7-11e7-a54c-00000aab0f6b&acdnat=1489517936_86d51a61f7f0b8d31ee3d70efd4f837c.

Ometto, Lino, Alessandro Cestaro, Sukanya Ramasamy, Alberto Grassi, Santosh Revadi, Stefanos Siozios, Marco Moretto, et al. 2013. “Linking Genomics and Ecology to Investigate the Complex Evolution of an Invasive Drosophila Pest.” *Genome Biology and Evolution* 5 (4): 745–57. https://doi.org/10.1093/gbe/evt034.

Padidam, Malla, Stanley Sawyer, and Claude M. Fauquet. 1999. "Possible Emergence of New Geminiviruses by Frequent Recombination." *Virology* 265 (2): 218–25. https://doi.org/10.1006/viro.1999.0056.

Palatini, Umberto, Reem A. Masri, Luciano V. Cosme, Sergey Koren, Françoise Thibaud-Nissen, James K. Biedler, Flavia Krsticevic, et al. 2020. "Improved Reference Genome of the Arboviral Vector Aedes Albopictus." *Genome Biology* 21 (1): 215. https://doi.org/10.1186/s13059-020-02141-w.

Papadopoulou, Anna, Ioannis Anastasiou, and Alfried P. Vogler. 2010. "Revisiting the Insect Mitochondrial Molecular Clock: The Mid-Aegean Trench Calibration." *Molecular Biology and Evolution* 27 (7): 1659–72. https://doi.org/10.1093/molbev/msq051.

Patterson, Jessica, Maura Sammon, and Manish Garg. 2016. "Dengue, Zika and Chikungunya: Emerging Arboviruses in the New World." *Western Journal of Emergency Medicine* 17 (6): 671–79. https://doi.org/10.5811/westjem.2016.9.30904.

Pettersson, John H.O., Jon Bohlin, Myrielle Dupont-Rouzeyrol, Ola B. Brynildsrud, Kristian Alfsnes, Van Mai Cao-Lormeau, Michael W. Gaunt, et al. 2018. "Re-Visiting the Evolution, Dispersal and Epidemiology of Zika Virus in Asia Article." *Emerging Microbes and Infections* 7 (1). https://doi.org/10.1038/s41426-018-0082-5.

Pettersson, John H.O., Vegard Eldholm, Stephen J. Seligman, Åke Lundkvist, Andrew K. Falconar, Michael W. Gaunt, Didier Musso, et al. 2018. "Erratum for Pettersson et Al., 'How Did Zika Virus Emerge in the Pacific Islands and Latin America?'" *MBio* 9 (2). https://doi.org/10.1128/mBio.00386-18.

Pettersson, John H.O., and Omar Fiz-Palacios. 2014. "Dating the Origin of the Genus Flavivirus in the Light of Beringian Biogeography." *Journal of General Virology* 95 (PART 9): 1969–82. https://doi.org/10.1099/vir.0.065227-0.

Pettis, Jeffery S, Elinor M Lichtenberg, Michael Andree, Jennie Stitzinger, Robyn Rose, and others. 2013. "Crop Pollination Exposes Honey Bees to Pesticides Which Alters Their Susceptibility to the Gut Pathogen Nosema Ceranae." *PloS One* 8 (7): e70182.

Peyton, E L, Scott R Campbell, Thomas M Candeletti, Michael Romanowski, and Wayne J Crans. 1999. "Aedes (Finlaya) Japonicus Japonicus (Theobald), a New Introduction into the United States."

Pfeffer, Martin, and Gerhard Dobler. 2010. "Emergence of Zoonotic Arboviruses by Animal Trade and Migration." *Parasites and Vectors*. https://doi.org/10.1186/1756-3305-3-35.

Phumee, Atchara, Rome Buathong, Rungfar Boonserm, Proawpilart Intayot, Nucharat Aungsananta, Akanitt Jittmittraphap, Yutthana Joyjinda, Supaporn Wacharapluesadee, and Padet Siriyasatien. 2019. "Molecular Epidemiology and Genetic Diversity of Zika Virus from Field-Caught Mosquitoes in Various Regions of Thailand." *Pathogens* 8 (1). https://doi.org/10.3390/pathogens8010030.

Pike, N, W Y S Wang, and A Meats. 2003. "The Likely Fate of Hybrids of Bactrocera Tryoni and Bactrocera Neohumeralis." *Heredity* 90 (5): 365–70.

Pimentel, David, Rodolfo Zuniga, and Doug Morrison. 2005. "Update on the Environmental and Economic Costs Associated with Alien-Invasive Species in the United States." *Ecological Economics* 52 (3 SPEC. ISS.): 273–88. https://doi.org/10.1016/j.ecolecon.2004.10.002.

Poinar, George, Thomas J. Zavortink, and Alex Brown. 2020. "Priscoculex Burmanicus n. Gen. et Sp. (Diptera: Culicidae: Anophelinae) from Mid-Cretaceous Myanmar Amber." *Historical Biology* 32 (9): 1157–62. https://doi.org/10.1080/08912963.2019.1570185.

Pollard, Daniel A, Venky N Iyer, Alan M Moses, and Michael B Eisen. 2006. "Widespread Discordance of Gene Trees with Species Tree in Drosophila: Evidence for Incomplete Lineage Sorting." *PLoS Genetics* 2 (10): e173.

Pombi, Marco, and Fabrizio Montarsi. 2020. "Mosquitoes (Culicidae)." In *Reference Module in Biomedical Sciences*. Elsevier. https://doi.org/10.1016/b978-0-12-818731-9.00019-7.

Posada, David, and Keith A. Crandall. 2002. "The Effect of Recombination on the Accuracy of Phylogeny Estimation." *Journal of Molecular Evolution* 54 (3): 396–402. https://doi.org/10.1007/s00239-001-0034-9.

Prasad, V, C A E Strömberg, A D Leaché, B Samant, R Patnaik, L Tang, D M Mohabey, S Ge, and A Sahni. 2011. "Late Cretaceous Origin of the Rice Tribe Provides Evidence for Early Diversification in Poaceae." *Nature Communications* 2 (1): 1–9.

Prum, Richard O, Jacob S Berv, Alex Dornburg, Daniel J Field, Jeffrey P Townsend, Emily Moriarty Lemmon, and Alan R Lemmon. 2015. "A Comprehensive Phylogeny of Birds (Aves) Using Targeted next-Generation DNA Sequencing." *Nature* 526 (7574): 569–73.

Rai, Karamjit S. 1991. "Aedes Albopictus in the Americas." *Annual Review of Entomology* 36 (1): 459–84.

Ramasamy, Sukanya, Lino Ometto, Cristina M Crava, Santosh Revadi, Rupinder Kaur, David S Horner, Davide Pisani, Teun Dekker, Gianfranco Anfora, and Omar Rota-Stabelli. 2016. "The Evolution of Olfactory Gene Families in Drosophila and the Genomic Basis of Chemical-Ecological

Adaptation in Drosophila Suzukii." *Genome Biology and Evolution* 8 (8): 2297–2311. https://doi.org/10.1093/gbe/evw160.

Rambaut, Andrew, Alexei J. Drummond, Dong Xie, Guy Baele, and Marc A. Suchard. 2018. "Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7." *Systematic Biology* 67 (5): 901–4. https://doi.org/10.1093/sysbio/syy032.

Rannala, Bruce, and Ziheng Yang. 1996. "Probability Distribution of Molecular Evolutionary Trees: A New Method of Phylogenetic Inference." *Journal of Molecular Evolution* 43 (3): 304–11.

Rasmussen, Sonja A., Denise J. Jamieson, Margaret A. Honein, and Lyle R. Petersen. 2016. "Zika Virus and Birth Defects — Reviewing the Evidence for Causality." *New England Journal of Medicine* 374 (20): 1981–87. https://doi.org/10.1056/nejmsr1604338.

Reeves, Lawrence E, Chris J Holderman, Erik M Blosser, Jennifer L. Gillett-Kaufman, Akito Y Kawahara, Phillip E Kaufman, and Nathan D. Burkett-Cadena. 2018. "Identification of Uranotaenia Sapphirina as a Specialist of Annelids Broadens Known Mosquito Host Use Patterns." *Communications Biology* 1 (1): 1–8. https://doi.org/10.1038/s42003-018-0096-5.

Reidenbach, Kyanne R., Shelley Cook, Matthew A. Bertone, Ralph E. Harbach, Brian M. Wiegmann, and Nora J. Besansky. 2009. "Phylogenetic Analysis and Temporal Diversification of Mosquitoes (Diptera: Culicidae) Based on Nuclear Genes and Morphology." *BMC Evolutionary Biology* 9 (1): 1–14. https://doi.org/10.1186/1471-2148-9-298.

Reinert, John F, Ralph E Harbach, and Ian J. Kitching. 2004. "Phylogeny and Classification of Aedini (Diptera: Culicidae), Based on Morphological Characters of All Life Stages." *Zoological Journal of the Linnean Society* 142 (3): 289–368. https://doi.org/10.1111/j.1096-3642.2004.00144.x.

Reinhold, Joanna M., Claudio R. Lazzari, and Chloé Lahondère. 2018. "Effects of the Environmental Temperature on Aedes Aegypti and Aedes Albopictus Mosquitoes: A Review." *Insects* 9 (4). https://doi.org/10.3390/insects9040158.

Reis, Izabel Cristina Dos, Gerusa Gibson, Tania Ayllon, Alessandre de Medeiros Tavares, Josélio Maria Galvão de Araújo, Elias da Silva Monteiro, Amanda Rodrigues Aguiar, et al. 2019. "Entomo-Virological Surveillance Strategy for Dengue, Zika and Chikungunya Arboviruses in Field-Caught Aedes Mosquitoes in an Endemic Urban Area of the Northeast of Brazil." *Acta Tropica* 197: 105061.

Reis, Mario Dos, Philip C.J. Donoghue, and Ziheng Yang. 2016. "Bayesian Molecular Clock Dating of Species Divergences in the Genomics Era."

Nature Reviews Genetics 17 (2): 71–80. https://doi.org/10.1038/nrg.2015.8.

Reisen, W K. 2016. "Update on Journal Policy of Aedine Mosquito Genera and Subgenera." *Journal of Medical Entomology* 53 (2): 249. https://doi.org/10.1093/jme/tjv169.

Reisen, William K, Ying Fang, and Vincent M Martinez. 2014. "Effects of Temperature on the Transmission of West Nile Virus by Culex Tarsalis (Diptera: Culicidae)." *Journal of Medical Entomology* 43 (2): 309–17.

Rezza, Giovanni. 2018. "Chikungunya Is Back in Italy: 2007--2017." *Journal of Travel Medicine* 25 (1): tay004.

Richards, Stephen. 2019. "Arthropod Genome Sequencing and Assembly Strategies." *Insect Genomics*, 1–14.

Richards, Stephen, and Shwetha C. Murali. 2015. "Best Practices in Insect Genome Sequencing: What Works and What Doesn't." *Current Opinion in Insect Science* 7: 1–7. https://doi.org/10.1016/j.cois.2015.02.013.

Richter, Sandy, Francine Schwarz, Lars Hering, Markus Böggemann, and Christoph Bleidorn. 2015. "The Utility of Genome Skimming for Phylogenomic Analyses as Demonstrated for Glycerid Relationships (Annelida, Glyceridae)." *Genome Biology and Evolution* 7 (12): 3443–62. https://doi.org/10.1093/gbe/evv224.

Rossi-Stacconi, Marco Valerio, Rupinder Kaur, Valerio Mazzoni, Lino Ometto, Alberto Grassi, Angela Gottardello, Omar Rota-Stabelli, and Gianfranco Anfora. 2016. "Multiple Lines of Evidence for Reproductive Winter Diapause in the Invasive Pest Drosophila Suzukii: Useful Clues for Control Strategies." *Journal of Pest Science* 89 (3): 689–700.

Rota-Stabelli, Omar, Mark Blaxter, and Gianfranco Anfora. 2013. "Drosophila Suzukii." *Current Biology* 23 (1): R8--R9.

Rota-Stabelli, Omar, Allison C Daley, and Davide Pisani. 2013. "Molecular Timetrees Reveal a Cambrian Colonization of Land and a New Scenario for Ecdysozoan Evolution." *Current Biology* 23 (5): 392–98. https://doi.org/10.1016/j.cub.2013.01.026.

Rota-Stabelli, Omar, Ehsan Kayal, Dianne Gleeson, Jennifer Daub, Jeffrey L Boore, Maximilian J Telford, Davide Pisani, Mark Blaxter, and Dennis V Lavrov. 2010. "Ecdysozoan Mitogenomics: Evidence for a Common Origin of the Legged Invertebrates, the Panarthropoda." *Genome Biology and Evolution* 2 (1): 425–40. https://doi.org/10.1093/gbe/evq030.

Rota-Stabelli, Omar, Lino Ometto, Gabriella Tait, Silvia Ghirotto, Rupinder Kaur, Francesco Drago, Josefa González, Vaughn M Walton, Gianfranco Anfora, and Marco Valerio Rossi-Stacconi. 2020. "Distinct Genotypes and Phenotypes in European and American Strains of Drosophila Suzukii:

Implications for Biology and Management of an Invasive Organism." *Journal of Pest Science* 93 (1): 77–89. https://doi.org/10.1007/s10340-019-01172-y.

Rota-Stabelli, Omar, and Maximilian J Telford. 2008. "A Multi Criterion Approach for the Selection of Optimal Outgroups in Phylogeny: Recovering Some Support for Mandibulata over Myriochelata Using Mitogenomics." *Molecular Phylogenetics and Evolution* 48 (1): 103–11. https://doi.org/10.1016/j.ympev.2008.03.033.

Ruchusatsawat, Kriangsak, Pattara Wongjaroen, Arisara Posanacharoen, Isabel Rodriguez-Barraquer, Somchai Sangkitporn, Derek A.T. Cummings, and Henrik Salje. 2019. "Long-Term Circulation of Zika Virus in Thailand: An Observational Study." *The Lancet Infectious Diseases* 19 (4): 439–46. https://doi.org/10.1016/S1473-3099(18)30718-7.

Russel, Patricio Maturana, Brendon J. Brewer, Steffen Klaere, and Remco R. Bouckaert. 2019. "Model Selection and Parameter Inference in Phylogenetics Using Nested Sampling." *Systematic Biology* 68 (2): 219–33. https://doi.org/10.1093/sysbio/syy050.

Saarela, Jeffery M, Sean V Burke, William P Wysocki, Matthew D Barrett, Lynn G Clark, Joseph M Craine, Paul M Peterson, Robert J Soreng, Maria S Vorontsova, and Melvin R Duvall. 2018. "A 250 Plastome Phylogeny of the Grass Family (Poaceae): Topological Support under Different Data Partitions." *PeerJ* 6: e4299.

Sæther, O A. 2000. "Phylogeny of Culicomorpha (Diptera)." *Systematic Entomology* 25 (2): 223–34.

Salminen, Mika O, Jean K Carr, Donald S Burke, and FRANCINE E McCUTCHAN. 1995. "Identification of Breakpoints in Intergenotypic Recombinants of HIV Type 1 by Bootscanning." *AIDS Research and Human Retroviruses* 11 (11): 1423–25.

Sánchez-Gracia, A, F G Vieira, and J Rozas. 2009. "Molecular Evolution of the Major Chemosensory Gene Families in Insects." *Heredity* 103 (3): 208–16.

Sanderson, Michael J. 1997. "A Nonparametric Approach to Estimating Divergence Times in the Absence of Rate Constancy."

Sanderson, Michael J. 2002. "Estimating Absolute Rates of Molecular Evolution and Divergence Times: A Penalized Likelihood Approach." *Molecular Biology and Evolution* 19 (1): 101–9.

Sardelis, MICHAEL R, and MICHAEL J Turell. 2001. "Ochlerotatus j. Japonicus in Frederick County, Maryland: Discovery, Distribution, and Vector Competence for West Nile Virus." *Journal of the American Mosquito Control Association-Mosquito News* 17 (2): 137–41.

Sayers, Eric W., Mark Cavanaugh, Karen Clark, James Ostell, Kim D. Pruitt, and Ilene Karsch-Mizrachi. 2019. "GenBank." *Nucleic Acids Research* 47 (D1): D94–99. https://doi.org/10.1093/nar/gky989.

Schaffner, F., J. M. Medlock, and W. Van Bortel. 2013. "Public Health Significance of Invasive Mosquitoes in Europe." *Clinical Microbiology and Infection* 19 (8): 685–92. https://doi.org/10.1111/1469-0691.12189.

Schaffner, Francis, Wim Van Bortel, and Marc Coosemans. 2004. "First Record of Aedes (Stegomyia) Albopictus in Belgium." *Journal of the American Mosquito Control Association* 20 (2): 201–3.

Schaffner, Francis, Sébastien Chouin, and Jacques Guilloteau. 2003. "First Record of Ochlerotatus (Finlaya) Japonicus Japonicus (Theobald, 1901) in Metropolitan France." *Journal of the American Mosquito Control Association* 19 (1): 1–5.

Schaffner, Francis, C Kaufmann, D Hegglin, and Alexander Mathis. 2009. "The Invasive Mosquito Aedes Japonicus in Central Europe." *Medical and Veterinary Entomology* 23 (4): 448–51.

Schaffner, Francis, and Alexander Mathis. 2014. "Dengue and Dengue Vectors in the WHO European Region: Past, Present, and Scenarios for the Future." *The Lancet Infectious Diseases* 14 (12): 1271–80.

Schaffner, Francis, Marie Vazeille, Christian Kaufmann, Anna-Bella Failloux, and Alexander Mathis. 2011. "Vector Competence of Aedes Japonicus for Chikungunya and Dengue Viruses." *European Mosquito Bulletin*, no. 29: 141–42.

Scholte, E J, and F Schaffner. 2007. "Waiting for the Tiger: Establishment and Spread of the Aedes Albopictus Mosquito in Europe Emerging Pests and Vector-Borne Diseases in Europe, Vol. 1." Wageningen Academic Publishers, Wageningen.[Google Scholar].

Scott, Thomas W, Priyanie H Amerasinghe, Amy C Morrison, Leslie H Lorenz, Gary G Clark, Daniel Strickman, Pattamaporn Kittayapong, and John D Edman. 2000. "Longitudinal Studies of Aedes Aegypti (Diptera: Culicidae) in Thailand and Puerto Rico: Blood Feeding Frequency." *Journal of Medical Entomology* 37 (1): 89–101.

Seidel, B, D Duh, N Nowotny, and F Allerberger. 2012. "First Record of the Mosquitoes Aedes (Ochlerotatus) Japonicus Japonicus (Theobald, 1901) in Austria and Slovenia 2011 and for Aedes (Stegomyia) Albopictus (Skuse, 1895) in Austria." *Entomol Zeitschr* 122: 223–26.

Seidel, Bernhard, Fabrizio Montarsi, Hartwig P Huemer, Alexander Indra, Gioia Capelli, Franz Allerberger, and Norbert Nowotny. 2016. "First Record of the Asian Bush Mosquito, Aedes Japonicus Japonicus, in Italy: Invasion

from an Established Austrian Population." *Parasites and Vectors* 9 (1): 284. https://doi.org/10.1186/s13071-016-1566-6.

Seidel, Bernhard, Norbert Nowotny, Tamás Bakonyi, Franz Allerberger, and Francis Schaffner. 2016. "Spread of Aedes Japonicus Japonicus (Theobald, 1901) in Austria, 2011--2015, and First Records of the Subspecies for Hungary, 2012, and the Principality of Liechtenstein, 2015." *Parasites & Vectors* 9 (1): 1–7.

Silva, Alexandre Freitas da, Laís Ceschini Machado, Marcia Bicudo de Paula, Carla Júlia da Silva Pessoa Vieira, Roberta Vieira de Morais Bronzoni, Maria Alice Varjal de Melo Santos, and Gabriel Luz Wallau. 2020. "Culicidae Evolutionary History Focusing on the Culicinae Subfamily Based on Mitochondrial Phylogenomics." *Scientific Reports* 10 (1): 1–14. https://doi.org/10.1038/s41598-020-74883-3.

Silva, Natalia Ingrid Oliveira, Lívia Sacchetto, Izabela Maurício De Rezende, Giliane De Souza Trindade, Angelle Desiree Labeaud, Benoit De Thoisy, and Betânia Paiva Drumond. 2020. "Recent Sylvatic Yellow Fever Virus Transmission in Brazil: The News from an Old Disease." *Virology Journal* 17 (1): 1–12. https://doi.org/10.1186/s12985-019-1277-7.

Silverj, Andrea, and Omar Rota-Stabelli. 2020. "On the Correct Interpretation of Similarity Index in Codon Usage Studies: Comparison with Four Other Metrics and Implications for Zika and West Nile Virus." *Virus Research* 286 (July): 198097. https://doi.org/10.1016/j.virusres.2020.198097.

Simão, Felipe A., Robert M. Waterhouse, Panagiotis Ioannidis, Evgenia V. Kriventseva, and Evgeny M. Zdobnov. 2015. "BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs." *Bioinformatics* 31 (19): 3210–12. https://doi.org/10.1093/bioinformatics/btv351.

Simmons, Mark P. 2012. "Misleading Results of Likelihood-Based Phylogenetic Analyses in the Presence of Missing Data." *Cladistics* 28 (2): 208–22. https://doi.org/10.1111/j.1096-0031.2011.00375.x.

Simon-Loriere, Etienne, and Edward C. Holmes. 2011. "Why Do RNA Viruses Recombine?" *Nature Reviews Microbiology* 9 (8): 617–26. https://doi.org/10.1038/nrmicro2614.

Smith, Katherine F, Dov F Sax, Steven D Gaines, Vanina Guernier, and Jean-François Guégan. 2007. "Globalization of Human Infectious Disease." *Ecology* 88 (8): 1903–10.

Smith, Let\'\icia B, Shinji Kasai, and Jeffrey G Scott. 2016. "Pyrethroid Resistance in Aedes Aegypti and Aedes Albopictus: Important Mosquito Vectors of Human Diseases." *Pesticide Biochemistry and Physiology* 133: 1–12.

Smith, Stephen A, Jeremy M Beaulieu, and Michael J Donoghue. 2010. "An Uncorrelated Relaxed-Clock Analysis Suggests an Earlier Origin for Flowering Plants." *Proceedings of the National Academy of Sciences* 107 (13): 5897–5902.

Smitz, Nathalie, Katrien De Wolf, Isra Deblauwe, Helge Kampen, Francis Schaffner, Jacobus De Witte, Anna Schneider, et al. 2021. "Population Genetic Structure of the Asian Bush Mosquito, Aedes Japonicus (Diptera, Culicidae), in Belgium Suggests Multiple Introductions." *Parasites and Vectors* 14 (1): 1–12. https://doi.org/10.1186/s13071-021-04676-8.

Sniegowski, Paul D., Philip J. Gerrish, Toby Johnson, and Aaron Shaver. 2000. "The Evolution of Mutation Rates: Separating Causes from Consequences." *BioEssays* 22 (12): 1057–66. https://doi.org/10.1002/1521-1878(200012)22:12<1057::AID-BIES3>3.0.CO;2-W.

Snir, Sagi. 2014. "On the Number of Genomic Pacemakers: A Geometric Approach." *Algorithms for Molecular Biology* 9 (26): 1–11. https://doi.org/10.1186/s13015-014-0026-0.

Soares, André E R, and Carlos G Schrago. 2015. "The Influence of Taxon Sampling on Bayesian Divergence Time Inference under Scenarios of Rate Heterogeneity among Lineages." *Journal of Theoretical Biology* 364: 31–39.

Soghigian, John, Theodore G. Andreadis, and Todd P. Livdahl. 2017. "From Ground Pools to Treeholes: Convergent Evolution of Habitat and Phenotype in Aedes Mosquitoes." *BMC Evolutionary Biology* 17 (1): 1–13. https://doi.org/10.1186/s12862-017-1092-y.

Spasojevic, Tamara, Gavin R Broad, Ilari E Sääksjärvi, Martin Schwarz, Masato Ito, Stanislav Korenko, and Seraina Klopfstein. 2021. "Mind the Outgroup and Bare Branches in Total-Evidence Dating: A Case Study of Pimpliform Darwin Wasps (Hymenoptera, Ichneumonidae)." *Systematic Biology* 70 (2): 322–39.

Stahl, Judith, Francesco Tortorici, Marianna Pontini, Marie Claude Bon, Kim Hoelmer, Cristina Marazzi, Luciana Tavella, and Tim Haye. 2019. "First Discovery of Adventive Populations of Trissolcus Japonicus in Europe." *Journal of Pest Science* 92 (2): 371–79. https://doi.org/10.1007/s10340-018-1061-2.

Stamatakis, Alexandros. 2014. "RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies." *Bioinformatics* 30 (9): 1312–13. https://doi.org/10.1093/bioinformatics/btu033.

Stebbins, G. L., and R. C. Lewontin. 1971. *COMPARATIVE EVOLUTION AT THE LEVELS OF MOLECULES, ORGANISMS, AND POPULATIONS. Darwinian, Neo-Darwinian, and Non-Darwinian Evolution*,. University of

California Press. https://doi.org/10.1525/9780520313897-002/HTML.

Steinbrink, Antje, Sina Zotzmann, Sarah Cunze, and Sven Klimpel. 2019. "Aedes Koreicus—a New Member of the Genus Aedes Establishing in Germany?" *Parasitology Research* 118 (3): 1073–76. https://doi.org/10.1007/s00436-019-06232-x.

Stephens, A E A, Darren J Kriticos, and Agathe Leriche. 2007. "The Current and Future Potential Geographical Distribution of the Oriental Fruit Fly, Bactrocera Dorsalis (Diptera: Tephritidae)." *Bulletin of Entomological Research* 97 (4): 369–78.

Suter, Tobias, Eleonora Flacio, Begoña Feijoó Fariña, Lukas Engeler, Mauro Tonolla, and Pie Müller. 2015. "First Report of the Invasive Mosquito Species Aedes Koreicus in the Swiss-Italian Border Region." *Parasites & Vectors* 8 (1): 1–4.

T. Ohta. 1973. "Slightly Deleterious Mutant Substitutions in Evolution." *Nature* 246: 96–98.

Talamas, Elijah J, Lubomir Masner, and Norman F Johnson. 2013. "Systematics of Trichoteleia Kieffer and Paridris Kieffer (Hymenoptera, Platygastroidea, Platygastridae)." *Journal of Hymenoptera Research*, no. 34: 1–80.

Tamura, Koichiro, and Masatoshi Nei. 1993. "Estimation of the Number of Nucleotide Substitutions in the Control Region of Mitochondrial DNA in Humans and Chimpanzees." *Molecular Biology and Evolution* 10 (3): 512–26.

Tancredi, Alessandra, Davide Papandrea, Michele Marconcini, Rebeca Carballar-Lejarazu, Mauricio Casas-Martinez, Eugenia Lo, Xiao-Guang Chen, Anna R Malacrida, and Mariangela Bonizzoni. 2020. "Tracing Temporal and Geographic Distribution of Resistance to Pyrethroids in the Arboviral Vector Aedes Albopictus." *PLoS Neglected Tropical Diseases* 14 (6): e0008350.

Taucher, Christian, Angelika Berger, and Christian W. Mandl. 2010. " A Trans - Complementing Recombination Trap Demonstrates a Low Propensity of Flaviviruses for Intermolecular Recombination ." *Journal of Virology* 84 (1): 599–611. https://doi.org/10.1128/jvi.01063-09.

Tavaré, Simon, and others. 1986. "Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences." *Lectures on Mathematics in the Life Sciences* 17 (2): 57–86.

Taylor, Mark J, Achim Hoerauf, and Moses Bockarie. 2010. "Lymphatic Filariasis and Onchocerciasis." *The Lancet* 376 (9747): 1175–85. https://doi.org/10.1016/S0140-6736(10)60586-7.

Tett, Adrian, Kun D. Huang, Francesco Asnicar, Hannah Fehlner-Peach, Edoardo Pasolli, Nicolai Karcher, Federica Armanini, et al. 2019. "The

Prevotella Copri Complex Comprises Four Distinct Clades Underrepresented in Westernized Populations." *Cell Host and Microbe* 26 (5): 666-679.e7. https://doi.org/10.1016/j.chom.2019.08.018.

Thawornwattana, Yuttapong, Daniel Dalquen, and Ziheng Yang. 2018. "Coalescent Analysis of Phylogenomic Data Confidently Resolves the Species Relationships in the Anopheles Gambiae Species Complex." *Molecular Biology and Evolution* 35 (10): 2512–27. https://doi.org/10.1093/molbev/msy158.

Theron, C D, A Manrakhan, and Christopher William Weldon. 2017. "Host Use of the Oriental Fruit Fly, Bactrocera Dorsalis (Hendel)(Diptera: Tephritidae), in South Africa." *Journal of Applied Entomology* 141 (10): 810–16.

Thomas, Jessica A., John J. Welch, Robert Lanfear, and Lindell Bromham. 2010. "A Generation Time Effect on the Rate of Molecular Evolution in Invertebrates." *Molecular Biology and Evolution* 27 (5): 1173–80. https://doi.org/10.1093/molbev/msq009.

Thomas, Stephanie Margarete, Ulla Obermayr, Dominik Fischer, Juergen Kreyling, and Carl Beierkuhnlein. 2012. "Low-Temperature Threshold for Egg Survival of a Post-Diapause and Non-Diapause European Aedine Strain, Aedes Albopictus (Diptera: Culicidae)." *Parasites and Vectors* 5 (1): 1–7. https://doi.org/10.1186/1756-3305-5-100.

Thorne, Jeffrey L, Hirohisa Kishino, and Ian S Painter. 1998. "Estimating the Rate of Evolution of the Rate of Molecular Evolution." *Molecular Biology and Evolution* 15 (12): 1647–57.

Thrall, Peter H, John G Oakeshott, Gary Fitt, Simon Southerton, Jeremy J Burdon, Andy Sheppard, Robyn J Russell, Myron Zalucki, Mikko Heino, and R Ford Denison. 2011. "Evolution in Agriculture: The Application of Evolutionary Approaches to the Management of Biotic Interactions in Agro-Ecosystems." *Evolutionary Applications* 4 (2): 200–215.

Toews, David P L, and Alan Brelsford. 2012. "The Biogeography of Mitochondrial and Nuclear Discordance in Animals." *Molecular Ecology* 21 (16): 3907–30.

Tolou, H. J.G., P. Couissinier-Paris, J. P. Durand, V. Mercier, J. J. de Pina, P. de Micco, F. Billoir, R. N. Charrel, and X. de Lamballerie. 2001. "Evidence for Recombination in Natural Populations of Dengue Virus Type 1 Based on the Analysis of Complete Genome Suquences." *Journal of General Virology* 82 (6): 1283–90. https://doi.org/10.1099/0022-1317-82-6-1283.

Tomasello, Danilo, and Patricia Schlagenhauf. 2013. "Chikungunya and Dengue Autochthonous Cases in Europe, 2007-2012." *Travel Medicine and Infectious*

Disease 11 (5): 274–84. https://doi.org/10.1016/j.tmaid.2013.07.006.

Tomori, Oyewale. 2004. *Yellow Fever: The Recurring Plague*. *Critical Reviews in Clinical Laboratory Sciences*. Vol. 41. https://doi.org/10.1080/10408360490497474.

Trifinopoulos, Jana, Lam Tung Nguyen, Arndt von Haeseler, and Bui Quang Minh. 2016. "W-IQ-TREE: A Fast Online Phylogenetic Tool for Maximum Likelihood Analysis." *Nucleic Acids Research* 44 (W1): W232–35. https://doi.org/10.1093/nar/gkw256.

Troczka, Bartek, Christoph T Zimmer, Jan Elias, Corinna Schorn, Chris Bass, T G Emyr Davies, Linda M Field, Martin S Williamson, Russell Slater, and Ralf Nauen. 2012. "Resistance to Diamide Insecticides in Diamondback Moth, Plutella Xylostella (Lepidoptera: Plutellidae) Is Associated with a Mutation in the Membrane-Spanning Domain of the Ryanodine Receptor." *Insect Biochemistry and Molecular Biology* 42 (11): 873–80.

Vargas, Roger I, William A Walsh, Dale Kanehisa, Eric B Jang, and John W Armstrong. 1997. "Demography of Four Hawaiian Fruit Flies (Diptera: Tephritidae) Reared at Five Constant Temperatures." *Annals of the Entomological Society of America* 90 (2): 162–68.

Ventura, Camila V., Mauricio Maia, Vasco Bravo-Filho, Adriana L. Góis, and Rubens Belfort. 2016. "Zika Virus in Brazil and Macular Atrophy in a Child with Microcephaly." *The Lancet* 387 (10015): 228. https://doi.org/10.1016/S0140-6736(16)00006-4.

Versteirt, Veerle, James E Pecor, Dina M Fonseca, Marc Coosemans, and Wim Van Bortel. 2012. "Confirmation of Aedes Koreicus (Diptera: Culicidae) in Belgium and Description of Morphological Differences between Korean and Belgian Specimens Validated by Molecular Identification." *Zootaxa* 3191 (1): 21–32.

Versteirt, Veerle, Francis Schaffner, Claire Garros, Wouter Dekoninck, M Coosemans, and Wim Van Bortel. 2009. "Introduction and Establishment of the Exotic Mosquito Species Aedes Japonicus Japonicus (Diptera: Culicidae) in Belgium." *Journal of Medical Entomology* 46 (6): 1464–67.

Vilaça, Sibelle Torres, Riccardo Piccinno, Omar Rota-Stabelli, Maëva Gabrielli, Andrea Benazzo, Michael Matschiner, Luciano S. Soares, Alan B. Bolten, Karen A. Bjorndal, and Giorgio Bertorelle. 2021. "Divergence and Hybridization in Sea Turtles: Inferences from Genome Data Show Evidence of Ancient Gene Flow between Species." *Molecular Ecology*, no. July: 1–15. https://doi.org/10.1111/mec.16113.

Vitousek, Peter M, Carla M D'Antonio, Lloyd L Loope, and Randy Westbrooks. 1996. "Biological Invasions as Global Environmental Change."

Wahlberg, Niklas, Elisabet Weingartner, Andrew D. Warren, and Sören Nylin. 2009. "Timing Major Conflict between Mitochondrial and Nuclear Genes in Species Relationships of Polygonia Butterflies (Nymphalidae: Nymphalini)." *BMC Evolutionary Biology* 9 (1): 1–16. https://doi.org/10.1186/1471-2148-9-92.

Wang, Lulan, Stephanie G. Valderramos, Aiping Wu, Songying Ouyang, Chunfeng Li, Patricia Brasil, Myrna Bonaldo, et al. 2016. "From Mosquitos to Humans: Genetic Evolution of Zika Virus." *Cell Host and Microbe* 19 (5): 561–65. https://doi.org/10.1016/j.chom.2016.04.006.

Wang, Xiaoming, Tong Liu, Yang Wu, Daibin Zhong, Guofa Zhou, Xinghua Su, Jiabao Xu, et al. 2018. "Bacterial Microbiota Assemblage in Aedes Albopictus Mosquitoes and Its Impacts on Larval Development." *Molecular Ecology* 27 (14): 2972–85.

Waterhouse, Robert M, Mathieu Seppey, Felipe A Simão, and Evgeny M Zdobnov. 2019. "Using BUSCO to Assess Insect Genomic Resources." In *Insect Genomics*, 59–74. Springer.

Weaver, Scott C. 2018. "Prediction and Prevention of Urban Arbovirus Epidemics: A Challenge for the Global Virology Community." *Antiviral Research* 156 (May): 80–84. https://doi.org/10.1016/j.antiviral.2018.06.009.

Weger-Lucarelli, James, Claudia Rückert, Nunya Chotiwan, Chilinh Nguyen, Selene M. Garcia Luna, Joseph R. Fauver, Brian D. Foy, et al. 2016. "Vector Competence of American Mosquitoes for Three Strains of Zika Virus." *PLoS Neglected Tropical Diseases* 10 (10): 1–16. https://doi.org/10.1371/journal.pntd.0005101.

Wei, Yun, Biao Cheng, Guoding Zhu, Danyu Shen, Jiangtao Liang, Cong Wang, Jing Wang, et al. 2017. "Comparative Physical Genome Mapping of Malaria Vectors Anopheles Sinensis and Anopheles Gambiae." *Malaria Journal* 16 (1): 1–10. https://doi.org/10.1186/s12936-017-1888-7.

Werner, Doreen, and Helge Kampen. 2013. "The Further Spread of Aedes Japonicus Japonicus (Diptera, Culicidae) towards Northern Germany." *Parasitology Research* 112 (10): 3665–68.

Werner, Doreen, and Helge Kampen. 2015. "Aedes Albopictus Breeding in Southern Germany, 2014." *Parasitology Research* 114 (3): 831–34.

Werner, Doreen, Dorothee E Zielke, and Helge Kampen. 2016. "First Record of Aedes Koreicus (Diptera: Culicidae) in Germany." *Parasitology Research* 115 (3): 1331–34.

White, Ian M, and Marlene M Elson-Harris. 1992. "Fruit Flies of Economic Significance: Their Identification and Bionomics." *Wallingford: CAB International*, 601.

Whitehorn, James, and Sophie Yacoub. 2019. "Global Warming and Arboviral Infections." *Clinical Medicine* 19 (2): 149.

Wibowo, Marsha C., Zhen Yang, Maxime Borry, Alexander Hübner, Kun D. Huang, Braden T. Tierney, Samuel Zimmerman, et al. 2021. "Reconstruction of Ancient Microbial Genomes from the Human Gut." *Nature* 594 (7862): 234–39. https://doi.org/10.1038/s41586-021-03532-0.

Widdel, Andrea K, Linda J McCuiston, Wayne J Crans, Laura D Kramer, and Dina M Fonseca. 2005. "Finding Needles in the Haystack: Single Copy Microsatellite Loci for Aedes Japonicus (Diptera: Culicidae)." *The American Journal of Tropical Medicine and Hygiene* 73 (4): 744–48.

Wiegmann, Brian M., Michelle D. Trautwein, Isaac S. Winkler, Norman B. Barr, Jung Wook Kim, Christine Lambkin, Matthew A. Bertone, et al. 2011. "Episodic Radiations in the Fly Tree of Life." *Proceedings of the National Academy of Sciences of the United States of America* 108 (14): 5690–95. https://doi.org/10.1073/pnas.1012675108.

Wilkerson, Richard C., Yvonne-Marie Linton, Dina M. Fonseca, Ted R. Schultz, Dana C. Price, and Daniel A. Strickman. 2015. "Making Mosquito Taxonomy Useful: A Stable Classification of Tribe Aedini That Balances Utility with Current Knowledge of Evolutionary Relationships." Edited by Zhong-Jian Liu. *PLOS ONE* 10 (7): e0133602. https://doi.org/10.1371/journal.pone.0133602.

Williams, Ashley B. 2014. "Spontaneous Mutation Rates Come into Focus in Escherichia Coli." *DNA Repair* 24: 73–79.

Wu, C I, and W H Li. 1985. "Evidence for Higher Rates of Nucleotide Substitution in Rodents than in Man." *Proceedings of the National Academy of Sciences* 82 (6): 1741–45. https://doi.org/10.1073/PNAS.82.6.1741.

Wu, Pa, Xi Yu, Penghua Wang, and Gong Cheng. 2019. "Arbovirus Lifecycle in Mosquito: Acquisition, Propagation and Transmission." *Expert Reviews in Molecular Medicine* 21. https://doi.org/10.1017/erm.2018.6.

Xie, Wangang, Paul O. Lewis, Yu Fan, Lynn Kuo, and Ming Hui Chen. 2011. "Improving Marginal Likelihood Estimation for Bayesian Phylogenetic Model Selection." *Systematic Biology* 60 (2): 150–60. https://doi.org/10.1093/sysbio/syq085.

Xu, Hongxing, Xinhai Ye, Yajun Yang, Yi Yang, Yu H Sun, Yang Mei, Shijiao Xiong, et al. 2021. "Comparative Genomics Sheds Light on the Convergent Evolution of Miniaturized Wasps." *Molecular Biology and Evolution*.

Xu, Pengjun, Roberto Feuda, Bin Lu, Haijun Xiao, Robert I. Graham, and Kongming Wu. 2016. "Functional Opsin Retrogene in Nocturnal Moth." *Mobile DNA* 7 (1): 1–6. https://doi.org/10.1186/s13100-016-0074-8.

Yadav, Pragya D., Bharati Malhotra, Gajanan Sapkal, Dimpal A. Nyayanit, Gururaj Deshpande, Nivedita Gupta, Ullas T. Padinjaremattathil, et al. 2019. "Zika Virus Outbreak in Rajasthan, India in 2018 Was Caused by a Virus Endemic to Asia." *Infection, Genetics and Evolution* 69 (December 2018): 199–202. https://doi.org/10.1016/j.meegid.2019.01.026.

Yang, Zhong-Qi, Yan-Xia Yao, Lan-Fen Qiu, and Zhong-Xin Li. 2009. "A New Species of Trissolcus (Hymenoptera: Scelionidae) Parasitizing Eggs of Halyomorpha Halys (Heteroptera: Pentatomidae) in China with Comments on Its Biology." *Annals of the Entomological Society of America* 102 (1): 39–47.

Yang, Ziheng. 1994. "Maximum Likelihood Phylogenetic Estimation from DNA Sequences with Variable Rates over Sites: Approximate Methods." *Journal of Molecular Evolution* 39 (3): 306–14.

Yang, Ziheng, Rasmus Nielsen, and Masami Hasegawa. 1998. "Models of Amino Acid Substitution and Applications to Mitochondrial Protein Evolution." *Molecular Biology and Evolution* 15 (12): 1600–1611. https://doi.org/10.1093/oxfordjournals.molbev.a025888.

Yang, Ziheng, and Bruce Rannala. 2012. "Molecular Phylogenetics: Principles and Practice." *Nature Reviews Genetics* 13 (5): 303–14. https://doi.org/10.1038/nrg3186.

Yeap, Heng Lin, Siu Fai Lee, Freya Robinson, Roslyn G Mourant, John A Sved, Marianne Frommer, Alexie Papanicolaou, Owain R Edwards, and John G Oakeshott. 2020. "Separating Two Tightly Linked Species-Defining Phenotypes in Bactrocera with Hybrid Recombinant Analysis." *BMC Genetics* 21 (2): 1–16.

Yeates, David K, Brian M Wiegmann, Greg W Courtney, Rudolf Meier, Christine Lambkin, and Thomas Pape. 2007. "Phylogeny and Systematics of Diptera: Two Decades of Progress and Prospects." *Zootaxa* 1668 (1): 565–90.

Yoder, Anne D, and Ziheng Yang. 2000. "Estimation of Primate Speciation Dates Using Local Molecular Clocks." *Molecular Biology and Evolution* 17 (7): 1081–90.

Young, Paul R. 2018. "Arboviruses: A Family on the Move." *Dengue and Zika: Control and Antiviral Treatment Strategies*, 1–10.

Yule, George Udny. 1925. "II.—A Mathematical Theory of Evolution, Based on the Conclusions of Dr. JC Willis, FR S." *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character* 213 (402–410): 21–87.

Zadra, Nicola, Annapaola Rizzoli, and Omar Rota-Stabelli. 2021. "Chronological Incongruences between Mitochondrial and Nuclear Phylogenies of Aedes

Mosquitoes." *Life* 11 (3): 1–17. https://doi.org/10.3390/life11030181.

Zanluca, Camila, Vanessa Campos Andrade De Melo, Ana Luiza Pamplona Mosimann, Glauco Igor Viana Dos Santos, Claudia Nunes Duarte dos Santos, and Kleber Luz. 2015. "First Report of Autochthonous Transmission of Zika Virus in Brazil." *Memorias Do Instituto Oswaldo Cruz* 110 (4): 569–72. https://doi.org/10.1590/0074-02760150192.

Zhang, Jinping, Feng Zhang, Tara Gariepy, Peter Mason, Dave Gillespie, Elijah Talamas, and Tim Haye. 2017. "Seasonal Parasitism and Host Specificity of Trissolcus Japonicus in Northern China." *Journal of Pest Science* 90 (4): 1127–41.

Zhang, Qian, Kaiyuan Sun, Matteo Chinazzi, Ana Pastore Y. Piontti, Natalie E. Dean, DIana Patricia Rojas, Stefano Merler, et al. 2017. "Spread of Zika Virus in the Americas." *Proceedings of the National Academy of Sciences of the United States of America* 114 (22): E4334–43. https://doi.org/10.1073/pnas.1620161114.

Zheng, Yuchi, Rui Peng, Masaki Kuro-O, and Xiaomao Zeng. 2011. "Exploring Patterns and Extent of Bias in Estimating Divergence Time from Mitochondrial DNA Sequence Data in a Particular Lineage: A Case Study of Salamanders (Order Caudata)." *Molecular Biology and Evolution* 28 (9): 2521–35. https://doi.org/10.1093/molbev/msr072.

Zhong, Guangfa, Jianhua Geng, How Kin Wong, Zaitian Ma, and Nengyou Wu. 2004. "A Semi-Quantitative Method for the Reconstruction of Eustatic Sea Level History from Seismic Profiles and Its Application to the Southern South China Sea." *Earth and Planetary Science Letters* 223 (3–4): 443–59.

Zimin, Aleksey V., Guillaume Marçais, Daniela Puiu, Michael Roberts, Steven L. Salzberg, and James A. Yorke. 2013. "The MaSuRCA Genome Assembler." *Bioinformatics* 29 (21): 2669–77. https://doi.org/10.1093/bioinformatics/btt476.

Zuckerkandl, Emile, and Linus Pauling. 1965. "Evolutionary Divergence and Convergence in Proteins." *Evolving Genes and Proteins*, 97–166. https://doi.org/10.1016/b978-1-4832-2734-4.50017-6.