Quantitative studies of infectious diseases transmission in populations : from pathogen genetic sequences to individual-level epidemiological data

Maylis Layan

THÈSE DE DOCTORAT DE SORBONNE UNIVERSITÉ

Spécialité Biostatistiques

École Doctorale Pierre Louis de santé publique

Présentée par **Maylis Layan**

Pour obtenir le grade de **Docteure de Sorbonne Université**

Quantitative studies of infectious diseases transmission in populations: from pathogen genetic sequences to individual-level epidemiological data

Études quantitatives de la transmission des maladies infectieuses en population en utilisant des séquences génétiques de pathogènes et des données épidémiologiques individuelles

Thèse présentée et soutenue publiquement le **8 décembre 2022**

Devant le jury composé de:

Mme Lulla Opatowski, Professeure des Universités, Présidente

Mme Morgane Rolland, Directrice de recherche, Rapporteure

M. Samuel Alizon, Directeur de recherche, Rapporteur

Mme Katie Hampson, Professeure des Universités, Examinateur

M. Guy Baele, Directeur de recherche, Examinateur

M. Hervé Bourhy, Professeur de l'Institut Pasteur, Co-directeur de thèse

M. Simon Cauchemez, Directeur de recherche, Directeur de thèse

Sorbonne Université
Bureau d'accueil, inscription des doctorants et base de données
Esc G, 2ème étage
15 rue de l'école de médecine
75270-PARIS CEDEX 06

Tél. Secrétariat : 01 42 34 68 35
Fax : 01 42 34 68 40
Tél. pour les étudiants de A à EL : 01 42 34 69 54
Tél. pour les étudiants de EM à MON : 01 42 34 68 41
Tél. pour les étudiants de MOO à Z : 01 42 34 68 51
E-mail : scolarite.doctorat@upmc.fr

To my dad.
To my mum.
To my brother.

I had diverged, digressed, wandered, and become wild. I didn't embrace the word as my new name because it defined negative aspects of my circumstances or life, but because even in my darkest days - those very days in which I was naming myself - I saw the power of the darkness. Saw that, in fact, I had strayed and that I was a stray and that from the wild places my straying had brought me, I knew things I couldn't have known before.

Cheryl Strayed
Wild: From Lost to Found on the Pacific Crest Trail

Acknowledgments

I sincerely thank the members of my jury, Morgane Rolland, Samuel Alizon, Katie Hampson, Lulla Opatowski, and Guy Baele, for their time, consideration, and commitment.

I would like to thank my supervisors, Hervé and Simon. Thank you for giving me the opportunity to do a PhD in quantitative studies in epidemiology, although this is not my original background, and for trusting me during this long journey.

Many thanks to all the past and present members of the mathematical modelling of infectious diseases unit at Pasteur with whom I shared very joyful moments. My special thanks go to Alessio, Birgit, and Nathanaël. Merci Alessio pour ta patience et tes conseils sur à peu près tous les langages de programmation. Le modèle de simulation des épidémies de rage en Rcpp et le modèle en C++ de transmission du SARS-CoV-2 dans les ménages n'auraient pas vus le jour sans ton aide. Birgit, thanks for giving me the confidence that former biologists can also develop models. Nathanaël, merci pour ton esprit critique et ta spontanéité dans la relecture de mes papiers.

My special thanks go to my Belgian collaborators, Guy Baele and Simon Dellicour. You rapidly made me feel part of the teeny-tiny world of phylodynamics, Guy by encouraging me on the endless simulation project and Simon by implicating me in the Cambodian project. Our long-term collaboration was not only scientifically enriching and fruitful, but also very supportive.

I would like to thank all the researchers implicated in the projects presented in this thesis. Laurent Dacheux, Philippe Lemey, Kirstyn Brunker, Laurence Ma, Cécile Troupin, Philippe Dussart, Véronique Chevalier, James Wood, Sowath Ly, and Veasna Duong for the study of RABV spread in Cambodia. Tal Gonen, Miki Goldenfeld, Lilac Meltzer, Mayan Gilboa, and Gili Regev-Yochay for the evaluation of vaccine effectiveness and the impact of physical distancing on SARS-CoV-2 transmission in Israeli households. Nicola Müller and Nicola De Maio for the implementation and the interpretation of the outputs from MASCOT, MASCOT-GLM, and BASTA in the project on sampling bias and discrete phylogeography. Niel Hens for the simulation study on the impact of heterogeneous contact patterns on the transmission of respiratory diseases in households.

I would like to thank the Evolutionary and Computational Virology team in Leuven for your wholehearted welcome during the COVID era as well as the lovely days at the MCEB conference in Switzerland: Kanika, Nena, Ine, Sam, Barney, and Jade.

Merci aux doctorants de Pasteur pour vos discussions enrichissantes, vos partages d'expérience et votre sincérité: Eve, Salam, Guillem, David, Andrew, Emile (merci d'avoir été présent pendant la rédaction), Simon, Wilfried, Sophie, Lucia, et Camille.

Merci à tou.te.s mes ami.e.s de France et d'ailleurs dont l'amitié m'apporte courage et force. Clara, pour ton amitié dont je ne compte plus les années. Alexandre, Camille, Claire, Cyril (merci pour ton soutien inconditionnel à Capri), Enzo, Flora, Gaétan (merci pour les cours particuliers de maths et de C++), Guillaume, Lucie, Matthieu, Rémi, bientôt dix ans que l'on se connaît et que vous êtes à mes côtés. Katia et Elena, merci pour tous vos conseils et votre soutien. Audrey, merci ex-voisine, je tâcherai de suivre

ton exemple vers la voie de la sagesse yogique. Julien et Antoine, merci pour votre énergie débordante. Lucie, merci grande-marraine pour toutes les activités extra-scolaires. Kanika for generously hosting me in Leuven and sharing so much.

Enfin, merci à mes parents et à mon frère qui me montrent au quotidien le chemin de la qpersévérance et de la résilience.

Scientific output

Published and accepted articles

- **Layan M**, Dellicour S, Baele G, Cauchemez S, Bourhy H. Mathematical modelling and phylodynamics for the study of dog rabies dynamics and control: A scoping review. PLOS Neglected Tropical Diseases. 2021; 15: e0009449. doi:10.1371/journal.pntd.0009449
- **Layan M**, Gilboa M, Gonen T, Goldenfeld M, Meltzer L, Andronico A, Hozé N, Cauchemez S, Regev-Yochay G. Impact of BNT162b2 Vaccination and Isolation on SARS-CoV-2 Transmission in Israeli Households: An Observational Study. American Journal of Epidemiology. 2022; 191: 1224–1234. doi:10.1093/aje/kwac042

Articles submitted to pre-print servers and under review

- **Layan M**, Müller NF, Dellicour S, De Maio N, Bourhy H, Cauchemez S, Baele G. Impact and mitigation of sampling bias to determine viral spread: evaluating discrete phylogeography through CTMC modeling and structured coalescent model approximations. bioRxiv. 2022; 2022.07.07.498932. doi:10.1101/2022.07.07.498932

Scientific conferences

Oral presentations

- Impact and mitigation of sampling bias in discrete phylogeography inference: an application to RABV spread in Morocco, online conference. Molecular Epidemiology and Evolutionary Genetics of Infectious Diseases **November 2021**. Online conference.
- Impact and mitigation of sampling bias in discrete phylogeography inference: an application to RABV spread in Morocco. Epidemics **December 2021**. Online conference.
- Impact and mitigation of sampling bias on discrete phylogeography: from CTMC modeling to structured coalescent model approximations. Mathematical and Computational Evolutionary Biology **June 2022**. Château d'Oex, Switzerland.

Posters

- Impact of BNT162b2 vaccination and isolation on SARS-CoV-2 transmission in Israeli households: an observational study. Epidemics **December 2021**. Online conference.
- Impact of heterogeneous contact patterns between household members on the transmission of viral respiratory infections. Rencontre "Action Coordonnée" Modélisation des maladies infectieuses **Novembre 2022**. Bordeaux, France.

Abstract

Among the methods for the quantitative study of infectious diseases transmission in host populations, molecular epidemiology that reconstructs pathogen phylogenies by using pathogen genetic sequences and mathematical modelling of infectious diseases that fits mechanistic models of disease transmission to epidemiological data such as case counts are of particular interest to epidemiologists. These two approaches rely on different data sources whose availability depends on the setting. They also rely on different concepts and models leading to complementary pictures of disease transmission. The main objective of this thesis is to better understand how viral infectious diseases such as rabies and COVID-19 circulate in host populations using respectively geolocated and timestamped viral genetic sequences and detailed epidemiological data at the individual level. The first part of this thesis focuses on rabies, a neglected tropical zoonosis, that is estimated to cause 59,000 human deaths per year mostly among rural and poor populations in Africa and Asia. Its causing agent, rabies virus (RABV), mainly circulates in domestic dog populations. Despite being a vaccine-preventable disease in both humans and dogs, rabies remains poorly studied and its circulation in dogs poorly understood. First, we reviewed from the literature all mathematical models and molecular epidemiology studies on dog rabies circulation to synthesize the contribution of both approaches to the understanding of rabies dynamics in dogs. Then, we described RABV spread in Cambodia, one of the most affected countries worldwide, using RABV genomes isolated from dogs and Bayesian continuous phylogeography methods. We used Cambodia as a model of endemic circulation of RABV and exemplified how phylogeography can help characterize circulation in such context. We found that introductions from foreign countries are not necessary to sustain transmission in Cambodia. However, these results are conditional on the sampling of the RABV genomes. To further understand how sampling affects Bayesian phylogeography methods, we performed a simulation study where we evaluated the performances of three Bayesian discrete phylogeography algorithms under increasing levels of bias, and tested whether alternative sampling strategies, and integration of incidence data improve the performances of the algorithms under biased sampling conditions. The second part of this thesis concentrates on SARS-CoV-2 transmission at one of the smallest population scale, households. This setting is particularly suitable to detailed follow-up of household members after introduction of a case, and thus, enables to evaluate how susceptibility and infectivity vary between individuals. First, we estimated BNT162b2 vaccine effectiveness against infection and against transmission if infected during the Alpha wave in Israel using a mathematical model of SARS-CoV-2 transmission in partially vaccinated households. We further explored how model misspecification in a context of differing contact patterns between adults and children would impact estimates of relative infectivity and susceptibility of children compared to adults. Overall, this thesis explores how molecular epidemiology and modelling contribute to the understanding of infectious diseases transmission at the population level and highlights the need for data integration.

Keywords: infectious diseases, molecular epidemiology, rabies, household studies, modeling, SARS-CoV-2.

Résumé

Parmi les méthodes pour l'étude quantitative de la transmission des maladies infectieuses dans les populations, les épidémiologistes ont récemment focalisé leur attention sur l'épidémiologie moléculaire qui vise à reconstruire la phylogénie des pathogènes en utilisant leurs séquences génétiques, et la modélisation mathématique des maladies infectieuses qui ajuste des modèles mécanistes de transmission des maladies à des données épidémiologiques telles que le nombre de cas. Ces deux approches se basent sur des données très différentes dont la disponibilité varie selon le contexte. Les concepts et les modèles qu'elles utilisent permettent d'explorer des facettes différentes de la transmission des maladies. L'objectif principal de cette thèse est de mieux comprendre comment les maladies virales comme la rage et la covid-19 circulent dans les populations hôtes en utilisant pour la première des séquences génétiques virales datées et géolocalisées, et pour la deuxième, des données épidémiologiques à l'échelle individuelle. La première partie de cette thèse s'intéresse à la rage, une zoonose tropicale négligée, responsable d'environ 59,000 morts chaque année principalement dans les populations pauvres et rurales d'Afrique et d'Asie. Son agent étiologique, le virus de la rage (RABV), circule principalement dans les populations canines domestiques dont les modes de transmission restent peu étudiés et mal compris malgré l'existence de vaccins efficaces chez l'homme et l'animal. Nous avons tout d'abord synthétisé dans une revue de la littérature l'apport relatif des modèles mathématiques et de l'épidémiologie moléculaire dans la compréhension des dynamiques de la rage chez le chien. Puis, nous avons décrit la circulation endémique de la rage au Cambodge, un des pays les plus affectés, à partir de génomes de la rage isolés chez le chien et analysés avec des méthodes de phylogéographie Bayésienne continue. Nous avons montré que les introductions depuis d'autres pays ne sont pas nécessaires au maintien de la circulation. Toutefois, ces résultats sont conditionnés par l'échantillonnage des génomes. Pour mieux comprendre leurs impacts sur les méthodes de phylogéographie Bayésienne, nous avons entrepris une étude de simulation dans laquelle nous avons comparé les performances de trois algorithmes de phylogéographie discrète face à un échantillonnage plus ou moins biaisé. Nous avons testé des stratégies d'échantillonnage alternatives et intégré des données épidémiologiques afin d'atténuer l'effet potentiel des biais d'échantillonnage sur la performance des trois algorithmes. La deuxième partie de la thèse se concentre sur la transmission du SARS-CoV-2 dans une des plus petites populations, les ménages. Cette configuration est particulièrement adaptée au suivi détaillé de l'ensemble des membres du foyer après l'introduction d'un cas et permet ainsi d'évaluer comment la susceptibilité et l'infectivité varient au niveau individuel. Dans un premier temps, nous avons estimé l'effectivité vaccinale contre l'infection et la transmission si infecté pendant la vague de variant Alpha en Israël grâce à un modèle de transmission dans des ménages partiellement vaccinés. Nous avons ensuite exploré comment l'hétérogénéité de contact dans les ménages, notamment entre les adultes et les enfants, impacte les estimations de l'infectivité et de la susceptibilité relatives des enfants par rapport aux adultes. En conclusion, cette thèse explore les contributions de l'épidémiologie moléculaire et de la modélisation pour la compréhension de la transmission des maladies infectieuses à différentes échelles de population et souligne la nécessité d'intégrer les données génétiques et épidémiologiques.

Mots-clés: maladies infectieuses, épidémiologie moléculaire, rage, études de ménages, modélisation, SARS-CoV-2.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AMR	Antimicrobial resistance
ARACON	Asian Rabies Control Network
BASTA	Bayesian structured coalescent approximation
BEAST	Bayesian evolutionary analysis sampling trees
BF	Bayes factor
BSSVS	Bayesian stochastic search variable selection
CDC	Centers for Disease Control and Prevention
CI	Confidence interval
COVID-19	Coronavirus disease 2019
CrI	Credible interval
Ct	Cycle threshold
CTMC	Continuous-time Markov chain
DALY	Disability-adjusted life years
DFAT	Direct fluorescent antibody test
DTA	Discrete trait analysis
ESS	Effective sample size
FAO	Food and Agriculture Organization of the United Nations
GARC	Global Alliance for Rabies Control
GLM	Generalized linear model
GTR	General-time reversible
HCW	Healthcare worker
HPD	Highest posterior density
IPC	Institut Pasteur du Cambodge
MASCOT	Marginal approximation of the structured coalescent
MCC	Maximum clade credibility
MCMC	Markov chain Monte Carlo
MERS-CoV	Middle East respiratory syndrome coronavirus
NGS	Next-generation sequencing
OIE	World Organization for Animal Health
ORV	Oral rabies vaccination
PCR	Polymerase chain reaction
PRISMA-ScR	Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
R	Reproduction ratio
R_0	Basic reproduction ratio

R_e	Effective reproduction ratio
RABV	Rabies virus
RR	Relative risk
RRW	Relaxed random walk
RT-PCR	Reverse transcription-polymerase chain reaction
SAR	Secondary attack rate
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SIR	Susceptible-Infectious-Recovered
TMRCA	Time to the most recent common ancestor
WIS	Weighted interval score
WHO	World Health Organization
WLDV	Weighted lineage dispersal velocity
WNV	West Nile virus

General introduction

1. Epidemics of infectious diseases are complex processes that must be described with models

1.1. What is an epidemic?

1.1.1. Definition

The Centers for Disease Control and Prevention (CDC) defines an **epidemic** as "an increase, often sudden, in the number of cases of a disease above what is normally expected in that population in that area" (Disease Control and Prevention, 2006). Infectious diseases that are caused by the transmission of harmful organisms called **pathogens** from infected individuals to uninfected ones can lead to epidemics when the number of new cases exceeds the baseline number of cases. While the desired baseline level is zero case, the baseline level corresponds to the observed endemic level in the area under study. Here, endemic defines any disease that is commonly present within a geographic area (Disease Control and Prevention, 2006). In a stricter acceptation, endemicity characterizes diseases that circulate within a geographic area without the need for external introductions. In infectious disease epidemiology, epidemic and endemic circulations are not exclusive types of transmission and should be rather considered as different levels of transmission.

The transmission process can affect a more or less large population and it can be studied at different levels of resolution. Four levels of transmission are commonly distinguished depending on the size of the affected population or the resolution of observation: **clusters** that are confined to a small community like a school or a church, **outbreaks** that affect a larger population in a limited geographic area, **epidemics** that are similar to outbreaks but affect larger geographic areas, and **pandemics** that have spread over several countries or continents (Disease Control and Prevention, 2006).

Many common diseases are infectious diseases. We can cite strep throat, measles, flu, toxoplasmosis, eradicated diseases such as smallpox, or the very recent coronavirus disease 2019 (COVID-19). They are caused by pathogens of different taxonomic origin, more specifically viruses, bacteria, protozoa, arthropods, helminths, and fungi. In this thesis, we primarily focus on viral diseases and examples of infectious diseases caused by other taxonomic groups are described for strict illustrative purposes.

1.1.2. Description of the transmission process

Despite the wide taxonomic diversity of pathogens, they all share the same basic **transmission cycle**. First, the pathogen infects its host, then it multiplies within the host, and it eventually propagates to other hosts (Fig. 1A). This basic transmission cycle applies to diseases with direct transmission such as airborne, respiratory (droplet), or sexually-transmitted diseases. Other routes of transmission generally lead to more complex transmission cycles. In the case of zoonoses (diseases that naturally transmit from animal populations to humans) and a variety of vector-borne diseases, multiple host species are involved and human is not always an essential link in the transmission cycle. For example, West Nile virus (WNV) disease is a vector-borne zoonosis that circulates in birds through mosquito bites but it can also infect mammal hosts among which humans and horses that are dead-ends of transmission (Fig. 1B). Rabies is another example of viral zoonosis whose causative agent, rabies virus (RABV), circulates in bat populations, and domestic and wild non-flying carnivores through bites. It can spill over to many other mammal species including cattle and humans with no further transmission (Fig. 1C). Sometimes, human is not necessary to pathogen persistence but, once introduced, the pathogen causes large epidemics. This is the case of Ebola virus that persists in animal reservoirs and leads to self-sustaining chains of transmission in human. Food and water transmission routes are other types of indirect transmission that can involve complex food supply systems or multiple host species, respectively.

1.1.3. Determinants of the transmission process

The transmission process is **multifactorial**. Pathogen, host, and environmental factors influence the success of transmission by controlling the transmissibility of the pathogen, the contagiousness of the infector, the susceptibility of exposed individuals, and the environmental stress on the pathogen (Fig. 2).

1.1.3.1. Pathogen determinants

Pathogen characteristics determine its routes of transmission, its ability to spread and to cross species barriers, and its evolutionary speed (Leung, 2021).

Pathogen molecular composition governs the sites of infection within the organism and the host range that can be infected. Viruses, for example, are compulsory intracellular parasites and first interact with host cells through their surface proteins before hijacking the inner cellular machinery to multiply. The success of this interaction depends on the physicochemical and tridimensional properties of viral surface proteins. Influenza viruses are a typical example of the impact of molecular composition on host range as their ecology is primarily driven by the shape and composition of the hemagglutinin antigen (Long et al., 2019).

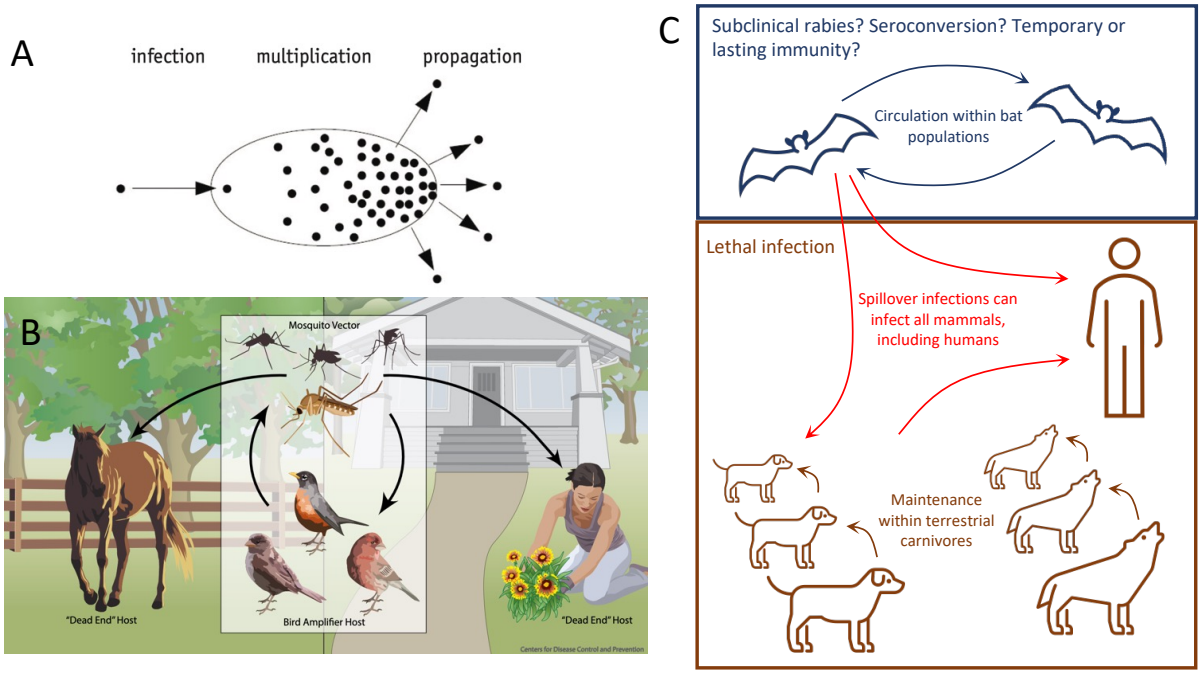


Figure 1: Transmission cycle of pathogens. (**A**) The three minimal steps of the transmission process. First, the pathogen infects an individual, then multiplies within the individual, and finally is shed into the environment (from Choisy 2010). (**B**) Transmission cycle of West Nile virus (WNV). WNV disease is a vector-borne disease affecting bird populations. Human and horse populations can be infected through mosquito bites but they are dead-end hosts (from the CDC). (**C**) Transmission cycle of rabies virus (RABV). Although domestic dogs are the main reservoir of RABV worldwide, other terrestrial carnivores and bat populations contribute to its circulation. Infections in humans are due to spillover infections from the animal reservoirs. Similar to WNV, human is a dead-end host of RABV and cannot naturally transmit the virus (adapted from Fisher et al. 2018).

Pathogen immune escape strategies determine reinfection patterns. Indeed, pathogens are recognized and fought by the immune system of the host upon infection, ultimately leading to pathogen clearance and the constitution of an effective immune memory that will prevent reinfection. Most of the vaccine-preventable childhood diseases (e.g., measles, rubella) induce sterilizing immunity conveying life-long protection against reinfection. However, many pathogens have developed an arsenal of strategies to escape the immune system and leave no- or short-lived immunity. This way, they can reinfect the same host leading to very different disease dynamics at the population level compared to pathogens that leave sterilizing immunity. Rapid evolution, pathogen diversity, and gene expression variability are common immune escape strategies (Bjørnstad, 2018). RNA viruses are fast-evolving pathogens that can adapt to the rapid build-up of population immunity during an epidemic, of which influenza viruses constitute a paradigmatic example. Influenza viruses evade pre-existing immunity in the host population by antigenic drift (gradual accumulation of mutations in antigens) causing seasonal influenza epidemics and imposing continuous update of flu vaccines. They are also of pandemic potential because sharpest changes of their antigens can arise from genome reassortment between different strains (antigenic shifts) inducing cross-species transmission (Kim et al., 2018). Another consequence of the fast evolution of RNA viruses is

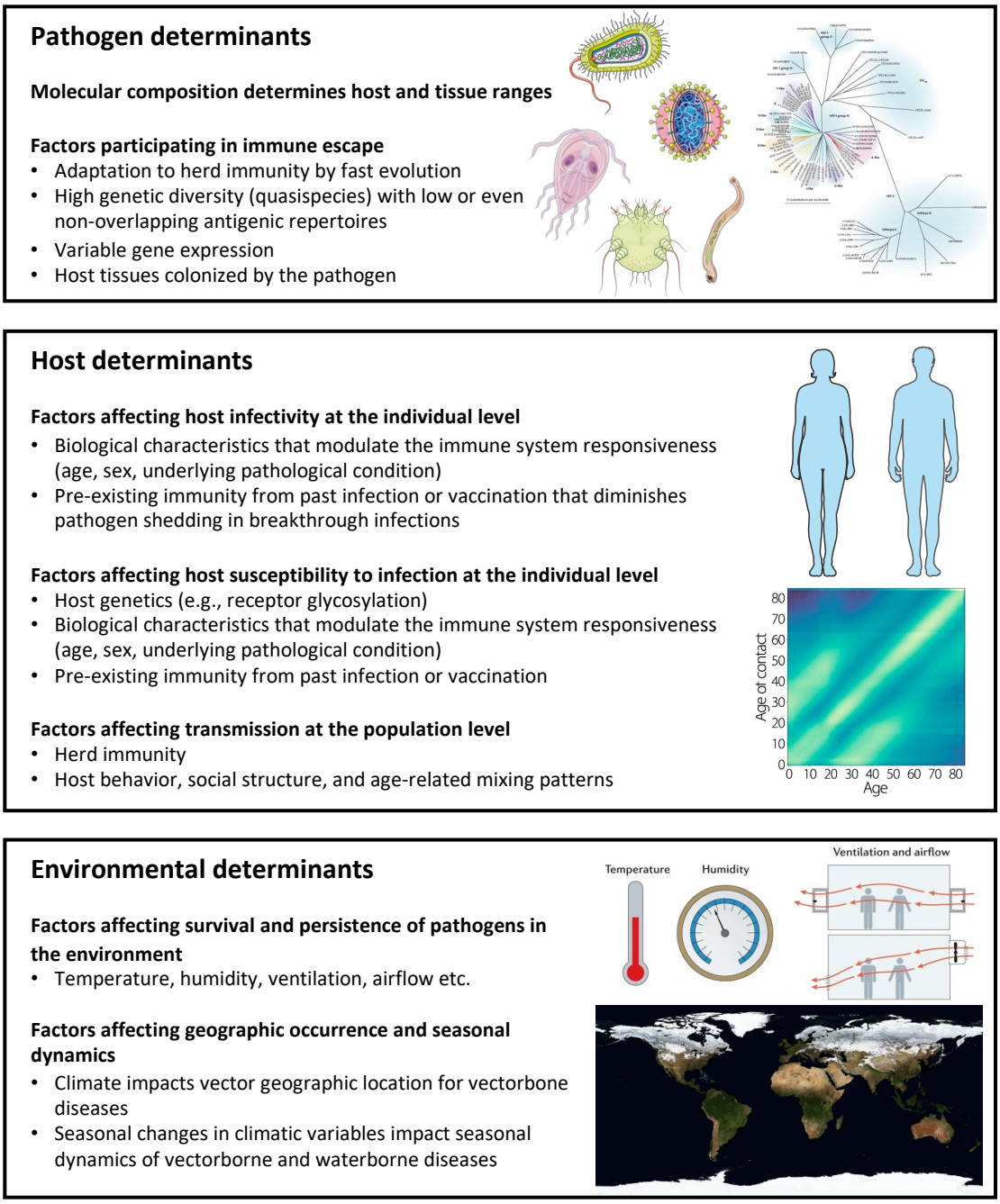


Figure 2: Pathogen, host, and environmental determinants of pathogen transmission. Pathogen, host, and environmental factors influence the success of the transmission process by affecting the host susceptibility, host infectivity, and pathogen persistence in the environment (adapted from Leung 2021; Mistry et al. 2021; Ariën et al. 2007).

their presence under diverse genomic variants (quasispecies) within the same host which also contributes to their fast adaptation (Vignuzzi et al., 2006). Numerous strains of *Plasmodium falciparum* (causative agent of malaria) co-circulate while having non-overlapping antigenic repertoires. Such pathogen diversity allows repeated infections of the same individual (Gupta et al., 1998). Finally, variable gene expression of surface proteins enables pathogen to escape the immune system and prevents the host to

build an effective immune memory as exemplified by *Nesseria gonorrhoeae* (Stern et al., 1984; Tettelin et al., 2000). It is important to note that the host tissue range of some pathogens drastically limits the host immune response. This is the case of RABV that hides from the host immune system by infecting the central nervous system. In such cases, the immune escape strategies based on rapid evolution, pathogen diversity, and gene expression variability have little impact on transmission dynamics in host populations.

1.1.3.2. Host determinants

Host characteristics modulate host infectivity and susceptibility at the individual level, as well as pathogen transmission at the population level (Leung, 2021).

The host immune response plays a critical role in transmission at both the individual and population levels. At the individual level, the host immune system controls infections and modulates disease severity. Its controlling capacities varies with multiple factors: sex, age (young children and elderly individuals are typically more susceptible to infections; Brodin and Davis 2017), and underlying disease conditions (individuals with heart disease or diabetes are more susceptible to severe forms of COVID-19; Sanyaolu et al. 2020). The immune system also shapes the individual response to vaccines that varies with age. Pre-existing immunity from past infections or vaccination can also modulate pathogen shedding in infector individuals, and consequently host contagiousness (Leung, 2021). More generally, host genetics govern susceptibility. For example, glycosylation patterns of host cell receptors determine which hosts can be infected by SARS-CoV and Middle East respiratory syndrome coronavirus (MERS-CoV; Jones et al. 2021). At the population level, pre-existing immunity induced by past epidemics or vaccination campaigns conveys indirect protection against transmission to susceptible individuals. This phenomenon is called herd immunity and determines the magnitude of spread and the shape of the epidemic curve (Anderson and May, 1985).

Host behavior and social structures are additional factors shaping transmission at the population level (Buckee et al., 2021). They encompass social contact and age-related mixing patterns, mass gatherings that may lead to superspreading events, and interactions with wildlife that modulate the risk of zoonotic emergence.

1.1.3.3. Environmental determinants

Environmental factors influence the survival and persistence of pathogens after their release in the environment. For example, respiratory pathogens are transmitted through fomites, droplets, or aerosols whose stability is impacted by temperature, humidity, ventilation, airflow etc. This mechanism explains

why most respiratory tract infections spread when temperature and humidity are low (Mäkinen et al., 2009). Environmental factors can also act indirectly on the geographic occurrence of diseases by delimiting the geographic range of the animal reservoirs or vector populations (Rocklöv and Dubrow, 2020).

Seasonal changes of environmental variables may lead to seasonal patterns of disease occurrence. For instance, seasonal rainfalls are associated with cholera outbreaks in sub-Saharan Africa and Southeast Asia (Emch et al., 2008; Perez-Saez et al., 2022), or more indirectly, they multiply the breeding sites of insect vectors in turn intensifying the circulation of vector-borne diseases (Rocklöv and Dubrow, 2020).

1.1.4. Epidemics are complex processes

Overall, the transmission process is complex because it is **multifactorial** but also **stochastic**, meaning due to random chance. Indeed, the introduction of a new pathogen in the population does not always lead to sustained transmission, epidemics can die out by chance. Besides, epidemics of infectious diseases can be **explosive** contrary to non-communicable diseases, especially when the pathogen is highly transmissible and the host population is dense with little pre-existing immunity.

1.2. Infectious diseases burden and threats

1.2.1. Burden of infectious diseases

The burden of infectious diseases has greatly changed over the ages. In the pre-modern world, life expectancy did not exceed 30 years in all regions of the world (Riley, 2005) as half of the children died from infectious disease before reaching adulthood (Volk and Atkinson, 2013). In parallel, humankind has been afflicted over the ages by pandemics like the Black Death that killed one third of the European population and tremendously impacted European geopolitics and history (Herlihy and Cohn, 1997). During the Enlightenment, most western countries experienced a slight but non-negligible increase of life expectancy thanks to income growth. A sharper increase occurred at the end of the 19th century thanks to the development of hygiene and sanitation that reduced infectious disease mortality. In the 20th century, life expectancy continued to increase thanks to the advent of vaccination, antibiotics, nutrition, medical practices, and health systems that all helped combat infectious diseases (Omran, 2005). Today, most countries have undergone the **health transition**: their populations are aging and more at risk of non-communicable diseases than infectious diseases (Vos et al., 2020). However, global disparities remain, as some low- and middle-income countries such as Kenya are still majorly affected by infectious diseases (Vos et al., 2020). Besides, health systems have to fight not only old pathogens like measles but also new infectious diseases risks: (*i*) **emerging zoonoses** that are a new array of infectious diseases of high epidemic potential that can spill over from vertebrate reservoirs to human, (*ii*) multi-resistant "superbugs"

that are responsible of **antimicrobial resistance** (AMR), and (*iii*) the geographic **expansion of endemic diseases** such as dengue, malaria, or tuberculosis (Bloom and Cadarette, 2019). Consequently, infectious diseases are still a major health problem in the world. For instance, lower respiratory tract infections, diarrheal infections, HIV/AIDS, tuberculosis, and malaria still ranked among the 15 leading causes of global disability adjusted life years (DALYs) in 2019 (Fig. 3A; Vos et al. 2020) and AMR was a leading cause of death, notably in low-resource settings (Murray et al., 2022).

A

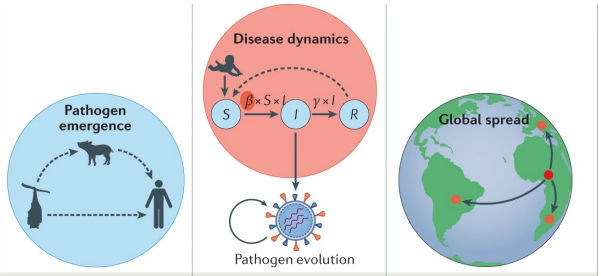
	Leading causes	Percentage of DALYs
1	Road injuries	6.6 (5.6 to 7.7)
2	Headache disorders	5.0 (0.6 to 10.9)
3	Self-harm	3.7 (3.1 to 4.5)
4	Depressive disorders	3.7 (2.6 to 5.0)
5	Interpersonal violence	3.5 (2.9 to 4.1)
6	Anxiety disorders	3.3 (2.3 to 4.4)
7	Low back pain	3.2 (2.2 to 4.3)
8	Dietary iron deficiency	2.6 (1.9 to 3.4)
9	HIV/AIDS	2.6 (1.9 to 3.5)
10	Diarrheal diseases	2.6 (1.9 to 3.6)
11	Neonatal diseases	2.3 (1.8 to 2.8)
12	Tuberculosis	2.1 (1.8 to 2.5)
13	Gynaecological diseases	1.9 (1.4 to 2.6)
14	Typhoid and paratyphoid	1.8 (0.8 to 3.3)
15	Maternal disorders	1.8 (1.5 to 2.2)
16	Malaria	1.8 (1.0 to 3.0)
17	Conduct disorder	1.8 (1.1 to 2.6)
18	Drug use disorders	1.6 (1.3 to 2.1)
19	Acne vulgaris	1.6 (1.0 to 2.4)
20	Idiopathic epilepsy	1.6 (1.2 to 2.1)
21	Congenital birth defects	1.5 (1.3 to 1.7)
22	Falls	1.4 (1.3 to 1.6)
23	Drowning	1.4 (1.2 to 1.7)
24	Lower respiratory infections	1.4 (1.2 to 1.7)
25	Age-related hearing loss	1.3 (0.9 to 1.8)

Communicable, maternal, neonatal, and nutritional diseases

Non-communicable diseases

Injuries

B



Climatic change	Drives range shifts for reservoir species	Affects transmission and susceptibility	Affects the geographical range of vectors
Technological change			
Transportation	Improved global surveillance		Air transit and high-speed rail affect pace and range of spread
Health care		Vaccination affects dynamics	Improved care reduces burden
Demographic change			
Population growth and land use	Increased contact with reservoir species	Population numbers affect evolution, birth rates affect dynamics	Larger population travelling
Urbanization	Depends on species	Density affects contact rate	Urban population more connected
Ageing	Imunosenescence affects spillover risk	Ageing population increases transmission	Possible larger burden

Figure 3: Global burden of infectious diseases and global changes increasing their risk (**A**) Leading 25 causes of disability-adjusted life years (DALYs) at all ages in the world in 2019 and percentage of total DALYs with its 95% confidence interval. Communicable, maternal, neonatal, and nutritional diseases are colored in red. Non-communicable diseases are colored in blue, and injuries in green (adapted from Vos et al. 2020). (**B**) Effects of global changes on disease emergence, dynamics, and spread (from Baker et al. 2021). Age-related hearing loss=age-related and other hearing loss.

1.2.2. Infectious diseases threats

Beyond their direct health consequences, infectious diseases pose social and economic risks. New pathogens of epidemic potential can overwhelm health systems which limit their capacity to deal with routine health issues as experienced during the COVID-19 pandemic when many non-infected patients were left on the side of the road (Rosenbaum, 2020). They can also delay biomedical research on other diseases by concentrating all the investigation efforts and resources as experienced during the COVID-19 pandemic (Rosenbaum, 2020). The fear of novel pathogens may refrain global travel and tourism which has huge economic implications as experienced by Brazil and Southeast Asian countries during dengue

peaks (Bärnighausen et al., 2013; Constenla et al., 2015). The protracted circulation of endemic infectious diseases also disrupt productivity and may considerably impact the gross domestic product (GDP). For example, a 10% reduction in malaria incidence is associated with a 0.3% higher growth of GDP in (sub)tropical countries (Gallup and Sachs, 2001). The impacts of infectious diseases may go further and even increase the vulnerability of weak health systems.

The three major infectious risks - emerging zoonoses, AMR, and endemic diseases expansion - will very likely gain ground due to ongoing demographic, climatic, and technological changes (Fig. 3; Baker et al. 2021). Demographic and urbanization growth, especially in countries with weak health systems, favor contagious transmission. Ageing is another critical challenge for infectious disease spread as immunosenescence (age-related alterations of the immune system) makes the elderly more susceptible to infectious diseases (Aw et al., 2007). On top of that, climate change contributes to the geographical expansion of vector-borne diseases by widening the habitats of disease-carrying vectors, as already occuring for dengue in western Europe (Lazzarini et al., 2020). Climate change, changes in land use, and intensive livestock farming that aims to satisfy the increasing demand for animal protein modify the human-animal interactions and increase the risk of zoonotic emergence. Finally, globalization through human transportation and global trade amplifies the pandemic risk of emerging zoonotic pathogens as well as antimicrobial resistant pathogens (Fig. 3B; Petersen et al. 2018; Bloom and Cadarette 2019; Baker et al. 2021).

In this rapidly evolving context, health systems need to adapt and reinforce epidemic preparedness, disease transmission surveillance, and response to infectious emergencies. Rapid access and analysis of epidemiological data is crucial to ensure a timely response. In parallel, research and development efforts are required to better understand disease spread, develop new therapeutics, and design control strategies to mitigate the impact of epidemics.

1.3. Quantitative approaches to study epidemics of infectious diseases

1.3.1. A modeling approach to integrate the complexity of the transmission process

Epidemics of infectious diseases constitute a growing threat but their dynamics and determinants are not directly accessible because they are complex processes. **Mathematical models**, that are simplified but informative representations of complex systems or processes formalized by equations, can help describe and anticipate epidemic processes by accounting for multiple determinants of disease spread and integrating multiple sources of information in a single framework (Keeling and Rohani, 2008).

1.3.2. Epidemiological modeling for the description of between-host transmission dynamics

Epidemiological models were the first **mechanistic** models of disease transmission, that is to say that they explicitly describe the transmission process under a set of assumptions. They first aimed at modeling disease transmission between individuals of the same host population, although they are more and more used to model within-host transmission. In this section, we will briefly present their history, conceptual framework, and underlying hypothesis, focusing on between-host transmission models.

1.3.2.1. Compartmental models, a milestone in epidemiological modeling

In their seminal work, Kermack et al. (1927) introduced the basics of the **Susceptible-Infectious-Recovered (SIR) model** that was later formalized by Dietz (1967). In this model, the host population is divided into three compartments: susceptible (not yet exposed to the pathogen), infectious (carry the pathogen and can spread it), and recovered (have successfully cleared the infection and are protected against reinfection). Individuals flow from one compartment to the next upon infection (S to I) or recovery (I to R) as represented in Fig. 4.

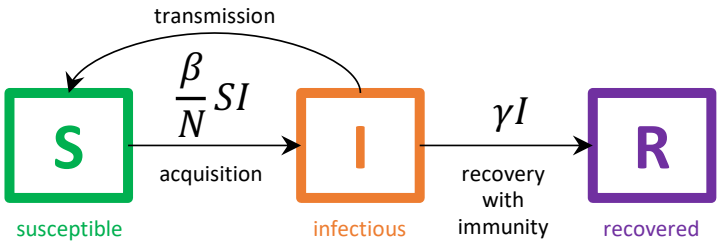


Figure 4: Flow diagram of the Susceptible-Infectious-Recovered (SIR) model. The host population is divided into three compartments: susceptible (S), infectious (I), and recovered (R) individuals. The total number of individuals is N. Susceptible individuals acquire infection upon contact with infectious individuals that transmit with force of infection β. Infectious individuals recover from the disease at rate γ and are protected from reinfection.

In mathematical terms, the model consists in a system of ordinary differential equations (ODEs):

$$\begin{aligned} \frac{dS}{dt} &= -\frac{\beta}{N}S(t)I(t) \\ \frac{dI}{dt} &= \frac{\beta}{N}S(t)I(t) - \gamma I(t) \\ \frac{dR}{dt} &= \gamma I(t) \end{aligned} \tag{1}$$

Where β is the transmission rate, γ the rate of recovery from infection, $S(t)$, $I(t)$, and $R(t)$ are the state variables representing the number of individuals as a function of time t in the susceptible, infectious, and

recovered compartments, respectively, and $N = S(t) + I(t) + R(t)$ is the total population size. Several assumptions underlie this model:

- The infectious period $\frac{1}{\gamma}$ is exponentially distributed,
- β and γ do not change over the course of the infection,
- Recovered individuals are immune from reinfection over their lifetime,
- Individuals from the same compartment contact each other at random which is known as homogeneous mixing, and
- The transmission rate β is invariant with respect to the population size N which is generally referred to as frequency-dependent transmission.

Some of these assumptions can be relaxed by adding compartments to account for transient immunity (Heaney et al., 2020; Hawkes et al., 2021), age-related or spatially heterogeneous contact patterns (Viboud et al., 2006), transmission from animal reservoirs (Allen et al., 2012; Hussaini et al., 2017), vector-borne transmission (Smith et al., 2012) etc. This modularity provides an ideal flexibility that allows to adapt the model to very diverse disease natural histories and account for multiple host-related determinants of spread. Most compartmental models make the assumption of frequency-dependent transmission. Alternatively, one can assume a density-dependent transmission where the transmission rate scales linearly with population density.

The analysis of compartmental models provides insights on the transmission process in the system under study. For simple compartmental models such as the SIR model presented in Fig. 4, analytical solutions that describe the steady states of the system can be easily derived. Otherwise, numerical integration is possible. By analyzing the dynamics of the SIR model, Kermack et al. (1927) showed that a pathogen can invade the host population only if the initial fraction of susceptible individuals $S(0)$ is less than $\frac{\gamma}{\beta}$. This threshold effect led to the definition of one of the most important metrics in infectious disease epidemiology, the reproduction number (R). The basic reproduction number R_0 is the average number of secondary cases arising from the primary case in an entirely susceptible population. In the SIR model, it is equal to $\frac{\beta}{\gamma}$, the inverse of the critical threshold (Keeling and Rohani, 2008). Thus, the pathogen spreads in a fully susceptible host population only if $R_0 > 1$. When part of the host population is protected against infection, we would rather use the effective reproduction ratio R_t that is the product of the basic reproduction ratio and the fraction of susceptible individuals at time $s(t)$. In the SIR model, this would be $\frac{\beta}{\gamma} s(t)$.

1.3.2.2. From deterministic to stochastic models

The compartmental model presented above is deterministic, hence it overlooks the stochastic nature of the transmission process. It can be translated, as well as any other compartmental model, in its **stochastic** equivalent using discrete Markov chains, continuous-time Markov chains (CTMCs), or stochastic differential equations (Allen, 2017). Numerical simulations of disease dynamics under these models can be produced using the Gillespie algorithm or the Euler-Maruyama method (Allen, 2017).

Stochastic models are useful to capture the variability of the epidemic profile notably when disease incidence is low, the population is small, or environmental variables or demographics strongly impact disease dynamics. In such cases, deterministic models are not a good approximation of disease dynamics because stochastic fluctuations may lead to disease extinction or introduce variance and covariance that influence disease transmission (Keeling and Rohani, 2008; Allen, 2017). Stochastic modeling is thus more appropriate when studying disease eradication as incidence is low, disease transmission in households that are very small populations, or zoonoses, vector-borne and waterborne diseases for which environmental variability is important (Keeling and Rohani, 2008; Allen, 2017).

1.3.2.3. Individual-based models to integrate heterogeneity at the individual model

Host heterogeneity can be integrated in compartmental models by specifying additional compartments, that is additional ODEs. However, when multiple layers of heterogeneities are to be accounted for, model formulation becomes rapidly cumbersome, or may even be limited. Alternatively, **individual-based models** (also called agent-based models) describe disease transmission by explicitly modeling the interactions between a finite number of fully characterized and autonomous individuals. Any attribute, constant or time-varying, related to social behavior, spatial location, and/or physiological traits governing susceptibility and contagiousness can be associated to the individuals. Hence, the history of every individual can be tracked, and at the same time, the impact of individual-level heterogeneity on disease transmission in a restricted population can be investigated (Willem et al., 2017). This category of models is very diverse and particularly adapted to the combined study of between- and within-host interactions and the analysis of targeted interventions such as targeted screening and vaccination in nosocomial infections (Smith et al., 2020). Over the past decade, individual-based models have been more widely used but important pitfalls subsist concerning calibration methods and goodness-of-fit measures which limits study reproducibility and affects the quality of inference (Hazelbag et al., 2020).

1.3.3. Phylodynamics, or epidemiology from the perspective of the pathogen

1.3.3.1. Definition of phylodynamics

Epidemiological models can easily integrate host and environmental determinants which is less the case for pathogen determinants because these models do not explicitly describe the evolutionary processes that pathogens undergo during transmission. Indeed, in the specific case of fast-evolving pathogens such as RNA viruses, evolutionary forces (mutation, migration, selection, and drift) not only shape pathogen genetic evolution and diversity, but they occur at the same time scale as the transmission process and can be influenced by it. For example, the genetic bottleneck at transmission shapes the viral diversity transmitted to other hosts and can be influenced by the seasonal or spatially heterogeneous dynamics of the host population. In such cases, evolution is measurable and genetic diversity that is informative about the transmission process (Drummond et al., 2003) can be exploited to investigate the relationship between pathogen evolution and epidemic and immunological processes. **Phylodynamics** studies this interplay between evolutionary, epidemiological, immunological, and sometimes even ecological processes in epidemics by combining quantitative methods from phylogenetics and population genetics (Grenfell et al., 2004). It is a recent discipline that emerged in the 2000s and has expanded rapidly. It completes the arsenal of quantitative tools already available for the study of infectious diseases spread and can be considered, in this sense, as a specific approach of epidemiological modeling. For better clarity, we refer to the models presented in Section 1.3.2 as epidemiological models as opposed to phylodynamics, and we use the term epidemiology in its widest acceptance by including epidemiological modeling and phylodynamics.

In the following paragraphs, we outline the basic modeling components of phylodynamics.

1.3.3.2. Substitution models link phylogenies and molecular sequences of pathogens

A **phylogeny** corresponds to the evolutionary relationships between sampled organisms. It is depicted as a phylogenetic tree whose external nodes correspond to sampled organisms (here pathogens isolated from cases), internal nodes correspond to their common ancestors, branches correspond to ancestral lineages, and branch length corresponds to the level of genetic divergence. Phylogenies are at the basis of phylodynamic analyses because their shape is an indicator of the underlying epidemic process (Fig. 5A; Grenfell et al. 2004; Volz et al. 2013).

Nucleotide sequences (DNA and RNA) are a good source of information of phylogenies because they gradually accumulate **substitutions** over time due to replication errors during pathogen multiplication (Lemey et al., 2009b). This means that any two sequences coming from the same ancestor but evolv-

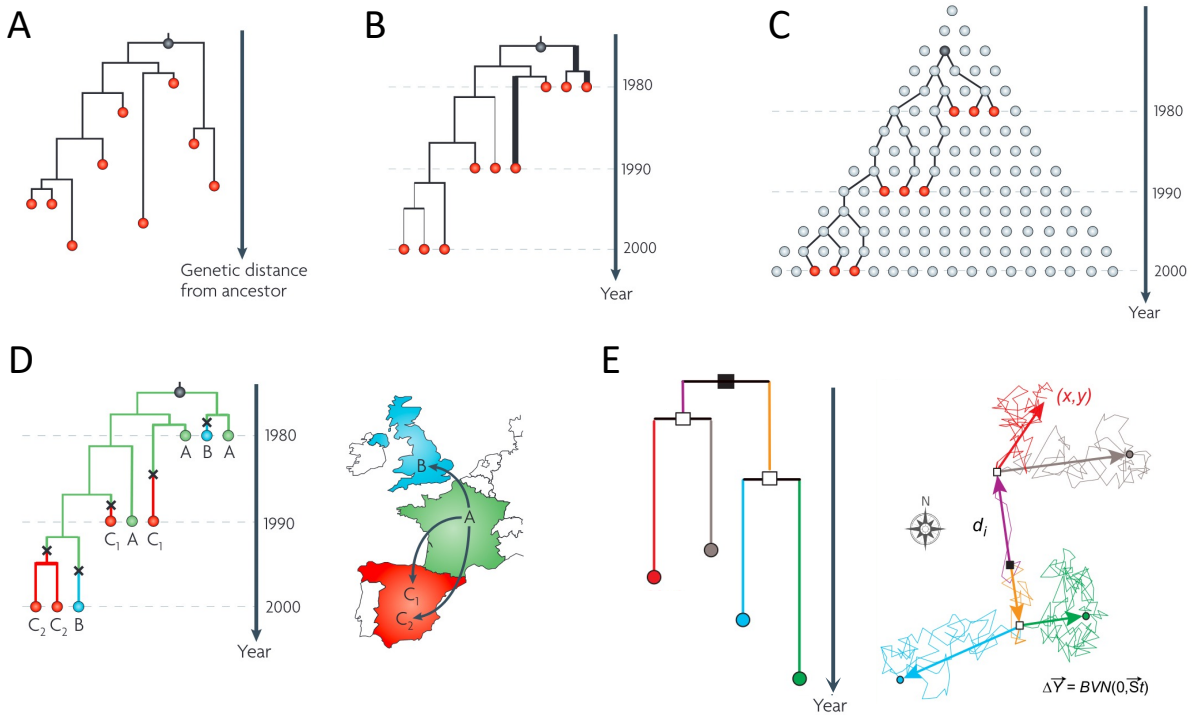


Figure 5: Phylodynamic models. (**A**) Rooted molecular phylogenies are estimated from alignments of pathogen genetic sequences using a nucleotide substitution model. Branch length corresponds to the genetic distance between an ancestor and its descendant, indicating their phylogenetic proximity. (**B**) By adding a molecular clock model to the inference framework, one can calibrate genetic divergence on calendar time, and thereby, date clade emergence and introduction events. (**C**) Demographic models stemming from population genetics describe pathogen population dynamics either backwards-in-time as depicted here for the coalescent model, or forwards-in-time for the birth-death model. These models allow the reconstruction of pathogen population growth and can be used to estimate the transmission potential of pathogens. (**D**) Discrete phylogeography enables to incorporate spatial heterogeneity and is particularly adapted for the study of human diseases. (**E**) Continuous phylogeography treats geospatial coordinates as continuous variables and models transmission without having to arbitrarily discretize space. It is particularly insightful when applied to pathogen circulation in animal populations (adapted from Pybus and Rambaut 2009 and Dellicour et al. 2021a).

ing independently eventually diverge by accumulating different substitutions. Stochastic approaches like CTMCs allow to model substitutions as random events and account for multiple substitutions per site (i.e., location in the genetic sequence). In CTMC models, the substitution process is entirely specified by the Q matrix that contains the instantaneous relative rates of change of each nucleotide along the sequence. These instantaneous rates depend on the mean instantaneous substitution rate, relative rate parameters, and nucleotide frequencies (Fig. 6). Depending on the assumptions made on these parameters, one can specify a wide range of models from the simple Jukes-Cantor 69 model where all instantaneous substitution rates are equal to the mean instantaneous substitution rate in the Q matrix, to the complex general-time reversible model (GTR) where all instantaneous substitution rates are different. Importantly, whatever the parameterization of the Q matrix, all CTMC models share the same underlying assumptions (Lemey et al., 2009b):

1. Markov property: at any given site in a sequence, the rate of change from base i to base j is

independent from the base that occupied that site prior base *i*.

2. Homogeneity: substitution rates are constant over time.

3. Stationarity: the relative frequencies of A, C, G, and T (π_A, π_C, π_G, π_T) are at equilibrium.

$$Q = \begin{array}{c} \begin{matrix} A & C & G & T \end{matrix} \\ \begin{pmatrix} -\mu(a\pi_C + b\pi_G + c\pi_T) & a\mu\pi_C & b\mu\pi_G & c\mu\pi_T \\ g\mu\pi_A & -\mu(g\pi_A + d\pi_G + e\pi_T) & d\mu\pi_G & e\mu\pi_T \\ h\mu\pi_A & i\mu\pi_c & -\mu(h\pi_A + j\pi_C + f\pi_T) & f\mu\pi_T \\ j\mu\pi_A & k\mu\pi_c & l\mu\pi_G & -\mu(i\pi_A + k\pi_C + l\pi_G) \end{pmatrix} \end{array}$$

Figure 6: Instantaneous rate matrix Q from the continuous-time Markov chain (CTMC) model applied to nucleotide substitution. Q is a square matrix whose entries represent the instantaneous substitution rates from one nucleotide to another. The instantaneous substitution rates are a function of the mean instantaneous substitution rate μ, the twelve relative substitution rates a, b, c, d, e, f, g, h, i, j, k, and l, and the four nucleotide frequencies π_A, π_C, π_G, and π_T. Diagonal elements are chosen so that the sum of each row is equal to zero. Rows and columns follow the order A, C, G, and T (from Lemey et al. 2009b).

1.3.3.3. Dating pathogen phylogenies using molecular clock models

Molecular clock models link genetic divergence with time which allows to scale the branch lengths of a phylogenetic tree on calendar time (Fig. 5B). The **strict molecular clock** model, introduced by Zuckerkandl and Pauling in the 1960s, assumes that nucleotide substitutions accumulate at a roughly constant rate (evolutionary rate) in the genetic sequences of pathogens (Zuckerkandl and Pauling, 1962; Zuckerkandl and Pauling, 1965). With this model, one can date the origin of an epidemic or the start of an infection, but it is restricted to cases where the evolutionary rate is the same among all lineages in the considered phylogeny although multiple factors are known to modulate it such as the underlying mutation rate, metabolic rates in a species, generation times, population sizes, or selective pressures (Bromham and Penny, 2003). **Relaxed molecular clock** models overcome this limitation by allowing evolutionary rates to vary through time or among lineages. In the local clock model, the number of rate changes along the phylogeny is prespecified. In the autocorrelated relaxed clock model, the evolutionary rate along one branch depends on the evolutionary rate of its ancestral branch. In the uncorrelated relaxed clock model, the evolutionary rates associated with each branch are independent and identically distributed, generally following a lognormal or gamma distribution (Ho and Duchêne, 2014).

Once informed by substitution and molecular clock models, pathogen phylogenies are approximatively a subtree of the whole transmission chain that connects the sampled individuals located at the tips of the phylogeny. Due to within-host evolution, transmission events in the transmission chain are more recent than ancestral nodes in the associated time-stamped phylogenetic tree (Plessis and Stadler, 2015).

1.3.3.4. Pathogen demographic models

In population genetics, the **coalescent theory** provides a conceptual framework that links population dynamics of sampled individuals to the topology of their phylogenetic tree (Fig. 5C). The most simple model is the Kingman coalescent in which the expected time at which two individuals coalesce into their common ancestor in a discrete and non-overlapping population of size N (the Wright Fisher population) is N generations going backwards in time (Kingman, 1982). Branch lengths in the phylogenetic tree thereby reflect pathogen population size: the longer the branches, the larger the population. In the Kingman coalescent, population size is constant over time, but extensions allow to account for time-varying dynamics. These refinements are either deterministic (Griffiths et al., 1994) like the exponential and logistic growth models, or non-parametric models (Pybus et al., 2000; Strimmer and Pybus, 2001; Drummond et al., 2003; Opgen-Rhein et al., 2005; Minin et al., 2008). In all these extensions, the overall pathogen population is supposed to be large compared to the number of sampled individuals (tree tips) and to evolve under neutral evolution. Although these assumptions are generally violated, an effective population size (Ne) that leads to the coalescent rate of an idealized population of size N can be derived. Interestingly, Volz et al. (2009) showed that the exponential coalescent model, i.e., a coalescent model where Ne follows an exponential growth, can be linked to the SIR model by expressing Ne as a function of the transmission rate and the prevalence. This formulation allows to calculate the reproduction number.

Coalescent models are deterministic and require sampled individuals to represent a small fraction of the total population. Thus, they are not adapted to model early epidemic dynamics or infection clusters because pathogen populations are small hence their dynamics are impacted by stochasticity, and sampling proportions are large (Pybus and Rambaut, 2009). **Birth-death approaches** on the contrary are stochastic and describe a forward-in-time process particularly adapted to model the start of epidemics (Plessis and Stadler, 2015). In these models, extant lineages either generate new lineages at birth rate λ, die at extinction rate μ, or are sampled at sampling rate ψ (Featherstone et al., 2022).

1.3.3.5. Describing spatial dynamics using phylogeography

The demographic models presented above assume that transmission occurs in a homogenous mixing host population, although host population is often structured due to age-structured contact patterns or spatial clustering. Phylogeographic approaches aim at filling this gap by modeling spatial diffusion along a phylogenetic tree between discrete populations (Fig. 5D) or in a continuous space (Fig. 5E).

Discrete phylogeography

There are three major approaches to modeling spatial dynamics between discrete populations: (*i*) the structured coalescent model, (*ii*) the multitype birth-death process, and (*iii*) the CTMC model. The

structured coalescent model corresponds to an extension of the coalescent model. It explicitly models how lineages coalesce within and migrate between subpopulations from present to past. Likewise, the **multitype birth-death process** is an extension of the birth-death process where extant lineages generate a new lineage in their population, generate a new lineage in another population, migrate to another population, are sampled, or die (Kühnert et al., 2016). Finally, the **CTMC** treats the geographic location of pathogens as a neutral trait that evolves just like nucleotides along the phylogenetic tree without explicitly modeling population demographics (Lemey et al., 2009a).

The applications of discrete phylogeography exceed spatial spread modeling as the discrete populations can represent different host species (De Maio et al., 2015; Dudas et al., 2018), host social groups (Stadler and Bonhoeffer, 2013), or host body compartments (Chaillon et al., 2014).

Continuous phylogeography

Continuous phylogeography provides a more realistic representation of the spatial diffusion process because it does not rely on the pre-specification of the number of discrete locations. In continuous phylogeography, pathogens' geographic coordinates (i.e., latitude and longitude) are continuous traits that evolve according to a Brownian diffusion process. To overcome the limiting assumption of constant diffusion rate along the phylogenetic tree, the **relaxed random walk model** uses the same approach as the uncorrelated molecular clock model by assuming that the diffusion rates associated with each tree branch are independent and identically distributed according to a discretized rate distribution (Lemey et al., 2010).

2. Challenges in estimating key parameters of the transmission process

Although the development of theoretical models is crucial to the advancement of epidemiology, calibrating and validating such models with empirical data is necessary to explore the submerged part of the transmission process and unravel the determinants of spread, estimate key epidemiological parameters, and thereby inform public health programs. **Statistical methods** that make the connection between models and data come in very different flavors. In epidemiology, statistical inference faces a major challenge related to the imperfect observation of the transmission process. We discuss here how imperfect epidemic data are, then we outline the different types of data that can be collected for outbreak evaluation, and finally, we present the statistical methods developed to overcome these challenges.

2.1. The transmission process is imperfectly observed

Unlike experimental sciences, epidemiology analyzes real-world data that result from complex processes occurring in uncontrolled conditions. This leads to two major challenges: (*i*) observed data are generally very limited which requires methodological development, and (*ii*) data collection and interpretation may be subject to bias.

By essence, the transmission process is **never fully observed**. The exact time of infection and all the factors that influence the transmission event are rarely directly observed or measured. Detailed transmission chains obtained by comprehensive contact tracing are an invaluable source of information from which key epidemiological parameters such as R_0 can be derived directly, but they are very difficult to obtain for most pathogens (Cauchemez et al., 2019). In some cases, like zoonotic pathogens, detailed transmission chains are not sufficient to characterize the transmission potential of the pathogen in human. Indeed, zoonotic pathogens stutter to extinction after spillover from the animal reservoir because they are not yet adapted to human (Lloyd-Smith et al., 2009). They do not cause enough cycles of transmission in human to allow the calculation of inter-human R_0. Nevertheless, Ferguson et al. (2004) demonstrated that there is a relation between inter-human R_0 and the distribution of the size of case clusters. This way, the authors could estimate the transmission potential of avian H5N1 influenza in human using very coarse data.

In addition, **multiple sources of bias** may impact data collection and interpretation (Cauchemez et al., 2019) generally resulting in under-reporting. During cluster investigation, some cases may be missed because investigated cases do not remember all their contacts, some of their contacts may not adhere to the investigation or may test false-negative due to imperfect diagnostic tests. Under-reporting is also an issue at larger scales like national routine surveillance systems that are rarely exhaustive because they rely on simple systems that collect a limited amount of information on each case. For example, influenza surveillance in France relies on a network of primary health-care providers that report weekly numbers of patients with flu-like illness (Debin et al., 2013). This syndromic surveillance system does not include all health-care providers and all affected patients do not necessarily seek care, thus only a fraction of the cases is reported. Models are thus necessary to estimate disease incidence. The extent of under-reporting also depends on the capacities of the surveillance systems that can vary in space and time (Moon et al., 2015; Mastin et al., 2017). In parallel, severe cases are more likely to be detected compared to mild, pauci-symptomatic, and asymptomatic cases leading to a selection bias. The detection of milder cases is crucial because they generally contribute to disease spread and their under-detection may lead to biased

forecasts. In the pyramid of severity, hospitalization and death data are generally more robust to under-detection, although attributing the cause of death is challenging in certain contexts like RABV infections when molecular testing is not available. Besides, hospitalization and death data do not necessarily reflect transmission patterns in the general population.

2.2. Data granularity gradient

Various data can be used to inform the transmission process. They range from detailed data at the individual level in contexts where epidemiological links between cases are more or less known to coarse data at the population level like aggregated number of cases for which epidemiological links between cases are generally unknown. Pathogen genetic sequences can be placed at either ends of the data granularity gradient depending on the way they are collected. When collected in an individual at different time points of infection, pathogen genetic data correspond to individual-level data. But, when they are collected in unrelated individuals at large spatial scales, they correspond to population-level data. In general, epidemiological and genetic data are accompanied with the background characteristics of cases related to their biology, behavior, and environment (Cori et al., 2017; Polonsky et al., 2019). In the following paragraphs, we present three types of data, individual-level epidemiological data, population-level epidemiological data, and pathogen genetic data at the population level, as well as the potential biases occurring at the collection, generation, and interpretation steps.

2.2.1. Individual-level data

Individual-level data consist in detailed data on cases related to their disease, demographics, behavior, and environment. In linelist data, each row corresponds to a case, and each column to an individual characteristic such as age, gender, location, symptom onset, detection date, clinical outcome, diagnostic test results, recent travel history, recent contacts with animals, dietary habits etc. (Polonsky et al., 2019). Useful statistics can be derived from linelists, including the case fatality rate (proportion of identified cases dying from the infection), demographics of the affected population (e.g., age, occupation), and case delays (e.g., times to hospitalization, recovery, and death). The collection of detailed individual-level information is feasible when the number of cases is limited like at the start of an epidemic or in cluster investigation. In other settings, data on uninfected individuals are collected to complete case data which allows the investigation of infection risk factors. Households are a good example of such setting in addition to being a great laboratory of disease transmission because the number of contacts is small and well-defined, and participant adherence is generally high ensuring data completeness (Cauchemez et al., 2019). When collected longitudinally, individual-level data on cases can inform case delays, viral

load dynamics that are generally used as a proxy of infectiousness, and antibody response dynamics that point to long-term protection against reinfection. Overall, individual-level data are highly valuable to tackle disease transmission dynamics and determinants.

Nevertheless, individual-level data do not directly capture the underlying transmission process, as infection times, sources, and dynamics are not perfectly known. For example, the high level of physical proximity between household members may hide the exact transmission chain, or another example, the temporal dynamics of viral load is informed by only few data points per individual. More generally, the number of participants and data completeness are limiting factors to achieve robust statistical inference and multiplying biological testings or questionnaires may exhaust participants. Consequently, study designs should find the right balance between data exhaustivity and participant adherence. In parallel, individual-level data often rely on molecular tests that are inherently imperfect, meaning that their sensitivity (true negative rate) and specificity (true positive rate) are lower than 100%. Serological tests are often subject to cross-reactivity issues, meaning that they not only detect past infections from the pathogen of interest but also from closely related pathogens. This is the case of many arboviruses (Hozé et al., 2021a). Finally, observational studies that aim to collect individual-level data are expensive and require interdisciplinary teams that define the ethical framework of the study, recruit and follow-up participants, perform biological tests, and analyze the generated data. That is why, individual-level data are generally more difficult to collect compared to population-level data.

2.2.2. Population-level data

Population-level data encompass any epidemiological data aggregated at the population level such as case and death counts, hospital and ICU admissions, seroprevalence, the number of vaccinated individuals etc. Aggregated data, except seroprevalence data, are generally collected through surveillance systems at national or regional levels and reflect past and present disease transmission dynamics. Their sole analysis already provides valuable information on the effective reproduction ratio or the impact of interventions. Additional data on climate, population mobility, and population behavior can help forecast (near) future transmission or make predictions under various intervention scenarios. The characteristics of the surveillance system determine the representativeness and the specificity of the collected data (Disease Control and Prevention, 2006). For instance, a case definition based on a diagnostic test is more specific than a case definition based on a set of symptoms, but it poses practical issues and representativeness is not guaranteed due to testing strategies that can over-represent specific populations and to testing behavior that varies with social class (Buckee et al., 2021). Furthermore, the representativeness

of surveillance system data is generally subject to selection bias as severe cases seek more care than mild cases.

Serosurveys aim to estimate seroprevalence, a measure of herd immunity (Hozé et al., 2021b). When stratified by age, seroprevalence can indicate the long-term history of pathogen circulation (Hens et al., 2010). This is particularly insightful in resource-limited countries with weak surveillance systems. However, there is a risk of misdiagnosis for some diseases due to cross-reactivity (Hozé et al., 2021a).

Recent studies have tried to bridge the gap between individual-level and population-level data by collecting detailed data on a very large number of individuals. For example, digital epidemiology has capitalized on the rapid growth of digital data, the widespread penetration of mobile phones and internet and social media usage to collect data on mobility for example, or on symptomatic search queries for syndromic tracking of influenza-like illnesses (Bansal et al., 2016; Salathé, 2018; Tarkoma et al., 2020). Digital epidemiology has strongly benefited from the developments of machine learning to efficiently analyze big data. Similar advancements were made by national surveillance systems that also try to link individual-level and population-level data. Countries with cutting edge health information systems (e.g. the UK, Israel) have matched national health registries with demographic databases during the SARS-CoV-2 pandemic and thereby could analyze transmission in tens of thousands households (Harris et al., 2021; Shah et al., 2021), unattainable numbers in traditional observational household studies. However, data granularity in big data remains limited especially when it comes to behavior.

2.2.3. Pathogen genetic sequences

Thanks to the decreasing costs of high-throughput next-generation sequencing, **pathogen genetic sequences** isolated from infected hosts now constitute an abundant source of information on disease spread. Pathogen genetic sequences are now so affordable that genomic surveillance is more insightful than traditional surveillance systems in resource-limited countries with weak surveillance capacities (Wilkinson et al., 2021). In addition, the time scale of data generation and analysis has greatly improved thanks to the development of genomic surveillance and the increased affordability of portable sequencers (Grubaugh et al., 2019a) allowing real-time analyses. The SARS-CoV-2 pandemic has led to a paradigm shift in terms of sampling: the number of sampled sequences increased from a few dozens to a few millions (Hill et al., 2021), the spatial coverage extended from a few regions or countries to the entire world, and the sampling strategy shifted from opportunistic to targeted (Hodcroft et al., 2021a). However, these changes mostly concern viral diseases of pandemic potential as other pathogens like bacteria and protozoa have much larger genomes which requires more intense sequencing efforts and prevents the democratization of ge-

nomic surveillance. Besides, during the SARS-CoV-2 pandemic, the global sequencing effort was highly heterogeneous which was a major issue for the study of cross-country transmission. Sequencing errors are another potential source of bias (Ma et al., 2019; Stoler and Nekrutenko, 2021). They are a central consideration at the beginning of an outbreak or when there is an intense sampling over a short time window because the pathogen did not have time to accumulate mutations. Finally, multiple sequence alignment tools can be a source of bias when they do not align correctly homologous sites (Bromham et al., 2018).

2.3. Ethical and scientific considerations related to epidemiological and genetic data

Epidemiological data should not only be relevant and reliable to appropriately inform models, they should also follow broader requirements related to **ethics** and **data sharing** (Cori et al., 2017). In terms of ethics, anonymity must be protected while ensuring a sufficient level of details, notably when it comes to individual-level data. In terms of data sharing, rapid data collection and sharing to the scientific community allow independent analyses, each with its own limitations, which fosters results comparison, scientific collaboration, and increases confidence in new findings. Phylodynamics comes from a place where open science really participated in the advancements of science. Public repositories of genetic sequences like NCBI GenBank where researchers publish their newly generated data have been used all over the world for a while now. Originally founded for influenza research, the GISAID Initiative, a public-private partnership, promotes the sharing of influenza virus and SARS-CoV-2 genomes along with their clinical and epidemiological data. Nextstrain is a third example. This open-source project that provides access and tools to analyze pathogen genomes has become a major actor of phylodynamic research for outbreak response (Hadfield et al., 2018). In epidemiological modeling, open access is less advanced but new tools like the Our World In Data global database on COVID-19 vaccination have recently emerged during the SARS-CoV-2 pandemic (Ritchie et al., 2020) paving the way to global epidemiological data sharing. Contrary to genetic data that are directly curated by researchers, epidemiological data in the Our World In Data global database are compiled from official sources including health ministries, government reports, and official social media accounts. All these data sharing tools, although impactful, pose practical challenges related to data quality and completeness. For genetic sequences, impaired processing may lead to incorrect mutations (Hodcroft et al., 2021a), while for epidemiological data official sources may not communicate frequently enough or extensively enough about vaccination rollouts or case numbers (Ritchie et al., 2020).

2.4. Development of dedicated statistical methods to analyze these data

Epidemiological and genetic data represent the tip of the iceberg, hence parameters of interest (e.g., reproduction number, proportion of asymptomatic cases) are rarely directly measurable. Models are needed to describe how the underlying transmission process results in the observed data, and statistical methods are needed to fit models to empirical data and estimate model parameters.

2.4.1. Bayesian statistics

Bayesian statistics are popular in epidemiology because they provide a flexible inference framework that can adapt to complex situations. Contrary to classical statistics, Bayesian statistics treat model parameters as random variables which explicitly captures parameter uncertainty in estimations. The most probable range of values of model parameters is estimated by updating existing information on value ranges with observed evidence. Simply put, in Bayesian statistics we combine what we know with what we see to update our knowledge. This approach is consistent with the way we think and learn facilitating result interpretation (Hoekstra et al., 2014).

Bayes' theorem is the foundation of Bayesian statistics. When a statistical model with parameters θ is fitted to observed data x, Bayes' theorem stipulates that the joint posterior probability density of the model parameters given the data $p(\theta|x)$ is a weighted combination of the data likelihood $p(x|\theta)$ (joint probability of the observed data as a function of model parameters) and the joint prior probability density of model parameters $p(\theta)$ that represents prior beliefs on the values that model parameters can take on. The denominator $\int_{\theta'} p(x|\theta')p(\theta')d\theta'$ in equation 2 corresponds to the total evidence, that is to say the likelihood marginalized over all parameter values. Its calculation is generally not necessary to evaluate the posterior (Gilks et al., 1995).

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{\int_{\theta'} p(x|\theta')p(\theta')d\theta'} \tag{2}$$

The quantity of data used to inform the model determines the relative importance of the likelihood and the prior at informing the posterior. While the prior is less influential with more data, its choice remains important and may even be a pitfall in Bayesian analysis. Whenever possible, prior choice should reflect real prior information. If such information is not available, one can opt for a prior that does not exert much influence on the posterior (Kruschke, 2014).

2.4.2. Markov chain Monte Carlo

In epidemic modeling, models are often too complex to derive the analytical form of the total evidence preventing the calculation of the posterior distribution. Alternatively, numerical approximation of the posterior distribution is achievable using **Markov chain Monte Carlo** (MCMC) methods. These methods sample from the posterior distribution while exploring as efficiently as possible the parameter space. The Metropolis-Hastings algorithm (Hastings, 1970) is the most simple and commonly used MCMC algorithm (Algorithm 1). In a nutshell, the algorithm explores the parameter space using a random walk, at each iteration a new value of the parameter is proposed x^{cand} based on the previous value $x^{(i-1)}$ using a proposal distribution q, and the relative change in the posterior density is evaluated to determine whether x^{cand} is accepted. x^{cand} is always accepted when it leads to a higher posterior density, and it is accepted with probability α otherwise. This algorithm is stochastic to avoid getting stuck in local maxima, but it requires some time (burn-in) before converging to the highest density regions. Exploration efficiency relies on the variance of the proposal distribution which might necessitate some tuning.

Algorithm 1 Metropolis-Hastings algorithm

Initialize $x^{(0)} \sim q(x)$
for iteration $i = 1, 2, \ldots$ **do**
  Propose: $x^{cand} \sim q(x^{(i)}|x^{(i-1)})$
  Acceptance Probability: $\alpha(x^{cand}|x^{(i-1)}) = min\left\{1, \frac{q(x^{(i-1)}|x^{cand})\pi(x^{cand})}{q(x^{cand}|x^{(i-1)})\pi(x^{(i-1)})}\right\}$
  $u \sim \text{Uniform}(u; 0, 1)$
  if $u < \alpha$ **then**
    Accept the proposal: $x^{(i)} \leftarrow x^{cand}$
  else
    Reject the proposal: $x^{(i)} \leftarrow x^{(i-1)}$
  end if
end for

Convergence can be assessed visually (chains of parameter values as a function of iteration should look like white noise), or dedicated metrics such as the effective sample size (ESS) can help assessing convergence. In phylodynamics, convergence is presumed when ESS is higher than 200.

2.4.3. Data augmentation

Models describing the epidemic process sometimes rely on **latent variables**, variables that are not directly observable but meaningful such as the infection time. In such cases, the likelihood is convenient to write with the latent variables but the augmented data posterior $p(\theta|x, y)$ can be calculated only if latent data y are imputed (Tanner and Wong 1987). The Metropolis-Hastings algorithm can be adapted to sample from the posterior while marginalizing over the augmented data by iterating the two following steps (Neal and Kypraios, 2015):

1. Update θ given x and y by sampling from $p(\theta|x,y)$
2. Update y given x and θ by sampling from $p(y|x,\theta)$

The reconstruction of transmission chains in household studies from symptom onsets or molecular tests is a case study of data augmentation. Indeed, viral respiratory diseases are characterized by short or even negative serial intervals which makes the reconstruction of transmission chains hazardous. By augmenting the observed symptom onsets with the unobserved infection dates, one can integrate over all possible transmission chains and account for observation uncertainty (Cauchemez et al., 2004). Data augmentation is also applicable in the case of missing or incomplete data.

2.4.4. BEAST: a milestone for phylodynamics

In phylodynamics, the Bayesian framework allows the estimation of past demographics and, if desired, of past spatial spread from sequence alignments while accounting for phylogenetic uncertainty. The likelihood of Bayes' theorem (equation 2) corresponds to the probability of observing the sequence data given a phylogenetic history that is characterized by the substitution process, tree topology, branch lengths, and node geographic locations. Depending on the molecular clock, demographic, and spatial models, the likelihood can take on multiple forms. Let us consider the simplest configuration where the likelihood decomposes into the substitution model and the tree model (Bromham et al., 2018). The substitution model is fully specified by the nucleotide substitution parameters, whereas the tree model is characterized by the tree prior (i.e., the demographic model) and the molecular clock model that are in turn specified by hyperparameters. This complex multi-layer model structure is known as a **hierarchical model**. This is the key mathematical structure that allows the joint inference of epidemiological and evolutionary processes. Such as any other Bayesian approach, prior choice in terms of prior parameter distribution, tree prior, and molecular clock model is crucial to achieve unbiased inference (Bromham et al., 2018). Concerning model choice, model selection procedures are available but not for all models and their implementation can be very tedious (Baele et al., 2016). During the MCMC procedure, location- and time-stamped phylogenetic trees are sampled from the tree posterior distribution, and after sampling, the sampled trees are generally summarized using the maximum clade credibility (mcc) tree. This tree corresponds to the sampled tree that maximizes the product of the posterior clade probabilities.

Bayesian evolutionary analysis sampling trees (BEAST) is a software package that achieves such Bayesian phylodynamics inference. It comes in two different flavors, BEAST 1 (Suchard et al., 2018) and BEAST 2 (Bouckaert et al., 2019), that share the same Metropolis-Hastings MCMC core algorithm (Drummond and Rambaut, 2007), but BEAST 2 is more modular as it allows the integration of third

parties extensions. BEAST has gained a lot of popularity since its early development because it can be used as an all-in-one evolutionary toolbox. It is now an essential tool in outbreak investigation and response (Plessis and Stadler, 2015; Gardy and Loman, 2018; Hill et al., 2021).

3. Applications and challenges of epidemiological and phylodynamic modeling

Modelling is a means of estimating key epidemiological parameters such as disease transmissibility, severity, or transmission heterogeneity which cast light on the underlying epidemiological processes. Beyond allowing the better understanding of transmission dynamics, modeling is also a flexible tool for scenario comparison, notably when it comes to the evaluation of control strategies, which is a great source of scientific evidence for public health authorities (Fig. 7; Cauchemez et al. 2019).

Epidemiological and phylodynamic modeling can be used restrospectively to untangle the relative contributions of different drivers of the epidemic process, evaluate the efficacy of past control measures, and test alternative control and prevention measures. Such retrospective studies are generally held between outbreak periods which is more suitable to model development and validation. In a prospective way, modeling assists in the analysis of surveillance data and is a valuable tool for disease emergence and public health planification through forecasts and comparison of control strategies in hypothetical scenarios (Fig. 7A).

Here, we describe the prospective and retrospective applications of epidemiological and phylodynamic modeling to better understand transmission dynamics and guide public decision-making (Fig. 7B). Finally, we highlight their complementarity and specific challenges.

3.1. Understanding transmission dynamics

3.1.1. Quantifying transmission risk factors at the individual-level

Individual-level data are essential to characterize case heterogeneity and identify factors underlying it. Pathogen genetic sequences can also be valuable. We present here key epidemiological quantities that can be estimated at the individual level as well as the design of the studies allowing their estimation.

Case delays related to the natural history of the disease are easily estimated by fitting parametric probability density functions such as the gamma, Weibull, and lognormal distributions, to interval-censored data on the timings of symptom onset and exposure. The incubation period (time from infection to

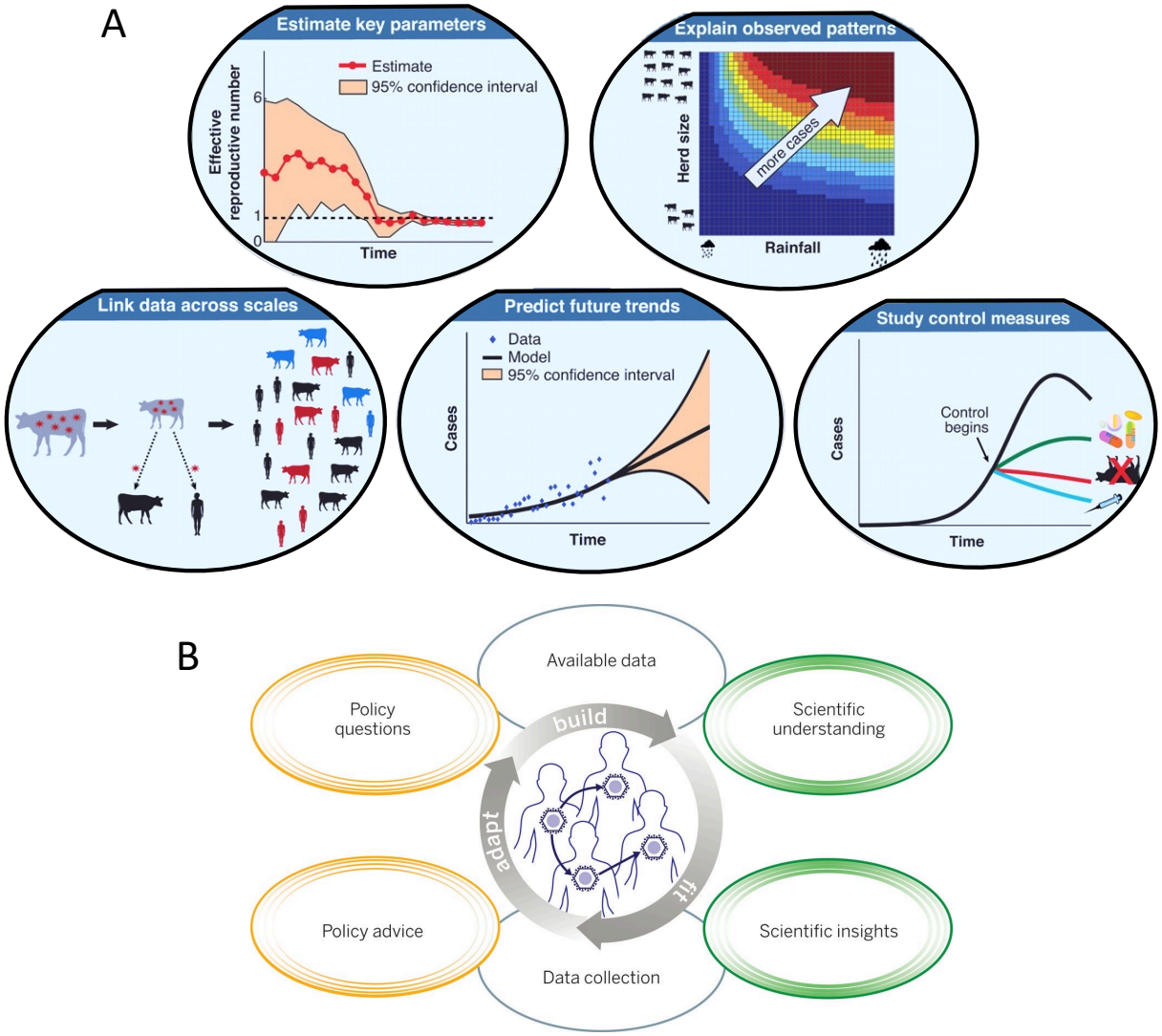


Figure 7: Applications of epidemiological and phylodynamic modeling. (**A**) Epidemiological and phylodynamic modeling allow to estimate key transmission parameters, link data across scales, explain observed patterns by unravelling the drivers of spread, predict future trends, and help design control measures (adapted from Lloyd-Smith et al. 2009). (**B**) Modelling is a powerful tool for public health by improving scientific understanding and advice policies. One can design a model to address policy questions. The model structure then relies on the current scientific understanding of the transmission process and the available relevant data. The model may require changes during the model validation and model fit steps. Sensitivity and uncertainty analyses may provide information on additional data that could be collected (adapted from Heesterbeek et al. 2015).

symptom onset) is typically estimated by fitting these density functions to individual data on the timing of exposure and symptom onset. The shape of the distribution gives an insight on individual heterogeneity. For example, the incubation period of rabies in dogs is highly variable with a mean of 22 days, but a non-negligible fraction of cases develop symptoms more than three months after infection (Hampson et al., 2009). The estimation of the incubation period can also guide the isolation and case management policies such as during the Monkeypox pandemic in 2022 (Miura et al., 2022). Two other delays are of particular interest: the serial interval (time between onsets of a case and their infector) derived from symptom onset data in transmission pairs, and the generation time (time between the dates of infection of

a case and their infector) that generally requires pre-existing estimates of the incubation period and relies on household data (Hart et al., 2022) or transmission pairs (Ferretti et al., 2020). These two delays are key to estimate time-varying reproduction numbers from incidence data (Ganyani et al., 2020). Although the serial interval may lead to biased estimates compared to the generation time, it is easier to estimate (Thompson et al., 2019).

Individual heterogeneity may not only concern the timing of the transmission process but also the magnitude of transmission. **Superspreading events** are an extreme case of transmission heterogeneity as they result from the unusual infection of many individuals by a single case. Detailed epidemiological data on clusters of cases allow to identify potential superspreading events but pathogen genetic sequences are much more informative to discriminate multiple independent introductions from within-cluster transmission, and homogeneous transmission from superspreading. Lemieux et al. (2021) analyzed the SARS-CoV-2 genomes of all confirmed early cases in Massachusetts between March and May 2020 along with cases from putative superspreading events. They reveal that one of the first lineages introduced in Massachusetts is associated with the largest superpreading event at an international business conference and led to sustained community transmission, circulation in homeless and higher-risk communities, and was exported internationally. This study exemplifies the role of superspreading events in explosive transmission that generally require targeted control policies (Lloyd-Smith et al., 2005). The low genetic diversity that characterizes cluster transmission and that is exploited to unravel superspreading events can also arise from over-sampling. In this case, it does not reflect epidemiological processes and can play as a confounding effect of superspreading events (Dearlove et al., 2017).

Finally, the study of age-related heterogeneity of respiratory disease transmission like influenza has greatly benefited from **household transmission modeling**. Children are shown to be more susceptible than adults to influenza infection but they are as infectious as adults (Cauchemez et al., 2004; Cauchemez et al., 2009; Cauchemez et al., 2014). Even when children have high pre-season haemagglutination-inhibition antibody titers indicating partial protection from past infections, they are more susceptible than adults (Cauchemez et al., 2014). These results have been replicated for multiple types of influenza viruses (seasonal influenza and 2009 pandemic influenza A) as well as in different settings (European and Asian households) highlighting that children are the entry door of influenza infections in households (Endo et al., 2019). This supports targeted influenza vaccination campaigns in school-aged individuals.

3.1.2. Quantifying transmission risk factors at the population-level

3.1.2.1. Dynamics related to the host

Host mobility and contact patterns influence disease spread at the population-level and may drive disease emergence in new host species or in new geographic areas. Phylodynamics can be used to point the shaping role of host mobility. For instance, molecular dating suggests that Zika was first introduced in the Americas when air passenger flows increased from endemic countries that were experiencing Zika outbreaks at the time (Faria et al., 2016). Phylogeographic techniques have also unraveled the modes of persistence of Ebola in the three most affected countries, Guinea, Liberia, and Sierra Leone, during the 2014 West African epidemic (Dudas et al., 2017). Transmission in Sierra Leone resulted from sustained local transmission after a single introduction, while the Guinean epidemic was fed by infrequent introductions from Libera and Sierra Leone. Interestingly, Dudas et al. (2017) also showed that younger populations were drivers of importations. Highlighting differential epidemic dynamics and drivers of spread is necessary to carry out appropriate control measures.

Metapopulation models explicitly incorporate the spatial structure of the host population which generally leads to the better understanding of disease spread. For example, the comparison of alternative metapopulation models of rabies transmission in dogs in Tanzania shows that the distance between villages is a strong predictor of disease spread whereas village size has little impact (Beyer et al., 2011). This supports gradual spatial diffusion in Tanzania rather than long-distance introductions suggested by landscape phylogeography approaches in North Africa (Dellicour et al., 2017). Similarly, age-structured SIR models parameterized with setting-specific contact patterns such as national contact heterogeneities are more realistic and better describe disease spread at large spatial scales as shown for influenza (Mistry et al., 2021).

Understanding the transmission dynamics between different **host reservoirs** in the context of zoonoses is necessary to assess the epidemic risk and the most adapted control measure. Depending on their evolutionary stage, zoonoses lead to either stuttering or sustained transmission in human (Lloyd-Smith et al., 2009) and discriminating these two categories of transmission is crucial in the context of emergence. Phylodynamic approaches supported the hypothesis of sustained inter-human transmission following a single introduction from the animal reservoir during the 2014 West African Ebola epidemic (Gire et al., 2014), whereas they clearly demonstrated stuttering chains originating from recurrent introductions from camel populations for MERS-CoV in Saudi Arabia (Dudas et al., 2018). Due to limited data in animal populations, it is particularly challenging to identify the most important animal reservoirs in disease

maintenance. Beyer et al. (2011) nevertheless argued using a metapopulation model that dog rabies does not originate from wildlife but from sustained and low-grade transmission in highly spatially-structured dog populations. In Iran, the picture is quite different as demonstrated by Dellicour et al. (2019). Using phylogeography, they showed that multiple viral lineages circulate independently in both domestic dogs and wildlife reservoirs.

3.1.2.2. Dynamics related to the pathogen

The analysis of pathogen genetic sequences provides valuable information on the role of pathogens' genetic diversity on their **transmission potential**, and thereby their epidemic growth. The relative transmissibility and **virulence** of new variants should be rapidly assessed when the first cases arise, such as during the COVID-19 pandemic, but it is not an easy task. Indeed, the evolutionary theory predicts that pathogens' fitness is maximal at intermediate levels of transmissibility and virulence. However, there is scarce empirical evidence of a trade-off between transmissibility and virulence, besides the drivers of this trade-off are unknown (Acevedo et al., 2019).

Practically speaking, new variants of SARS-CoV-2 are identified using phylogenetic approaches (Rambaut et al., 2020) and their transmissibility is quantified by estimating variant-specific growth rate or effective reproduction numbers from variant frequency data (Volz et al., 2021; Davies et al., 2021). These methodologies allowed to identify the repeated emergence of new variants during the pandemic and monitor their relative transmissibility. Most of them were shown to be more transmissible (Campbell et al., 2021), but virulence was significantly greater only for the Delta variant.

Other diseases are subject to the emergence of new strains with higher transmissibility. For example, Wymant et al. (2022) have recently shown that a new subtype-B HIV strain is currently circulating in the Netherlands. Patients infected with this subtype exhibit higher viral loads and a faster decline of CD4 cells indicative of a greater virulence. Phylodynamic analyses suggest that the new subtype emerged at the end of the 20th century and is more transmissible compared to other transmission clusters. Similar to the Delta variant of SARS-CoV-2, the new subtype-B HIV strain is more transmissible and more virulent.

3.1.2.3. Dynamics related to the environment

The incorporation of environmental factors in epidemiological or phylodynamic modeling allows to identify and sometimes quantify the role of **ecological processes** on the dynamics of infectious diseases. For example, cholera is a waterborne bacterial infection generally causing outbreaks after floods or monsoons in tropical areas. Koelle et al. (2005) have confirmed this climatic forcing by comparing environmental changes to time-varying transmission rates estimated from a four-decade time-series of cholera cases in

Bangladesh with a model that accounts for the number of susceptible individuals, immunity decay, and seasonal transmission. Landscape phylogeography can also unravel key landscape features in infectious disease circulation. For example, mean annual temperature is strongly and positively associated with WNV lineage dispersal velocity and variations in viral genetic diversity in the USA (Dellicour et al., 2020). Counter-intuitively, migratory bird flyways are not identified as correlates of viral spread which suggests that local mosquito and bird populations play a major role in disease circulation.

3.2. Modelling for public decision-making

3.2.1. Early-warning signals and surveillance

In addition to better understanding the transmission process, modeling can exploit surveillance data to estimate basic and effective reproduction numbers to inform public health authorities on the current and near future epidemiological situation.

At the start of an epidemic, $\boldsymbol{R_0}$ informs on the transmission potential of the disease, in other words, whether large and explosive outbreaks should be expected, which is highly valuable in a context of high uncertainty such as disease emergence. During the 2014 West African Ebola outbreak, Althaus (2014) estimated R_0 for Sierra Leone, Guinea, and Liberia, by fitting a simple SEIR model to incidence data. The maximum likelihood estimate for Sierra Leone was almost twice as high as for the two other countries. However, the estimation procedure required fixing the average duration of the incubation and infectious periods to previous estimates from an outbreak in Congo in 1995. Alternatively, birth-death approaches that rely on pathogen genetic sequences can jointly estimate R_0, the incubation period, and the infectious period. This way, Stadler et al. (2014) have estimated a similar R_0, however the median estimate of the infectious period was not in agreement with other estimates (Team, 2014), and sample size was too small to estimate the incubation period with high certainty. Overall, both approaches showed that transmission was more intense in Sierra Leone compared to Guinea and Liberia.

Rapidly after disease introduction, the depletion of susceptible individuals due to past infection or control measures should be accounted for in the calculation of the reproduction number that becomes the **effective reproduction number $\boldsymbol{R_e}$**. Monitoring R_e over an epidemic provides evidence of changes in transmission over time due to changes in pathogen transmissibility, host mobility, or the implementation of control measures. Bourhy et al. (2016) have quantified the instantaneous R_e of dog rabies in an endemic African city, Bangui, over more than one decade. R_e rarely exceeded the critical value of 1, even during epidemic waves, indicating no self-sustained transmission and suggesting that control measures

were effective. This analysis was done retrospectively but R_e can be estimated in real-time which can help planning outbreak response. Indeed, outbreaks usually increase pressure on health systems, especially on hospitals. Public health authorities can use predictions on the maximal number of beds for outbreak management to organize patient care in advance without overwhelming the rest of the health system. Predictions for response planning are particularly useful during pandemic times or in settings with little care capacities. For example, Andronico et al. (2017) have assessed a few weeks ahead using a compartmental model that the Zika outbreak in Martinique would require at most eight intensive care beds and seven ventilators for individuals with Guillain-Barré syndrome. R_e can also be estimated using phylodynamic modeling. A structured birth-death model fitted to the first Omicron genomes sequenced in South Africa has allowed the estimation of R_e that was up to 3.6 indicating intense community-transmission and the need for rapid response. This was further supported by continuous phylogeography suggesting rapid spatial spread (Viana et al., 2022).

The COVID-19 pandemic has facilitated and accelerated the development of two new types of surveillance systems: **genomic surveillance** and **wastewater-based surveillance**. Genomic surveillance is either deployed at the national level (Wilkinson et al., 2021) or targets higher-risk groups like heathcare workers (HCWs) and patients at hospitals (Meredith et al., 2020). National-level genomic surveillance allows to follow the emergence of new variants, and targeted genomic surveillance informs clinical, infection control, and hospital management teams to improve infection-control interventions in hospital settings (Meredith et al., 2020). However, genomic surveillance like surveillance systems based on case numbering remains costly. An alternative strategy relies on the monitoring of pathogen genetic material in wastewater resulting from fecal shedding and reflecting community-level transmission. The COVID-19 pandemic has really stimulated wastewater-based surveillance as well as the development of methodologies based on compartmental models (McMahan et al., 2021; Nourbakhsh et al., 2022) or deconvolution (Huisman et al., 2022a) to robustly estimate R_e from wastewater pathogen load.

3.2.2. Evaluating (non)pharmaceutical interventions

When the main drivers of disease spread are known and a transmission model that accurately captures the epidemic process is available, modeling and more specifically epidemiological modeling can be used to design and evaluate *in silico* the potential impact of **vaccination** and **nonpharmaceutical interventions**. Nonpharmaceutical interventions are also known as community mitigation strategies and encompass social distancing measures (e.g., quarantine, curfews, lockdowns, closure of places where people gather), travel bans, and combined testing and quarantine measures. Intervention evaluation can be done retro-

spectively to validate their implementation, or prospectively to guide decision-making.

Predicting the impact of **vaccination** and designing the best campaign to avoid spread provide incentives to governments or health authorities to take action, continue their control efforts, or strengthen them, particularly in the context of neglected tropical diseases. Mancy et al. (2022) reconstructed transmission chains from contact tracing data in dog populations in Serengeti district, Tanzania, and showed that with negligible vaccination coverage rabies would very probably circulate 10 to 30 weeks in dog populations after a single introduction. This argument strongly supports current vaccination efforts. The effect of vaccination can also be evaluated at the individual level. For example, Tsang et al. (2019) evaluated direct and indirect vaccine effectiveness against influenza infection in households and showed that even when other household members are vaccinated the individual benefit of vaccination still stands.

Community-level social distancing interventions have been evaluated worldwide during the COVID-19 pandemic. Compartmental models are an essential tool to do so. They allowed among others to quantify the effect of curfew (Andronico et al., 2021), school closures (Flaxman et al., 2020), or age-stratified lockdowns (Roche et al., 2020). More simple approaches that rely on the estimation of the instantaneous R_e allowed to dissect the most effective control strategies at a time when governments adapted control interventions on a monthly basis (Davies et al., 2020b; Huisman et al., 2022b). Testing strategies were also in the radar but studied with more complex tools like agent-based models. For example, Pullano et al. (2021) showed that most cases were not detected during the first wave of the pandemic in France, and that detection rates were considerably heterogeneous across regions. Quilty et al. (2021) investigated the potential beneficial impact of testing to reduce quarantine and highlighted that decreased test and trace delays and increasing adherence would improve the effectiveness of combined testing and quarantine measures.

So far, very few studies attempted to use phylodynamics to evaluate nonpharmaceutical interventions. Dellicour et al. (2018a) developed a methodology to assess the impact of hypothetical interventions using estimates of a discrete phylogeographic model. This way, they showed that preventing short-distance transmission would have drastically impeded the geographic expansion and epidemic size of the 2014 West Africa Ebola outbreak. However, these results do not translate easily into real world intervention measures.

3.3. Complementarity of epidemiological and phylodynamic modeling

In addition to relying on data of different nature, epidemiological and phylodynamic modeling provide complementary approaches for the study of infectious diseases spread. Phylodynamics opens the door

to the analysis of pathogen-related drivers of the epidemic process and brings an evolutionary standpoint to epidemiology. One can estimate viral fitness, study adaptive evolution to the host, and discriminate local transmission from imports. It also reliably infers transmission routes in cluster investigation, and thereby, cross-validates transmission chains or even identifies new epidemiological links. However, epidemiological modeling remains the tool of choice to perform real-time analyses, forecast epidemics, and test hypothetical intervention scenarios. Consequently, the evaluation of (non)pharmaceutical interventions to guide decision-making remains more amenable and detailed with epidemiological modeling.

Importantly, the COVID-19 pandemic has illustrated the need for interdisciplinary studies that combine genetic and epidemiological data analyzed with classical statistical analyses, epidemiological, and phylodynamic modeling to gain broad insights on the drivers and characteristics of the epidemic process (Viana et al., 2022; Müller et al., 2021).

3.4. Assessing inference bias, a key challenge in epidemiological and phylodynamic modeling

As in any statistical analysis, epidemiological and phylodynamic modeling can be affected by **statistical biases** that lead to a difference between the true parameter value and the expected value obtained with the estimator. Epidemiologists tend to use this term to indicate any phenomenon that leads to a systematic difference between the true parameter value and the one estimated from the data. These phenomena can arise at data collection when sampled data do not reflect the underlying epidemic process. This particular phenomenon is called sampling bias. Statistical biases can also arise at the model specification level when models do not account for a process that drives epidemic dynamics or are overparameterized, meaning that they try to estimate many correlated parameters. In this section, we focus on the inference challenges related to statistical biases in phylogeography and household transmission models.

3.4.1. Potential impact of sampling bias in phylogeography

The availability of pathogen genetic sequences was a limiting factor in phylodynamic analyses for a long time. Sample collection was mostly opportunistic, relying on existing surveillance systems and not planned as part of a grounded road map. Opportunistic sampling certainly allows rapid and low cost data collection and generation, but it also implies a high risk of sampling bias with sequences that are not reflective of outbreak dynamics (Hill et al., 2021). For example, some areas may be over- or under-sampled, and in extreme cases, they may even be left unsampled while they experience intense circulation. In other situations, many samples are available but all of them cannot be sampled due to limited financial resources as it is still the case for neglected diseases, or there is a known sampling

bias as this was the case during the COVID-19 pandemic (Hodcroft et al., 2021a). Samples to analyze should be selected according to a rationale. So far, subsampling strategies used external data such as incidence (Candido et al., 2020; Lemey et al., 2020) or hospitalization (Dellicour et al., 2021b) data to guide sample composition. Importantly, subsampling allows to save computational resources as Bayesian phylogeographic analyses are computationally-demanding and very large data sets do not necessarily provide more accurate results (Magee and Scotch, 2018).

Some studies have investigated the impact of spatial sampling bias on phylogeographic analyses. In discrete phylogeography, CTMC has been argued to be sensitive to spatial sampling bias whereas the structured coalescent model is robust to bias and its performances are maximized when the sample composition is even (De Maio et al., 2015). However, there is no comprehensive understanding of how sampling bias acts on discrete phylogeography inference, neither the ways this impact could be mitigated. In continuous phylogeography, the picture seems more clear. Indeed, Brownian Motion phylogeography is strongly affected by the lack of sampling in certain areas, but this can be mitigated by adding sequence-free cases to inform the model (Kalkauskas et al., 2021). On the other hand, the Λ-Fleming-Viot process, another model of continuous phylogeography that is more rarely used, is robust to spatial sampling bias but is more adapted to endemic long-term transmission (Kalkauskas et al., 2021).

So far, three strategies have been used to assess or mitigate the impact of sampling bias. First, multiple models are tested on the same data set and results are compared. This has been particularly implemented for discrete phylogeographic analyses in which researchers compared CTMC to the structured coalescent model (Faria et al., 2017; Brynildsrud et al., 2018; Yang et al., 2019; Mavian et al., 2020; Dudas et al., 2018). The best case scenario is when results across methods are in good agreement (Faria et al., 2017; Brynildsrud et al., 2018; Yang et al., 2019; Mavian et al., 2020). When this is not the case, the experimenter expectations may strongly orient model choice. Second, at least one model is tested on multiple subsamples. However, this approach is feasible only when enough samples are available, and there is currently no evidence to prefer one subsampling strategy over the others. Third, statistical methodologies are adapted to account for potential unrepresentative sampling. For example, Chaillon et al. (2020) corrected the metric of statistical support of migration rates between discrete locations in CTMC to accomodate for spatial sampling frequency and identify migration rates that participate the most to the epidemic process and are not just supported due to more intense sampling. In a recent study, Guindon and De Maio (2021) explicitely modelled the sampling strategy and showed that the assumptions on the sampling strategy may radically impact the reconstruction of past spatial dynamics.

In brief, **spatial sampling bias** is a key challenge in phylogeographic analyses. There is still a need to better characterize its impact. New methodologies should be developed in parallel to account for the potential impact of bias.

3.4.2. Model misspecification and overparameterization

3.4.2.1. In discrete phylogeography

The number of parameters to estimate in a discrete phylogeographic analysis exceeds the amount of information contained in the spatial location of the analyzed sequences that is condensed in a single observation. In CTMC, $n(n-1)$ parameters from the asymmetric rate matrix are to be estimated, with n the number of discrete locations included in the analysis. In the structured coalescent model, n deme sizes are to be estimated in addition to the $n(n-1)$ migration rates. There is very low chance that all migration events occurred in a given sample (Lemey et al., 2009a; Bloomquist et al., 2010). Bayesian stochastic search variable selection (BSSVS) circumvents this overparameterization issue by associating an indicator variable that has a Dirac delta distribution (i.e., equals to 0 or 1) to each migration parameter and by exploring in the MCMC all combinations of parameters that explain the most the diffusion process. However, this inference workaround does not necessarily imply correct inference as Gascuel and Steel (2020) have recently shown that it is not possible to estimate both migration rate parameters and node locations with the CTMC model.

The structured coalescent model faces another model misspecification challenge, the ghost deme issue (Ewing and Rodrigo, 2006; Beerli, 2004). Ghost demes are discrete locations that are not specified in the structured coalescent model despite intervening in the diffusion process. Accounting for ghost demes when they actually don't exist and not accounting for them when they actually exist lead to biased estimations of migration rates. To model spatial diffusion with a structured coalescent model, one should first make sure that the epidemic occurs in a closed environment and that all discrete locations affected by the epidemic were sampled. The impact of ghost demes in CTMC has not been tested yet, but would very certainly lead to biased inference.

In the end, the **inference abilities** of discrete phylogeographic models are little known and the impact of model misspecification remains poorly characterized.

3.4.2.2. In models of household transmission

Household transmission models generally hypothesize homogeneous mixing between household contacts (Longini et al., 1988; Cauchemez et al., 2004; Prunas et al., 2022; Dattner et al., 2021) although

Goeyvaerts et al. (2018) showed that household members do not necessarily contact each other at random. So far, there has been no study on the potential impact of misspecification of within household contact patterns in the estimation of transmission parameters. Furthermore, contact patterns are expected to vary across settings and contact heterogeneities cannot be easily deduced.

Thesis objectives

This thesis aims at better understanding viral disease spread using either highly detailed individual-level data or coarse pathogen genetic sequences data at large spatial scale that correspond to the two ends of the data granularity gradient presented in Section 2.2 of the introduction. The objective is not only to gain insights from empirical data by using adapted quantitative approaches, but also to compare the evidence that can be drawn from individual-level data and pathogen genetic sequences. Another aim of this thesis is to investigate in a more theoretical framework the limitations of the quantitative methodologies used to analyze these data, epidemiological modeling for individual-level data and phylodynamic approaches for pathogen genetic sequences, notably concerning missing data and sampling bias. These objectives are addressed by focusing on two case studies of viral diseases: RABV in domestic dog populations and SARS-CoV-2 in households. RABV circulation in animal populations has been studied since the developments of epidemiological modeling to understand its spread and help design effective control strategies but it remains a neglected tropical disease for which limited resources and data are available. Dogs are the main reservoir in Africa and Asia but the drivers of RABV spread in these populations remain poorly characterized. Concerning SARS-CoV-2, it has caused a major pandemic and has led to massive lockdowns and vaccination campaigns around the world. Its high transmission potential coupled to its fast evolution necessitated rapid understanding of the drivers of spread and rapid assessment of control strategies. It has led to an unprecedented scientific production in virology, immunology, epidemiology, evolutionary biology, medical anthropology etc. Although the modes of transmission of RABV and SARS-CoV-2 differ, similar quantitative analytical tools can be applied to better understand the drivers of their spread at the individual and population levels. This thesis is divided into five chapters that address specific questions related to the disease (rabies or COVID-19), methodological choice (phylodynamics or epidemiological modeling), and setting (large spatial scale or individual-level).

First, I review the epidemiological and phylodynamic modeling approaches applied to dog rabies using the rigorous methodology of a scoping review (**Chapter 1**). Indeed, the scoping review approach extends the scientific methodology of systematic reviews for article search, selection, and analysis to a heterogeneous scientific production. In this review, I explore the uses of modeling approaches and synthesize their contributions to the understanding of RABV circulation and control. I highlight their complementarity in a context of limited resources and potentially very biased sampling, without omitting their limitations, and I propose future directions to broaden their inputs.

Second, I reconstruct possible transmission chains from empirical data at the country level to characterize RABV endemicity in Cambodia (**Chapter 2**), and at the household level to evaluate COVID-19 vaccine effectiveness (**Chapter 3**). In Chapter 2, I analyze a novel RABV genome data set that covers the whole country by using phylogeographic approaches to assess the role of intra-country transmission versus external introductions from neighboring countries, compare circulation in Cambodia to other settings, and investigate the landscape features associated with RABV spread. In Chapter 3, I estimate the secondary attack rates of SARS-CoV-2 stratified by age, vaccination, and isolation status in Israeli households. I further apply an epidemiological model to quantify the transmission probability, as well as vaccine effectiveness against transmission and infection.

Third, I investigate the limitations of phylodynamic (**Chapter 4**) and epidemiological (**Chapter 5**) modeling using simulations when the determinants of the epidemic process are unknown. In Chapter 4, I assess the impact of spatial sampling bias on discrete phylogeographic reconstruction when viral spatial dynamics are unknown. To do so, I simulate RABV epidemics in dog populations in Morocco and proceed to biased or unbiased sampling. Then, I reconstruct spatial dynamics using different discrete phylogeographic algorithms before assessing their relative inference performance. I also showcase how discrete phylogeographic algorithms lead to contrasting results on empirical data sets. In Chapter 5, I investigate how a major assumption in models of disease transmission in households impacts the inference of transmission heterogeneity at the individual-level. Indeed, household transmission models generally assume homogeneous mixing between household contacts which is not necessarily true. In this study, I use simulated epidemics under a heterogeneous mixing scenario to quantify the inference bias of age-dependent transmission.

Chapter 1

Epidemiological and phylodynamic modeling of dog rabies

Rabies is a fatal yet vaccine-preventable disease. In the last two decades, domestic dog populations have been shown to constitute the predominant reservoir of rabies in developing countries, causing 99% of human rabies cases. Despite substantial control efforts, dog rabies is still widely endemic and is spreading across previously rabies-free areas. Developing a detailed understanding of dog rabies dynamics and the impact of vaccination is essential to optimize existing control strategies and develop new ones. In this scoping review, we aimed at disentangling the respective contributions of epidemiological models and phylodynamic approaches to advancing the understanding of rabies dynamics and control in domestic dog populations. We also addressed the methodological limitations of both approaches and the remaining issues related to studying rabies spread and how this could be applied to rabies control. Through a detailed search of the PubMed, Web of Science, and Scopus databases, we identified a total of n = 59 relevant studies using epidemiological models (n = 30, referred to mathematical models hereafter), phylodynamic inference (n = 22) and interdisciplinary approaches (n = 7). We found that despite often relying on scarce rabies epidemiological data, mathematical models investigated multiple aspects of rabies dynamics and control. These models confirmed the overwhelming efficacy of massive dog vaccination campaigns in all settings and unraveled the role of dog population structure and frequent introductions in dog rabies maintenance. Phylodynamic approaches successfully disentangled the evolutionary and environmental determinants of rabies dispersal and consistently reported support for the role of reintroduction events and human-mediated transportation over long distances in the maintenance of rabies in endemic areas. Potential biases in data collection still need to be properly accounted for in most of these analyses. Finally, interdisciplinary studies were determined to provide the most comprehensive assessments through hypothesis generation and testing. They also represent new avenues, especially concerning the reconstruction of local transmission chains or clusters through data integration. Despite advances in rabies knowledge, substantial uncertainty remains regarding the mechanisms of local spread, the role of wildlife in dog rabies maintenance, and the impact of community behavior on the efficacy of control strategies including vaccination of dogs. Future integrative approaches that use phylodynamic analyses and mechanistic models within a single framework could take full advantage of not only viral

sequences but also additional epidemiological information as well as dog ecology data to refine our understanding of rabies spread and control. This would represent a significant improvement on past studies and a promising opportunity for canine rabies research in the frame of the One Health concept that aims to achieve better public health outcomes through cross-sector collaboration.

1. Introduction

1.1. Background

Rabies is a viral zoonosis affecting the central nervous system of mammals that is almost always fatal to humans. Domestic dogs represent the main reservoir of rabies virus (RABV) worldwide. After infection, that occurs through the bite of an infected animal, RABV travels through the nerves to the spinal cord and brain. This incubation period lasts on average 3 weeks in dogs (Hampson et al., 2009). Symptoms appear when RABV reaches the brain and can last up to ten days (Hampson et al., 2009). There are two classical forms of rabies, furious or paralytic (World Health Organization (WHO), 2018). While dogs affected by the furious form become aggressive and attack with no trigger, dogs affected by the paralytic form develop paralysis and have difficulty swallowing leading to hypersalivation.

Dogs are responsible for 99% of human rabies cases (World Health Organization (WHO), 2018). In-depth understanding of dog ecology and host-pathogen interactions is necessary to characterize rabies dynamics and design appropriate control measures. Rabies is a vaccine-preventable disease in both human and canine populations, and dog vaccination is the most cost-effective control measure (Anothaisintawee et al., 2019). Strong evidence is available for the efficacy of dog rabies elimination programs in endemic areas (Cleaveland and Hampson, 2017; Lembo et al., 2010; Arechiga Ceballos et al., 2014; Cleaveland et al., 2014; Cleaveland et al., 2018), notably in South America where massive dog vaccination campaigns in the 1980s alleviated the burden of canine rabies. Regardless, there has been only little improvement of the global burden since the successes in South America. Dog rabies is still endemic in Africa, Asia, and the Middle East (Hampson et al., 2015; Mbilo et al., 2021).

In 2015, the World Health Organization (WHO), the Global Alliance for Rabies Control (GARC), the World Organization for Animal Health (OIE) and the Food and Agriculture Organization of the United Nations (FAO) launched a comprehensive framework targeting the global elimination of dog-mediated human rabies by 2030 (World Health Organization (WHO) et al., 2018). Effective One Health interventions such as the improvement of the current prophylaxis in both humans (Hampson et al., 2019b; Hampson et al., 2019b) and dogs should enable reaching this goal. Despite valuable efforts in several endemic countries (Mbilo et al., 2021; World Health Organization (WHO) et al., 2019; Tiembré et al., 2018), control strategies have not stopped rabies from circulating due to inadequate political, economic, and social responses. Weak interest from veterinary services, lack of sustainable resources and political neglect (Welburn et al., 2017) prevent most endemic countries to reach the 70% vaccination coverage

recommended by the WHO (Mbilo et al., 2021). Moreover, rabies infections continue to spread, notably in previously rabies-free areas in countries such as Indonesia (Townsend et al., 2013b; Dibia et al., 2015; Mahardika et al., 2014) and the Philippines (Tohma et al., 2014; Tohma et al., 2016). In this resource-limited context, in-depth knowledge of the mechanisms underlying rabies dynamics (environmental drivers of spread, impact of dog density, impact of dog behavior, etc.) would be a key asset to limiting the spread of this vaccine-preventable disease, notably by aiding to design more effective vaccination campaigns that are robust to resurgence in the long-term. The development of novel methodologies to better understand rabies epidemiology and transmission dynamics therefore constitutes a promising avenue of research.

1.2. Objectives

In this scoping review, we focused on the insights of two quantitative approaches applied to the study of rabies: mathematical modeling of infectious diseases and phylodynamics. The former is a field of research that exploits epidemiological data to unravel the spread of diseases in populations, assess the impact of interventions, support policy making, and optimize control strategies. The latter studies the interactions between epidemiological, immunological, and evolutionary processes from the analysis of viral genetic sequence data (Volz et al., 2013). Within phylodynamics, phylogeographic inference specifically aims at reconstructing the dispersal history and dynamics of viral lineages in space and time. Here, we assessed the uses and respective contributions of both approaches, as well as their limitations and the remaining knowledge gaps concerning rabies dispersal and control in domestic dog populations.

2. Methods

2.1. Search strategy

This review follows the guidelines of the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) statement for scoping reviews (Tricco et al., 2018). In this review, we screened PubMed, Web of Science and Scopus databases on the 2nd of June, 2020 using the following combination of terms [“rabies” AND (“dog” OR “canine”) AND (“modelling” OR “modeling” OR “phylogeography” OR “phylodynamics”) AND “dynamics”] along with the “all fields” option and without restriction on publication year. The “all fields” option enabled to apply the search terms for their appearance in the title, abstract and keywords. Only English-written papers published in scientific journals were considered. All data were searched and screened by the same researcher

(ML). The search strategy identified 65, 94 and 768 publications in PubMed, Web of Science and Scopus databases respectively, which corresponded to 797 unique records. In addition, references of selected publications were screened manually, leading to the identification and inclusion of two additional studies (Coleman and Dye, 1996; Kitala et al., 2002) that were not identified during the searching step. Finally, the paper of Colombi et al. (Colombi et al., 2020), which was not identified in the databases nor in the references, was also included (Fig. 8).

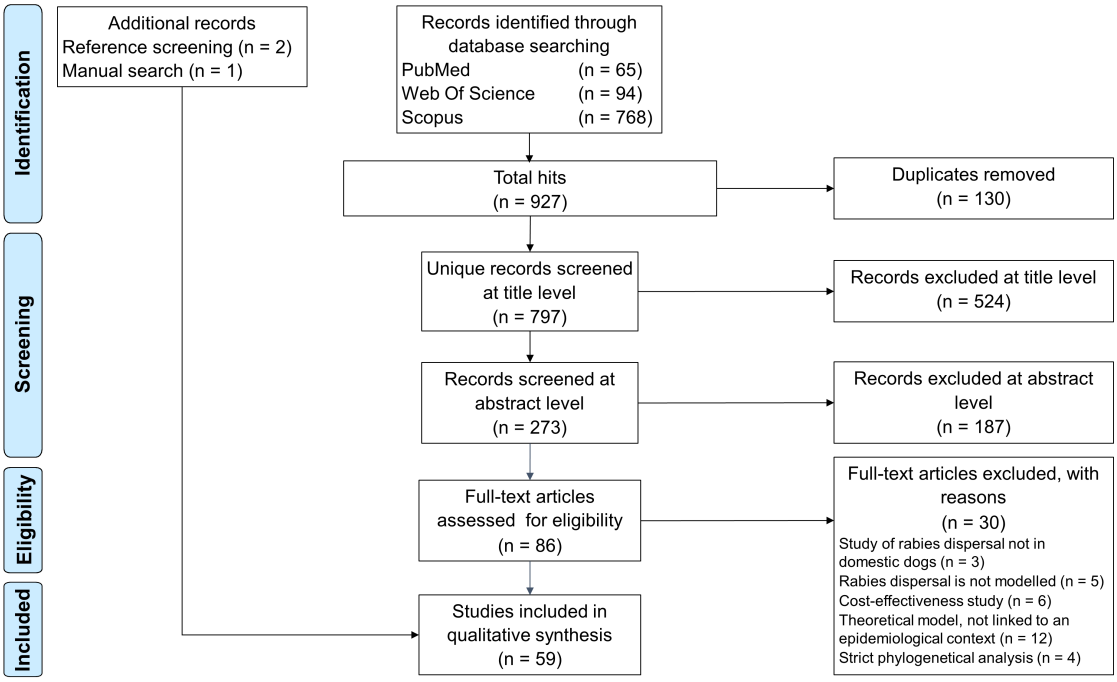


Figure 8: PRISMA-ScR Flow Diagram showing the number of identified and selected records along the multi-stage selection process. Scopus accounted for most of the records as it retrieved 71% (n = 46) of PubMed records and 79% (n = 74) of Web of Science records.

2.2. Selection of studies

In total, 797 records were included and processed manually in a multi-stage procedure. At each selection step, a conservative approach was taken to ensure the best sensitivity level. Firstly, studies were selected based on their title using the following inclusion criteria: mathematical models of dog and human rabies assessing the impact of control strategies, the risk of rabies importation, the drivers of rabies spread or models estimating epidemiological parameters, cost-effectiveness studies, phylodynamic studies including RABV isolated from dogs, and broad studies on new phylodynamic or mathematical models. Indeed, rabies has often been used as a model disease in phylodynamics and mathematical modeling, and a reference to rabies might not appear directly in the title or the abstract. The following exclusion criteria were used: reviews, studies strictly on wildlife rabies, dog ecology and population dynamics, conservation biology, and evolutionary analyses for diagnostic purposes. Secondly, studies were selected based on their

abstract with a refined set of exclusion criteria to exclude statistical analyses of epidemiological data, cost-effectiveness studies with no focus on rabies dynamics, experimental rabies cross-species transmission which did not incorporate a modeling aspect and studies on the evolutionary processes of RABV. Finally, studies went through a full-text reading step to verify that their content matched our selection criteria. At this step, theoretical models which were not grounded in a specific epidemiological context were excluded (Fig. 8).

2.3. Data extraction and analysis

Selected studies were classified into three categories based on their methodology: mathematical models, phylodynamic and interdisciplinary studies. Most phylodynamic studies identified in this review correspond to phylogeographic analyses, where the main focus is on inferring the spread of a pathogen over time using location data associated with the available sequence data. The interdisciplinary category covers studies either integrating epidemiological and genetic data in a unified modeling framework or mixing modeling approaches with phylodynamics. Data were systematically charted in an Excel spreadsheet designed to retrieve: i) the main modeling strategy with its assumptions; ii) the data source; iii) remarks about potential bias of the data in relation to the underlying evolutionary and epidemiological processes; iv) the qualitative and quantitative results concerning the dynamics of dog rabies; and v) if performed, the sensitivity analysis determining the robustness of the methodology to parameter values or potential biases.

3. Results

3.1. General characteristics of selected studies

Our selection procedure identified 59 studies that meet our selection criteria with 30 mathematical models (Townsend et al., 2013b; Kitala et al., 2002; Coleman and Dye, 1996; Colombi et al., 2020; Zhang et al., 2012; Fitzpatrick et al., 2012; Beyer et al., 2012; Townsend et al., 2013a; Dürr and Ward, 2015; Ferguson et al., 2015; Chen et al., 2015; Sparkes et al., 2016; Leung and Davis, 2017; Laager et al., 2018; Kadowaki et al., 2018; Laager et al., 2019; Wilson-Aggarwal et al., 2019; Beyene et al., 2019; Abdul Taib et al., 2019; Huang et al., 2019; Hudson et al., 2019a; Brookes et al., 2019; Carroll et al., 2010; Ortega et al., 2000; Hampson et al., 2007; Zinsstag et al., 2009; Zhang et al., 2011; Beyer et al., 2011; Hou et al., 2012), 22 phylodynamic studies (Dibia et al., 2015; Tohma et al., 2014; Bourhy et al., 2008; Lemey et al., 2009a; Talbi et al., 2009; Meng et al., 2011; Hayman et al., 2011; Carnieli et al.,

2011; Yu et al., 2012; Mollentze et al., 2013; Guo et al., 2013; Carnieli et al., 2013; Horton et al., 2015; Brunker et al., 2015; Yao et al., 2015; Troupin et al., 2016; Zhang et al., 2017; Ma et al., 2017; Dellicour et al., 2017; Brunker et al., 2018b; Wang et al., 2019; Dellicour et al., 2019), and 7 interdisciplinary studies (Tohma et al., 2016; Talbi et al., 2010; Mollentze et al., 2014; Bourhy et al., 2016; Zinsstag et al., 2017; Cori et al., 2018; Tian et al., 2018), all published between 1996 and 2020 (Fig. 8 and Fig. 9A-B). Mathematical models were first published followed by phylodynamic and interdisciplinary studies (Fig. 9B). This timeline can be explained by the recent developments of Bayesian phylodynamic, and in particular phylogeographic, models in BEAST (Baele et al., 2016; Suchard et al., 2018; Bouckaert et al., 2019). Africa and Asia are the most studied continents in the three methodological categories, while China accounts for most of the Asian studies (Fig. 9C). Oceania is not represented in the interdisciplinary and phylodynamic categories since it is a rabies-free area (Fig. 9A).

3.2. Topics addressed by the studies

Phylodynamic studies are homogeneous in terms of methodologies (essentially phylogeographic studies) and research goals. They predominantly focus on unraveling the dispersal dynamics of rabies at the regional and country levels (n = 16; Dibia et al. 2015; Tohma et al. 2014; Bourhy et al. 2008; Lemey et al. 2009a; Talbi et al. 2009; Meng et al. 2011; Hayman et al. 2011; Carnieli et al. 2011; Yu et al. 2012; Carnieli et al. 2013; Brunker et al. 2015; Yao et al. 2015; Ma et al. 2017; Dellicour et al. 2017; Wang et al. 2019; Dellicour et al. 2019). In four of them, the authors deciphered the role of lineage introduction in rabies maintenance or emergence (Mollentze et al., 2013; Guo et al., 2013; Horton et al., 2015; Zhang et al., 2017). In recent years, researchers have been trying to identify external factors impacting the spatial dynamics of RABV spread (n = 5; Brunker et al. 2015; Yao et al. 2015; Dellicour et al. 2017; Brunker et al. 2018b; Dellicour et al. 2019; Fig. 9D and Table A1). Contrary to phylodynamic studies, the modeling category gathers a diverse panel of models with aims that cover the implementation of new mathematical methodologies (n = 2; Hudson et al. 2019b; Ortega et al. 2000), the characterization of rabies dynamics (n = 11; Zhang et al. 2012; Fitzpatrick et al. 2012; Ferguson et al. 2015; Chen et al. 2015; Abdul Taib et al. 2019; Huang et al. 2019; Brookes et al. 2019; Hampson et al. 2007; Zinsstag et al. 2009; Zhang et al. 2011; Hou et al. 2012), the identification of factors driving the resurgence or maintenance of rabies (n = 9; Townsend et al. 2013b; Kitala et al. 2002; Colombi et al. 2020; Sparkes et al. 2016; Leung and Davis 2017; Laager et al. 2018; Laager et al. 2019; Wilson-Aggarwal et al. 2019; Hudson et al. 2019a), the assessment of control strategies efficacy (n = 18; Townsend et al. 2013b; Kitala et al. 2002; Coleman and Dye 1996; Fitzpatrick et al. 2012; Beyer et al. 2012; Townsend et al. 2013a;

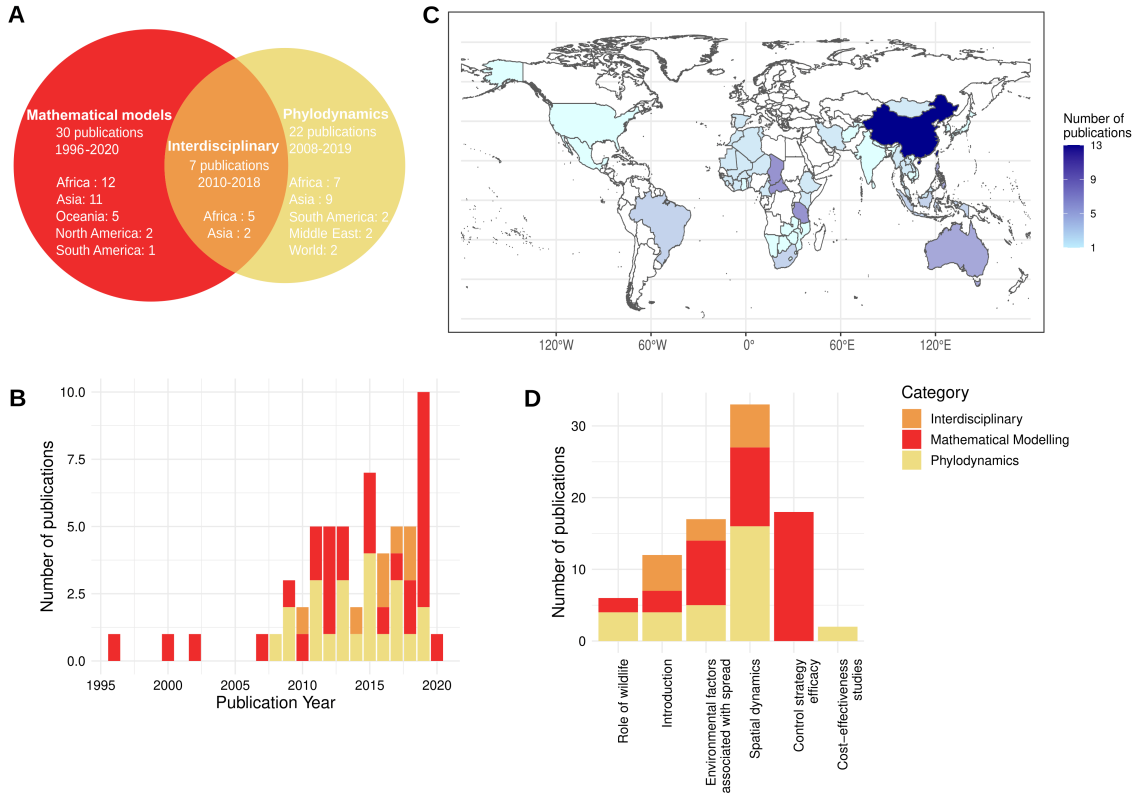


Figure 9: General characteristics of the selected dog rabies studies. (**A**) Classification of the included publications with the total number of studies, the publication time span, and the number of publications per continent of study. Asia and Africa account for up to 78% of the included studies. (**B**) Number of publications per year and per methodological category. Mathematical models were the first studies to be published followed by phylodynamic and interdisciplinary studies. (**C**) Number of publications per country of study. Each publication was attributed to one or multiple countries based on the origin of the RABV genetic sequences, rabid case data or dog ecology data. For phylodynamic studies, countries were not considered if their genetic data were included only in regular phylogenetic tree reconstructions. Similarly, two studies which described rabies dynamics at the global scale (Bourhy et al., 2008; Troupin et al., 2016) were not considered in this figure. In our collected records, China accounts for most Asian studies. Spain appears on the map because Ceuta and Melilla, which are Spanish enclaves in North Africa, are represented in two data sets of RABV genetic sequences (Dellicour et al., 2017; Talbi et al., 2010). (**D**) Number of studies per topic and methodological category. The World Bank, https://datacatalog.worldbank.org/dataset/world-bank-official-boundaries,CC-BY4.0.

Ferguson et al. 2015; Sparkes et al. 2016; Leung and Davis 2017; Laager et al. 2018; Kadowaki et al. 2018; Hudson et al. 2019a; Hudson et al. 2019b; Brookes et al. 2019; Carroll et al. 2010; Zhang et al. 2011; Beyer et al. 2011; Hou et al. 2012), the risk assessment of rabies introduction and the evaluation of outbreak preparedness in rabies-free areas (n = 3; Dürr and Ward 2015; Kadowaki et al. 2018; Hudson et al. 2019b), and cost-effectiveness studies (n = 2; Beyene et al. 2019; Zinsstag et al. 2009; Fig. 9D and Table A1). Finally, interdisciplinary studies mainly focused on rabies dynamics in endemic areas (n = 6; Tohma et al. 2014; Talbi et al. 2010; Mollentze et al. 2014; Bourhy et al. 2016; Cori et al. 2018; Tian et al. 2018), and the identification of environmental factors influencing rabies spread and maintenance such as recurrent reintroductions (n = 3; Talbi et al. 2010; Zinsstag et al. 2017; Tian et al. 2018). Two of these used genetic and epidemiological data of dog rabies in a unified modeling approach (Mollentze et al.,

2014; Cori et al., 2018), whereas the others analyzed sequences through regular phylogenetic approaches and completed their analysis with a mathematical model (Talbi et al. 2010; Bourhy et al. 2016; Zinsstag et al. 2017; Tian et al. 2018; Tohma et al. 2016; Fig. 9D and Table A1).

3.3. Potential sources of bias in the data

Data source (active/passive surveillance), resolution (number and length of RABV sequences, incidence per country/region, etc.) and representativeness influence the level of evidence of the studies on the underlying epidemiological and evolutionary processes. In particular, recorded cases collected through passive surveillance systems are expected to underestimate the disease burden and to be potentially spatiotemporally biased (Hampson et al., 2015; De la Puente-León et al., 2020). Similarly, genetic sequences collected from publicly available databases such as GenBank often lack precise metadata (e.g., sampling time and location) and/or are of short length.

In our text corpus of phylodynamic and interdisciplinary studies, passive surveillance systems and GenBank represent the main sources of RABV genetic sequence data (Tables A2 to A4). By combining these two data sources, researchers have generally managed to increase the spatiotemporal coverage of their data set. This however does not guarantee a good representativeness of the epidemic process. Active surveillance was mostly used to collect dog specimens from animal markets in China (n = 2; Yu et al. 2012; Guo et al. 2013) and thorough contact tracing after biting events in China and Tanzania (n = 2; Brunker et al. 2015; Zhang et al. 2017). On average, the data sets analyzed in these studies contained 183 sequences spanning from approximatively 3% to 100% of the RABV genome length. Short sequences containing the N gene constitute the most common type of data. They are less informative than whole genomes which were only generated and analyzed in recent years across four studies (Brunker et al. 2015; Troupin et al. 2016; Brunker et al. 2018b; Dellicour et al. 2019; Table A2).

In studies from the modeling and interdisciplinary categories, authors generally simulated rabies epidemics (n = 24; Tohma et al. 2016; Kitala et al. 2002; Colombi et al. 2020; Beyer et al. 2012; Townsend et al. 2013a; Dürr and Ward 2015; Ferguson et al. 2015; Chen et al. 2015; Sparkes et al. 2016; Leung and Davis 2017; Laager et al. 2018; Kadowaki et al. 2018; Wilson-Aggarwal et al. 2019; Abdul Taib et al. 2019; Hudson et al. 2019b; Hudson et al. 2019a; Brookes et al. 2019; Carroll et al. 2010; Ortega et al. 2000; Hampson et al. 2007; Zhang et al. 2011; Beyer et al. 2011; Hou et al. 2012; Talbi et al. 2010), and thus predominantly relied on publicly available estimates of the natural history of rabies, dog demographics and dog ecology (Tables C3 and C4). When models were fitted to incidence data (n = 13; Mollentze et al. 2014; Bourhy et al. 2016; Zinsstag et al. 2017; Cori et al. 2018; Tian et al. 2018;

Townsend et al. 2013b; Coleman and Dye 1996; Zhang et al. 2012; Fitzpatrick et al. 2012; Laager et al. 2019; Beyene et al. 2019; Huang et al. 2019; Zinsstag et al. 2009), human and/or dog case data from passive surveillance systems were used, or bite incidence data from thorough active surveillance. In general, there was a lack of data on dog rabies cases (available in 10 studies; Townsend et al. 2013b; Coleman and Dye 1996; Zhang et al. 2012; Fitzpatrick et al. 2012; Laager et al. 2019; Zinsstag et al. 2009; Mollentze et al. 2014; Bourhy et al. 2016; Zinsstag et al. 2017; Cori et al. 2018; Tian et al. 2018) and estimates on dog demographics and ecology integrating the local specificities of host ecology were available in only seven studies (Fitzpatrick et al., 2012; Laager et al., 2019; Beyene et al., 2019; Huang et al., 2019; Zinsstag et al., 2009; Zinsstag et al., 2017; Tian et al., 2018). Access to local data is crucial since differences in rabies spread (Fitzpatrick et al., 2012) and dog carrying capacities (Beyene et al., 2019) were estimated between areas of the same country. We would expect these differences to be more pronounced across different countries. To overcome the lack of epidemiological data on dog rabies, one study used serological data (from vaccination campaigns) to model the dynamics of rabies (Ortega et al., 2000), and another study (Kadowaki et al., 2018) based its analyses on historical records in Japan from the 1950s. Similarly, most Australian studies (Dürr and Ward, 2015; Hudson et al., 2019b; Hudson et al., 2019a; Brookes et al., 2019) took the perspective of dog ecology data since Australia is free of rabies. This way, the authors explored the impact of dog population structure and dog roaming behavior on rabies dynamics.

3.4. Description of the models

In studies using phylodynamic approaches, the geographical dispersal of rabies was studied using either parsimony (n = 4; Bourhy et al. 2008; Talbi et al. 2009; Meng et al. 2011; Yu et al. 2012), Bayesian discrete phylogeography (n = 18; Dibia et al. 2015; Tohma et al. 2014; Tohma et al. 2016; Lemey et al. 2009a; Hayman et al. 2011; Carnieli et al. 2011; Mollentze et al. 2013; Guo et al. 2013; Horton et al. 2015; Brunker et al. 2015; Yao et al. 2015; Zhang et al. 2017; Ma et al. 2017; Wang et al. 2019; Dellicour et al. 2019; Talbi et al. 2010; Tian et al. 2018; Brunker et al. 2015), or Bayesian continuous phylogeography (n = 6; Carnieli et al. 2013; Dellicour et al. 2017; Brunker et al. 2018b; Dellicour et al. 2019; Bourhy et al. 2016; Tian et al. 2018; Tables A2 to A4). All Bayesian phylogeographic studies were carried out in BEAST 1 (Suchard et al., 2018) with discrete trait analysis (DTA) to perform a phylogeographic reconstruction based on discrete/discretized sampling locations (e.g. provinces or countries) or with continuous trait analysis to perform a phylogeographic reconstruction based on spatially-explicit sampling location data (latitude and longitude coordinates). Several methodologies take advantage of

such phylogeographic inferences to investigate the impact of external factors on the dispersal of viruses: a generalized linear model (GLM) extension of DTA developed by (Lemey et al., 2014) to test predictors of dispersal transition frequencies among discrete locations which was implemented by Brunker et al. (2018b); and post hoc statistical approaches developed by Dellicour et al. (Dellicour et al., 2019; Dellicour et al., 2016b; Dellicour et al., 2018a) to investigate the impact of environmental factors on the dispersal velocity, direction, or frequency of viral lineages in continuous phylogeographic frameworks which were applied in four rabies studies (Dellicour et al., 2017; Brunker et al., 2018b; Dellicour et al., 2019; Tian et al., 2018). Finally, Zinsstag et al. (2017) were the only authors to implement a birth-death model in BEAST 2 (Bouckaert et al., 2019) to reconstruct R_e along vaccination campaigns and compare it to estimates obtained with a modeling approach (Table A4).

Compared to phylodynamics, mathematical models display a large diversity of specifications and parameterizations. Compartmental models (n = 18; Tohma et al. 2016; Kitala et al. 2002; Coleman and Dye 1996; Zhang et al. 2012; Fitzpatrick et al. 2012; Sparkes et al. 2016; Leung and Davis 2017; Beyene et al. 2019; Abdul Taib et al. 2019; Huang et al. 2019; Carroll et al. 2010; Ortega et al. 2000; Hampson et al. 2007; Zinsstag et al. 2009; Zhang et al. 2011; Hou et al. 2012; Zinsstag et al. 2017; Tian et al. 2018) are the most represented models, followed by agent-based (n = 8; Townsend et al. 2013b; Dürr and Ward 2015; Ferguson et al. 2015; Laager et al. 2018; Kadowaki et al. 2018; Hudson et al. 2019b; Hudson et al. 2019a; Brookes et al. 2019) and metapopulation (n = 5) models (Colombi et al., 2020; Beyer et al., 2012; Chen et al., 2015; Laager et al., 2019; Beyer et al., 2011). Other model types such as network models or branching processes are also represented (Townsend et al. 2013a; Wilson-Aggarwal et al. 2019; Mollentze et al. 2014; Bourhy et al. 2016; Cori et al. 2018; Tables A3 and A4). The development of new dog rabies models builds upon the literature since 15 models out of the 37 identified were adapted from previously published dog rabies or wildlife rabies models (Tables A3 and A4). This is the case notably for compartmental models which correspond to the simplest models of rabies dynamics. Metapopulation, agent-based, and other model types are more complex, in that these approaches often integrate spatial dynamics of dog rabies (Colombi et al., 2020; Dürr and Ward, 2015; Chen et al., 2015; Laager et al., 2018; Kadowaki et al., 2018; Laager et al., 2019; Wilson-Aggarwal et al., 2019; Hudson et al., 2019b; Hudson et al., 2019a; Talbi et al., 2010; Mollentze et al., 2014; Cori et al., 2018).

Population structure can be integrated in any modeling framework under the form of contact heterogeneity, age-structured populations, roaming behavior, or individual heterogeneity. In compartmental models, population structure is integrated either as a set of strata (stray dogs, owned free-roaming dogs, owned

confined dogs) interacting together (Sparkes et al., 2016), or by specifying a structured next-generation matrix from which R_0 is generally derived (Leung and Davis, 2017). Such models are also referred to as multi-host models and may integrate other hosts: humans (Chen et al., 2015; Beyene et al., 2019; Abdul Taib et al., 2019; Zinsstag et al., 2009; Zhang et al., 2011; Hou et al., 2012; Juan Zhang et al., 2012), cattle (Beyene et al., 2019), wildlife (Fitzpatrick et al., 2012; Huang et al., 2019). In agent-based and network models, population structure is defined at the individual level using spatial kernels (Townsend et al., 2013b; Colombi et al., 2020; Dürr and Ward, 2015; Ferguson et al., 2015; Kadowaki et al., 2018; Hudson et al., 2019b), individual contact rates (Dürr and Ward, 2015; Laager et al., 2018; Brookes et al., 2019), vaccination status (Dürr and Ward, 2015; Kadowaki et al., 2018), life span, infectious period (Townsend et al., 2013b; Ferguson et al., 2015; Brookes et al., 2019), etc.

3.5. Sensitivity analyses

Sensitivity analyses are commonly used to assess the robustness of inference to both data representativeness and model specifications, and to identify the most influential parameters on specific model outputs. In our text corpus, no sensitivity analyses were found to be carried out in phylodynamic studies which can be attributed to the relatively small number of sequences analyzed in those studies. In contrast, sensitivity analyses were commonly performed in mathematical models, either to unravel the key parameters influencing rabies dynamics or to verify the robustness of the results to model assumptions. Dog ecology parameters such as birth rate and carrying capacities are often reported as key parameters on rabies dynamics predictions although they are not estimated using local data. Transmission rates are also determinant in model predictions (Table A3). In spatially explicit studies, mobility parameters also have a strong impact on model inferences. Finally, the impact of under-reporting was tested only in interdisciplinary studies, two of which reported a strong impact of the reporting rate on model inference (Tohma et al., 2016; Cori et al., 2018) whereas the other two were robust to a change in this parameter (Bourhy et al. 2016; Zinsstag et al. 2017; Table A4).

3.6. Insights into dog rabies dynamics and its drivers from phylodynamic and modeling studies

Phylogeographic analyses have aimed to unravel the spatial dynamics of dog rabies at the global and regional scales and showed that dog RABV lineages cluster spatially at the global scale, except for one lineage, referred to as the cosmopolitan lineage, which is largely distributed across the world (Bourhy et al., 2008). At the regional and country scales, there is co-circulation of dog-related lineages, notably in

China (Meng et al., 2011; Yu et al., 2012; Yao et al., 2015; Zhang et al., 2017; Wang et al., 2019), in the Middle East (Dellicour et al., 2019; Horton et al., 2015), as well as in Western and Central Africa (Talbi et al., 2009). However, each lineage exhibits a strong geographical structure. In the case of country-specific lineages, various studies have suggested that transboundary movements are not a major force of rabies dispersal (Tohma et al., 2014; Lemey et al., 2009a; Talbi et al., 2009; Mollentze et al., 2013; Guo et al., 2013; Dellicour et al., 2017). All study categories unraveled the role of human-mediated movements in rabies spread. Overall, phylogeographic analyses provided evidence for the effect of anthropogenic factors: major roads are associated with rabies dispersal in North Africa (Talbi et al., 2010), and RABV lineages tended to preferentially circulate within populated areas in North Africa (Dellicour et al., 2017) and the Middle East (Dellicour et al., 2019). Other factors are associated with rabies spread in Yunnan (China, Tables 1 and A6). These results may reflect the intimate link between rabies dynamics, host ecology and dog-human interactions. Mathematical models highlighted the short length of canine rabies transmission chains (Mollentze et al., 2014; Cori et al., 2018; Ferguson et al., 2015) and unraveled the importance of long-range human movements in disease spread (Colombi et al., 2020; Chen et al., 2015). Finally, interdisciplinary approaches highlighted the crucial role of long-distance transmission events likely due to humans in rabies dynamics in North Africa (Talbi et al., 2010) and also showed that main roads act as barriers to dog rabies dispersal in an urban setting in Africa (Laager et al., 2018).

Table 1: Estimated parameters in phylodynamic models. The sampling window and the spatial scale of the studies are highly variable. Thus, it is not possible to directly compare the velocity and diffusion coefficients amongst the different study settings.

Location	Sampling window	Viral lineages	RABV sequence	Migration rate (migration.year^{-1}, 95% HPD)	Velocity[a] (km.year^{-1}, 95% HPD)	Diffusion coefficient (km^2.year^{-1}, 95% HPD)	Factors facilitating viral spread[b]	Factors impeding viral spread[b]	Reference
Bangui, Central African Republic	1986-2012	Africa 1 Africa 2	N P M G intergenic G-L	-	v = 0.9 (0.65 - 1.2)	-	-	-	Bourhy et al. 2016
Serengeti district, Tanzania	2004-2013	Africa 1b	Whole-genome	-	v = 4.46 (3.22 – 5.88) Coefficient of variation M = 3.10	-	Dog presence	Elevation Rivers	Brunker et al. 2018b
Yunnan province, China	2008-2015	SEA-1 SEA-2 SEA-3	N G	-	v = 57.5 (39.2 - 85.1) $v_{weighted}$ = 23.4 (2.4 - 32.6)	D = 1733 (1082 - 2928) $D_{weighted}$ = 1064 (116 - 1638)	Forest coverage (but with a tendency to spread towards areas associated with relatively low forest coverage)	-	Tian et al. 2018
North and Northeast regions, Brazil	2002-2005	-	N	-	$v_{overall}$ = 12.88[c] v_{dogs} = 30.5[c] $v_{cerdocyonthous}$ = 9.0[c]	-	-	-	Carnieli et al. 2013
Algeria	2001-2008	Africa 1	N	-	$v_{great\ circle\ distances}$ = 26 (18 - 34)	-	Major roads	-	Talbi et al. 2010

Table 1 continued from previous page

			P intergenic G-L		$v_{road\ distances}$ = 33 (23 - 43)				
Algeria	2001-2008	Africa 1	N P intergenic G-L	-	$v_{wavefront} \sim 15^{c}$	D = 2874 (1900 - 5420) $D_{weighted}$ = 1305 (1086 - 1574)	Grasslands Urban areas	Elevation	Dellicour et al. 2017
Morocco	2004-2008	Africa 1	N P intergenic G-L	-	$v_{great\ circle\ distances}$ = 42 (26 – 58) $v_{road\ distances}$ = 51 (34 - 72)	-	Major roads	-	Talbi et al. 2010
Morocco	2004-2008	Africa 1	N P intergenic G-L	-	$v_{wavefront} \sim 22^{c}$	D = 2874 (1900 - 5420) $D_{weighted}$ = 1305 (1086 - 1574)	Grasslands Urban areas	Elevation	Dellicour et al. 2017
Iran	2008-2015	-	Whole-genome	-	v = 55.5 (38.9 - 142.4) $v_{weighted}$ = 18.1 (16.3 - 20.8)	D = 2676 (1935 - 5066) $D_{weighted}$ = 1643 (1356 - 2325)	(Tendency to spread towards and preferentially circulate within accessible areas associated with relatively higher human population density)	(Tendency to avoid circulating in barren vegetation areas and to avoid spreading towards grasslands)	Dellicour et al. 2019
China	1983-2016	Arctic-like 2 Central Asian 1 SEA-1 to 5	N	5.81e-3 (3.92e-3 - 7.77e-3)	-	-	-	-	Wang et al. 2019

Table 1 continued from previous page

Abbreviations: HPD, Highest Posterior Density; SEA-1, South-East Asia 1; SEA-2, South-East Asia 2; SEA-3, South-East Asia 3.
The sampling window and the spatial scale of the studies are highly variable. Thus, it is not possible to directly compare the velocity and diffusion coefficients amongst the different study settings.

[a] Depending on the study, estimates of RABV lineage velocity or diffusivity were obtained by estimating different dispersal statistics. Talbi et al. (2010) reconstructed for each branch of the phylogenetic tree the expected number of migrations between two locations using a discrete phylogeographic model. The authors multiplied these estimates by the great-circle distance between the two locations, and thus, obtained the expected distance travelled within the time elapsed on each branch. Carnieli et al. (2013), Bourhy et al. (2016), Brunker et al. (2018b), Tian et al. (2018), and Dellicour et al. (2019) estimated the mean branch velocity using continuous phylogeographic reconstructions. Finally, Dellicour et al. (2017) estimated the temporal evolution of the wavefront velocity that corresponds to the distance between the reconstructed epidemic origin and the maximal epidemic wavefront. While the mean branch velocity (v) and diffusion coefficient (D) are estimates of the dispersal velocity and of the diffusion coefficient averaged over all tree branches, respectively, their weighted average counterparts involve a weighting by branch time resulting in lower-variance estimates (Dellicour et al., 2019).

[b] Depending on the study, the impact of environmental factors on dispersal of viral lineages were investigated using different approaches. (Talbi et al., 2010) simulated random or conditional dispersal of RABV in northern Africa along phylogenetic trees reconstructed by phylogeographic inference and compared simulated dispersal patterns with the observed spread. Brunker et al. (2018b) parameterized a generalized linear model (GLM) in a discrete phylogeographic framework with resistance distances derived from landscape data between clusters of rabies cases. Dellicour et al. (2017) and Tian et al. (2018) assessed which environmental factors are associated with RABV velocity using continuous phylogeographic inference and post hoc statistical analyses. Dellicour et al. (2019) and Tian et al. (2018) also identified factors associated with the direction of spread using phylogeographic reconstructions and subsequent post hoc analyses.

[c] 95% Highest Posterior Density (HPD) intervals are not specified in the original publications.

Phylodynamic studies showed that introduction through infected dog movement is the major force of rabies spread towards disease-free areas, as Indonesia (Townsend et al., 2013b; Dibia et al., 2015; Mahardika et al., 2014) and the Philippines (Tohma et al., 2014; Tohma et al., 2016) have recently experienced, and also represents a driver of rabies spread in endemic areas where frequent reintroductions counteract local rabies elimination after vaccination campaigns (Bourhy et al., 2016; Zinsstag et al., 2017). In these settings, phylodynamic analysis constitutes a powerful tool to confirm introduction events (Bourhy et al., 2016; Zinsstag et al., 2017; Talbi et al., 2010; Mollentze et al., 2013; Hayman et al., 2011; Tohma et al., 2014). Multiple mathematical models have also shown that frequent reintroductions drive rabies persistence in endemic areas (Ferguson et al., 2015; Laager et al., 2019; Mollentze et al., 2014; Cori et al., 2018).

Population structure constitutes another driving force of rabies maintenance as explored in simulation studies integrating dog ecology data in Australian (Dürr and Ward, 2015; Hudson et al., 2019b; Hudson et al., 2019a; Brookes et al., 2019), Japanese (Kadowaki et al., 2018), Tanzania (Beyer et al., 2012; Beyer et al., 2011) and Chadian (Laager et al., 2018) settings. Rabies-induced behavioral changes were shown to contribute to rabies persistence in small dog populations (Brookes et al., 2019) as well as differential roaming behavior, contact rates between dog strata and the structure of contact networks (Dürr and Ward, 2015; Leung and Davis, 2017; Laager et al., 2018; Kadowaki et al., 2018; Brookes et al., 2019).

The contribution of wildlife to canine rabies spread and maintenance is rarely addressed in phylodynamic studies because viruses isolated from wildlife specimens often correspond to dog-related lineages (Tohma et al., 2014; Hayman et al., 2011; Yao et al., 2015; Zhang et al., 2017; Wang et al., 2019) or because of insufficient sampling efforts when it comes to wildlife (Yu et al. 2012; Table A1). Nevertheless, specific RABV lineages were shown to circulate both in wildlife and domestic dogs in the Middle East and Tanzania with complex interspecies transmissions (Horton et al., 2015; Troupin et al., 2016; Brunker et al., 2018b; Dellicour et al., 2019). A phylodynamic study at the global scale showed that host shifts from dogs to wildlife with adaptation to the new host were common in RABV history (Troupin et al., 2016), which may explain why different lineages circulate in dogs and wild foxes in Brazil (Carnieli et al., 2013), in dogs and ferret badgers in Asia (Huang et al., 2019) and in dogs and mongooses in South Africa (Troupin et al., 2016) with rare interspecies transmission events. By incorporating direct interspecies transmission, mathematical modeling studies showed that dog population contributes to sustained rabies circulation in wildlife instead of the other way around (Huang et al., 2019; Fitzpatrick et al., 2012). Similarly, the proximity to wildlife was shown to not impact rabies spread in dogs in the model

of Beyer et al. (2012).

Finally, mathematical models and phylodynamics provide convenient estimates of a range of parameters on rabies dispersal dynamics (lineage dispersal velocities, diffusion coefficients; in Table 1), rabies evolutionary processes and dog ecology. For example, the evolutionary rate was homogeneously estimated to be between 1×10^{-4} and 5×10^{-4} substitutions per site per year across RABV genes and lineages, except for the Asian lineage which is estimated to evolve faster (Fig. 10A). The time to the most recent common ancestor (TMRCA) is also frequently estimated in phylodynamic studies (Table A2) which is generally more recent than suggested by historical records. R, the expected number of secondary infections, is often estimated by fitting case data to mathematical models (Fig. 10B) or by computing its value based on the choice of parameters value (Table A7). Its estimate ranges from 0.80 to 3.36 according to the setting but it is generally estimated to be between 1 and 2, corresponding to a low-grade transmission with frequent stochastic extinctions. Other parameters such as the dog-to-dog transmission rate, the introduction rate or the dog carrying capacity are also frequently estimated (Table A7).

3.7. Effective control strategies

Interdisciplinary and modeling studies generally assessed the impact of past or potential control strategies to eliminate dog rabies. The specifications of the explored control strategies depended on the economic situation of the country in which the study was supposed to be performed, as well as the model type. Dog vaccination was the most studied control measure (n = 28; Townsend et al. 2013b; Kitala et al. 2002; Coleman and Dye 1996; Zhang et al. 2012; Fitzpatrick et al. 2012; Beyer et al. 2012; Dürr and Ward 2015; Ferguson et al. 2015; Chen et al. 2015; Sparkes et al. 2016; Leung and Davis 2017; Laager et al. 2018; Kadowaki et al. 2018; Laager et al. 2019; Beyene et al. 2019; Abdul Taib et al. 2019; Huang et al. 2019; Hudson et al. 2019a; Brookes et al. 2019; Carroll et al. 2010; Ortega et al. 2000; Hampson et al. 2007; Zinsstag et al. 2009; Zhang et al. 2011; Beyer et al. 2011; Hou et al. 2012; Zinsstag et al. 2017; Tian et al. 2018), whereas culling (n = 7; Dürr and Ward 2015; Sparkes et al. 2016; Hudson et al. 2019b; Carroll et al. 2010; Zinsstag et al. 2009; Zhang et al. 2011; Hou et al. 2012), dog confinement or movement ban (n = 4; Dürr and Ward 2015; Ferguson et al. 2015; Kadowaki et al. 2018; Hudson et al. 2019b), control of dog birth rate (n = 4; Abdul Taib et al. 2019; Carroll et al. 2010; Zhang et al. 2011; Juan Zhang et al. 2012) and community behavior (n = 1; Ferguson et al. 2015) were rarely modelled. Culling was shown to be effective in two compartmental model studies (Carroll et al., 2010; Hou et al., 2012) while vaccination was generally found to be the most effective strategy. Vaccination coverage strongly depends on the setting: 90% or complete dog vaccination coverages are recommended in rabies-free areas with

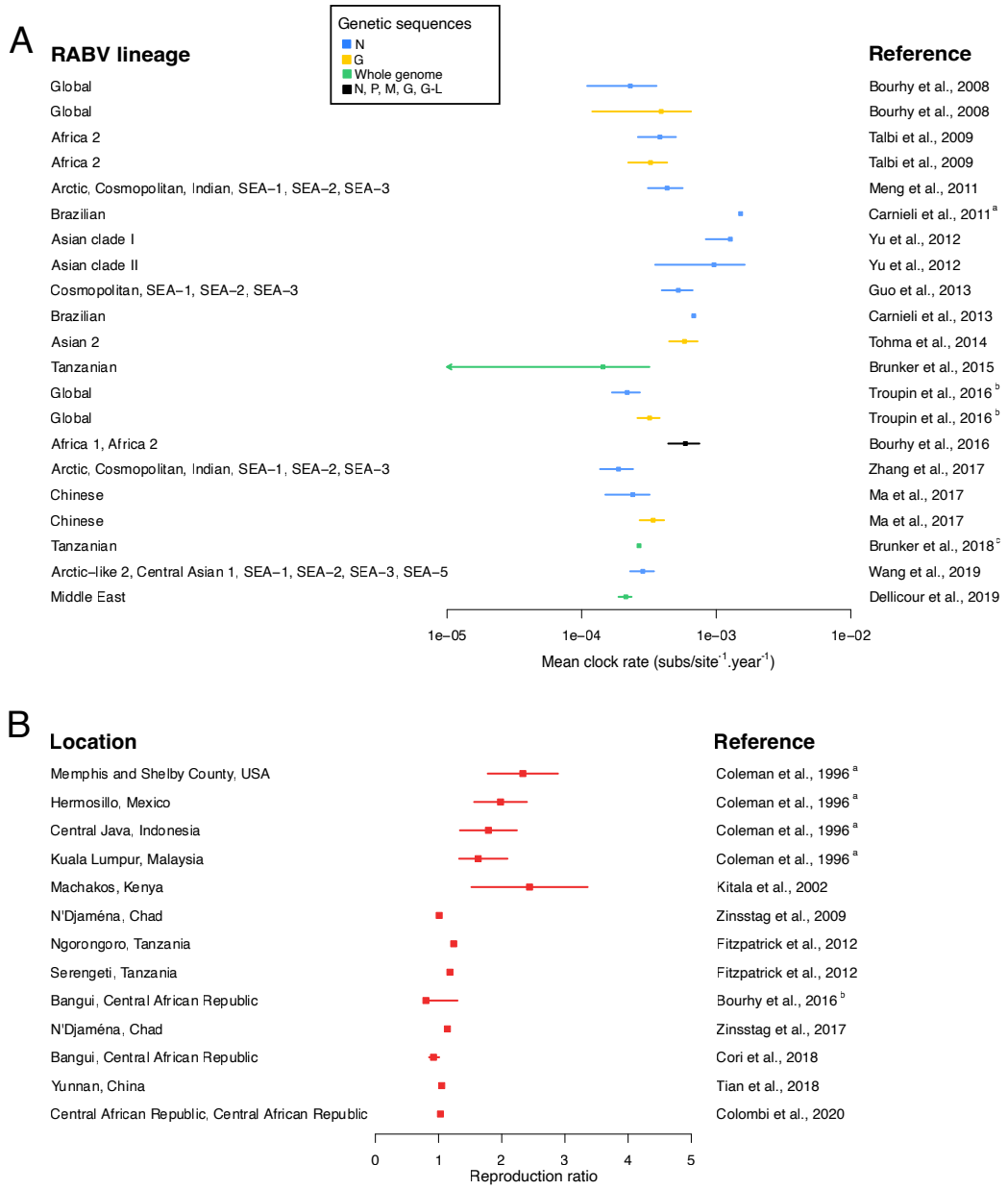


Figure 10: Estimates of the mean evolutionary rate of RABV and the reproduction ratio of canine rabies in the included studies. (**A**) Bayesian credibility intervals (mean and 95% Highest Posterior Density, HPD) of the mean evolutionary rate of canine RABV per genetic sequence and RABV lineage. [a]The estimate corresponds to the upper bound of the 95% HPD. [b]The dot corresponds to the median and the interval to the 95% HPD interval. [c]The 95% HPD was not specified in the original publication. (**B**) Estimates of the reproduction ratio of dog rabies per control strategy or geographical location. The dot corresponds to the mean and the interval to the 95% confidence interval unless stated otherwise. [a]The interval corresponds to the standard error. [b]The authors estimated the effective reproduction ratio along time. Here, the value range of the median monthly point estimate is depicted.

high surveillance and control capacities whereas lower coverages associated with complementary strategies are recommended in endemic areas (Table 2). Nevertheless, the efficacy of vaccination strategies is mitigated by new introductions due to neighboring transmission or long-distance movements mediated by humans (Colombi et al., 2020; Townsend et al., 2013a; Ferguson et al., 2015; Laager et al., 2019; Bourhy et al., 2016; Zinsstag et al., 2017; Knobel et al., 2013), notably in low vaccinated pop-

ulations (Chen et al., 2015). In this case, reactive vaccination strategies (Townsend et al., 2013b) or dog movement bans (Colombi et al., 2020) constitute alternative effective measures. However, Ferguson et al. (2015) evaluated the impact of new introductions in vaccinated areas, and concluded that vaccination coverages were robust to rabies introduction in their specific setting. Similarly, Beyer et al. (2011) suggested that the spatial structure of dog population had more impact than rabies introduction on the efficacy of vaccination campaigns. In terms of vaccination coverage, successful vaccination campaigns should target homogeneous coverage since hidden pockets of rabies transmission might jeopardize control efforts (Townsend et al., 2013b; Kitala et al., 2002; Townsend et al., 2013a; Ferguson et al., 2015). In terms of campaign frequency, the efficacy of pluriannual compared to annual vaccination campaigns is difficult to evaluate as it results from many factors including the number of vaccination pulses, the time interval between each pulse, dog birth rate and the introduction rate of infectious animals (Kitala et al., 2002; Beyer et al., 2012; Carroll et al., 2010).

Table 2: Recommended control strategies in mathematical modeling studies. The efficacy of control strategies on dog rabies dynamics has been addressed in only a subset of the currently available mathematical modeling studies. Studies presented in this table compared several control strategies or different dog vaccination coverages on rabies elimination prospects. The optimal control strategy inherently depends on the epidemiological context (endemic or introduction in previously rabies-free areas), the setting (local surveillance and vaccination capacities), the assumptions of the dog rabies model and the control strategies tested by the researchers. Here, we report the strategies recommended by the authors which include quantitative and qualitative criteria such as the estimated impact of public awareness on rabid dog detection and management. Three studies (Carroll et al., 2010; Leung and Davis, 2017; Townsend et al., 2013a) are not grounded in a specific geographical area. Using simulated scenarios, they test the impact of control strategies according to the time to detection (Townsend et al., 2013a), dog population structure (Leung and Davis, 2017) and the use of immunocontraceptives (Carroll et al., 2010).

Epidemiological context	**Recommended control strategy**	**Specificities of the recommended control strategy**	**Location**	**Reference**
Introduction in previously rabies-free areas	**Reactive dog vaccination**	Followed by a 2-year monitoring period		Townsend et al. 2013a
		Until all targeted dogs are vaccinated	Northern Peninsula Area and Elcho Island, Australia	Dürr and Ward 2015
	90% dog vaccination coverage		Northern Australia and New South Wales, Australia	Sparkes et al. 2016
			Kubin, Saibai and Warraber divisions, Australia	Brookes et al. 2019
	Targeted dog vaccination campaigns	Vaccination of free-roaming dogs	Northern Peninsula Area, Australia	Hudson et al. 2019a
	Integrated approach	Mandatory dog vaccination Dog owner awareness Dog registration Capture of free-roaming dogs Quarantine of imported animals	Ibaraki and Hokkaido prefectures, Japan	Kadowaki et al. 2018
Endemic areas	**90% dog vaccination coverage**		Lemuna-bilbilo and bishoftu districts, Ethiopia	Beyene et al. 2019

Table 2 continued from previous page

	75% dog vaccination coverage	Stray dog management	Guangdong, China	Hou et al. 2012
		Annual vaccination (or biannual vaccination with a 60% coverage)	Machakos district, Kenya	Kitala et al. 2002
	70% dog vaccination coverage		N'Djaména, Chad	Zinsstag et al. 2009
		Even coverage	Bali, Indonesia	Townsend et al. 2013b
			Serengeti and Ngorongoro districts, Tanzania	Fitzpatrick et al. 2012
	≥50% dog vaccination coverage	≥50% fertility control coverage		Carroll et al. 2010
			Sarawak state, Malaysia	Abdul Taib et al. 2019
	Even dog vaccination coverage		Region IV, Philippines	Ferguson et al. 2015
	Targeted dog vaccination campaigns	Frequent dog vaccination campaigns targeting the reduction in metapopulation risk	Serengeti district, Tanzania	Beyer et al. 2012
		Stray dog vaccination coverage based on dog population composition		Leung and Davis 2017
		Vaccination based on social and roaming behaviors Public awareness Locally reactive interventions Reporting of 60% of cases by the surveillance system	N'Djaména, Chad	Laager et al. 2018
	Dog population management	Dog vaccination Public awareness	China	Zhang et al. 2012
		Massive dog vaccination campaigns in urban areas Dog movement bans	Central African Republic	Colombi et al. 2020
		Dog vaccination	China	Zhang et al. 2011

Recent studies (Beyer et al., 2012; Leung and Davis, 2017; Laager et al., 2018; Kadowaki et al., 2018; Hudson et al., 2019b; Hudson et al., 2019a) proposed targeting at-risk dog populations, such as explorers and roaming dogs, to improve the efficacy of vaccination campaigns (Table 2). However, the sensitivity analysis of Laager et al. (2019) showed that population structure did not impact the efficacy of vaccination strategies. There are conflicting results concerning stray dog vaccination which was either less efficient than owned dog vaccination (Hou et al., 2012) or dependent on population composition (Leung and Davis, 2017).

Several studies also suggested an impact of dog birth rate reduction on the incidence of rabies (Abdul Taib et al., 2019; Huang et al., 2019; Carroll et al., 2010; Zhang et al., 2011; Kitala et al., 2002; Zhang et al., 2012). However, the cost and feasibility of dog population management strategies such as sterilization render this unfeasible in many settings (Taylor et al., 2017b). Dog confinement, which is

generally spontaneously put in place by local communities during rabies outbreaks, may improve elimination prospects but, when implemented, the level of confinement is not sufficient to reach elimination (Colombi et al., 2020; Dürr and Ward, 2015; Ferguson et al., 2015). Concerning the rabies burden in humans, some studies recalled the importance of public awareness (Table 2) and proper PEP coverage to reduce the number of human cases, even though it does not impact rabies circulation in dogs (Zhang et al., 2012; Laager et al., 2018; Kadowaki et al., 2018; Huang et al., 2019). All these findings confirmed and justified the strategic plan that provides a phased, all-inclusive, intersectoral approach to eliminate human deaths from rabies recently launched by United Against Rabies, in a collaboration between four partners: WHO, FAO, OIE and GARC (Cleaveland and Hampson, 2017).

4. Discussion

4.1. Insights on rabies epidemiology and control

In this review, we assessed the respective contributions of mathematical modeling and phylodynamics to the understanding of rabies spread and control in dog populations. Contrary to phylodynamic studies, mathematical modeling approaches were multi-faceted and mainly addressed the efficacy of control strategies and, less frequently, the drivers of rabies spread. They revealed the crucial role of frequent introductions and the potential role of dog population structure in disease dispersal and maintenance, as well as the overwhelming efficacy of dog vaccination campaigns over other control strategies. Certain studies also estimated key parameters of rabies dynamics and dog ecology, such as dog birth rate or dog carrying capacity. On the other hand, phylodynamic studies mostly focused on the description of viral dynamics at the global, regional, and local scales, and recently tested which environmental factors are impacting RABV spread. These approaches consistently unraveled the occurrence of long-distance movements suspected to be human-mediated and confirmed the role of humans in rabies dispersal dynamics in Africa and the Middle East. A third kind of studies either combined phylodynamics and mathematical modeling or presented new models integrating epidemiological and genetic data. In the former approach, hypotheses on rabies spread were generated and tested in the same epidemiological context, and thus, confirmed the impact of introductions and human movements in a low-grade transmission process characterized by small clusters and frequent stochastic extinctions. The latter approaches aimed at reconstructing local transmission chains or clusters, opening new horizons on data integration and the study of rabies (Fig. 11). Unfortunately, a large number of endemic countries is still not, or poorly studied. Data collection and/or model formulation are still needed in Russia, and most of Africa,

and South-East Asia.

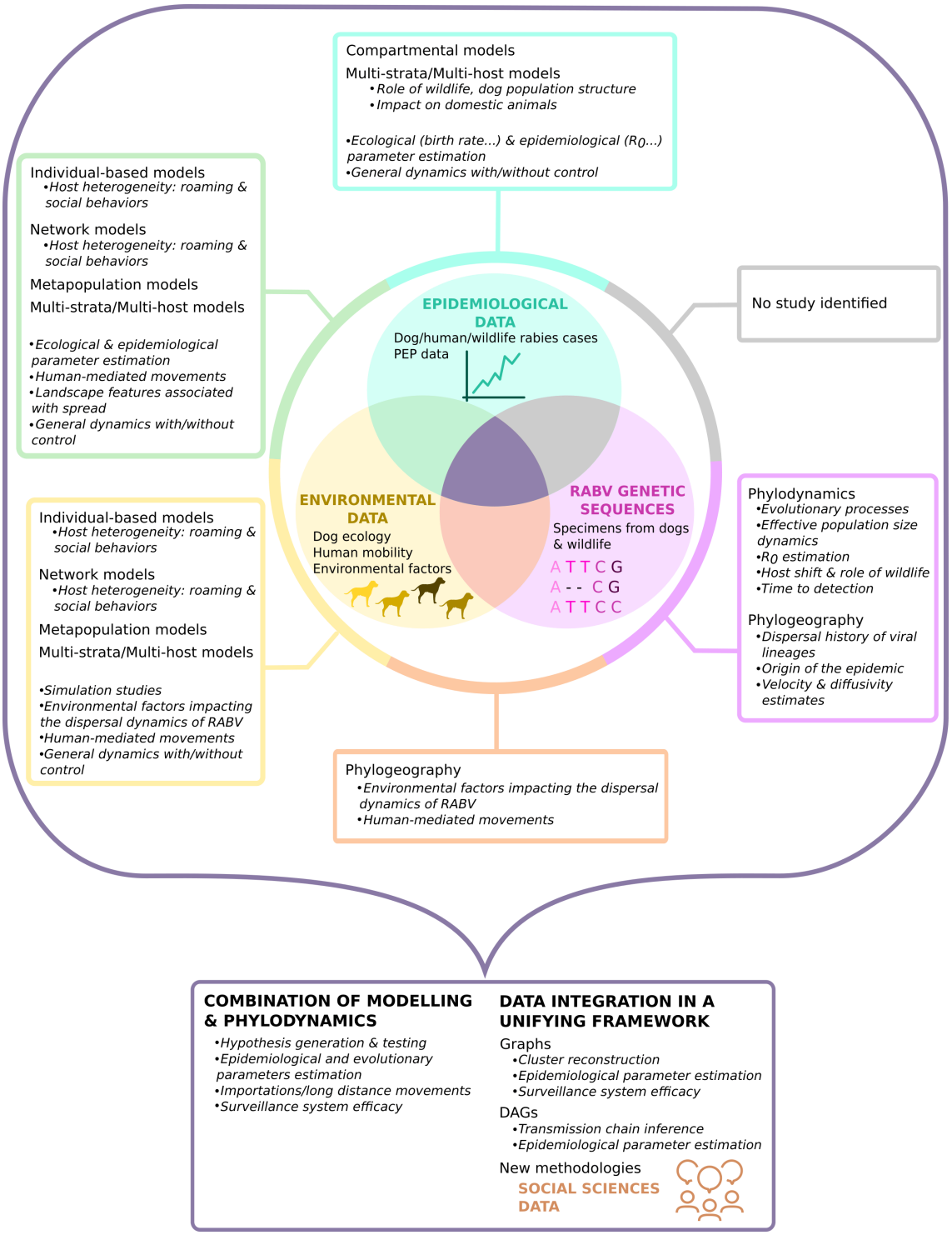


Figure 11: Visual summary of the uses of epidemiological data, environmental data and RABV genetic sequences for the study of rabies dynamics and control. Epidemiological data, environmental data, RABV genetic sequences and social sciences data are highlighted in cyan, yellow, pink, and brown, respectively. The section corresponding to models combining epidemiological data and RABV genetic sequences only is colored in grey since no study that meets this criterion has been identified using our search strategy. Models and their contributions to the understanding of rabies spread and control are detailed in the colored tags. Models using multiple types of data are colored with the intersection color of the corresponding data types. In our text corpus, few studies combined epidemiological, ecological, and genetic data in a unified framework.

The limitations of our review should be acknowledged. In preliminary analyses, we noticed a high variability in record selection according to the combination of search terms, and certainly due to the ambiguous use of specific terms such as phylodynamics in the literature. Since the studies selected in this review are mainly in line with previous reviews (Brunker et al., 2018a; Rattanavipapong et al., 2019; Fisher et al., 2018), we argue that we retrieved a large part of the available studies on rabies dynamics and control.

4.2. Open questions in rabies epidemiology and control

In this review, we summarized the findings of mathematical modeling and phylodynamics on the factors that impact rabies spread. Nevertheless, the full picture of rabies epidemiology remains unclear. First, the role of dog roaming behavior, and dog contact networks in dog rabies spread should be further investigated. In this review, we identified seven studies (Sparkes et al., 2016; Leung and Davis, 2017; Laager et al., 2018; Wilson-Aggarwal et al., 2019; Hudson et al., 2019a; Brookes et al., 2019; Carroll et al., 2010), all situated in Australia and Africa, showing that highly connected dogs or free roaming dogs participate in a large part in the spread of the disease. By specifically targeting this type of dogs, vaccination campaigns could be more effective according to Leung and Davis (2017), Laager et al. (2018), and Hudson et al. (2019a). Yet only one study combined contact data with epidemiological data (Laager et al., 2018). The ecological and behavioral drivers of rabies circulation in domestic dogs are still not fully understood. If stray dogs do constitute a major driver of rabies dispersal, this could have direct implications on the field concerning stray dog population management for example.

Additionally, the role of wildlife and other host species remains unclear (Rupprecht et al., 2019). Even though the circulation of dog rabies seems predominant in dog populations, there are too few studies addressing the dynamics of RABV in wildlife and dogs. Furthermore, the interactions between dogs and other carnivore species are expected to change from location to location. Indeed, the interactions between dog populations and wild carnivores depend on the abundance of wild populations and the frequency of contacts between the dog reservoir and wildlife (Fitzpatrick et al., 2012; Huang et al., 2019). Better understanding the role of wildlife could also have direct implications on local policies such as increasing public awareness, notably in rural areas or strengthening wildlife surveillance systems for rabies.

At a broader scale, the spatial dynamics of rabies are still poorly understood. Urban areas were first thought to be hubs of rabies transmission but recent studies have shown that rabies could be eliminated temporarily at the city-level through mass dog vaccination campaigns (Laager et al., 2019; Bourhy et al.,

2016; Zinsstag et al., 2017). These case studies suggest that urban areas are not hubs of rabies transmission but part of the complex spatial heterogeneity of dog ecology and dog movement. By exploring the dynamics of dog rabies circulation in urban, peri urban and rural areas, rabies research could see an improved understanding of rabies ecology. This could have direct implications on the design of vaccination campaigns, by prioritizing vaccination campaigns in hubs of rabies transmission, followed by locations with intermediate and low transmission.

Finally, there is extensive evidence of the efficacy of dog vaccination to control the spread of rabies in both human and dog populations. We showed in this review that higher coverages are recommended in rabies-free areas than in endemic areas, however, the practicalities of vaccination campaigns are rarely addressed. As a neglected tropical disease, rabies control programs are designed and deployed in resource-limited contexts. Thus, high, and even intermediate vaccination coverages cannot be achieved at once over a large area. The periodicity, the spatial prioritization, the targeted populations, and the association with other control strategies (dog population management, dog movement ban...) are interesting modalities that could be tested in models and could substantially improve resource allocation.

4.3. Future directions of mathematical modeling and phylodynamics for rabies research

There is an evident lack of extensive and adequate databases possibly due to restricted data collection, data accessibility and/or data analysis capacity in resource-limited settings ("Aiming for elimination of dog-mediated human rabies cases by 2030" 2016; Hampson et al., 2019a). This constitutes the main weakness of mathematical modeling and phylodynamic studies that we identified in this review (Table 3). Epidemiological and ecological (census data, population structure, contact networks) data are needed to account for local specificities in terms of modeling interactions between rabies virus (RABV), dog reservoir, domestic animals, wildlife reservoir and human populations. Similarly, there is a need for longer RABV genetic sequences and more thorough sampling to discriminate fine-scale migration events and better characterize the interactions between RABV lineages (Brunker et al., 2015; Brunker et al., 2018a; Brunker et al., 2020). Improving operational data collection is nevertheless challenging. Genomic surveillance relies on laboratory infrastructures, supply chains and expertise, all of which are costly and generally lacking in low- and middle-income countries. New portable sequencing technologies combined with bioinformatics workflows could accelerate capacity building through portability and affordability (Brunker et al., 2020; Gigante et al., 2020). In parallel, potential sampling bias effects should not be overlooked (Ishikawa et al., 2019; Lemey et al., 2009a) since they may hide a part of dis-

Table 3: Strengths and weaknesses of phylodynamics and mathematical modeling studies identified in this review for the study of rabies.

	Strengths	Weaknesses
Phylodynamics	• Homogeneous methodology which facilitates the comparison of rabies dynamics in different areas • Recent advances in phylogeographic models	• Small data sets and short genetic sequences • Studies generally remain descriptive in terms of environmental factors contributing to rabies spread • Large room to apply other models (such as models implemented in BEAST 2) • The potential impact of reporting biases is barely addressed
Mathematical modeling	• Diversity of models that explore multiple aspects of rabies spread • Assessment of rabies control strategies efficacy • Integration of the waning of vaccine-induced immunity	• Mostly simulation studies, models are rarely fitted to dog rabies data • Mostly deterministic models with strong assumptions (homogeneous mixing of dogs, absence of dog population structure, absence of individual heterogeneity) • No direct comparison of rabies dynamics due to the diversity of models

ease dynamics such as silent spread in deprived rural areas. Additionally, many endemic countries with high human incidence (Russia, Malaysia, Cambodia, Burma, Niger, Mozambique, etc.) (Hampson et al., 2015) remain largely unstudied using quantitative approaches. This represents an opportunity for data collection, rabies dynamics characterization and control strategy optimization. Besides filling knowledge gaps, improving the availability of epidemiological, ecological, and genetic data offers an opportunity to strengthen countries' veterinary surveillance capacities (Welburn et al., 2017) and enhance the impact assessment of control strategies, two pillars of the 2030 strategic elimination plan.

Other data types such as social sciences data could help identify knowledge gaps and refine control measures to be tested further using mathematical models. For example, there is little quantitative evidence of the impact of community response on the efficacy of control measures (Rupprecht et al., 2019), although it is key to human rabies prevention (Hasanov et al., 2018; Bardosh, 2014) and it is expected to change over rabies outbreaks and affect rabies dynamics. By bridging the two disciplines, alternative control strategies that are both effective and adapted to community preferences could be designed (Degeling et al. 2018; Fig. 11).

Finally, novel methodologies combining genetic, epidemiological, and environmental data in a comprehensive analysis framework are promising tools for the rabies field. Indeed, the interdisciplinary studies identified in this review exploited the complementarity of genetic and epidemiological information to efficiently generate and test hypotheses on the mechanisms of rabies dynamics (Talbi et al., 2010; Mollentze et al., 2014; Cori et al., 2018; Tian et al., 2018; Tohma et al., 2016), and the limitations of control strategies (Bourhy et al., 2016; Zinsstag et al., 2017). These new avenues represent a significant improvement on past studies and a promising opportunity for canine rabies research in the frame of a One

Health concept that aims to achieve better public health outcomes through cross-sector collaboration.

4.4. Conclusions

In this review, we highlighted the need for more epidemiological, ecological, and genetic data to better characterize rabies dynamics and to get practical information on the drivers of disease transmission. We think that the development of new methodologies integrating genetic and epidemiological data, or the combined use of mathematical models and phylodynamics, constitutes a promising approach that could ultimately contribute to the improvement of the efficacy of control measures including vaccination campaigns and help optimizing the allocation of resources in a context of limited funding.

Chapter 2

Phylodynamics to characterize RABV endemic circulation in Cambodia

In epidemiology, endemicity characterises sustained pathogen circulation in a geographical area, which involves a circulation that is not being maintained by external introduction. Because it could potentially shape the design of public health interventions, there is an interest in fully uncovering the endemic pattern of a disease. Here, we use a phylogeographic approach to investigate the endemic signature of RABV circulation in Cambodia. Cambodia is located in one of the most affected regions in the world, but RABV circulation between and within Southeast Asian countries remains understudied in the area. Our analyses are based on a new comprehensive data set of 199 RABV genomes collected between 2014 and 2017 as well as previously published Southeast Asian RABV sequences. We show that most Cambodian sequences belong to a distinct clade that has been circulating almost exclusively in Cambodia. Our results thus point toward rabies circulation in Cambodia that does not rely on external introductions, which could have concrete implications in terms of mitigation strategy in the region. More globally, our study illustrates how phylogeographic investigations can be performed to assess viral endemicity in a context of relatively limited data.

1. Introduction

From an epidemiological perspective, endemicity corresponds to the sustained circulation of a pathogen in a geographical area (Disease Control and Prevention, 2006), implying that it does not need external introductions. Determining how the circulation of a pathogen is maintained at a local scale can have direct epidemiological implications on its surveillance and control strategies. For instance, the evaluation of the origin of dog rabies cases, either due to importations from neighboring areas or due to cryptic local transmission chains, has been crucial to demonstrate that vaccination campaigns at the city level can interrupt rabies transmission for several months, as exemplified in N'Djamena (Zinsstag et al., 2017) and Bangui (Bourhy et al., 2016). In that context, phylogeographic approaches extracting information from pathogen genomic sequences constitute powerful tools to discriminate importation from local circulation of viral lineages. Phylogeographic investigations can however be impacted by heterogeneous sampling effort and sampling bias (De Maio et al., 2015; Kalkauskas et al., 2021; Liu et al., 2022; Layan et al., 2022), confronting researchers with the challenge being to mitigate or to take into account their effects on the outcomes and conclusions of those analyses (Faria et al., 2017; Vrancken et al., 2020; Guindon and De Maio, 2021).

Rabies is a fatal zoonosis affecting mammals with dogs constituting the main reservoir of its causal agent, the rabies virus (RABV). It has been estimated that dog-related rabies causes approximately 59,000 human deaths per year worldwide (Hampson et al., 2015) despite the existence of effective dog and human vaccination, as well as human postexposure prophylaxis (World Health Organization (WHO), 2018). The global burden mostly affects African and Asian countries where resources are limited to sustain long-term control efforts. In such contexts, extensive surveillance data are often scarce and molecular epidemiological approaches can therefore constitute useful and complementary tools to unravel the dynamics of viral circulation (Layan et al., 2021). Over the past decade, several studies have explored the patterns of RABV circulation in diverse endemic settings. While circulation is spatially-clustered in northern African countries and likely driven by human-mediated dispersal events (Talbi et al., 2010; Dellicour et al., 2017), multiple viral lineages co-circulate in Iran due to the co-occurrence of dog and wildlife reservoirs and the geographical location of Iran being at the crossroads of Asian, Middle-East, and European countries (Dellicour et al., 2019). In Tanzania, RABV circulation is restricted to dogs and maintained at the local scale by well-interconnected dog populations (Brunker et al., 2018b). The mechanisms underlying viral maintenance are context-specific and may involve frequent introductions from neighboring countries, long-distance human-mediated movements, wildlife reservoirs, sustained trans-

mission in highly and heterogeneously mixing dog populations, etc. Understanding these mechanisms could assist countries in prioritizing long-term surveillance efforts or adopting new regulations on dog trade and movement.

Cambodia has one of the highest dog bite and human rabies incidences in the world (Ly et al., 2009; Ponsich et al., 2016; Chevalier et al., 2021), with a rabies circulation that remains uncontrolled due to the lack of a national prevention and surveillance program, which together maintains a cycle of neglect (Li et al., 2019). Fortunately, Cambodia has integrated the Asian Rabies Control Network (ARACON) in 2018 (Coetzer et al., 2018), and recent actions by stakeholders of rabies surveillance and control have been implemented to move towards the elimination goal. For example, a recent study has evaluated the implementation of a hospital-based dog bite case management (Tazawa et al., 2022). To improve vaccination accessibility, the Institut Pasteur du Cambodge (IPC) has recently created two additional vaccination centers in Battambang (2018) and Kampong Cham (2019), specifically targeting poor and rural populations, and organized a large communication campaign to raise awareness about rabies transmission and prevention (www.pasteur-kh.org/rabies-prevention-centers). IPC has also been instrumental for the development and promotion of a shortened rabies post exposure regimen that is now recommended by WHO (Cantaert et al., 2019; Tarantola et al., 2019). Still, several knowledge gaps persist concerning the epidemiological situation in Cambodia due to the scarcity of epidemiological and RABV genetic data. To our best knowledge, only one study has described dog-related RABV lineages in Cambodia (Mey et al., 2016), which mostly belong to the SEA-1 lineage (Troupin et al., 2016) along with RABV sequences isolated in Thailand, Vietnam, and Laos. This first phylogenetic study suggests that there is a substantial circulation of dog RABV between Southeast Asian countries.

While rabies is known to sustainably circulate in Cambodia, the importance of external RABV introduction events from surrounding countries as well as the dispersal dynamics of the virus at the country scale have not yet been investigated. In the present study, we propose to implement and apply phylogeographic approaches to characterize those aspects and evaluate the endemicity pattern of RABV in Cambodia, with the goal to improve our understanding of the important public health burden associated with rabies in that region. For this purpose, we generate and analyze a new data set of RABV genomes isolated from infected dogs in Cambodia. Specifically, we aim to exploit this comprehensive data set to (*i*) assess the contribution of introduction events from neighboring countries, (*ii*) evaluate the dynamics of spatial spread at the country level, (*iii*) formally compare the dispersal capacity of the virus with other regions of the World, and (*iv*) test the impact of key environmental factors on local viral circulation.

2. Methods

2.1. Samples collection

As part of the rabies surveillance in Cambodia, brain samples from suspected rabid dogs were sent to the Virology unit at IPC where rabies diagnosis activity has been performed routinely since 1998 (Mey et al., 2016). Dog's heads were usually referred to the IPC lab from people who were consulting for post-exposure prophylaxis treatments following animal bite. The animal samples were tested by a standard direct fluorescent antibody test (DFAT) as previously described (Duong et al., 2016). Positive brain samples collected from 2012 to 2017 were selected for sequencing as described below. Before submitting samples for sequencing, all selected samples were tested using a RT-qPCR for detecting RABV (Dacheux et al., 2016).

2.2. Procedure to select samples to sequence

627 positive brain samples were collected between 2012 and 2017 across Cambodia by the IPC, with the possibility to sequence 208 of them. At the beginning of this study, all 33 samples collected in 2017 had already been sequenced, and we subsequently aimed at selecting 2012-2016 samples while maximizing the spatio-temporal coverage in the perspective of phylogeographic analyses. For this purpose, we implemented a Markov process to select 35 samples per year while maximizing the sum D of great-circle distances along the edges of a Delaunay triangulation network connecting all selected samples. The algorithm started by selecting all samples along the spatial limits of the minimum convex hull polygon built around all the samples, as well as a random set of sequences within that polygon. At each iteration i, the algorithm then changed one selected sample per year, randomly replaced by another sample collected the same year, and re-computes the sum D_i on the resulting Delaunay triangulation network. Similarly to the SAMOVA algorithm developed by Dupanloup et al. (2002) for population genetic clustering analyses, the new selection is accepted with probability $p = 1$ if D increased since the last iteration, and with the following probability if D did not increase:

$$p = e^{(D_i - D_{i-1})S^A} \tag{2.1}$$

where S is the number of iterations performed so far in the process, and A is an arbitrary constant controlling the speed of what is called a "cooling process", here implemented to avoid becoming trapped at a local optimum (Dupanloup et al., 2002). The Markov process was run for 100,000 iterations, and

we performed 1,000 independent repetitions of the algorithm starting each time with a different initial subsampling.

2.3. Full genome sequencing

Total RNA extraction done on brain samples was performed according to Marston et al. (2013) with modifications. Briefly, brain samples were crushed into 1 ml of TRI Reagent (Sigma). After addition of 200 μl of chloroform and centrifugation at 12,000 rpm for 15 min at 4°C, the aqueous phase was collected and added to an equal volume of ethanol 70% before RNA extraction and purification using RNeasy Mini Kit (Qiagen). Genomic host DNA was depleted from purified RNA by a DNAse I treatment using the DNA-free Kit (Ambion) and following the manufacturer's instructions. Depleted RNA was finally purified using Agencourt RNAclean XP beads (Beckman Coulter) at a ratio of 1:1.8, following the manufacturer's instructions and eluted with 20 μl of nuclease-free water.

Genome sequences were obtained using next generation sequencing (NGS) as previously described (Dacheux et al., 2019; Luo et al., 2021) with minor modifications. After RNA extraction, a ribosomal RNA depletion step was first carried out with 2-4 μg of RNA and 1 μl of Terminator 5'-Phosphate-Dependent Exonuclease (Epicentre Biotechnologies), in addition to 2 μl of buffer A and 0.5 μl of RNAsin Ribonuclease inhibitor (Promega). After being adjusted to 20 μl with nuclease-free water, the mix was incubated for 1h at 30°C. The depleted RNA was then purified using Agencourt RNAclean XP beads (Beckman Coulter) at a ratio of 1:1.8, following the manufacturer's instructions. Reverse transcription in complementary DNA (cDNA) of purified RNA was then done using the SuperScript III First-Strand Synthesis SuperMix kit (Invitrogen) according to the manufacturer's instructions. For this step, 8 μl of RNA was first incubated at 65°C for 5 min with 1 μl of Annealing Buffer (Invitrogen) and 1 μl of 50 μM of random hexamers (Invitrogen), then placed on ice. The complementary step was performed with the addition of 10 μl of 2X First-Strand Reaction Mix and 2 μl of Superscript III Reverse transcriptase / RNaseOUT Enzyme Mix for a final volume of 22 μl. The mix was incubated at 25°C for 10 min then at 50°C for 50 min. Finally, inactivation of the enzymes was performed after incubation at 85°C for 5 min. Afterward, double-stranded DNA (dsDNA) was synthesized in a reaction mix containing 20 μl of fresh cDNA, 10x Second-Strand Synthesis Reaction Buffer (New England Biolabs), 3 μl of 10 mM dNTP mix (Invitrogen), 1 μl (10 U) of Escherichia coli DNA ligase (New England Biolabs), 4 μl (40 U) of E. coli DNA polymerase I (New England Biolabs), 1 μl (5 U) of E. coli RNase H (New England Biolabs), and 43 μl of nuclease-free water, after incubation at 16°C for 2 h. The total volume (80 μl) of dsDNA was finally purified for each virus, using a ratio of 1:1.8 of AMPure XP beads (Beckman

Coulter) following the manufacturer's instructions. Finally, dsDNA libraries were constructed using the Illumina Nextera Kit (Illumina), which includes TruSeq Nano kit and TruSeq DNA UD Indexes. After nine amplification cycles, purified pooled libraries were sequenced on an Illumina NextSeq500 instrument using Flowcell High Output and running paired-end 2x150 cycles. The NGS data were analyzed using *de novo* assembly and mapping (both using CLC Assembly Cell, Qiagen) with a dedicated workflow built on the Galaxy platform of Institut Pasteur (Troupin et al., 2016; Dacheux et al., 2019). The quality and the accuracy of the final genome sequences were checked after a final mapping step of the original cleaned reads and visualized using Tablet (Milne et al., 2013). The partial genome sequences (11879-11882 nt) of the 199 rabies viruses were submitted to GenBank under the accession numbers OP515147-OP515345.

2.4. Investigating the signature of endemicity

Sequence alignments were generated with MAFFT v7.467 (Katoh and Standley, 2013). To analyze the distribution of Cambodian lineages within a global RABV phylogeny, maximum likelihood phylogenetic inference was performed using the program IQ-TREE v2.0.6 (Minh et al., 2020), with 1,000 bootstrap replicates and the substitution model (GTR+I+Γ4) selected by the IQ-TREE's ModelFinder tool. The discrete phylogeographic analysis used to investigate the signature of endemicity was performed with the discrete diffusion model (Lemey et al., 2009a) implemented in the software package BEAST 1.10 (Suchard et al., 2018). For this discrete phylogeographic reconstruction aiming at analyzing the introduction events into Cambodia, we only considered two discrete locations "Cambodia" and "other countries" - and included all RABV N genes available in GenBank on September 2021 from Indonesia (n = 2), Laos (n = 22), Malaysia (n = 17), Myanmar (n = 5), Philippines (n = 63), Thailand (n = 8), and Vietnam (n = 22). For this analysis, we used a constant population size coalescent prior, and we modeled the branch-specific evolutionary rates according to a relaxed molecular clock with an underlying log-normal distribution (Drummond et al., 2006), and the nucleotide substitution process according to a GTR+I+Γ4 parameterization (Tavare, 1986). We used the default prior distributions in BEAST 1.10 for the substitution, molecular clock, demographic, and phylogeographic models. The Markov chain Monte-Carlo (MCMC) algorithm was run for more than 19×10^7 generations and sampled every 12,000 generations. We used the program Tracer 1.7 (Rambaut et al., 2018) for assessing the convergence/mixing properties and calculating effective sampling sizes (ESS), which were all >200 for continuous model parameters. We then used the program TreeAnnotator 1.10 (Suchard et al., 2018) to obtain a maximum clade credibility (MCC) tree. Patristic distances reported in Fig. 12 were computed with the function "distTips" of

the R package "adephylo" (Pavoine et al., 2008). We compared the prior and posterior distributions in Appendix B.

2.5. Spatially-explicit phylogeographic reconstruction

Continuous phylogeographic analyses were performed with the relaxed random walk (RRW) diffusion model (Lemey et al., 2010; Pybus et al., 2012) implemented in BEAST 1.10 (Suchard et al., 2018) coupled with the high-performance computing library BEAGLE 3 (Ayres et al., 2019). Branch-specific evolutionary rates and the nucleotide substitution process were again modeled according to a relaxed molecular clock with an underlying log-normal distribution and to a GTR+Γ4 parameterization, respectively. We used a gamma distribution to model the among-branch heterogeneity in diffusion velocity, and a flexible skygrid coalescence model for the tree topology, the latter enabling the inference of changes in the effective viral population size over time. The Markov chain Monte-Carlo (MCMC) algorithm was run for 15×10^7 generations for the analysis based on the full genome data set and for 4×10^8 generations for the analysis based on the N gene sequence data set, sampling every 15×10^4 generations in both cases. We again used the program Tracer 1.7 to assess the convergence/mixing properties (and to ensure that ESS values associated with relevant parameters were all >200), as well as the program TreeAnnotator 1.10 to obtain and annotate the MCC tree. We used the R package "seraphim" (Dellicour et al., 2016a) to extract the spatiotemporal information embedded within posterior trees, to visualize the continuous phylogeographic reconstructions, and to estimate the weighted lineage dispersal velocity (WLDV), the latter being defined as follows:

$$v_{weighted} = \frac{\sum_{i=1}^{n} d_i}{\sum_{i=1}^{n} t_i} \tag{2.2}$$

where d_i is the geographic distance traveled along the phylogeny branch, and t_i the duration of that branch, respectively. As detailed in the Results section, we also computed WLDV estimates while considering increasing cut-off values defining the maximal geographic distance that can be traveled by a phylogenetic branch to be included in the estimation.

2.6. Landscape phylogeographic analyses

We applied two previously described analytical procedures to investigate the impact of environmental factors on the dispersal location (Dellicour et al., 2019) and velocity (Dellicour et al., 2017) of viral lineages. Both analytical procedures are based on the comparison between 1,000 spatially-annotated trees

sampled from the post-burn-in posterior distribution of trees inferred by a continuous phylogeographic analysis, hereafter referred to as "inferred trees", and the same tree topologies along which we simulated a RRW diffusion process (Dellicour et al., 2020), hereafter referred to as "simulated trees". These RRW simulations were performed with the "simulatorRRW1" function of the R package "seraphim" and based on the sampled precision matrix parameters estimated by the phylogeographic analyses (Dellicour et al., 2018a). For each tree, the RRW simulation started from the position inferred at the root node and was constrained such that the simulated node positions remained within the study area, which was here defined by the minimum convex hull built around all node positions, avoiding non-accessible sea areas. The purpose of these simulations is to obtain spatially-annotated trees corresponding to the trees inferred by continuous phylogeography but along which we generated a new diffusion process that has not been impacted by environmental factors.

We first investigated whether viral lineages tended to avoid or preferentially circulate within areas associated with particular environmental conditions. For this purpose, we extracted and subsequently averaged the environmental values at the tree node positions to obtain, for each environmental factor, a posterior distribution of mean environmental values at tree node positions. We then compared values obtained through inferred trees and their corresponding simulated trees using an approximated Bayes factor (BF) support (Suchard et al., 2005): $\text{BF} = \frac{p_e/(1-p_e)}{0.5/(1-0.5)}$. To test if a particular environmental factor e tended to attract viral lineages, p_e was defined as the frequency at which the environmental values from inferred trees were greater than values from simulated trees; and to test if a particular environmental factor e tended to repulse viral lineages, p_e was defined as the frequency at which the environmental values from inferred trees were lower than values from simulated trees. We considered BF values > 20 as strong statistical supports (Kass and Raftery, 1995).

Second, we investigated to what extent viral lineage dispersal velocity was impacted by environmental factors acting as conductance or resistance factors. For each branch in the inferred and simulated trees we calculated an "environmental distance" using the path model implemented in the program Circuitscape (McRae, 2006). This path model is based on circuit theory and allows accommodating uncertainty in the travel route. An environmental distance was first computed from each environmental raster, and then from a uniform "null" raster whose cell values are all set to 1. The environmental distance is a spatial distance that is weighted according to the values of the underlying environmental raster, and therefore constitutes a proxy for geographical distance when computed on the null raster. Each environmental variable was considered twice: once as a potential conductance factor that facilitates movement, and once

as a potential resistance factor that impedes it. For each environmental variable, we also generated and tested several distinct rasters by transforming the original raster cell values with the following formula: $v_t = 1 + k(v_o/v_{max})$, where v_t and v_o are the transformed and original cell values, and v_{max} the maximum cell value recorded in the raster. The rescaling parameter k here allows the definition and testing of different strengths of raster cell conductance or resistance, relative to the conductance/resistance of a cell with a minimum value set to 1, which corresponds to the null raster. For each environmental variable, we generated three distinct rasters using the following values for rescaling factor k: k = 10, 100, and 1000. For these analyses, we estimated the statistic Q defined as the difference between the coefficient of determinations obtained (*i*) when branch durations are regressed against the environmental distances computed on an environmental, and (*ii*) when branch durations are regressed against the environmental distances computed on the null raster. We estimated a Q statistic for each environmental raster and each of the 1,000 trees sampled from the posterior distribution. An environmental factor was only considered as potentially explanatory if both its distribution of regression coefficients and its associated distribution of Q values were positive (Jacquot et al., 2017), i.e. with at least 90% of positive values. In this case, the statistical support associated with the resulting Q distribution was compared with the corresponding null distribution of Q values obtained when computing environmental distances for phylogenetic branches of simulated trees. Similar to the procedure used for the investigation of the impact of environmental factors on the dispersal locations of viral lineages, the comparison between inferred and simulated distributions of Q values was formalized by approximating a Bayes factor support.

3. Results

3.1. Samples selection and sequencing

We selected and sequenced 208 RABV specimens from the 627 brain isolates of confirmed infected dog cases collected by the IPC between 2012 and 2017. To select the samples to sequence, we developed and applied a Markov process aiming at maximizing the spatio-temporal coverage of the resulting sampling, thus optimizing it for phylogeographic investigations of the dispersal of viral lineages across the study area (Fig. B1).The final data set included 199 new full genome sequences as the PCR amplification or next-generation sequencing (NGS) was unsuccessful for nine specimens. Throughout the study, we either analyzed this full genome data set or a data set made of all Cambodian RABV nucleoprotein (N) genes available in GenBank and for which we managed to retrieve GPS coordinates associated with the sampling location. While the full genome data set offers the best phylogenetic and molecular clock

signal, the N gene data consists of a larger number of sequences (n = 354) with a broader spatial and temporal coverage.

3.2. Investigating the endemic signature of RABV circulation in Cambodia

We first conducted a maximum likelihood phylogenetic inference to visualize the position of Cambodian lineages within an overall RABV phylogeny (Fig. B2). This analysis was based on our data set of Cambodian N gene sequences as well as a balanced subset of RABV N gene sequences retrieved from GenBank and collected in all continents. In the resulting tree, Cambodian sequences appear uniquely related to the Southeast Asian lineages, suggesting that, contrary to Vietnam (Gigante et al., 2020), neither the Cosmopolitan lineage nor the Chinese lineages circulate in Cambodia (Fig. B2).

We subsequently performed a discrete phylogeographic analysis to delineate Cambodian RABV clades resulting from a distinct introduction event into the country (Fig. 12). In this analysis, we included all dog related RABV N genes from Southeast Asian countries (i.e. Indonesia, Laos, Malaysia, Myanmar, Philippines, Thailand, and Vietnam) available in GenBank, and only considered two discrete locations: "Cambodia" and "other countries", the latter encompassing all the other Southeast Asian countries. The vast majority of the Cambodian sequences (351 out of 354) were grouped within a unique large clade also containing three sequences sampled in Vietnam. We estimated that this clade was introduced in Cambodia around 1971, with a 95% highest posterior density (HPD) ranging from 1963 to 1979 (Fig. 12B-C). Furthermore, we identified that, within our sampling, only three other Cambodian sequences resulted from distinct introduction events into the country (Fig. 12B). Our discrete phylogeographic analysis indicated that those three sequences resulted from two different introduction events in Cambodia that did not seem to have led to substantial onward local transmission (Fig. 12C). These results clearly point towards an endemic RABV circulation in Cambodia, i.e. a circulation that does not seem maintained or fueled by external introductions.

As a result of the comprehensive data set of genomic sequences generated in the present study, Cambodia is oversampled compared to its surrounding countries, which might affect the outcome and conclusions of the discrete phylogeographic analysis aiming at identifying distinct introduction events. In other words, the lower number of RABV genomes available for the surrounding countries could potentially lead to an underestimation of the lineage introductions into Cambodia. In that context, we performed further investigations to assess if the large clade of Cambodian sequences delineated in the discrete phylogeographic inference could indeed correspond to a clade reflecting local rabies circulation not or barely fed

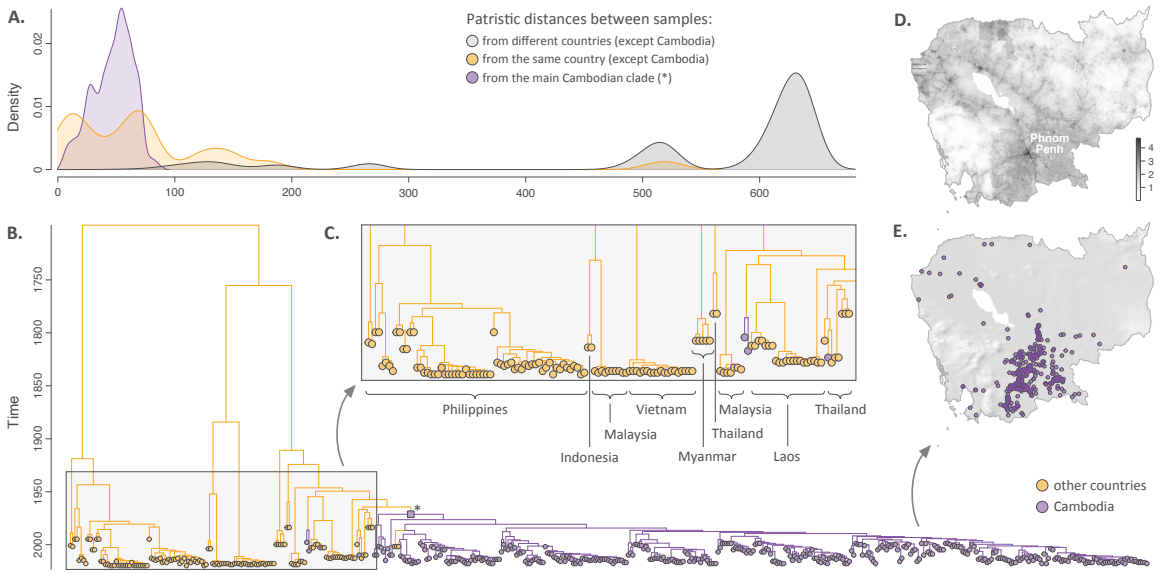


Figure 12: Investigating the endemic signature of RABV circulation in Cambodia. (**A**) Comparison between patristic distances (in years) computed on the Maximum clade credibility (MCC) at the inter-countries level (excluding Cambodia; in gray), at the intra-country level (excluding Cambodia; in orange), and within the main Cambodian clade (*; in purple). (**B**) MCC tree obtained from the preliminary discrete phylogeographic inference based on the N gene sequences and considering only two different discrete locations: "Cambodia" and "other countries". Tip nodes and phylogeny branches are colored according to their sampling location and inferred ancestral location, respectively. (*) refers to the main Cambodian clade inferred by the discrete phylogeographic analysis, whose introduction in Cambodia is estimated to be around 1971 (95% HPD = [1963-1979]). (**C**) Zoom on one particular part of the MCC tree corresponding to the terminal branches outside the main Cambodian clade. The country of origin of sampled sequences are displayed below the tip nodes. (**D**) Map of log10-transformed human population density in Cambodia. (**E**) Sampling map of all N gene sequences used for the discrete phylogeographic as well as in the subsequent analyses. Comparison between patristic distances computed on the Maximum clade credibility (MCC) at the inter-countries level (excluding Cambodia; in gray), at the intra-country level (excluding Cambodia; in orange), and within the main Cambodian clade (*; in purple).

by viral exchanges from neighbouring countries. For this purpose, we visually compared the patristic distances computed for each pair of sequences coming (*i*) from the same country (except Cambodia), (*ii*) from two different countries (except Cambodia), as well as (*iii*) for all pairs of sequences belonging to the major Cambodian clade (Fig. 12A). The purpose of this comparison was to assess if the pairwise patristic distances computed within the major Cambodian clade corresponded to the range of intra-country patristic distances computed for the surrounding countries. Our results confirmed that this is the case and the distribution of inter-countries patristic distance did not overlap with the patristic distances computed within that clade (Fig. 12A). Interestingly however, we noticed small overlaps between the distribution of intra-country and inter-countries patristic distances. First, a small proportion of inter-countries patristic distances were lower than 300, which corresponded to a range of distance values mostly computed for intra-country patristic distances. Second, a small proportion of intra-country patristic distances were close or higher than 500, this time mostly corresponding to a range of distance values computed for

inter-countries patristic distances (Fig. 12A). While in the first case it corresponded to pairs of sequences collected in different countries but relatively close to each other in the phylogenetic tree, in the second case it corresponded to pairwise patristic distances between samples collected in the same country but belonging to distinct clades within the tree. In the latter case, it corresponded to pairs of sequences resulting from distinct transmission chains that entered into the country by independent introduction events from aboard.

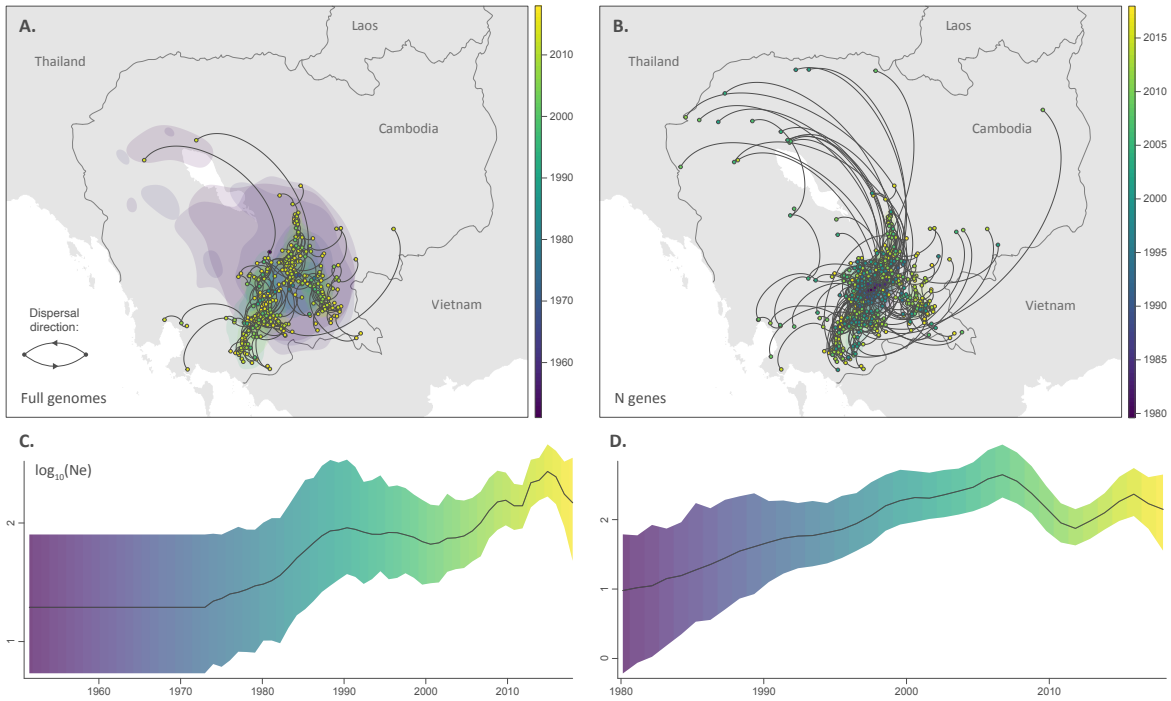


Figure 13: Dispersal and genetic diversity history of RABV lineages in Cambodia, as inferred by a spatially-explicit phylogeographic analysis jointly performed with a skygrid reconstruction. The analyses were either based on the full genomes (panels A and C) or on the N genes (panels B and D). (**A-B**) For both continuous analyses, we mapped the maximum clade credibility (MCC) tree whose nodes are colored from purple (the time of the most recent common ancestor, TMRCA) to yellow (most recent sampling time). MCC trees are superimposed on 80% highest posterior density (HPD) polygons reflecting phylogeographic uncertainty and colored according to the same time scale. (**C-D**) Skygrid reconstructions of the recent evolution of the effective size of the viral population (Ne) based on the full genomes and N genes data sets, respectively. The solid lines correspond to the median estimates and the surrounding polygons colored according to the time scale of the corresponding continuous phylogeographic inference correspond to the 95% HPD interval.

3.3. Unravelling the dispersal dynamics of RABV in Cambodia

We performed a continuous phylogeographic inference to analyze the dispersal dynamics of RABV lineages within Cambodia. In this analysis, we only included the Cambodian sequences belonging to the main Cambodian clade identified by our discrete phylogeographic analysis (Fig. 12B). The continuous phylogeographic reconstructions obtained from the analysis of full genomes and N genes are presented in Fig. 13A and 13B respectively. Compared to the full genome data set, the N genes data set covered

a larger geographical area with more sequences at the Thai border, as well as a larger time period with sequences isolated between January 1998 and December 2017. For both continuous phylogeographic analyses, the uncertainty associated with the inference of ancestral locations was reported by overlapping shaded polygons colored according to their temporal occurrence. The same color scale was used to color the tree nodes according to their time of occurrence (inferred for the internal nodes and corresponding to the collection date for tip nodes). For the N genes data set, the uncertainty polygons were much smaller and even hidden by the tree nodes reported on the map. The smaller phylogeographic uncertainty associated with this data set is likely related to the higher number of samples included in the analysis. Both phylogeographic reconstructions, but the reconstruction based on the N genes in particular, highlight the Phnom Penh area as a crossroad of RABV lineage circulation. Indeed, most of the samples collected distantly from the area of Phnom Penh do not seem to be directly connected with each other within the phylogeny. In other words, those more remote samples seem to arise from distinct transmission chains emanating from the Phnom Penh area.

Secondly, we used the continuous phylogeographic reconstructions to estimate the weighted dispersal velocity of RABV lineages within Cambodia: 12.6 km/year (95% HPD = [11.0-14.3]) for the analysis based on the N genes, but only 6.7 km/year (95% HPD = [6.1-7.4]) for the analysis based on the full genomes. The fact that the two analyses led to different WLDV estimates despite a similar overall phylogeographic pattern (Fig. 13A-B), might be due to the wider geographical sampling of the N genes. Indeed, wider sampling might increase the probability to sample long-distance lineage dispersal events that, on average, are more likely to include fast dispersal events pulling the WLDV estimate to higher values (see below for further discussion on this aspect). The phylogeographic inferences were performed using the skygrid coalescent model (Gill et al., 2013) for the tree prior, which allowed us to get an estimation of the recent evolution of the effective viral population size. Our skygrid reconstructions indicated that there was a global increase of the effective RABV population size in the last 25-30 years (Fig. 13C-D).

3.4. Impact of long-distance lineage dispersal events on the estimation of dispersal velocity

With the objective to assess whether RABV lineage dispersal was slower in Cambodia relative to other geographic areas, we compared these estimates with the WLDV estimates previously obtained and reported for other RABV data sets (Fig. 14). As introduced above, we suspected that the spatial extent of a study area might have an impact on the frequency at which long-distance dispersal events were sampled,

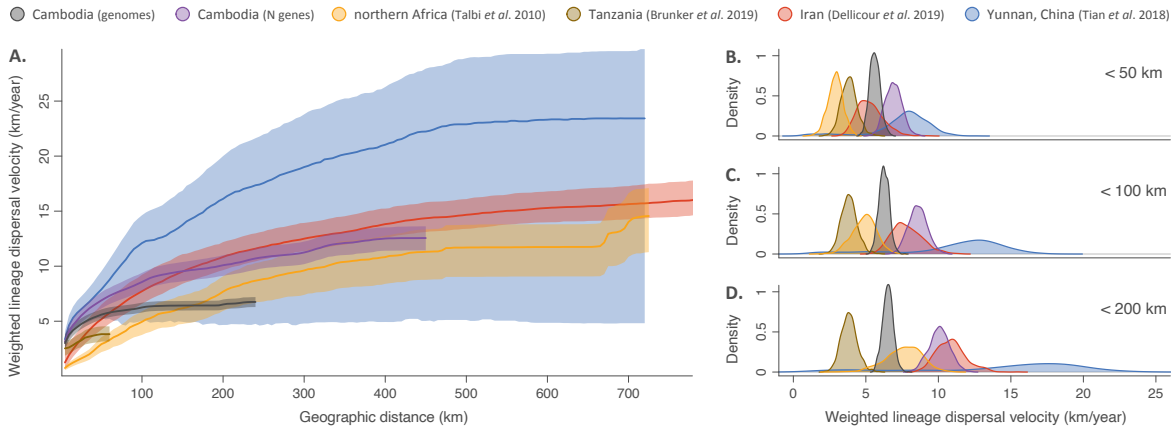


Figure 14: Comparison of RABV lineage dispersal velocity among different data sets. Specifically, we compare the weighted lineage dispersal velocity estimated for the following data sets: RABV in Cambodia (for both the full genomes and N genes data sets), northern Africa (Talbi et al., 2010), Tanzania (Brunker et al., 2018b), Iran (Dellicour et al., 2019), and Yunnan province in China (Tian et al., 2018). (**A**) Evolution of the weighted lineage dispersal velocity when increasing the maximal geographic distance used as the cut-off value to select the lineage dispersal events for the estimation. Solid curves and shaded polygons represent the median and 80% highest posterior density (HPD) interval estimates, respectively. (**B-D**) Weighted lineage dispersal velocity estimates for different maximal geographic distance cut-off values (50, 100, and 200 km, respectively).

which in turn could affect the dispersal velocity estimate due to a higher probability of long-distance dispersal events to be associated with a higher dispersal velocity. Under this assumption, such a higher probability could potentially be explained by two non-exclusive hypotheses. First, short-distance lineage dispersal events are, on average, more likely to capture local circulation of a transmission chain, which would involve a higher probability to infer phylogenetic branches associated with a short geographic distance. Second, long-distance dispersal events could be fast because they were human-mediated, which has been suspected in the case of RABV dispersal in other regions of the world (Talbi et al., 2010; Dellicour et al., 2017).

In our comparisons of WLDV estimates, we then tried to take into consideration the spatial extent of the respective study areas. For this purpose, we implemented a comparative approach in which we estimated the WLDV metric associated with each data set while increasing a cut-off value defining the maximal geographic distance that can be traveled by a phylogenetic branch retained for the WLDV estimation (Fig. 14A). Overall, this analysis confirmed that when progressively increasing this cut-off, we indeed observe an increase in the WLDV value estimated for a given data set. However, the WLDV estimated for the full genomes data set barely increased with the cut-off value, which is likely due to the overall low frequency of long-distance lineage dispersal events inferred for that data set. Our WLDV estimates for the N genes data set appear however quite close to the estimates obtained for a RABV data set from Iran (Dellicour et al., 2019), and to some extent to the estimates obtained for a RABV data set from

northern Africa (Talbi et al., 2010). For these three data sets, WLDV estimates tend to reach a WLDV plateau roughly ranging from 10 to 15 km/year when the cut-off value reaches 400 km (Fig. 14A; see also Fig. 14B-D for more direct comparisons based on specific cut-off values). In the light of those results, the dispersal velocity of RABV lineages does not seem particularly slower (nor faster) in Cambodia as compared with other regions of the world for which some equivalent estimates are currently available.

3.5. Investigating the impact of environmental factors on RABV dispersal

Finally, we took advantage of the spatially-explicit reconstructions of the dispersal history of RABV lineages in Cambodia to perform landscape phylogeographic analyses, which consisted in investigating the impact of environmental factors on the dispersal of viral lineages (Dellicour et al., 2018b). Specifically, we conducted two categories of analyses that aim at testing associations between environmental factors (Fig. 15) and (*i*) the dispersal velocity of viral lineages (Dellicour et al., 2017), or (*ii*) the dispersal location of viral lineages (Dellicour et al., 2019). The environmental factors included the main land use factors occurring in the study area (forests, savannas, grasslands, croplands, water areas), two global climate variables (mean annual temperature, annual precipitation), elevation, and human population density (see Table B1 for the source of the environmental data).

The first analyses led to some consistent results obtained when analyzing the full genomes and N genes data sets: they indicated that RABV lineages tended to avoid circulation in areas associated with relatively higher forest and savanna coverages (Bayes factors [BFs] > 20), and to preferentially circulate in areas associated with a relatively higher cropland coverage, human population density, and annual mean temperature (BFs > 20; Table B2). The analyses based on the N genes data set alone also return some support (BFs > 20) for a tendency of RABV lineages to avoid circulating in areas associated with relatively higher elevation and annual precipitation (Table B2). On a cautionary note, because their outcomes are strongly influenced by the sampling effort and pattern, those first analyses are somehow more a description of the environmental conditions related to the dispersal locations of inferred viral lineages than an actual test of the impact of those conditions on the dispersal (Dellicour et al., 2019). Consistently between the full genomes and N genes data sets, the second analyses of the dispersal velocity of viral lineages did not highlight any supported association between the environmental layers tested as either conductance or resistance factors and the heterogeneity in the lineage dispersal velocity (Tables B3 and B4).

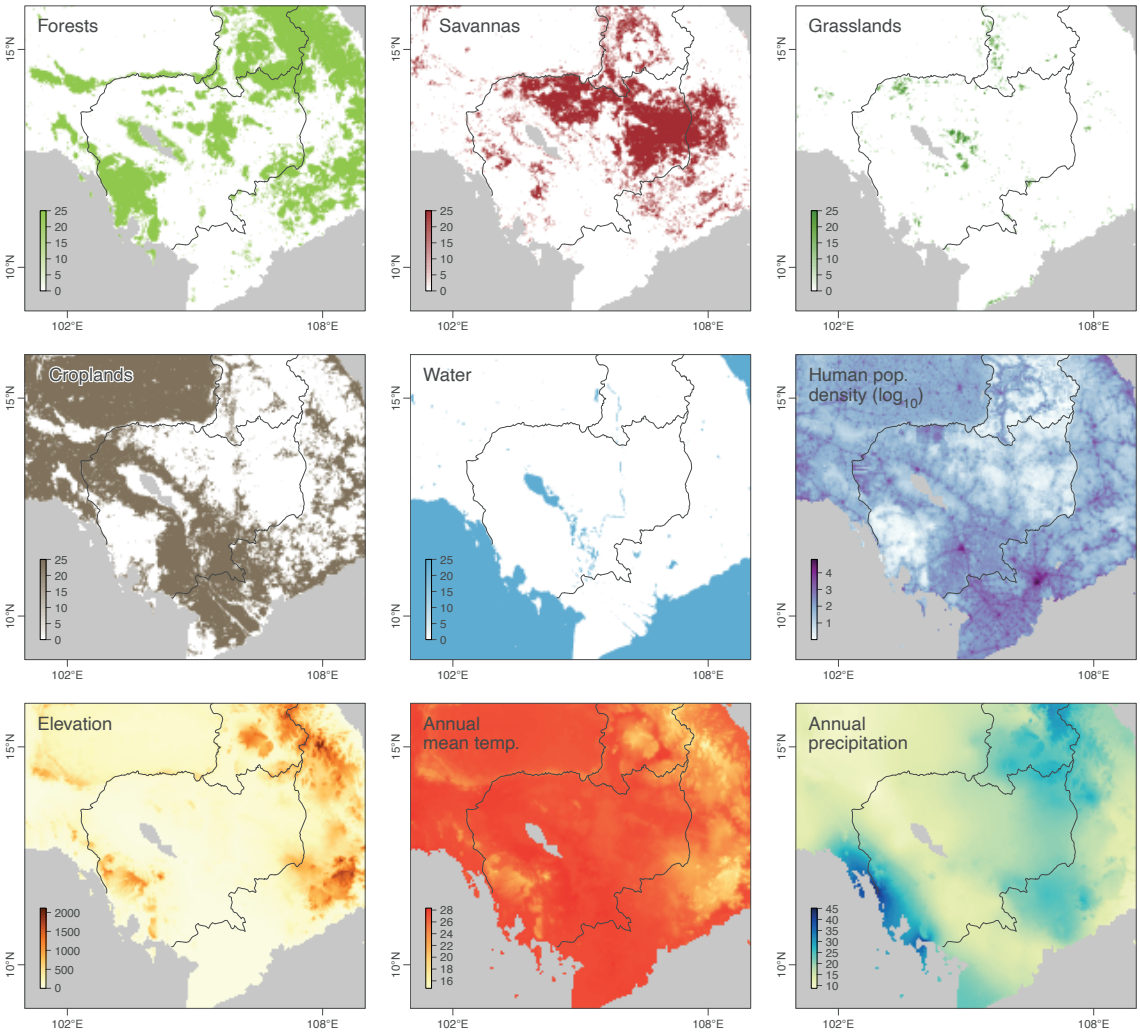


Figure 15: Environmental variables tested for their impact on the dispersal dynamic of RABV lineages in Cambodia. Elevation is reported in meters, mean annual temperature is reported in Celsius degrees, and annual precipitation is reported in meters per year.

4. Discussion

By quantifying external introductions and assessing their role relative to local transmission, the degree of endemic circulation of a disease can be determined in a given geographic area. As illustrated in the present study, phylogeographic approaches allow to objectify to what extent the local circulation of fast-evolving pathogens such as RNA viruses is fueled by external introduction events. Such approaches have for instance been broadly applied to tackle similar questions in the context of the COVID-19 pandemic (Attwood et al., 2022). It is worth noting that modeling approaches that rely on contact tracing data can also be used to study the role of imported cases to local dynamics, but data collection requires extensive active surveillance efforts (Lembo et al., 2008; Hampson et al., 2009; Mancy et al., 2022).

From a public health perspective, continuous RABV circulation in dog populations and frequent human infections correspond to sufficient criteria to consider the disease endemic in several African and Asian countries. However, RABV circulation may not solely rely on within-country transmission, and molecular epidemiological studies have shown that RABV often crosses country borders, sometimes likely driven by human movements (Lemey et al., 2009b; Talbi et al., 2010; Zhang et al., 2017). In addition, foreign introductions can lead to the local establishment of new lineages resulting in the co-circulation of multiple independent lineages in the long-term. The Indonesian island Bali exemplifies how rare events of RABV introductions can lead to the establishment of endemic transmission (Susilawathi et al., 2012). Foreign introductions could also be a driver of sustained RABV circulation if transmission chains rapidly go extinct. Indeed, a recent modeling study of contact tracing data in dogs showed that RABV circulates at a very low level in dog populations in the Serengeti district in Tanzania and most transmission chains rapidly die out, which would lead to the interruption of transmission without recurrent introductions from neighboring areas (Mancy et al., 2022). A similar observation was made in Bangui where control measures successfully eliminated rabies over several months but each time resurged due to introductions from neighboring areas (Bourhy et al., 2016; Colombi et al., 2020). In brief, the mechanisms of sustained circulation very likely depend on the geographical scale as well as geographic context of transmission at which it occurs.

Our results point out that RABV in Cambodia represents a remarkable case of endemicity with presumably very few external introductions feeding the local circulation of the virus. Notably, the vast majority of the Cambodian sequences analyzed in our study cluster within the same clade that we estimated to be circulating almost exclusively in Cambodia. This suggests that Cambodia has few exchanges of RABV lineages with its neighbors, a surprising pattern as Meng et al. (2011) predicted Cambodia to be an important source of rabies cases in China, Laos, and Thailand. Their results have however not been replicated in later studies indicating that they might result from small sample size and/or biased sampling. Furthermore, our sampling does not cover the northeastern region of Cambodia close to Laos and that is home for a particular dog population mostly used for hunting and relatively isolated from the rest of the country. We thus cannot exclude undetected introduction events from Laos to this population, and which would very certainly remain undetected in the rest of the Cambodian dog population. Nevertheless, the limited number of introduction events contrasts with the epidemiological situation in Vietnam (Nguyen et al., 2017), Laos (Ahmed et al., 2015), and Thailand (Benjathummarak et al., 2016) where multiple RABV lineages co-circulate and where some viral lineages are very close to Chinese lineages. It would mean that in those countries, RABV introductions and maintenance of these introduced lineages was probably

more frequent and successful. We could also have expected more successful importation events of RABV from neighboring countries to Cambodia due to dog trade for meat consumption that occurs across the whole Mekong region (Chevalier et al., 2021; Vu et al., 2021). While we might not have detected introductions through dog trade because sample collection has not focused on slaughterhouses, traded dogs are slaughtered and may not contribute to rabies transmission anyway (Chevalier et al., 2021). Finally, even if we employed a procedure to maximize the spatio-temporal coverage of our sampling, we note that the resulting selection of RABV samples remains more dense in densely-populated areas, in particular the area of Phnom Penh, which at least partially reflects a surveillance bias (Fig. 12D-E). While we cannot discard the hypothesis that a higher number of samples collected near the borders shared with neighboring countries might have allowed us to identify additional introduction events, we here argue that if those undetected introduction events had implanted active transmission chains in Cambodia, we would have likely detected them with the investigations we performed.

Surprisingly, we estimate a lower lineage dispersal velocity from the continuous phylogeographic reconstruction based on the full genomes compared to the equivalent reconstruction based on the N genes. We have further investigated this difference by comparing lineage dispersal velocity across multiple studies that analyzed dog RABV sequences using the same RRW diffusion model (Talbi et al., 2010; Brunker et al., 2018b; Tian et al., 2018; Dellicour et al., 2019). When the lineage dispersal velocity is estimated on lineage dispersal events of less than 50 km, estimates appear relatively similar across settings. On the contrary, lineage dispersal velocity estimates start to differ between data sets when increasing this distance cut-off value. This pattern suggests that at small spatial scales RABV roughly circulates at the same pace, which would remain coherent with a viral circulation primarily driven by infected dog movements. Notably, the vast majority of dogs, even when owned, are free-roaming in Cambodia (Chevalier et al., 2021) as in other countries (Muinde et al., 2021; Warembourg et al., 2021; De la Puente-Arévalo et al., 2022; Sparkes et al., 2022). Our results also suggest that at larger spatial scales, it becomes more difficult to compare the lineage dispersal velocity estimated from different data sets. As introduced above, larger sampling areas are expected to be associated with higher probabilities to sample long-distance dispersal events that will, on average, more likely correspond to fast dispersal events than short-distance dispersal events. Therefore, comparing the lineage dispersal velocity of data sets sampled from study areas of very different spatial extents might potentially lead to an artifactual conclusion of different dispersal velocities. Overall, we would thus recommend to use a similar approach based on maximal distance cut-off values when aiming to compare lineage dispersal velocity statistics estimated from data sets collected across various spatial scales.

While our landscape phylogeographic analyses have allowed us to characterize the environmental conditions in which the inferred lineages tended to disperse within Cambodia, they have not led to any supported evidence that one or several studied environmental factor(s) could partly explain the heterogeneity in viral lineage dispersal velocity. It could imply that the velocity of RABV dispersal in Cambodia is at least not drastically shaped by the environmental factors tested in our study, but this could also result from an insufficient sampling of lineages circulating in some Cambodian regions associated with different environmental conditions. In the latter case, insufficient sampling in these areas could, for instance, limit the statistical power of the tests aiming at detecting an association between the spatial variation of an environmental factor and the dispersal velocity of viral lineages (Dellicour et al., 2016b). More generally, phylogeographic approaches are known to be sensitive to sampling bias (De Maio et al., 2015; Guindon and De Maio, 2021; Kalkauskas et al., 2021; Liu et al., 2022). One way to mitigate the impact of sampling bias on phylogeographic reconstructions is to subsample locations according to local incidence data. During the COVID-19 pandemic, this kind of subsampling approach has been widely implemented in the context of phylogeographic analyses targeting SARS-CoV-2 (Dellicour et al., 2021b; Hodcroft et al., 2021b; Lemey et al., 2021), but it would in our case require the availability of dog rabies incidence data and sufficient numbers of genomic sequences from each location to subsample from. Unfortunately, neither condition was met in the context of the present study. While not based on incidence data, our samples selection is however based on a procedure implemented as a Markov process trying to maximize the spatio-temporal coverage of the selected samples to sequence, which corresponds to an optimal sampling pattern when aiming to capture the dispersal history of lineages through phylogeographic inference.

Even though we did not identify external introductions as drivers of RABV circulation in Cambodia, they remain likely and could still lead to re-importations of active transmission chains as observed in the neighboring countries. However, considering that dog rabies incidence in Cambodia appears to be primarily driven by within-country transmission, the deployment of an ambitious rabies control policy should focus on campaigns of free and regular dog vaccination at the national level (Tarantola et al., 2015; Sor et al., 2018), promotion of owner awareness (Sor et al., 2018; Chevalier et al., 2021), improvement of surveillance and diagnostic capabilities (Duong et al., 2016), improved multisectoral collaboration in a one health approach, and expansion of post-exposure prophylaxis (Li et al., 2019).

Chapter 3

Impact of vaccination on household transmission of SARS-CoV-2 in Israel

Several studies have characterized the effectiveness of vaccines against SARS-CoV-2 infections. However, estimates of their impact on transmissibility remain limited. Here, we evaluated the impact of isolation and vaccination (7 days after the second dose) on SARS-CoV-2 transmission within Israeli households. From December 2020 to April 2021, confirmed cases were identified among health-care workers of the Sheba Medical Centre and their family members. Recruited households were followed up with repeated PCR for at least 10 days after case confirmation. Data were analyzed using a data augmentation Bayesian framework. A total of 210 households with 215 index cases were enrolled; 269 out of 667 (40%) susceptible household contacts developed a SARS-CoV-2 infection. Of those, 170 (63%) developed symptoms. Compared with unvaccinated and unisolated adult/teenager (aged >12 years) contacts, vaccination reduced the risk of infection among unisolated adult/teenager contacts (relative risk (RR) = 0.21, 95% credible interval (CrI): 0.08, 0.44), and isolation reduced the risk of infection among unvaccinated adult/teenager (RR = 0.12, 95% CrI: 0.06, 0.21) and child contacts (RR = 0.17, 95% CrI: 0.08, 0.32). Infectivity was reduced in vaccinated cases (RR = 0.25, 95% CrI: 0.06, 0.77). Within households, vaccination reduces both the risk of infection and of transmission if infected. When contacts were unvaccinated, isolation also led to important reductions in the risk of transmission.

1. Introduction

SARS-CoV-2 is a highly transmissible virus that was first detected in Wuhan China in December 2019 (Lu et al., 2020; Izda et al., 2021). It is the cause of COVID-19, which has spread through the world, leading to a pandemic that had infected at least 250 million people and caused more than 5 million deaths worldwide by November 10, 2021 (Mathieu et al., 2021). The advent of novel COVID-19 vaccines has been an important breakthrough in the management of the pandemic. To determine how vaccination may modify epidemic dynamics, it is essential to estimate its effectiveness with respect to infection, transmission, and disease severity. Multiple studies have shown that COVID-19 vaccines are effective at reducing both the risk of infection (Dagan et al., 2021; Tande et al., 2022; Pawlowski et al., 2021; SE, 2021; Martínez-Baz et al., 2021) and the risk of developing severe symptoms (Dagan et al., 2021; Martínez-Baz et al., 2021; Haas et al., 2021; Goldberg et al., 2021) in the general population.

Documenting vaccine impact on transmission is more challenging, stemming from the difficulty of thoroughly documenting chains of transmission and accounting for the ways different types of contacts may lead to different risks of transmission (Bi et al., 2020). Households represent the perfect environment to evaluate factors affecting transmission such as vaccination because the probability of SARS-CoV-2 transmission among household members is high, ranging between 14% and 32% (Madewell et al., 2020; Lei et al., 2020; Thompson et al., 2021). Beyond the evaluation of vaccine effectiveness, understanding how vaccines affect household transmission is also important to determine how recommendations should evolve with vaccines. For example, should isolation precautions be maintained in partially vaccinated households (World Health Organization (WHO), 2021b)? A number of studies have shown that vaccines provide indirect protection against household transmission (Shah et al., 2021; Salo et al., 2021; Harris et al., 2021; Prunas et al., 2022; Gier et al., 2021). However, none of these studies evaluated how isolation affected the outcome, and for some of the studies (Shah et al., 2021; Salo et al., 2021; Harris et al., 2021; Prunas et al., 2022), the passive nature of surveillance may have led to underestimating household transmission rates.

During the first months of 2021, Israel underwent its third pandemic wave due to the rise of the Alpha variant that quickly accounted for 90% of infections (Munitz et al., 2021). Concomitantly, vaccination was extended to all adults older than age 16 years, making Israel one of the first countries to reach high vaccination coverage in their population, with 60% of the total population being vaccinated by March 22, 2021 (Mathieu et al., 2021). During this period, we followed SARS-CoV-2 transmission in the households of 12,518 HCWs of the Sheba Medical Center, the largest medical center in Israel. Here, we

describe dynamics of transmission in these households and evaluate the impact vaccination and isolation measures had on these dynamics.

2. Material and Methods

2.1. Study design and study population

All HCWs, regardless of their vaccination status, were required to use an electronic questionnaire to report daily any COVID-19 related symptom they, or a member of their household, had. SARS-CoV-2 PCR testing was readily available, and HCWs were encouraged to be tested for any mild symptom or suspected exposures. All HCWs were instructed to notify the infection prevention and control unit if one of their household members was SARS-CoV-2 positive. All SARS-CoV-2–detected HCWs as well as those with a positive SARS-CoV-2 household member were immediately contacted as part of the epidemiologic investigation for contact tracing and were provided with instructions regarding isolation precautions. All unvaccinated household members (i.e., those that did not receive the 2 vaccine doses at least 7 days before the detection of the COVID-19 patient) were required to perform 2 PCR tests in the 10 days after the diagnosis of the positive COVID-19 patient. Vaccinated household members were encouraged to perform 2 PCR tests during the 10 days after detection. Household members were not required to test a second time if they had a positive test (Table C1). Unvaccinated HCW contacts were isolated at home, whereas vaccinated HCWs were instructed to perform a PCR test every day they reported to the hospital for work.

Between December 31, 2020, and April 26, 2021, the HCWs who were SARS-CoV-2–positive or reported a positive household member were contacted at least 10 days after detection and were offered enrollment in the study. Those who agreed, and gave their consent, answered a telephone interview.

2.2. Data and sample collection

Data collected during the phone interview included the age and gender of the HCW's household members, their vaccination status, information about prior COVID-19 infections, their COVID-19 PCR test dates and results, their symptoms (i.e., fever, cough, myalgia, headache, congestion, diarrhea, vomiting, anosmia, or ageusia), the number of rooms and bathrooms in the household, and the degree to which isolation precautions were adhered to (Section 2 in Appendix C). At the time of the study, only individuals 16 years old or older were eligible for vaccination. The household member who had the first

positive PCR test was defined as the index case. When multiple household members had a positive PCR test on the same day, they were defined as co-index cases. We defined complete isolation as complete separation in sleeping and eating between household contacts and index case(s) (i.e., they did not spend any time in the same room) and whether a separate bathroom was provided for the index case(s). Partial isolation was defined if one of the above was violated, but masks were continuously used, and eating was consistently separate.

For HCWs, nasopharyngeal swabs were collected by trained personnel, and reverse-transcriptase quantitative PCR analysis was performed using the Allplex 2019-nCov RT-qPCR assay (Seegene Inc., Seoul, South Korea) and expressed by cycle threshold (Ct). Other household members reported the results of their COVID-19 test(s) performed by their health-care providers.

2.3. Clinical outcome

Confirmed SARS-CoV-2 infections were defined by a positive PCR test (i.e., with a Ct value lower than 40). Symptomatic cases were defined as confirmed cases with the presence of at least 1 symptom from among the following: fever, cough, myalgia, headache, congestion, diarrhea, vomiting, anosmia, or ageusia. Contacts who reported at least 1 of the above-mentioned symptoms but were not confirmed because they performed no PCR test (n = 6) or a single test at inclusion (n = 2) were also considered as symptomatic cases. Asymptomatic cases were defined as confirmed cases who did not report any symptom over the follow-up period of the household.

2.4. Statistical analysis

We evaluated transmission in households using 2 metrics: the secondary attack rate (SAR), defined as the proportion of susceptible household contacts that are infected after the index case is detected (Liu et al., 2020), and the person-to-person probability of transmission, defined as the per-capita probability that an infected individual transmits to a susceptible household contact. The first metric includes tertiary (i.e., household contacts infected by a household member that is not the index case) and community cases (i.e., household contacts infected in the community) contrary to the second metric. In both cases, we assumed that individuals who reported past infection of SARS-CoV-2 confirmed by PCR over the year preceding the detection of the household index case (n = 20) were protected from infection and therefore, did not count as susceptible household contacts.

Baseline characteristics of the index cases and household contacts were described according to their vaccination status. All individuals older than 12 years were considered as adults/teenagers. We calculated

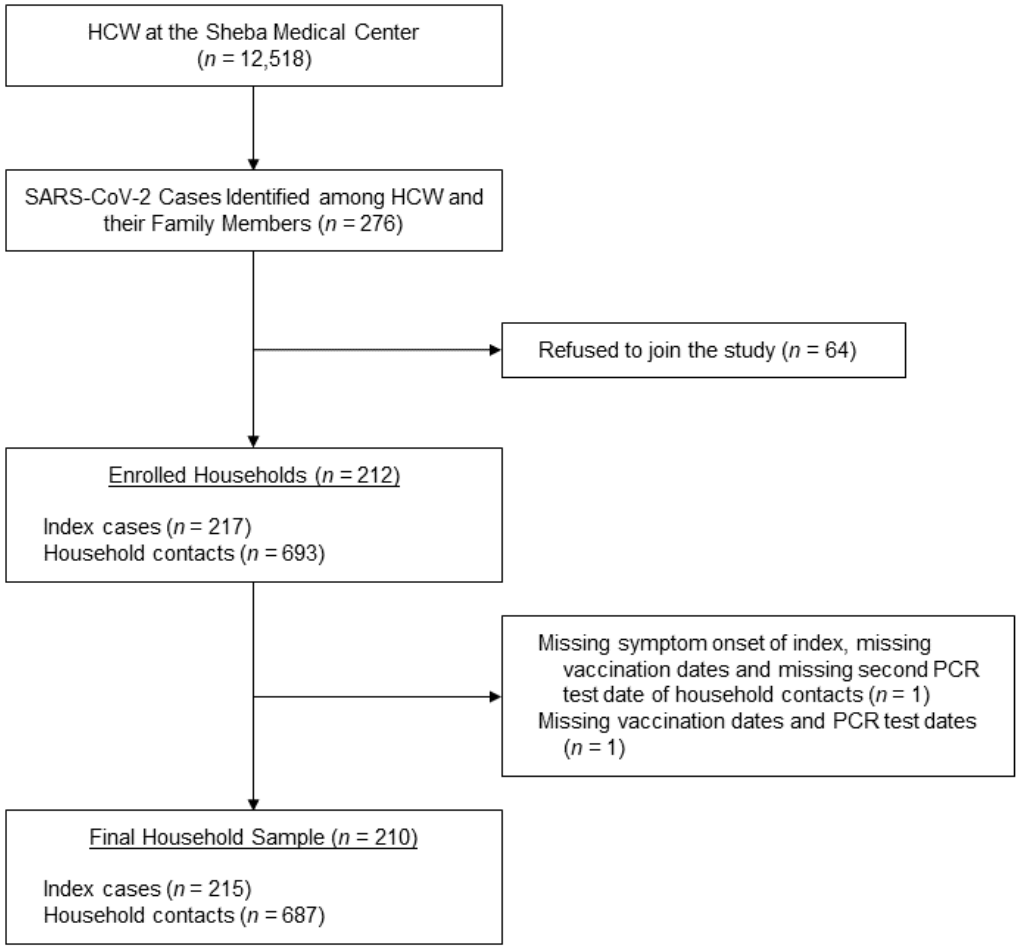


Figure 16: Flow chart of the households included in our analysis, Ramat Gan, Israel, 2020–2021. HCW, health-care worker; PCR, polymerase chain reaction; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

the SAR for different categories of household contacts: unisolated and unvaccinated adults/teenagers, unisolated and vaccinated adults/teenagers, isolated and unvaccinated adults/teenagers, vaccinated and isolated adults/teenagers, unisolated children, and isolated children. Here, isolation corresponds to complete or partial isolation between household contacts and the index case. We also defined the SAR of vaccinated and unvaccinated index cases as the proportion of infected household contacts in households with vaccinated or unvaccinated index cases, respectively. In a sensitivity analysis, the SAR calculation was restricted to households in which a single index case was identified (Table C2). We also report the 95% confidence interval of the SAR. We developed a statistical model to evaluate the effect of age, isolation precautions, BNT162b2 vaccination, and household size on SARS-CoV-2 transmission dynamics in households (Section 4 in Appendix C). The model uses the sequence of symptom onset dates and positive molecular test dates to estimate the person-to-person risk of transmission within the household while accounting for the community hazard of infection (i.e., household contacts infected outside the

household) and the possibility of tertiary transmissions (i.e., household contacts infected by a member of the household that is not the index case) (Cauchemez et al., 2004). The person-to-person risk of transmission is decomposed into the baseline person-to-person risk of infection depending on household size, the relative infectivity of the infector depending on their vaccination status (reference group: unvaccinated cases), and the relative susceptibility of the infectee depending on their age, isolation behavior, and vaccination status. The relative susceptibility is estimated separately for unisolated children, isolated children, isolated and unvaccinated adults/teenagers, unisolated and vaccinated adults/teenagers, and adults/teenagers that are both isolated and vaccinated, considering the group of adults/teenagers that are unisolated and unvaccinated as the reference group. None of the children were vaccinated at the time of the study. This formulation accommodates the potential confounding effects between the 3 variables characterizing household contacts (i.e., being vaccinated, being isolated, or being a child). We assumed that individuals whose isolation behavior was missing ($n = 6$) did not comply with isolation precautions.

Model parameters were estimated using Bayesian Markov chain Monte Carlo sampling with data augmentation (Cauchemez et al., 2004) (Section 5 in Appendix C). Data were augmented with the probable date of infection of confirmed cases. For symptomatic cases, the date of infection was reconstructed from the date of symptom onset, using the probabilistic distribution of the incubation period (McAloon et al., 2020). For asymptomatic cases, we assumed that the date of infection could occur up to 10 days prior to their molecular detection based on a meta-analysis (Cevik et al., 2021).

Since the study was conducted during the vaccine rollout, participants were enrolled at varying stages of their vaccination process. We assumed that vaccines reach their full effect 7 days after receiving a second dose (Dagan et al., 2021; Haas et al., 2021; Goldberg et al., 2021). Cases were therefore considered vaccinated if their symptom onset (or if unknown, the date of their first positive PCR test) occurred ≥ 7 days after the second dose. Similarly, household contacts were considered vaccinated if their exposure to the index case (starting with symptom onset or, in its absence, from the date of first positive PCR of the index case) occurred ≥ 7 days after the second dose. In a sensitivity analysis, we investigated how parameter estimates changed under the assumption that vaccination is effective ≥ 15 days after the first dose. We also assessed how estimates changed when the analysis was restricted to households in which all negative contacts had performed at least 1 or 2 PCR tests in the 10 days following the detection of the index case. In the baseline scenario, we assumed that asymptomatic cases are 40% less infectious than symptomatic cases based on a meta-analysis (Byambasuren et al., 2020), and we investigated whether assuming the same level of infectivity in asymptomatic and symptomatic cases modified our estimates.

Finally, in our baseline analysis, we chose a log-normal with log-mean = 0 and log-standard deviation = 1 prior distribution for the relative infectivity and relative susceptibility parameters and explored smaller and larger values (log-standard deviation = 0.7 or 2) in a sensitivity analysis.

We compared the observed and expected distributions of the number of cases per household size to assess the goodness-of-fit of the model (Table C3). We report the posterior median and the 95% credible interval (CrI) of estimated parameters. We also report the posterior probability that isolated and vaccinated adult/teenager contacts are less susceptible than vaccinated adult/teenager contacts that do not isolate. To measure the strength of evidence of a reduced susceptibility in isolated contacts among vaccinated ones, we report the associated Bayes factor. Here, it directly corresponds to the posterior odds of a reduced susceptibility in isolated contacts among vaccinated ones. Additional details are available in Sections 1 to 6 in Appendix C.

2.5. Ethics

The study was approved by the Sheba Medical Center institutional review board committee (approval #8130-21).

3. Results

All 12,518 HCWs employed by the Sheba Medical Center were eligible to join the study. Between December 19 and April 28, 2021, 91% of the Sheba Medical Center personnel received both doses of the BNT162b2 vaccine, and a rapid and significant decrease in newly detected cases was observed among HCWs.

From December 31, 2020, to April 26, 2021, 276 SARS-CoV-2 cases were identified among HCWs of the Sheba Medical Center and their household members (Fig. 16). Of these, 212 agreed to participate, gave their consent, and were enrolled in the study with their household members. Two households were excluded due to missing vaccination status, dates of PCR test, and/or symptom onset. In total, we analyzed data from 210 households with 215 index cases, including 4 co-index cases, and their 687 household contacts. The median household size was 4 (interquartile range, 3–5). Mean age was 32 years among index cases (Table 4) and 27 years among household contacts (Table 5). Age was missing for 5 adult/teenager contacts, and isolation behavior was missing for 6 contacts. There was a slight over-representation of females among index cases (58%), and 191 index cases (89%) were adults/teenagers, of whom 15 (8%) were vaccinated. None of the 24 child index cases were vaccinated. Among the 494

Table 4: Characteristics of the Index Cases According to Age.

Characteristic	Adult/teenager index cases[a] (N = 191)			Child index cases (N = 24)			All index cases (N = 215)			
	No.	%	Median (IQR)	No.	%	Median (IQR)	No.	%	Median (IQR)	Median (IQR)
Male sex	76	40		14	58		90	42		
Age, years[b]	36 (14)			6 (4)			32 (16)			
Symptom status										
Symptomatic	172	90		10	42		182	85		
Asymptomatic	19	10		14	58		33	15		
Vaccination										
Vaccinated	15	8		N/A	N/A		15	7		
Days from 2nd dose to detection			44 (13 – 59)			N/A			44 (13 - 59)	44 (13 - 59)

Abbreviations: IQR, interquartile range; N/A, not applicable.
[a] Individuals aged >12 years were considered adults/teenagers.
[b] Values are expressed as mean (standard deviation).

adult/teenager household contacts, 125 (25%) were vaccinated. Of these, 83 (17%) also complied with isolation precautions. Among the 369 unvaccinated adult/teenager contacts, 259 (70%) isolated during the study. None of the 193 child household contacts were vaccinated and 47% of them (n = 90) isolated during the study period (Table 5). In the following, we refer to susceptible contacts (i.e., contacts that did not report SARS-CoV-2 infection over the preceding year) as contacts.

A total of 269 out of 667 (40%) household contacts developed a SARS-CoV-2 infection. Of those, 170 (63%) developed symptoms (Table 5). The SAR varied with the characteristics of the contacts. Among the 105 adult/teenager contacts who were unisolated and unvaccinated, 80 (76%) were infected by SARS-CoV-2 (Table 6). This proportion dropped to 28% (11 out of 40) among those who were unisolated and vaccinated, 29% (71 out of 245) among those who were isolated but unvaccinated, and 11% (9 out of 83) among those who were isolated and vaccinated; 65% (66 out of 101) of child contacts who were unisolated got infected by SARS-CoV-2. This proportion declined to 33% (29 out of 87) for isolated child contacts. The proportion of asymptomatic cases varied from 26% (46 out of 174) among adult/teenager contact cases to 56% (53 out of 95) among child contact cases (Table 5).

The SAR also varied with the vaccination status of the index case regardless of the contacts' characteristics. Among the 622 household contacts whose index case was unvaccinated, 261 (42%) developed a SARS-CoV-2 infection (Table 6). This proportion dropped to 19% (8 out of 42) among household contacts whose index case was vaccinated. Finally, the SAR was relatively invariant with household size: 31%, 40%, 32%, and 32% for households of size 2, 3, 4, and 5, respectively (Fig. C1).

Table 5: Characteristics of the Household Contacts According to Age.

Characteristic	Adult/teenager household contacts[a] (N = 494)			Child household contacts (N = 193)			All household contacts (N = 687)		
	No.	%	Median (IQR)	No.	%	Median (IQR)	No.	%	Median (IQR)
Male sex	242	49		109	56		351	51	
Age, years[b]	36 (17)[c]			6 (4)			27 (20)		
Infection and symptom status									
Past infection	16	3		4	2		20	3	
Not infected	304	62		94	49		398	58	
Symptomatic	127	26		41	21		168	24	
Asymptomatic	46	9		53	27		99	14	
Symptomatic (missing onset)	1	0		1	1		2	0	
Vaccination									
Vaccinated	125	25		N/A	N/A		125	18	
Days from 2nd dose to exposure			23 (14 - 36)			N/A			23 (14 - 36)
Isolation									
Partial	115	23		32	17		147	21	
Complete	227	46		58	30		285	41	
Missing	5	1		1	1		6	1	

Abbreviation: IQR, interquartile range; N/A, not applicable.
[a] Individuals aged >12 years were considered adults/teenagers.
[b] Values are expressed as mean (standard deviation).
[c] Missing age for 5 adult/teenager contacts.

Our statistical model makes it possible to perform a multivariate analysis of the drivers of SARS-CoV-2 transmission in households. We estimate that, relative to adult/teenager contacts who were unisolated and unvaccinated, the relative risk of being infected was 0.21 (95% CrI: 0.08, 0.44) among adult/teenager household contacts who were vaccinated but unisolated (Fig. 17A, Table C4). It was 0.12 (95% CrI: 0.06, 0.21) among household contacts who did isolate and were unvaccinated, and 0.07 (95% CrI: 0.03, 0.16) among household contacts who were both isolated and vaccinated. Isolation might reduce the risk of infection among vaccinated contacts (96% posterior probability, Bayes factor = 23) with a relative risk of 0.34 (95% CrI: 0.11, 1.14). Relative to adult/teenager contacts who were unisolated and unvaccinated, the relative risk of infection was 0.50 (95% CrI: 0.32, 0.77) for child contacts that did not isolate, and 0.17 (95% CrI: 0.08, 0.31) for those that did. We estimate that the risk of transmission from vaccinated cases was 0.25 (95% CrI: 0.06, 0.77) times that of unvaccinated cases (Fig. 17B and Table C4).

Overall, we estimate that, in a household of size 4, the person-to-person probability of SARS-CoV-2 transmission is 61% (95% CrI: 48, 72) between an unvaccinated case and an unvaccinated and unisolated adult/teenager. This probability drops to 4% (95% CrI: 1, 16) between 2 vaccinated adults/teenagers who do not follow isolation rules (Fig. 18 and Table C5). The person-to-person probability of transmission

Table 6: Observed Secondary Attack Rates According to the Type of Contact.

Type	No. of infected contacts	No. of susceptible contacts	SAR	
			%	95% CI
Contacts[a]				
Unisolated and unvaccinated adult/teenager	80	105	76	67, 84
Isolated and unvaccinated adult/teenager	71	245	29	23, 35
Unisolated but vaccinated adult/teenager	11	40	28	15, 44
Isolated and vaccinated adult/teenager	9	83	11	5, 20
Unisolated child	66	101	65	55,75
Isolated child	29	87	33	24, 44
Index[b]				
Vaccinated	8	42	19	9, 34
Unvaccinated	261	622	42	38, 46

Abbreviations: CI, confidence interval; SAR, secondary attack rate.

[a] Isolation is missing for 1 child contact and for 5 adult contacts.

[b] The last 2 rows correspond to the SAR among the household contacts of vaccinated (n = 14 households) and unvaccinated index cases (n = 195 households). One household was excluded from this analysis because its co-index cases did not have the same vaccination status.

from an unvaccinated case to a child who does not isolate is 37% (95% CrI: 27, 48). This probability drops to 11% (95% CrI: 3, 31) if the case is vaccinated and to 14% (95% CrI: 7, 25) if the child contact is isolated.

In general, our estimates of relative susceptibility and relative infectivity were robust to model assumptions (Fig. 19). When the analysis was restricted to households in which all contacts performed at least 1 or 3 PCR tests in the 10 days following the recruitment of the index case, the relative susceptibility of vaccinated adult/teenager contacts who did not isolate was slightly higher compared with the baseline scenario. It increased from 0.21 (95% CrI: 0.08, 0.44) in the baseline scenario to 0.28 (95% CrI: 0.09, 0.66) in the analysis with at least 1 PCR and 0.32 (95% CrI: 0.09, 0.83) with at least 2 PCR tests (Table C4). In the alternative scenarios, the number of individuals included was substantially lower, increasing CrIs (Figs. C3 and C4 and Tables C6 to C9). Similarly, the relative susceptibility of vaccinated adult/teenager contacts who did isolate increased from 0.07 (95% CrI: 0.03, 0.16) in the baseline scenario to 0.12 (95% CrI: 0.04, 0.28) in the analysis with at least 1 PCR, and 0.13 (95% CrI: 0.04, 0.32) in the one with at least 2 PCR tests. Consequently, the posterior probability that isolated and vaccinated adult/teenager contacts were less susceptible than vaccinated adult/teenager contacts that did not isolate dropped from 96% to 88% with 1 PCR and 89% with 2 PCR tests. Still, the statistical support was high with a Bayes factor equal to 7 and 8, respectively. Relative infectivity and relative susceptibility were slightly sensitive to their prior distribution (Table C10). When the log-standard deviation increased,

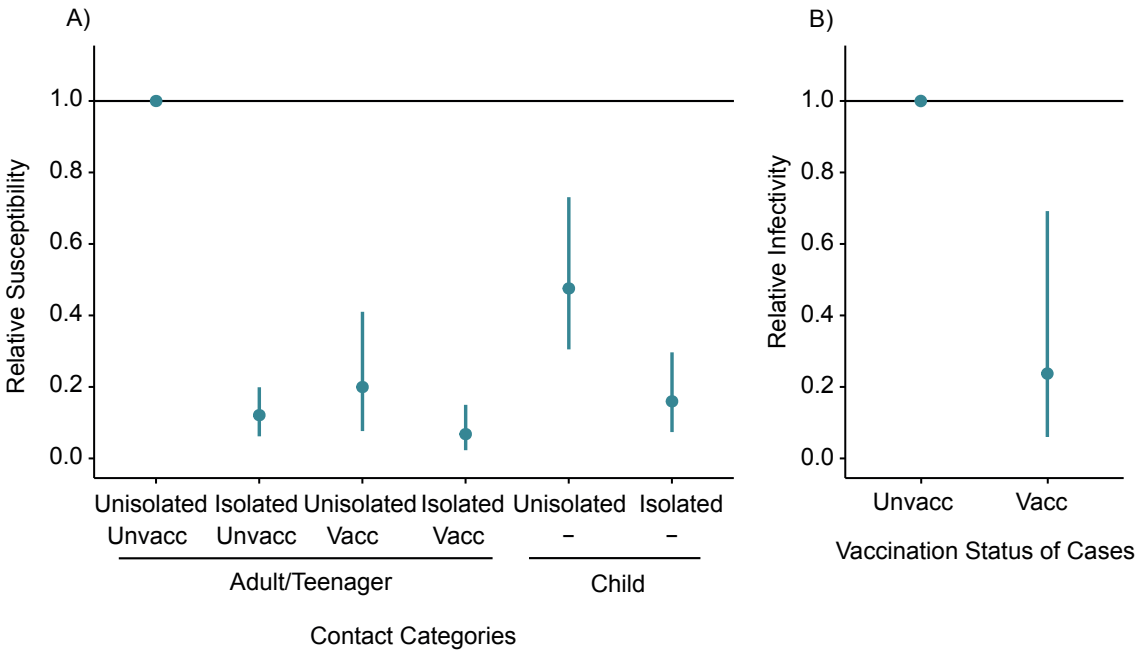


Figure 17: Estimates of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) transmission parameters within households, Ramat Gan, Israel, 2020–2021. (**A**) Estimated relative susceptibility of isolated and unvaccinated adults/teenagers, unisolated but vaccinated adults/teenagers, isolated and vaccinated adults/teenagers, unisolated children, and isolated children. The reference group is the group of adults/teenagers that were unisolated and unvaccinated. (**B**) Estimated relative infectivity of vaccinated cases compared with unvaccinated cases. The posterior median and its associated 95% Bayesian credible interval are reported.

estimates were pulled towards lower values.

4. Discussion

We evaluated the impact of BNT162b2 vaccination on case infectivity and the mitigating effect of age, isolation from the index case, and BNT162b2 vaccination on susceptibility to infection in household settings. Our approach accounts for infections in the community, potential tertiary infections within the households, the reduced infectivity of asymptomatic cases, potential misidentification of the index case(s), and varying follow-up periods between households.

In our analysis, the SAR in unvaccinated adult/teenager contacts who did not isolate was estimated at around 76%, which is substantially higher than previous estimates obtained in household settings (Madewell et al., 2020; Lei et al., 2020; Thompson et al., 2021; Harris et al., 2021; Jing et al., 2020; Li et al., 2021). In meta-analyses (Madewell et al., 2020; Lei et al., 2020; Thompson et al., 2021), the average SAR ranged between 14% and 32%; however, in some studies, it could be as high as 90% (Lei et al., 2020). Most of these studies date back to the time when historical lineages were still dominant. In contrast, our study took place when the Alpha variant represented up to 90% of infections in Israel

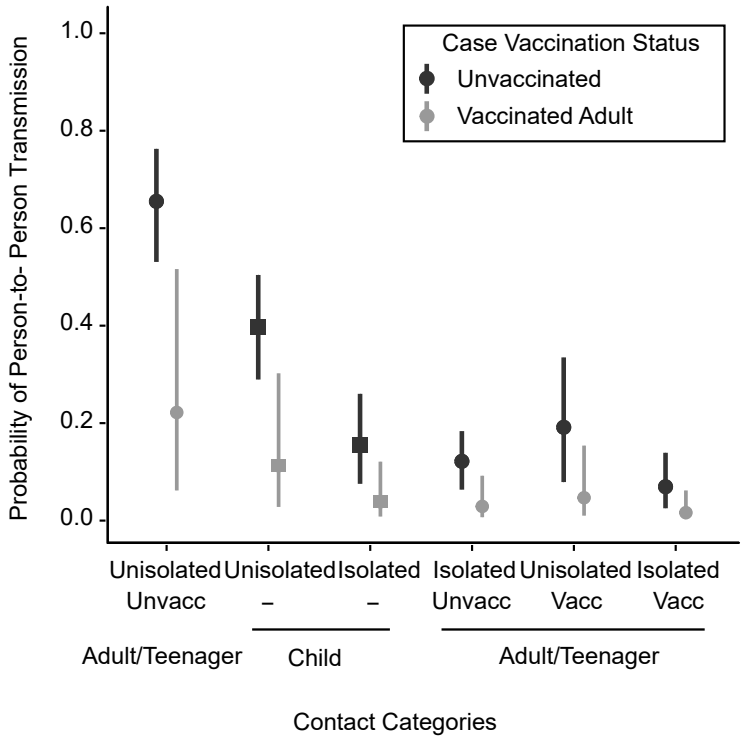


Figure 18: Person-to-person probability of transmission within households according to the characteristics of the case and of the contact, Ramat Gan, Israel, 2020–2021. Estimated person-to-person probability of transmission within households of size 4, decomposed by the age, isolation behavior, and vaccination status of the contact as well as the vaccination status of the case. The posterior median and its associated 95% Bayesian credible interval are reported.

(Munitz et al., 2021). Our higher estimate could be at least partly explained by the fact that the Alpha variant is substantially more transmissible than historical lineages (Munitz et al., 2021; Kissler et al., 2021; Davies et al., 2020b; Volz et al., 2021).

In agreement with previous reports, we found that children are less susceptible to SARS-CoV-2 infections than adults/teenagers (Madewell et al., 2020; Lei et al., 2020; Thompson et al., 2021; Viner et al., 2021). We further estimated that, 7 days after their second dose, vaccinated adults/teenagers benefit from a 79% reduction in the risk of infection compared with unvaccinated adults/teenagers. We show, consistent with previous studies (Munitz et al., 2021; Pritchard et al., 2021), that BNT162b2 vaccination is highly effective against infection by the Alpha variant. In general population studies, vaccine effectiveness for symptomatic infections ranged from 57% 14 days after the first dose (Dagan et al., 2021) to 89% (Dagan et al., 2021), and 97% 7 days after the second dose (Haas et al., 2021). For asymptomatic infections, vaccine effectiveness against infection was 79% 10 days after the first dose (Tande et al., 2022) and 94% 14 days after the second dose (SE, 2021). Our estimate of vaccine effectiveness in household settings is lower than those obtained in the general population. This is consistent with estimates obtained in households (Prunas et al., 2022; Gier et al., 2021; Pritchard et al., 2021) and might in part be explained by

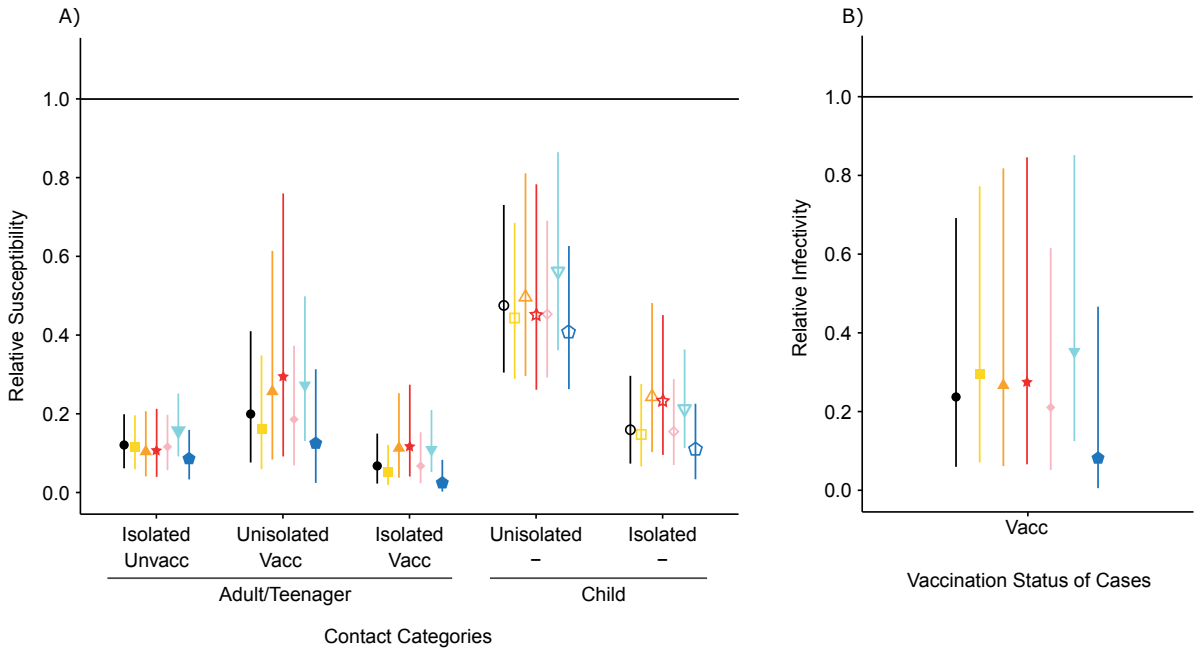


Figure 19: Impact of model assumptions on the estimation of the relative susceptibility and relative infectivity parameters, Ramat Gan, Israel, 2020–2021. (**A**) Estimates of the relative susceptibility of household contacts for the baseline and sensitivity analysis scenarios. (**B**) Estimates of the relative infectivity of vaccinated cases compared with unvaccinated ones for the baseline and sensitivity analysis scenarios. In the baseline scenario (black circle), we assumed that vaccination was effective from 7 days after the second dose, the relative infectivity of asymptomatic cases compared with symptomatic cases was equal to 60%, and the log-standard deviation of the relative infectivity and relative susceptibility prior distributions was equal to 1. Sensitivity analysis scenarios: yellow square, vaccination is effective $\geq$ 15 days after the dose; orange triangle, 1 polymerase chain reaction (PCR) test for all negative contacts; red star, 2 PCR tests for all negative contacts; pink diamond, 100% infectivity of asymptomatic cases; blue inverted triangle, relative parameter prior with log-standard deviation = 0.7; blue pentagon, relative parameter prior with log-standard deviation = 2. The posterior median and its associated 95% Bayesian credible interval are reported.

the elevated contact rates in households that may favor transmission. Additionally, studies in the general population are less suitable to detect all asymptomatic cases compared with the household setting. This might lead general population studies to overestimate vaccine effectiveness against asymptomatic infections if vaccinated contacts are less often tested than unvaccinated ones. On another note, we estimate a vaccine effectiveness against transmission of 75% (95% CrI: 23, 94), which is in line with other studies in household settings (Harris et al., 2021; Prunas et al., 2022; Gier et al., 2021).

To our knowledge, this is the first study estimating the effect of isolation on SARS-CoV-2 transmission in households that are partially vaccinated. We showed that isolation precautions markedly reduce the overall infection risk in both adult/teenager and child contacts even when considering partial physical distancing measures. We estimated a similar reduction of infection in adult/teenager contacts that were vaccinated but did not isolate. There was a signal in the data that isolation also benefited vaccinated individuals, although credible intervals were larger, and further investigations are required to confirm

this finding.

Our study has several limitations. First, household studies such as ours may be affected by multiple sources of bias. On the one hand, we may overestimate the SAR if we are more likely to detect households with multiple cases. On the other hand, we might underestimate it if some asymptomatic, or paucisymptomatic, cases are missed during follow-up. Second, we estimated an important reduction of infectivity in vaccinated cases with 2 doses compared with unvaccinated cases as previously shown (Harris et al., 2021; Prunas et al., 2022; Gier et al., 2021; Regev-Yochay et al., 2021). However, this is associated with important uncertainty due to the small number of cases (15 vaccinated index cases and 21 vaccinated secondary cases). Thus, more data are needed to reduce the size of credible intervals. Third, we assumed that vaccination was effective from 7 days after the second dose (or 15 days after the first dose in our sensitivity analysis; see Table C11). In practice, the effect of the vaccine is likely to be progressive, which might push down estimates of effectiveness since individuals with early partial protection would be considered to be unvaccinated. However, excluding households with the early-vaccinated index cases did not affect our estimates (Fig. C5 and Table C12). The limited number of households does not make it possible to dissociate early vs. full protection conferred by the vaccine nor to investigate the infectivity of children relative to adults/teenagers. Fourth, testing instructions were different for vaccinated and unvaccinated household contacts, as well as HCWs and non-HCWs. Most vaccinated contacts were HCWs at the Sheba Medical Center who complied with testing instructions to go back to work, leading to high testing rates in vaccinated individuals, with 67% having at least 2 PCR tests and 70% having 1 positive PCR or at least 2 PCR tests in the 10 days following case detection (Table C1). Among unvaccinated contacts, 49% had at least 2 PCR tests and 79% had 1 positive PCR or at least 2 PCR tests in the 10 days following case detection. This higher testing rate is notably due to the high proportion of single positive tests (30%). These differential testing behaviors and positivity rates between vaccinated, unvaccinated, HCW, and non-HCW contacts make it difficult to anticipate the directionality of a potential bias. When restricting our evaluation to households where all negative contacts were tested at least once or twice, estimates remained relatively similar to the baseline values. In the analysis with at least 2 tests for all negative contacts, we observed a slight reduction in the point estimate for vaccine effectiveness against infection that remained difficult to interpret given the very broad credible intervals (17%–91%). Sixth, the measurement of isolation precautions, vaccination status, and symptoms are based on the declaration of participants, and thus may be subject to recall bias. More importantly, the measurement of isolation precautions and vaccination status can be subject to overreporting, as they represent a socially desirable behavior. The timing and evolution of isolation precautions were not measured, and thus not integrated

in our model. Nevertheless, our estimate of isolation effectiveness is consistent with a 10-day period of quarantine in modeling studies (Ashcroft et al., 2021), and our estimates of vaccination effectiveness are also consistent with the literature as mentioned above. Finally, we estimate vaccine effectiveness against infection and transmission in a context where the Alpha variant was dominant. These estimates are very likely to be different for the Delta variant (Eyre et al., 2021) that was first reported in October 2020 and rapidly became dominant worldwide (World Health Organization (WHO), 2021a).

To conclude, vaccination with 2 doses substantially reduces the risk of transmission and the risk of infection in households. Isolation from the index case while sleeping and eating provides a high level of protection to unvaccinated household members, whether they are adults/teenagers or children. Household contacts of COVID-19 patients should ideally isolate, or at least refrain from significant contact, with household cases. This may also be the case for vaccinated household members, although larger studies are required to confirm this finding.

Chapter 4

Impact and mitigation of sampling bias in discrete phylogeography

Bayesian phylogeographic inference is a powerful tool in molecular epidemiological studies that enables reconstructing the origin and subsequent geographic spread of pathogens. Such inference is, however, potentially affected by geographic sampling bias. Here, we investigated the impact of sampling bias on the spatiotemporal reconstruction of viral epidemics using Bayesian discrete phylogeographic models and explored different operational strategies to mitigate this impact. We considered the CTMC model and two structured coalescent approximations (BASTA and MASCOT). For each approach, we compared the estimated and simulated spatiotemporal histories in biased and unbiased conditions based on simulated epidemics of RABV in dogs in Morocco. While the reconstructed spatiotemporal histories were impacted by sampling bias for the three approaches, BASTA and MASCOT reconstructions were also biased when employing unbiased samples. Increasing the number of analyzed genomes led to more robust estimates at low sampling bias for CTMC. Alternative sampling strategies that maximize the spatiotemporal coverage greatly improved the inference at intermediate sampling bias for CTMC, and to a lesser extent, for BASTA and MASCOT. In contrast, allowing for time-varying population sizes in MASCOT resulted in robust inference. We further applied these approaches to two empirical data sets: a RABV data set from the Philippines and a SARS-CoV-2 data set describing its early spread across the world. In conclusion, sampling biases are ubiquitous in phylogeographic analyses but may be accommodated by increasing sample size, balancing spatial and temporal composition in the samples, and informing structured coalescent models with reliable case count data.

1. Introduction

Over the past decade, Bayesian discrete phylogeographic inference has greatly benefited viral epidemiological studies in unraveling the origin and subsequent spread of viral epidemics (Faria et al., 2019; Lemey et al., 2020; Lu et al., 2021), the spatial processes driving viral spread (Müller et al., 2021), and environmental and human-related factors associated with viral spread (Dudas et al., 2017; He et al., 2022; Lemey et al., 2014). BEAST is a popular Bayesian phylodynamics software package commonly used in the analysis of time-stamped viral molecular sequences. It offers different discrete phylogeography approaches: a popular and computationally efficient discrete phylogeographic inference approach that makes use of continuous-time Markov chain (CTMC) modeling (Lemey et al., 2009a), also known as the discrete trait analysis or DTA, and the structured coalescent model under its exact and approximated forms (Vaughan et al., 2014; De Maio et al., 2015; Müller et al., 2018). CTMC models migration between discrete locations in the same way as nucleotide substitutions are modeled. In other words, geographical locations are modeled as a neutral trait that evolves on top of the tree from the root to the tips. As such, CTMC modeling does not explicitly model the branching process that gave rise to the tree. In contrast, the structured coalescent model - which is an extension of the coalescent model to a structured population - is a tree-generating model that explicitly models how lineages coalesce within and migrate between subpopulations from present to past. Two computationally efficient approximations of the structured coalescent model are available in BEAST2: the Bayesian structured coalescent approximation (BASTA) (De Maio et al., 2015) and the marginal approximation of the structured coalescent (MASCOT) (Müller et al., 2018). Currently, they both assume constant prevalence through time for each deme/population, while the CTMC approach does not (Lemey et al., 2009a).

Bayesian discrete phylogeography approaches are complementary to mathematical modeling and epidemiological studies, and particularly informative when epidemiological data are scarce. In such contexts, viral genetic sequences are expected to compensate for the lack of epidemiological data. However, genetic samples may constitute a biased snapshot of the underlying viral spread, especially when isolated through passive surveillance systems. The impact of such sampling bias on discrete phylogeographic inference has been discussed and examined ever since. Indeed, CTMC estimates were suspected to be biased towards the most sampled location (Lemey et al., 2009a) and, later, sampling heterogeneity was shown to inform the posterior, and more specifically the migration parameters, which is not the case for BASTA (De Maio et al., 2015). In BASTA, sampling evenness is not informative as such and the estimated migration rates are more correlated to the true values under simulated biased and unbiased

conditions compared to CTMC (De Maio et al., 2015). As a result, BASTA has been argued to be more robust to sampling bias (De Maio et al., 2015). Nevertheless, the structured coalescent model is known to be sensitive to unsampled locations, known as ghost demes (Beerli, 2004; Ewing and Rodrigo, 2006; De Maio et al., 2015). In parallel, several studies tested alternative strategies to mitigate the potential effects of sampling bias, mostly focusing on CTMC as it was shown to be potentially less robust to sampling bias compared to the structured coalescent model. Downsampling that was tested early on but was limited to large data sets (Yang et al., 2019; Lemey et al., 2014) rapidly became a prerequisite in any SARS-CoV-2 data analysis study due to the large number of available sequences and the high sampling heterogeneity between countries (Hodcroft et al., 2021a). However, (Magee and Scotch, 2018) showed that inference accuracy rapidly plateaus when using up to 25-50% of the sequence data available. Other studies aimed at improving inference accuracy by integrating additional reliable epidemiological data. For example, CTMC was extended to incorporate information on the recent migration events using individual travel records (Lemey et al., 2020; Hong et al., 2021). More recently, a simulation study focused on quantifying the impact of sampling bias on the predicted location of internal nodes, the prediction of migration events that lead to large local spread as well as on the estimation of migration rates in a maximum likelihood framework (Liu et al., 2022). The authors showed that prediction accuracy actually depends on multiple factors: the underlying migration rate, the magnitude of sampling bias and the magnitude of traveler sampling. Importantly, they observed a lower relative accuracy with biased samples and when samples overrepresent travelers. Concerning the structured coalescent model, (Müller et al., 2019) informed the deme population sizes with reliable case count data from the 2014 Ebola epidemic in Sierra Leone using MASCOT. This allows modeling time-varying population sizes instead of assuming constant population sizes over time. Sampling bias is also a concern in continuous phylogeography analyses in which other mitigation approaches were tested. Recently, (Dellicour et al., 2021b) downsampled SARS-CoV-2 genomic records from New York City based on hospitalisations rather than case counts to analyze representative samples irrespective of testing effort and strategy, (Kalkauskas et al., 2021) incorporated sequence-free samples from unsampled areas, and (Guindon and De Maio, 2021) explicitly modeled sampling strategy in the data likelihood.

Whereas numerous studies tested strategies to deal with sampling bias, the impact of sampling bias on discrete phylogeographic reconstructions remains insufficiently characterized. Here, we compare the performance of the different phylogeographic methods using simulated viral epidemics using a stochastic metapopulation model, based on RABV epidemics in dogs in Morocco. We investigated the impact of sampling bias on the spatiotemporal reconstruction of these viral epidemics using CTMC, BASTA, and

MASCOT, with the latter two assuming populations to stay constant over time. Next, we explored different approaches to mitigate sampling bias, maximizing the spatial and/or temporal coverage of the sample, and informing the deme sizes under MASCOT with the true (time-varying) case count data per location. The latter is to test to what degree biases originating from assuming constant population sizes over time can be mitigated by allowing them to vary over time. Finally, we applied the three algorithms to two empirical data sets: a data set of RABV sequences isolated in the Philippines islands between 2004 and 2010, and a global data set of SARS-CoV-2 genomes of the early spread of the pandemic.

2. Material and Methods

2.1. Simulation study

2.1.1. Simulation of viral transmission chains using a metapopulation model

In order to address the impact of spatial sampling bias on discrete phylogeographic inference, we performed a detailed simulation study. Sampling bias concerns all diseases, but it is even more challenging to address in the context of zoonotic diseases for which most of the transmission process is unobserved. We grounded our study in the context of dog rabies in North Africa where transmission processes are relatively well-documented. It was notably shown that rabies transmission relies on human movement over long distances. We simulated rabies epidemics in dog populations according to realistic scenarios using a stochastic, discrete-time and spatially-explicit model implemented in R using the Rccp package (Eddelbuettel and Balamuta, 2018). We divided the Moroccan dog population into three or seven subpopulations corresponding to arbitrary regions (see the section below on the parametrization of the mobility matrix, Fig. D22). We divided each subpopulation into three compartments: susceptible, exposed, and infectious individuals (Fig. 20A). At each discrete time step, we drew newborns and dead individuals in the susceptible compartment from Poisson distributions with respective means the birth rate b and the death rate d. We defined the force of infection $\Lambda_{i,t}$, i.e. the per-capita rate of infection of susceptible individuals in region i on day t, as:

$$\Lambda_{i,t} = \frac{\beta}{H_i}\left(I_{i,t-1} + \sum_{j \neq i} C_S \nu_{j \to i} I_{j,t-1}\right) \tag{4.1}$$

where β is the transmission rate of rabies scaled by H_i, i.e. the human population size in region i, $\nu_{j \to i}$ is the per-capita mobility rate of individuals moving from region j to region i, $I_{j,t-1}$ is the number of

infectious individuals in region i on day $t-1$, and C_S is a scale factor (see below for more information). Exhaustive dog census data were not available and it is well known that human-mediated movement plays a major role in the spread of rabies in North Africa (Talbi et al., 2010; Dellicour et al., 2017), thus we assumed that dog populations were proportional to human populations (Table D2). We scaled the rabies transmission rate by population size to ensure that the force of infection is density-independent as previously documented on rabies (Morters et al., 2013). We used the scale factor C_S to monitor the proportion of inter-region infections. Its value was arbitrarily chosen so that 1% of infection events occurred between regions, and the basic reproduction ratio is approximately equal to 1.05 within and between regions. At each time step, we drew the number of newly exposed individuals in each region from Poisson distributions with a mean specified by the number of susceptible individuals in region i on day $t-1$ ($S_{i,t-1}$) multiplied by the force of infection in region i on day t ($\Lambda_{i,t}$). Once an individual $e_{j,t}$ entered the exposed compartment, it was uniquely identified. The location of its infector was drawn from a multinomial distribution with the following probabilities:

$$P(e_{j,t}\ infected\ by\ I_{i,t-1}) = \frac{\nu_{i\to j} I_{i,t-1}}{\sum_k \nu_{k\to j} I_{k,t-1}} \tag{4.2}$$

Once the location of the infector was drawn, the ID of the infector was randomly sampled from the set of infectors present in the location. All infectious individuals in each region had the same probability of infection. The incubation period of exposed individuals was drawn from a gamma distribution with shape 2 and rate 11.055 (Hampson et al., 2009) and its infectious period was drawn from a discretized gamma distribution adapted from (Hampson et al., 2009) so that it could not exceed 15 days (World Health Organization (WHO), 2018). Finally, the life span was drawn from an exponential distribution with rate d. If natural death occurred before the end of the incubation or infectious periods, the individual was removed prematurely. Otherwise, the individual went through the exposed and infectious compartment before dying from rabies (Table D2).

We initiated all simulations with the introduction of a single index case in Region 3 (Fig. D22). According to (Darkaoui et al., 2017), there are on average 400 confirmed animal cases per year in Morocco which is certainly an underestimation (Broban et al., 2018). We assumed a 20% reporting rate of dog cases in Morocco (Taylor et al., 2017a), and thus retained epidemics with at least 60,000 cases over a 20 to 30-year period (Fig. 20C). We analyzed the results for 50 simulations.

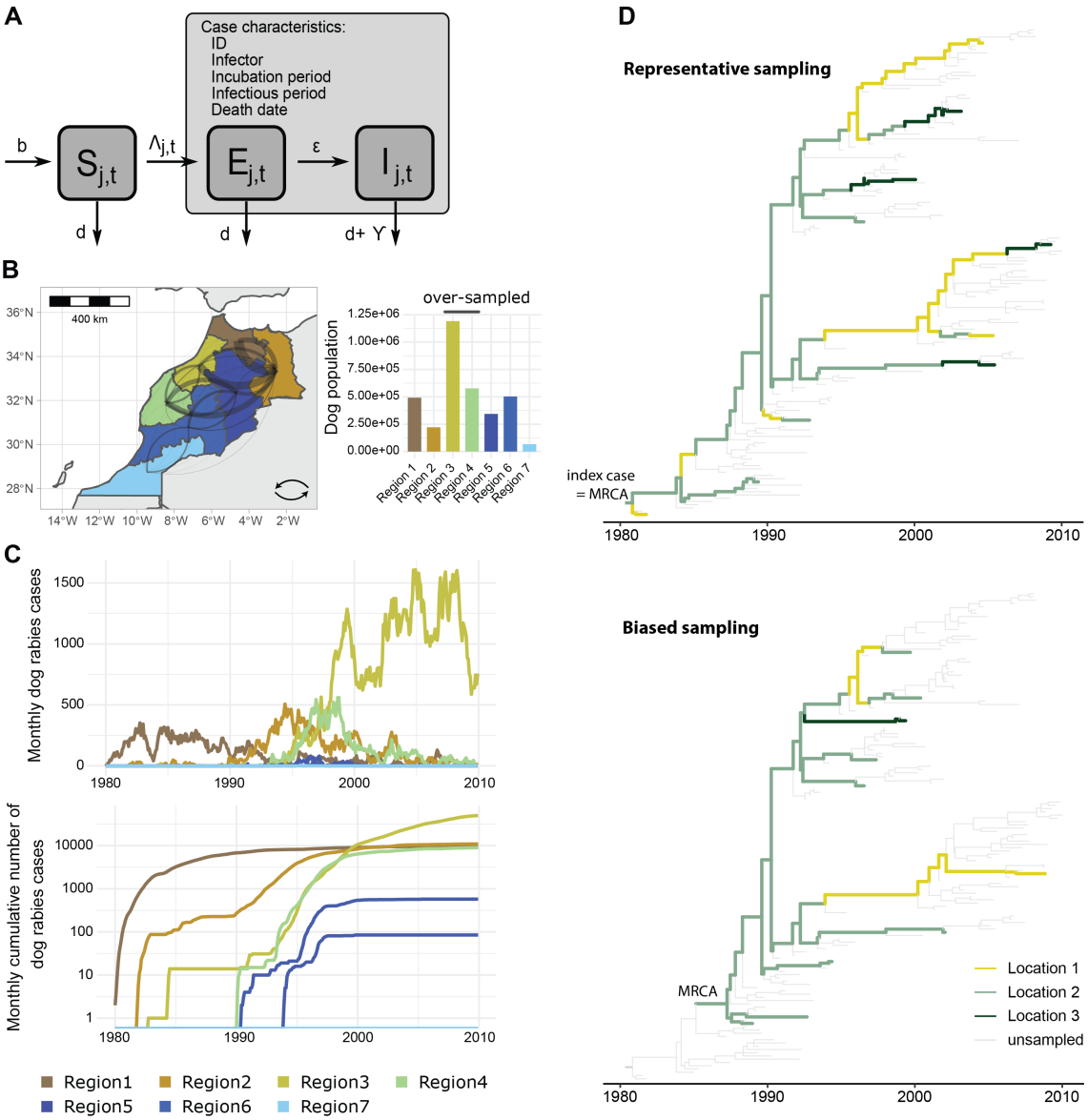


Figure 20: Rabies virus (RABV) epidemic simulation framework. We simulate realistic epidemics by emulating the scenario of RABV spread in dog populations in Morocco. (**A**) Metapopulation model of rabies spread in dogs. In each geographical location j, the dog population is divided into three compartments: susceptible, exposed but yet not infectious, and infectious individuals. Individuals are born at rate b and die from natural causes at rate γ. The rate of infection corresponds to the per-capita force of infection $\lambda_{j,t}$ that aggregates the force of infection from infectors in location j and all the other locations. Individuals become infectious at rate ε. We identify all infected individuals and simulate their infector, incubation period, infectious period and date of death. (**B**) Connectivity between the seven arbitrary Moroccan regions estimated by the radiation model and estimated dog population size per region. Curvature indicates flux direction. (**C**) Example for one simulation of the prevalence (first row) and cumulative number (second row) of rabid cases per month and location. (**D**) Graphical illustration of the potential impact of sampling bias on the reconstruction of the phylogenetic relationships between viral samples over an epidemic, assuming no intra-host evolution.

2.1.2. Parametrization of the between-region mobility matrix

To avoid computational difficulties and over-parameterization of the different discrete phylogeographic models, we aggregated the fifteen official Moroccan regions retrieved from the GADM data set (`http://www.gadm.org`) into three or seven locations (in two simulated scenarios, respectively) that are based

on human demographics and ecological features (Fig. D22). Dog mobility was defined across locations by fitting a radiation model to a raster of human population distribution (WorldPop, n.d.) using the R package movement (Golding et al., 2015). In the radiation model, commuting is determined by the job seeking behavior modeled as an absorption and radiation process (Simini et al., 2012). The average commuting flux $T_{i,j}$ from location i to location j with population m_i and n_j, respectively is:

$$\langle T_{i,j} \rangle = T_i \frac{m_i n_j}{(m_i + s_{i,j})(m_i + n_j + s_{i,j})} \tag{4.3}$$

with $s_{i,j}$ the total population in the circle of radius $r_{i,j}$ centred at i (excluding the source and destination population).

We used a model of human mobility as it has been shown that humans play a major role in dog rabies spread and maintenance in North Africa, especially across long distances (Talbi et al., 2010; Dellicour et al., 2017). We preferred the radiation model over the gravity model for two reasons: the radiation model has been shown to outcompete the gravity model at local and large scales (Simini et al., 2012), and it presents the advantage of having no free parameter(s). In our study, we inferred the average daily number of commuters between raster cells of 20 km with more than 1,000 inhabitants per km^2. The size of the cells corresponds approximately to the municipality level, and the density threshold corresponds to the urban density in Morocco. The number of commuters was then aggregated at the location level.

2.1.3. Evolutionary model of RABV genomes associated with cases

Simulation studies that analyze the accuracy of phylogeographical techniques often use the inference model as the simulation model (De Maio et al., 2015; Müller et al., 2017; Kalkauskas et al., 2021). Here, we took an epidemiological perspective by simulating rabies epidemics using a metapopulation model and by inferring the spatiotemporal history of rabies from RABV sequences and not from phylogenetic trees. After simulating rabies epidemics as described above, RABV genomes associated with each case were simulated according to the HKY model (Hasegawa et al., 1985). We simulated in R sequence evolution forwards-in-time along the transmission chains which were used in the same way as a phylogeny. We opted for a simple evolutionary process in which selection, gene partition, and site heterogeneity were not considered. Parameter values are listed in Table D2. The genome of the index case is a real canine rabies genome of 13 kb length isolated in Morocco in 2013 (GenBank Accession Number KF155001.1) (Marston et al., 2013).

2.1.4. Sampling schemes of viral sequences

The aim of the study is to determine the impact of sampling bias on phylogeographic inference and how alternative sampling schemes may mitigate the effects of such sampling bias. To address the former issue, we sampled either uniformly (uniform) or with a sampling bias favoring viral sequences from highly populated locations (Regions 3 and 4). In the latter scenario, sequences from Regions 1, 2, 5, 6, and 7 had a weight equal to one, whereas Regions 3 and 4 had a weight equal to 2.5, 5, 10, 20 or and 50 (biased-2.5, biased-5, biased-10, biased-20, and biased-50, respectively, Fig. 20D). To mitigate the potential effects of sampling bias, we tested a different setup reproducing a surveillance system. In this setup, a biobank of 5,000 sequences were drawn from each epidemic with a weight of one for Regions 1, 2, 5, 6, and 7, and a weight of 10 or 20 for Region 3 and 4. Subsets of sequences were sampled from the biobank either uniformly (uniform surv.), by maximizing the spatial coverage (max per region), or by maximizing the spatiotemporal coverage (max per region and per year). For all sampling schemes, a large sample of 500 sequences and a nested sample of 150 sequences were drawn over the entire epidemic except for the first year, as we assumed that the spread of the virus would remain undetected at the start of the epidemic as observed in other settings (Townsend et al., 2013b).

2.2. Discrete phylogeographic analysis in BEAST

2.2.1. Generation of BEAST XML files and phylogeography inference set up

Tailored XML template files for the BASTA and MASCOT structured coalescent models, as well as for the discrete trait analysis (CTMC) model, were edited using the lxml Python package to add sequence alignments along with their metadata. Bayesian phylogeographic analyses were performed using BEAST v1.10.5 (Suchard et al., 2018) for the CTMC model (Lemey et al., 2009a), and BEAST v2.6.4 (Bouckaert et al., 2019) for MASCOT v2.2.1 (Müller et al., 2018) and BASTA v3.0.1 (De Maio et al., 2015), making use of the BEAGLE library v3.1.1 (Ayres et al., 2012). We assumed an HKY substitution model with a strict molecular clock. Population dynamics in the CTMC model followed a constant population size prior. We chose this prior since the model of population dynamics is not expected to impact migration history inference and the constant population size model is often chosen for the analysis of endemic diseases. For the BASTA and MASCOT structured coalescent models, all demes were set to have equal size due to numerical issues leading to a computation time of over 70 hours per million iterations (data not shown). For both models, asymmetric migration matrices were inferred and BSSVS was used to avoid over-parametrization. The detailed list of prior distributions is available in Table D3 for each inference framework.

If deme sizes are set to be equal in the structured coalescent model but the actual population dynamics vary through time, the model tends to explain population dynamics by migration dynamics. In our simulations, the incidence changed dramatically over time and location (Fig. 20C), thus the inference by the structured coalescent model is expected to improve when accounting for time-varying population dynamics. To test this hypothesis, we used monthly incidence data from our simulations as a predictor of the deme sizes by using a GLM in MASCOT (Müller et al., 2019). We tested this alternative parametrization (MASCOT-GLM) in the following conditions: uniform, biased-2.5, biased-5, biased-10, biased-20, biased-50, uniform surv. 10, and uniform surv. 20.

These different BEAST analyses were run for at least 20 and 40 million steps, and sampled every 2,000 and 4,000 steps for small and large alignments, respectively. In total, 8,800 XML files were run for this study, for a total of an estimated 1,500 hours of computation on multi-core CPUs across different computing infrastructures (Table D4).

2.2.2. Analysis of phylogeographic inference output

For each BEAST analysis, adequate mixing was assessed based on the ESS values of the continuous parameters. We calculated ESS values using a Python function adapted from Tracer v1.7.2 (Rambaut et al., 2018). When at least one continuous parameter had an ESS value below 200, chains were resumed to reach at most 120 million iterations. Analyses that exhibited ESS values lower than 200 at this point were discarded (Tables D5 and D6). Due to the higher computational burden of BASTA, the ESS cut-off was reduced to 100. We discarded a 10% burn-in in the selected chains. The combined posterior tree distributions were summarized into MCC trees using TreeAnnotator for BASTA and the CTMC, and the Python library dendropy (Sukumaran and Holder, 2010) for MASCOT and MASCOT-GLM. Summary statistics, ESS values, Bayes factors (BFs) on migration rates (Lemey et al., 2009a), root state probabilities, dates of lineage introduction, and lineage migration counts were calculated in Python before plotting the results in R using the ggplot2 package (Wickham, 2016).

2.2.3. Performance analysis

To assess the accuracy of the phylogenetic reconstruction, the time to the most recent common ancestor (TMRCA) of every pair of sampled tips was computed on both the MCC tree and the simulated transmission chain, and these outcomes were subsequently compared using the Pearson correlation coefficient (Fig. 20A). In addition, we evaluated the impact of sampling bias and alternative sampling strategies on the estimation of the total migration counts, lineage migration counts, and dates of first lineage introduction into each sampled location using five metrics:

- Kendall's tau correlation: a rank-correlation measure that is less sensitive to outliers compared to Pearson's correlation coefficient

$$\tau = \frac{(no.\ concordant\ true/simulated\ value\ pairs)(no.\ discordant\ true/simulated\ value\ pairs)}{\binom{n}{2}} \tag{4.4}$$

- Calibration

$$calibration_{95\%} = \frac{1}{n}\sum_{i=1}^{n} 1_{\{\theta_i \in HPD_{95\%}(D_i)\}} \tag{4.5}$$

- Mean relative bias

$$MRB = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\theta_i}(\hat{\theta}_i - \theta_i) \tag{4.6}$$

- Mean relative 95% highest posterior density (HPD) width

$$width_{95\%} = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\theta_i}(HPD_{97.5\%}(D_i) - HPD_{2.5\%}(D_i)) \tag{4.7}$$

- Weighted interval score (WIS): a generalization of the absolute error accounting for estimation uncertainty. We present the formula of the WIS and refer to the original article for further details, notably on the interval score (Bracher et al., 2021).

$$WIS_{\alpha_{\{0:K\}}}(F,y) = \frac{1}{K+1/2} \times (w_0 \times |y-m| + \sum_{k=1}^{K} \{w_k \times IS_{\alpha_k}(F,y)\}) \tag{4.8}$$

We denote θ_i the true value of the parameter, D_i the parameter posterior distribution, $\hat{\theta}_i$ the median estimate, $HPD_{95\%}$ the 95% HPD, K the number of prediction intervals included in the calculation of the WIS, y the observed outcome by forecast F, m the predictive median on the $(1-\alpha_k)\times 100$ prediction interval, IS_{α_k} the interval score on the $(1-\alpha_k)\times 100$ prediction interval and w_k its weight. The mean relative bias and the mean relative 95% HPD width are defined when the true value is not zero. However, the total migration counts and the lineage migration counts for some pairs of locations can be null in our simulations whereas the algorithms infer a non-null median. These cases were not considered in the calculation of the mean relative bias and the mean relative 95% HPD width. We reported their numbers in the caption of the corresponding figures.

2.3. Data analysis

2.3.1. RABV expansion in the Philippines

We extended our comparative analysis of the CTMC, BASTA, and MASCOT by analyzing a set of RABV genetic sequences using the three approaches. In total, 233 sequences corresponding to the RABV glycoprotein gene were sampled in the Philippines from 2004 and 2010 (Saito et al., 2013). In the original discrete phylogeographic analysis, the authors studied viral spread across 11 out of the 17 Philippines regions and showed that the genetic diversity was highly spatially-structured, notably at the island level (Tohma et al., 2014). Here, we evaluated spread across the six sampled islands (Luzon, Catanduanes, Oriental Mindoro, Cebu, Negros Oriental, and Mindanao) to compare the reconstructions on a highly structured data set and limit the number of demes that considerably slow down BASTA and MASCOT. We assumed an HKY nucleotide substitution model with an among-site rate heterogeneity modeled by a discretized gamma distribution (Yang, 1994), and an uncorrelated relaxed molecular clock with an underlying lognormal distribution (Drummond et al., 2006). For the CTMC, we assumed a constant size coalescent model for the viral demographics as in the original analysis. For MASCOT and BASTA, current implementations assume a constant population size model for the viral demographics within demes. A detailed description of the priors is reported in Table D7. For each algorithm, we combined three post-burnin independent chains of 50 million iterations each.

2.3.2. The early dynamics of SARS-CoV-2 worldwide spread

Tracking viral disease spread in animal populations faces many challenges, and to our knowledge, no reliable incidence data are available for zoonoses such as rabies. In this context, MASCOT-GLM cannot readily be used. We analyzed the early worldwide spread of SARS-CoV-2 to compare the inferences of the CTMC, BASTA, MASCOT, and MASCOT-GLM. (Lemey et al., 2020) analyzed this data set to characterize SARS-CoV-2 spread across 44 location states by incorporating individual travel histories of sampled individuals to help correct for sampling bias and unsampled locations. By using the carefully obtained results of (Lemey et al., 2020) as a reference, we can evaluate how the four algorithms are impacted by sampling bias.

The data set comprises 282 SARS-CoV-2 genomic sequences sampled in the five continents from December 24th, 2019 to March 4th, 2020. We assumed an HKY nucleotide substitution model with a proportion of invariant sites, an among-site rate heterogeneity modeled by a discretized gamma distribution, and a strict molecular clock. For the CTMC, we assumed an exponential growth model for the viral demographics. MASCOT and BASTA assume a constant size model for the viral demes demographics.

Contrary to the original study, we analyzed migration between six discrete locations: Africa, Americas, Asia, China, Europe, and Oceania. For MASCOT-GLM, we used the daily number of confirmed cases at the continent level from Our World In Data (Ritchie et al., 2020), or from the (World Health Organization (WHO), n.d.) as a predictor of the deme sizes. The former is referred to as MASCOT-WID, and the latter as MASCOT-WHO. We smoothed the number of new confirmed cases using a seven-day moving average. The detailed description of the priors is reported in Table D8. We combined three post-burnin independent chains of 50 or 100 million iterations for each inference.

3. Results

3.1. Simulation framework

We simulate RABV epidemics across three or seven locations using a stochastic metapopulation model (Fig. 20A) whose connectivity matrix is parameterized using human population mobility that we estimated by fitting the radiation model of (Simini et al., 2012) to human population density data from (WorldPop, n.d.) (Fig. 20B). As each location is associated to a specific deme/population, we refer to the two simulation frameworks as the three or seven demes framework for the remainder of the text. We simulate 50 epidemics that start with the introduction of a single case and lead to at least 60,000 cases over a 30-year period (Fig. 20C). On top of the transmission chains, we simulate viral genomes for each case and then subsample starting one year after the introduction of the index case either 150 or 500 sequences in a biased or unbiased fashion (Fig. 20D). We then perform Bayesian discrete phylogeographic analysis on the geolocated and time-stamped sequence alignments before comparing the true and reconstructed evolutionary and migration histories for each discrete phylogeographic approach. Importantly, the vast majority of samples in the three demes framework contain at least one sequence of each deme which is not the case for the seven demes framework for which sampling bias often leads to unsampled locations, also called “ghost” demes.

Robust estimation of the phylogeny and genetic parameters with respect to sampling bias While the focus of our simulation study is on reconstructing the spatial spread, we first assess the potential impact of sampling bias on estimating the phylogeny itself, as well as the evolutionary parameters (Fig. 21A). The phylogeny of the simulated pathogen is not impacted by sampling bias when using CTMC, BASTA, and MASCOT (Fig. 21A and S1). In addition, the average evolutionary rate (Fig. 21C), the stationary nucleotide frequencies (Figs. D2 to D5), and the ratio of transition-transversion rates (Fig. D6) are all well estimated at any level of sampling bias.

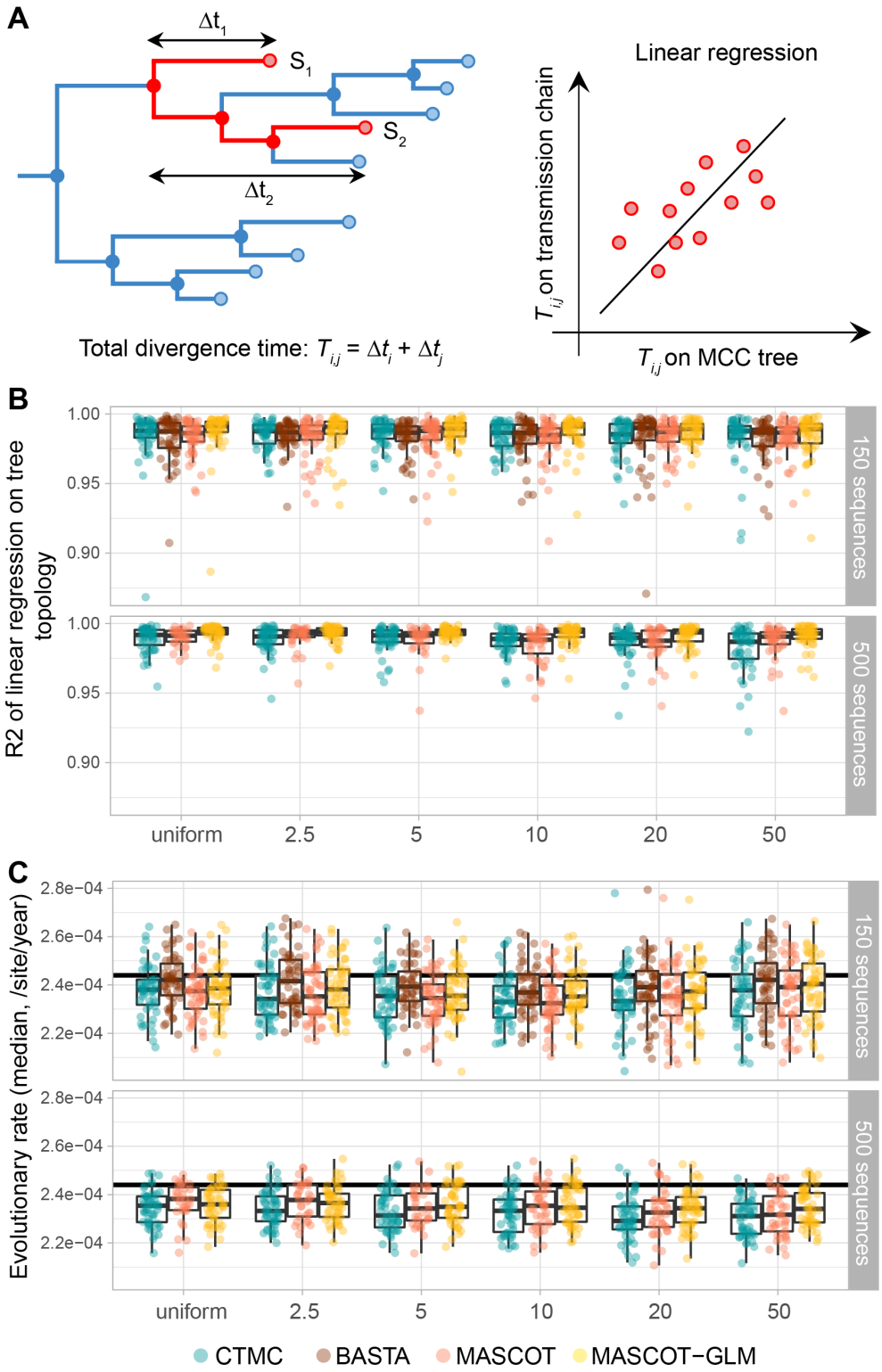


Figure 21: Estimation of genetic and phylogenetic parameters under spatially-biased sampling conditions. (**A**) Comparison of the simulated transmission chain and the estimated maximum clade credibility (MCC) tree topologies. For the estimated and simulated topologies, we computed the total divergence time between every pair of sampled tips. We compared the two using linear regression. (**B**) Pearson's determination coefficient of the pairwise divergence time between the simulated transmission chain and the MCC tree. (**C**) Estimation of the evolutionary rate. Each dot corresponds to the median estimate of the evolutionary rate in one simulation (n = 50 per sampling protocol and sample size).

3.2. Spatiotemporal history reconstruction in (un)biased conditions

As the inferred spatiotemporal histories of lineages cannot be compared in a unique simple way between the different approaches, we use four types of summary statistics: (*i*) the total migration counts - corresponding to Markov jumps in the case of CTMC and their equivalent for BASTA and MASCOT - that account for multiple migration events along the tree branches (Fig. 22), (*ii*) the lineage migration counts (Fig. D7), (*iii*) the lineage introduction dates into the sampled locations (Fig. 23), and (*iv*) the root location (Fig. 24). We evaluate the performance of the phylogeographic models using five metrics: the correlation between true and estimated values, the proportion of estimated parameters for which the true value is in the 95% highest posterior interval (HPD) that we refer to as the calibration, the mean relative bias, the mean relative 95% HPD width, and the weighted interval score (WIS). The WIS is a generalization of the absolute error accounting for estimation uncertainty (Bracher et al., 2021). The smaller the WIS, the better the inference. It is widely used to evaluate epidemic forecasts and favors estimates that are slightly biased but with a narrow confidence interval compared to estimates without bias but very large uncertainty (Bracher et al., 2021).

First, we assess the reconstruction of the spatial process in the absence of sampling bias. In the unbiased/representative (uniform) scenario, CTMC correctly estimates the four types of parameters. Indeed, the correlation between the true and estimated parameter values is high, and the WIS is close to zero. BASTA and MASCOT show no correlation for the total migration counts on uniform samples and higher WIS compared to CTMC (Fig. 22A and E) indicating biased median estimates and higher uncertainty around the point estimate. The correlation is over 0.5 when we consider the lineage migration counts under the three and seven demes frameworks. This suggests that BASTA and MASCOT only partly recover the global migration process in the absence of sampling bias (Figs. D7 and D15). Overall, CTMC outperforms BASTA and MASCOT when the sampling is representative of the true underlying transmission process, as BASTA and MASCOT only recover the big picture of the migration process.

Secondly, we evaluate how phylogeographic algorithms perform under increasing levels of bias. While CTMC satisfyingly estimates the total migration counts in the absence of sampling bias, the correlation and the calibration drop rapidly with bias and the mean relative 95% HPD width tends to decrease suggesting that bias strongly impacts CTMC estimates (Fig. 22A-C). Nevertheless, the WIS and the mean relative bias remain smaller than those of BASTA and MASCOT, even at high levels of bias. Consequently, CTMC leads to median estimates that are closer to the true values but with 95% HPDs that are too narrow. It leads to a biased picture of the geographical process with some transition events that are

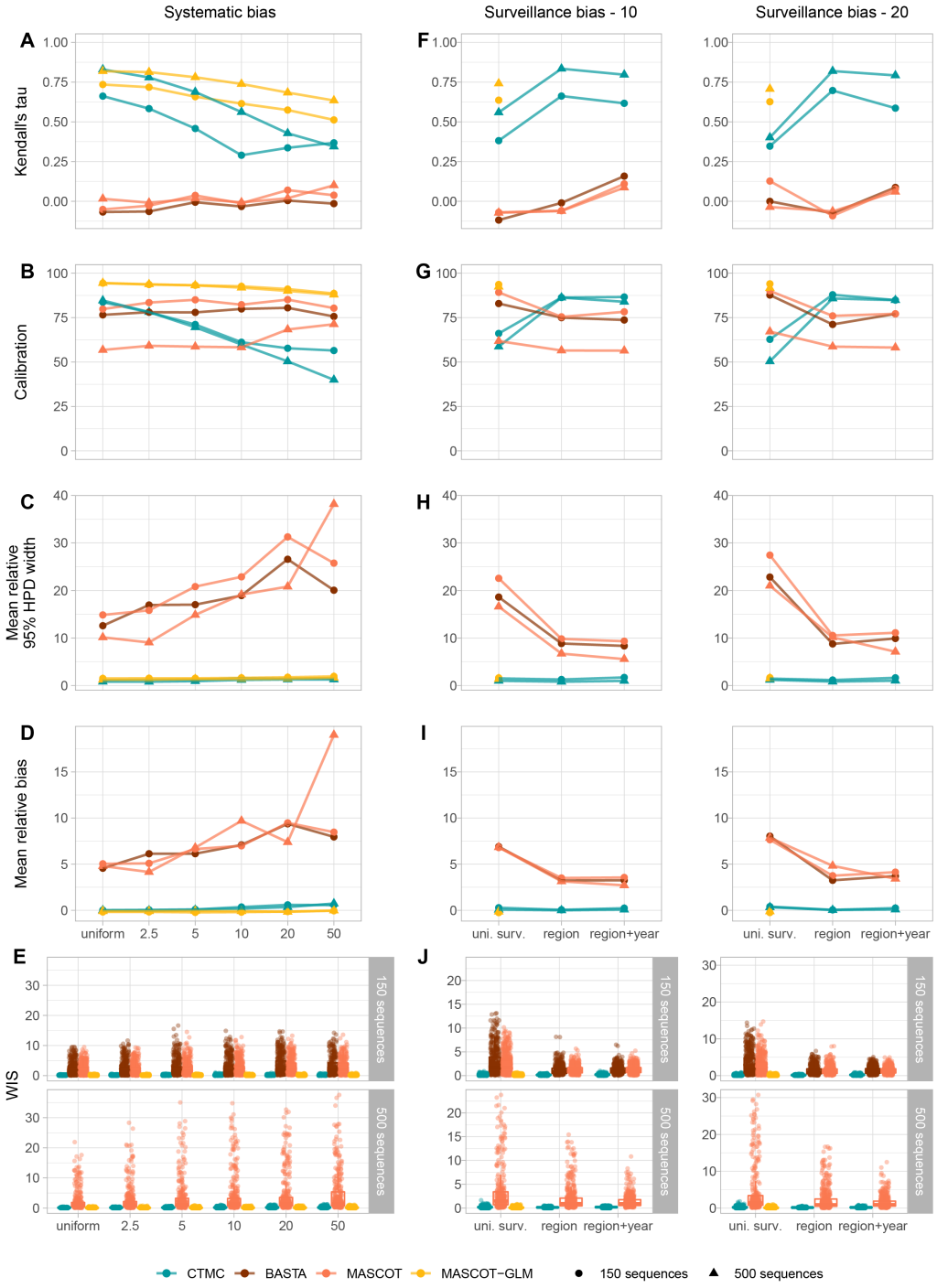


Figure 22: Impact and mitigation of spatial bias on the estimation of the total migration counts. (**A-E**) Impact of the increasing levels of spatial bias on the correlation, the calibration, the mean relative 95% highest posterior density (HPD) width, the mean relative bias, and the WIS between the simulated and the estimated total migration counts. (**F-J**) Mitigation of the impact of spatial bias on the correlation, the calibration, the mean relative 95% HPD width, the mean relative bias, and the WIS between the simulated and estimated total migration counts by using alternative sampling strategies. In the left and right columns, samples are drawn from biobanks with an underlying bias of 10 and 20, respectively. Overall, the algorithms correctly estimate the total migration counts when the correlation and the calibration are high (close to 1 and 100, respectively) and when the mean relative 95% HPD width, the mean relative bias, and the WIS are close to zero. Finally, the mean relative bias and the mean relative 95% HPD width are not defined when the true value is null. We removed 612 out of 3,600 and 380 out of 3,600 simulated migration events in the small and large samples, respectively, due to null true values.

drastically underestimated (Fig. D7). BASTA and MASCOT less accurately estimate the total migration counts with no correlation between simulated and estimated values. They are also less confident with an average 95% HPD width that is ten to thirty times higher compared to CTMC. This uncertainty is exacerbated in large samples analyzed with MASCOT in the seven demes framework, for which almost 30% (87 out of 300) of the chains have low ESS values often due to bimodal structured coalescent posterior density. Additionally, BASTA and MASCOT partly recover the global migration process (lineage migration counts) even at high levels of bias since correlation and calibration are not impacted by bias (Fig. D8). When we consider transmission dynamics between three demes, BASTA and MASCOT yield higher correlation levels than in the seven demes scenario (Figs. D15 and D16). Overall, the WIS indicate better performance of CTMC over BASTA and MASCOT.

When it comes to the estimation of lineage introduction dates, BASTA seems to outcompete CTMC and MASCOT under the three demes framework (Fig. D17) but not under the seven demes framework (Fig. 23A-E). In the three demes framework, the uncertainty around the median estimate remains high for BASTA and MASCOT, and the correlation and the calibration are barely affected by bias for BASTA, contrary to CTMC and MASCOT. In the seven demes framework, correlation is low for both BASTA and MASCOT but not affected by bias. CTMC performs poorly with a sharp decrease in both correlation and calibration in the three demes framework, and a slighter decrease in the seven demes framework. It also tends to estimate more ancient lineage introduction dates compared to BASTA and MASCOT in both frameworks. Of note, samples of 500 sequences displayed a higher correlation than samples of 150 sequences at low and intermediate levels of bias for CTMC (conditions 2.5, 5 and 10 in Fig. 23A and S8A).

Finally, we analyze the potential impact of sampling bias on root location estimation (Fig. 24A and S18A). Of note, the location probability of the true root location is very heterogeneous among the 50 simulated epidemics when there is no or little sampling bias, notably for the two approximations of the structured coalescent model. Root location prediction by CTMC is affected by sampling bias, notably in the three demes framework (Fig. D18), which is in agreement with previous findings (De Maio et al., 2015). As for the other parameters, BASTA and MASCOT perform less well compared to CTMC, at any level of bias. On the other hand, sampling bias moderately worsens their estimates. They also perform relatively better in the three demes framework.

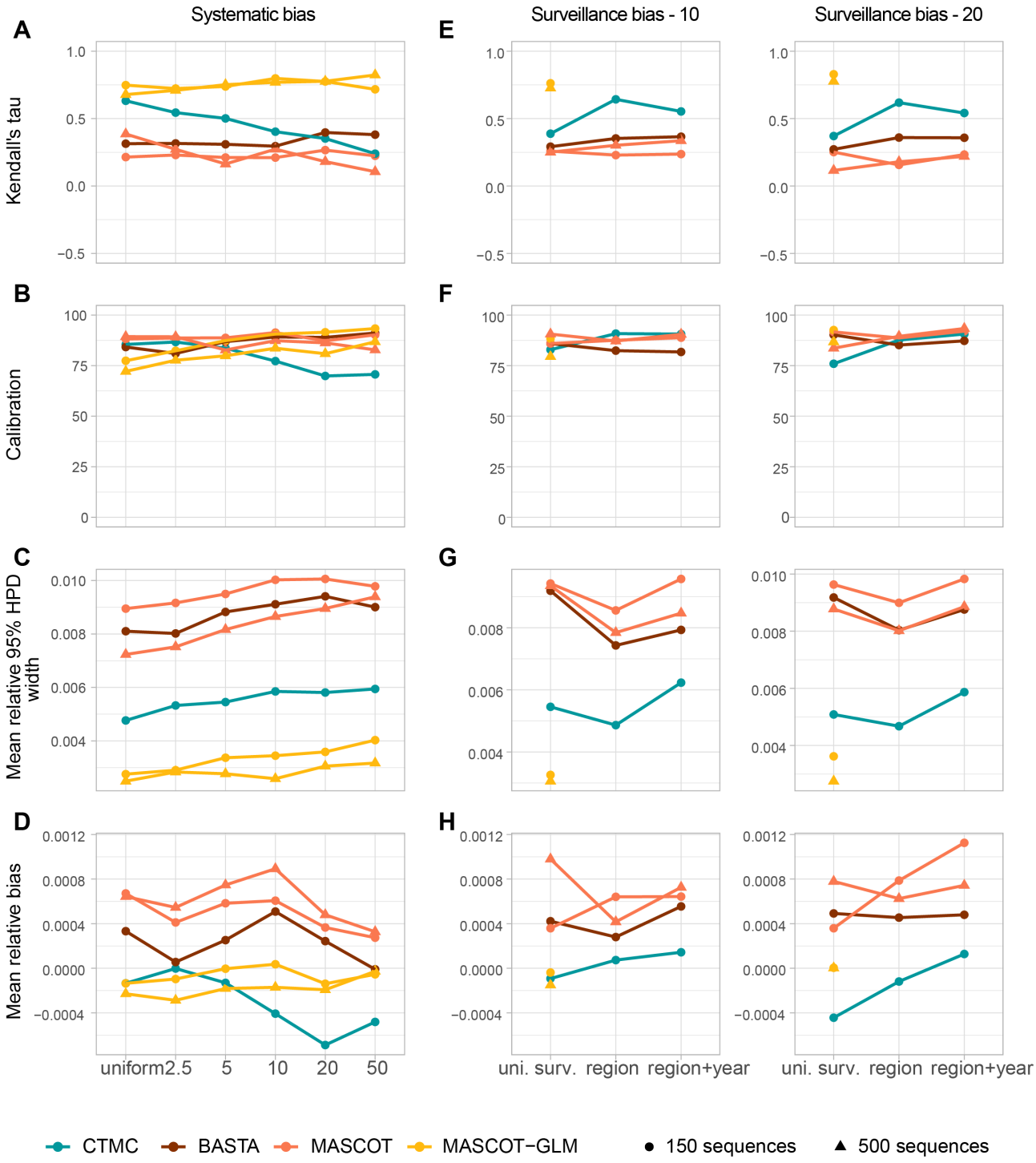


Figure 23: Impact and mitigation of spatial bias on the estimation of the lineage introduction dates. (**A-D**) Impact of the increasing levels of spatial bias on the correlation, the calibration, the mean relative 95% highest posterior density (HPD) width, and the mean relative bias between the simulated and the estimated introduction dates. Uniform samples are representative of the simulated spatiotemporal dynamics of the virus. Samples 2.5, 5, 10, 20 and 50 samples biased towards Regions 3 and 4. Samples 2.5 and 5 correspond to low levels of bias, samples 10 and 20 to intermediate levels of bias and sample 50 to high levels of bias. E-H: Mitigation of the impact of spatial bias on the correlation, the calibration, the mean relative 95% HPD width, and the mean relative bias between the simulated and estimated introduction dates by using alternative sampling strategies. In the left and right columns, samples are drawn from biobanks with an underlying bias of 10 and 20, respectively.

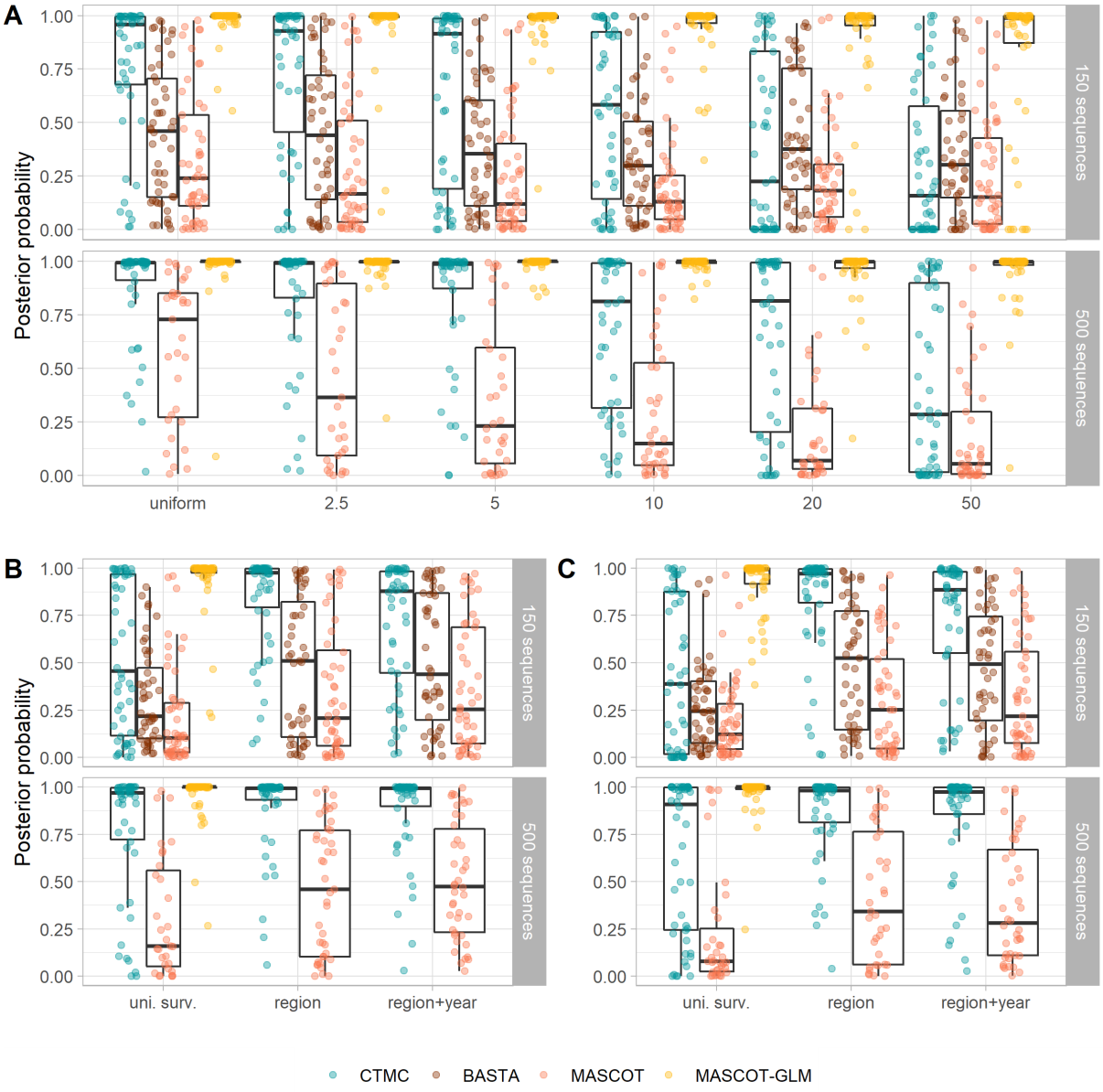


Figure 24: Impact and mitigation of spatial bias on the estimation of the root location. (**A**) Posterior probability of estimating the true root state for increasing levels of biais. (**B-C**): Mitigation of the effects of spatial biais using alternative sampling strategies under a surveillance bias of 10 and 20, respectively. Each dot corresponds to the median root state posterior probability in one simulation (n = 50 per sampling protocol and sample size).

3.3. Sample balancing mitigates the impact of sampling bias

We test alternative sampling strategies in order to mitigate the impact of sampling bias. Large and biased samples of 5,000 sequences that we refer to as biobanks were generated, then discrete phylogeographic analyses were carried out on subsamples of 150 or 500 sequences, which aimed at reproducing real life situations. For example, researchers may have access to numerous viral specimens from biobanks but cannot analyze all of them due to computational limitations, potential underlying biased sampling that may lead to spurious results, or financial limitations.

Similar to the analyses on systematically biased samples, the estimation of the total migration counts (Fig. 22F-J), lineage migration counts (Fig. D6.F-J), lineage introduction dates (Fig. 23E-H), and root location posterior probabilities (Fig. 24B-C) is strongly impacted in biased subsamples (uniform surv.) for the three approaches. By maximizing the spatial (region) or the spatiotemporal coverage (region+year), the correlation of lineage migration counts increased substantially for the CTMC even when the underlying sampling bias was high (weight=20, i.e. sequences from oversampled regions are 20 times more likely to be samples), and to a lesser extent for BASTA and MASCOT. Calibration remained high for BASTA and MASCOT, as shown earlier (Fig. 22G and Fig. 23F) while it considerably improved for CTMC. Estimates of the lineage migration counts by BASTA and MASCOT are improved in the region and region+year conditions compared to the uniform surv. condition, illustrated by a decreased mean relative 95% HPD width and decreased WIS. Still, performance remained lower than for CTMC. In the three demes framework, we obtain even stronger improvements in terms of correlation and decreased mean relative bias for BASTA and MASCOT (Fig. D15). Overall, subsampling strategies that maximize the spatial or spatiotemporal coverage considerably improved the inference of the geographical spread by the CTMC, and improved inference under BASTA and MASCOT to a lesser extent.

3.4. True incidence data as a predictor of the time-varying deme sizes mitigate sampling bias in MASCOT

Due to the lack of statistical power (data not shown), we have forced all deme sizes to be equal in BASTA and MASCOT and to be constant over time, the latter being currently the default assumption of both structured coalescent models. This hypothesis is potentially impactful given that deme sizes are directly related to the migration history in the structured coalescent model (De Maio et al., 2015; Müller et al., 2018). To relax this assumption and allow for time-varying effective population sizes, we next use the monthly incidence data from our simulations as a predictor of the deme sizes over time in the generalized linear model (GLM) extension of MASCOT and denote the resulting model as MASCOT-

GLM. This approach is only available for MASCOT, we can therefore not perform the same analysis for BASTA.

By accommodating for the time variations of deme sizes, the correlation, mean relative 95% HPD width, mean relative bias, and WIS are markedly improved with MASCOT-GLM compared to BASTA and MASCOT for the total migration counts (Fig. 22A-E), lineage migration counts (Fig. D7A-E), and lineage introduction dates (Fig. 23A-E) under biased and unbiased spatial sampling. In the absence of spatial sampling bias (uniform), the mean relative bias for the total migration counts decreases from 5% for BASTA and MASCOT to -0.2% for MASCOT-GLM (Fig. 22D), and the correlation between simulated and inferred total migration counts increases approximately from 0.2 to 0.75. Inference performance is improved for all migration parameters even under highly biased sampling conditions. For example, the mean relative bias of total migration counts remains close to zero for MASCOT-GLM, while it increases up to 38% for BASTA. In addition to the strong correlation and the low mean relative bias between simulated and estimated values, the uncertainty around the true value is low compared to BASTA and MASCOT. We obtain similar results in the biased subsamples of the surveillance sampling protocols (uniform surv.).

3.5. Analysis of the spread of RABV in the Philippines

As a case study to compare the performance of the three algorithms, we analyze the spread of RABV in dog populations between six Philippine islands (Luzon, Catanduanes, Oriental Mindoro, Cebu, Negros Oriental, and Mindanao) using 233 sequences of the RABV glycoprotein gene isolated between 2004 and 2010 (Saito et al., 2013; Tohma et al., 2014). Discrete phylogeography is particularly adapted here to model transmission in animal populations across an archipelago. In this data set, sampling is highly heterogeneous across the different islands: Luzon represents up to 65% of the total data set while Oriental Mindoro is represented only by a single sequence (Fig. D19). This heterogeneity is very unlikely to be representative of the underlying transmission but rather due to case underreporting outside Luzon.

Previous studies on RABV in the Philippines suggested that although the circulating lineages likely circulate independently in the main islands (Saito et al., 2013; Tohma et al., 2014), inter-island transmission events can lead to sustained circulation in previously rabies-free islands (Tohma et al., 2016). Importantly, the patterns of spatial spread were evaluated using the CTMC at a finer spatial scale (Tohma et al., 2014). Here, the CTMC model also predicts a highly spatially-structured phylogeny with few migration events between islands. It reconstructs four island-specific clades located in Catanduanes, Luzon, Mindanao, and Negros Oriental with high node and location posterior support (Fig. 25A). BASTA

and MASCOT also predict the Catanduanes, Mindanao, and Negros Oriental clades with high node and location posterior support (Fig. 25B-C). However, the migration history of the Luzon clade is more uncertain with potential intense migrations between Luzon and Oriental Mindoro islands, the most and least sampled islands, respectively. As shown in the simulations, CTMC might be overconfident compared to BASTA and MASCOT but the uncertainty of the two approximations of the structured coalescent model might be related to the pseudo-ghost demes, i.e. locations for which very few sequences are available. As we don't have information regarding the number of cases over time, we could not apply MASCOT-GLM to this data set.

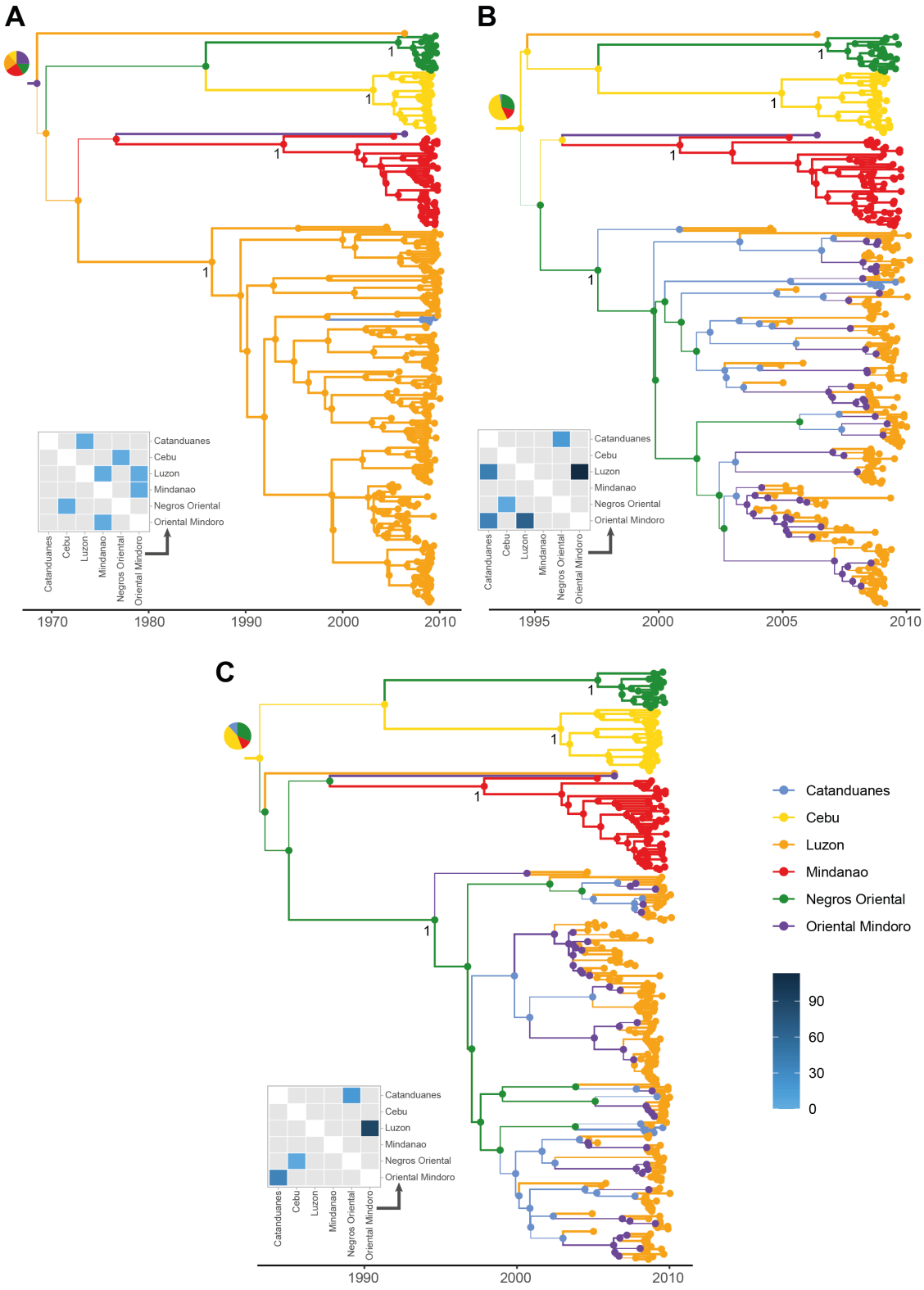


Figure 25: Maximum clade credibility (MCC) trees and median total migration counts estimated on the rabies data set. (**A-C**) MCC trees and median number of total migration counts estimated on the rabies data set by CTMC, BASTA, and MASCOT, respectively. Branch width is proportional to the maximal ancestral location probability predicted by the algorithms, and branches are colored by the maximal ancestral location. Posterior support of the Negros Oriental, Catanduanes, Mindanao, and Luzon island lineages are reported. Pie charts displayed at root nodes represent the posterior probability distribution of the root location. Median estimates of the total migration counts are reported as heatmaps. Gray tiles correspond to transitions associated with a migration rate that is not statistically supported, i.e., with a Bayes factor lower than 3.

3.6. Analysis of the early spread of SARS-CoV-2 across the world

In the context of zoonotic diseases, surveillance systems mostly rely on disease monitoring in human populations. Thus, there are typically no reliable estimates of the number of new cases in wild animal populations and, depending on the country and the species considered, domestic animal populations. However, for pathogens infecting the human population, such estimates are typically widely available, as is the case for SARS-CoV-2 and as has also been shown used previously when studying Dengue virus, HIV and West Nile virus (Gill et al., 2016; Dellicour et al., 2020).

To compare the phylogeographic reconstructions of the four algorithms tested above, we analyze a data set of SARS-CoV-2 genomic sequences from the early stage of the pandemic (Lemey et al., 2020). In the original study, the initial wave of SARS-CoV-2 infections was investigated using a novel travel history-aware extension of the CTMC model, which we here refer to as CTMC-TRAVEL.

We again use the CTMC model, BASTA and MASCOT as well as MASCOT-GLM to analyze the data set. MASCOT-GLM is informed using the seven-day moving average of case count data either from Our World In Data (Ritchie et al., 2020) or from the (World Health Organization (WHO), n.d.). MASCOT-GLM is then referred to as MASCOT-WID and MASCOT-WHO, respectively (Fig. D20). Due to the low number of mutations accumulated in the SARS-CoV-2 genome at the start of the pandemic, the posterior support of internal nodes for each algorithm is low and the tree topology very uncertain (Morel et al., 2021). Besides, we do not intend to reconstruct the origins of SARS-CoV-2 which in any case cannot be addressed solely with phylogeographic analyses (Pipes et al., 2021). That is why our comparison focuses on the posterior support of four clades originally identified by (Lemey et al., 2020): clades A, A.1, B.1, and B.4. Whereas clades A.1, B.1, and B.4 are predicted with high posterior support by all algorithms, clade A is predicted with a satisfying posterior support only by CTMC (Fig. D21). In general, CTMC and MASCOT-WHO predictions are closer to the original predictions than the other algorithms, in terms of tree topology (Fig. D21.A) and of total migration counts (Fig. 26). As previously shown, BASTA and MASCOT lead to more uncertain ancestral migration histories with the extreme case of BASTA for which the posterior evolutionary rate and the structured coalescent density are bimodal. We report two maximum clade credibility (MCC) trees for BASTA, corresponding to the two modes of the evolutionary rate and structured coalescent density (Fig. D21.B-C and Fig. D22). The first mode of BASTA infers a tree topology and a migration history that are similar to CTMC and CTMC-TRAVEL. For example, the predicted location of the MRCA of the B.4 lineage is China for CTMC-TRAVEL, CTMC, and the 1st mode of BASTA, whereas it is located in Oceania by MASCOT and the 2nd mode of BASTA (Table D1).

For the latter two reconstructions, most of the ancestral branches were not inferred to occur in China and, similarly to the RABV data set, these approaches predict the least sampled locations (Africa and Oceania) to play a major role in the transmission process.

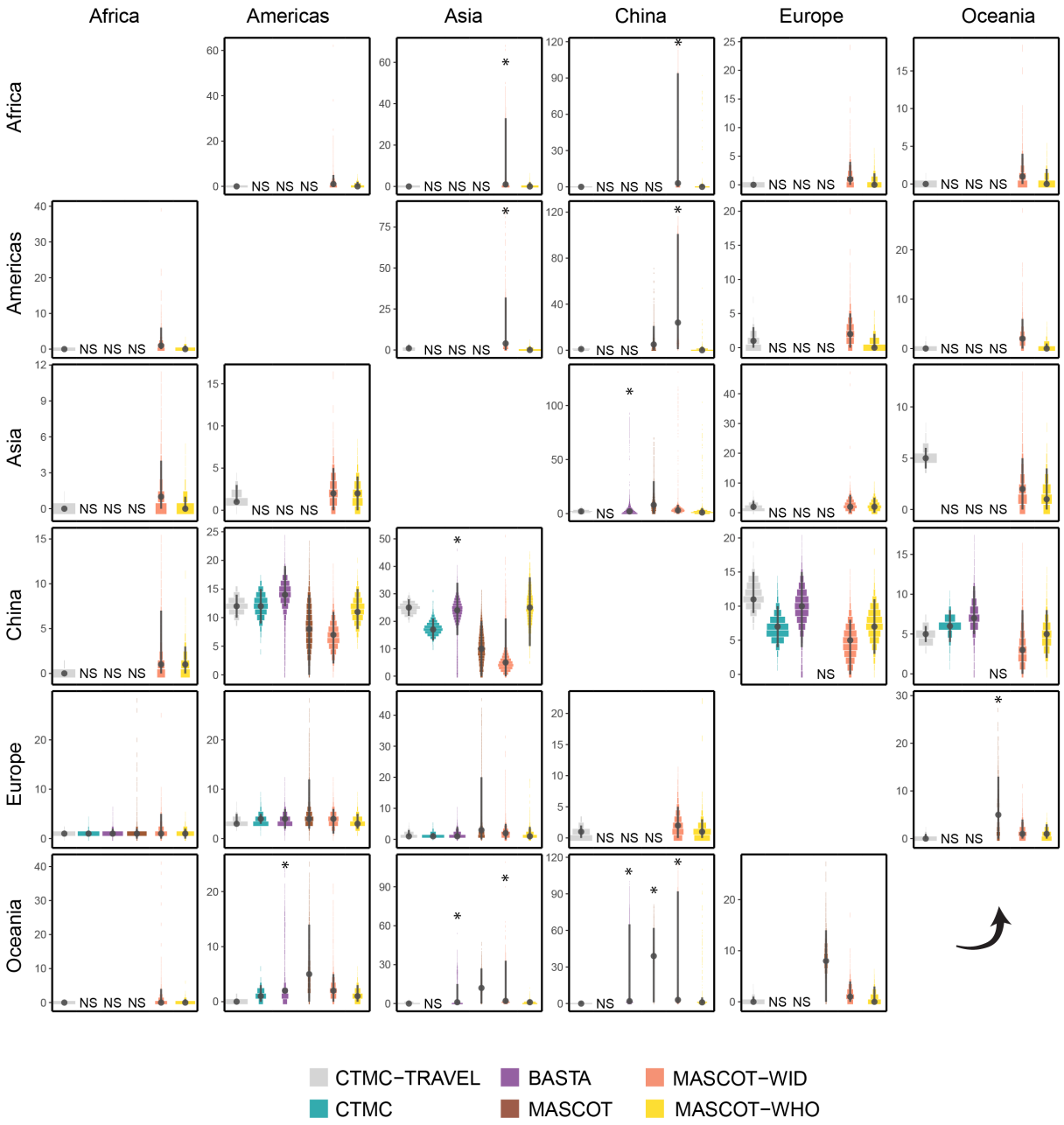


Figure 26: Posterior distributions of the total migration counts estimated on the SARS-CoV-2 data. Source locations are displayed by rows and destination locations by columns. For CTMC, BASTA and MASCOT, posterior distributions of the total migration counts with a Bayes factor (BF) < 3 are not depicted but marked as non-significant (NS). We identify bimodal marginal posterior distributions with a (*) and we report for each posterior distribution the median and 95% HPD. We normalize the width of the violin plots so that the cumulative density is equal to one.

We next incorporated incidence data from either Our World In Data or the WHO into MASCOT-GLM. Interestingly, the reconstructions differed strongly between the two data sets for incidence. While

MASCOT-WID predictions are uncertain with multimodal total migration counts (Fig. 26) and do not reflect the original spread from China (Fig. D21.E), MASCOT-WHO estimated migration counts that are close to the estimates of CTMC-TRAVEL (Fig. 26) and its MCC tree is in agreement with the origin of the pandemic (Fig. D21.E). Importantly, the two data sets differ strongly in how well early cases are covered (Fig. D20), with the WHO data set being more representative of the incidence over time. Overall and as also suggested by our simulations, while the structured coalescent model, in principle, allows to mitigate sampling biases, it can itself be highly biased when the wrong population dynamics are assumed.

4. Discussion

Sampling bias is a key challenge in phylodynamic inference (Frost et al., 2015), as in discrete phylogeography. In its early developments, the evaluation of the impact of sampling bias on Bayesian discrete phylogeography models was restricted by the availability of whole genomes (Lemey et al., 2009a). The SARS-CoV-2 pandemic has led to a paradigm shift as genomic surveillance became part of routine surveillance systems around the world (Hodcroft et al., 2021a). Here, we evaluated the impact of sampling bias on discrete phylogeography inference using simulated and empirical data to provide insightful knowledge on how sampling bias affects such inference and how it could be mitigated.

4.1. Inference performance in absence of sampling bias

In our simulation study, genetic parameters (i.e., average evolutionary rate, stationary nucleotide frequencies, ratio of transition-transversion rates) are correctly estimated and tree topologies match the corresponding simulated transmission chains for all approaches. In addition, CTMC leads to high correlation between simulated and estimated spatiotemporal parameters as well as low relative and absolute error in absence of sampling bias. Overall, CTMC reconstructs the spatiotemporal histories well and its estimates are more accurate in large samples. BASTA and MASCOT do not correctly infer the spatiotemporal parameters in the seven demes framework but correlation between simulated and estimated total migration counts is slightly improved in the three demes framework while remaining lower than CTMC. This could result from three different causes. First, we assumed that all deme sizes are equal and constant over time in BASTA and MASCOT, the former to avoid overparameterization and the latter being the only available assumption in current implementations. Such a parameterization is more appropriate in the case of endemic circulation with limited time-varying dynamics such as local extinctions. However, large variations in time and local extinctions occur in our simulations meaning that we had to assume

incorrect population dynamics in BASTA and MASCOT. This is confirmed by the better performance of MASCOT-GLM in uniform samples that accommodates for the true population dynamics. Secondly, we would expect BASTA and MASCOT to perform better on "even" samples that contain approximately as many sequences of each sampled location (De Maio et al., 2015). In our simulation study, uniform sampling does not imply an even representation of sampled locations. Indeed, locations where the virus has not circulated much are less represented. Such an effect is more pronounced in the seven demes framework than the three demes framework and we effectively observe poorer performances of BASTA and MASCOT in the seven demes framework. Finally, the structured coalescent model is known to be sensitive to ghost demes, i.e. unsampled locations (Beerli, 2004; Ewing and Rodrigo, 2006; De Maio et al., 2015). As we considered the sampling process to be naive of the number of affected locations, locations where the virus has not circulated much may remain unsampled. This is true for the seven demes framework only for which we observe poorer performance of BASTA and MASCOT compared to the three demes framework. However, the impact of ghost deme inclusion and potential misspecification on the estimation of the migration patterns remains unclear. While two studies showed that accounting for ghost demes in the structured coalescent model improves the inference of deme size (Beerli, 2004; Ewing and Rodrigo, 2006), Ewing and Rodrigo (2006) also showed that adding just a few sequences from the ghost deme leads to the overestimation of the migration rate.

4.2. Inference performance under sampling bias

We show that CTMC, BASTA, and MASCOT are impacted by spatial sampling bias in different ways. CTMC performance is dramatically impaired with increasing levels of sampling bias. This is directly linked to the geographical sampling frequencies that inform the likelihood of CTMC (De Maio et al., 2015). It also tends to be overconfident, and this overconfidence worsens with stronger sampling bias as previously shown (De Maio et al., 2015). However, the impact of sampling bias can be mitigated by either using large samples at low levels of sampling bias or controlling for sampling bias by balancing sample composition (region and region+year subsamples) at intermediate levels of sampling bias. These results were well-replicated in a simpler framework of transmission between three locations which rules out the confounding effect of the simulation complexity and unsampled locations on our results (see section 4 of the Supplementary Materials).

BASTA and MASCOT do not accurately estimate the total migration counts nor the lineage introduction dates in biased and unbiased conditions. Nevertheless, the overall migration process evaluated by the lineage migration counts is relatively well captured with a correlation around 0.5 that is not impacted by

sampling bias contrary to CTMC in both the three demes and seven demes scenarios. We show that the approximations of the structured coalescent model are generally less confident than CTMC which is in agreement with a previous study (De Maio et al., 2015) and their uncertainty around median estimates increases with sampling bias. We also show that sample composition impacts the inference of BASTA and MASCOT in the three demes framework since correlation levels are strongly improved and bias and uncertainty are reduced for all spatiotemporal parameters in "even" samples (region and region+year), despite the underlying surveillance bias. Still, BASTA and MASCOT estimates display lower correlation with the simulated values, higher uncertainty and higher relative and absolute bias compared to CTMC. In the seven demes framework, the results are less clear which may be due to the presence of ghost demes. Interestingly, BASTA seems to outperform CTMC and MASCOT in the inference of the lineage introduction dates in the three demes framework. This result was however not replicated in the seven demes framework.

While structured coalescent methods potentially allow mitigating sampling biases as previously shown (De Maio et al., 2015), assuming incorrect population dynamics very likely introduces biases. Structured coalescent models currently assume constant population sizes in all demes, and often require the additional assumption of equal population sizes to reach convergence and attain proper mixing. When the true underlying population dynamics are complex with large differences between populations, the models cannot estimate the population sizes with low uncertainty and compensate for this issue in the estimation of the migration rates, so ultimately in the migration history. We addressed this issue by modeling population dynamics more accurately using a GLM approach whenever the required incidence data to do so were available. Indeed, using incidence data to inform population dynamics in MASCOT counteracts the impact of sampling bias even at high levels. This result also underlines the fact that location sampling frequencies do not inform the structured coalescent model when population dynamics are known (De Maio et al., 2015). It also shows that the inclusion of ghost demes is not necessary when the true population dynamics are incorporated in the model. Overall, our results showcase the importance of considering the assumptions of population dynamics on the ancestral state reconstruction in structured coalescent model approximations (Table 7).

4.3. Analysis of empirical RABV and SARS-CoV-2 data sets

We further compare the approaches on real data sets of RABV and SARS-CoV-2. As dog case counts were not available for RABV, we compare only CTMC, BASTA, and MASCOT. CTMC predicts a highly spatially-structured migration process whereas BASTA and MASCOT predict a non-parsimonious sce-

nario. We observe similar results for the SARS-CoV-2 data set. As we have set equal deme sizes in BASTA and MASCOT but a single tip is sampled for Oriental Mindoro in the RABV data set and Africa in the SARS-CoV-2 data set, the two algorithms compensate for location underrepresentation by estimating high backwards-in-time migration rates to the underrepresented location (Oriental Mindoro and Africa). Our results are in line with previous studies reporting strong differences between CTMC and the structured coalescent model on real data sets (De Maio et al., 2015; Dudas et al., 2018). However, there is also evidence in the literature of a good agreement between the two types of models (Faria et al., 2017; Brynildsrud et al., 2018; Yang et al., 2019; Mavian et al., 2020). Such similarities can result from sample composition (at least ten sequences per location in (Yang et al., 2019)), the parameters used for comparison (probability of clade ancestral location in (Faria et al., 2017)), prior information (information on the root location in (Brynildsrud et al., 2018)), or the underlying transmission dynamics. Besides, these studies focused on the overall migration process which corresponds to the lineage migration counts in our simulation study and we showed that the overall migration process is roughly estimated at any level of bias. In brief, we show on real data sets that singletons may be inferred as drivers of the migration process in an unparsimonious way by structured coalescent model approximations. This result supplements a previous study on the impact of the inclusion of few ghost deme sequences on the inference of migration rates (Ewing and Rodrigo, 2006), however their impact remains unclear and deserves close consideration.

Interestingly, the posterior density of the structured coalescent model in BASTA is bimodal for the SARS-CoV-2 data set. Its major mode corresponds to a past migration history close to CTMC-TRAVEL and our expectations of SARS-COV-2 spread at the start of the pandemic, whereas the minor mode corresponds to the non-parsimonious scenario. Such bimodality was not observed for MASCOT in the SARS-CoV-2 analysis. This difference in estimation is not unexpected since the two structured coalescent model approximations are different. However, it is not clear which characteristics of the two algorithms would lead to different behaviors. Another possibility relies on the choice of operators that determine how well the two approximations explore the parameter and tree space in which case MASCOT should lead to a bimodal posterior density on the long run.

Table 7: Summary of the simulation study.

	CTMC	**BASTA**	**MASCOT**	**MASCOT-GLM**
Model type	Continuous-time Markov chain	Approximation of the structured coalescent with equal deme sizes	Approximation of the structured coalescent with equal deme sizes	Approximation of the structured coalescent with time-varying deme sizes

Inference in unbiased spatial sampling conditions				
Tree topology	Correct topology estimated	Correct topology estimated	Correct topology estimated	Correct topology estimated
Migration parameters	Correct estimation	Biased and uncertain estimates	Biased and uncertain estimates	Correct estimation
Inference in biased spatial sampling conditions				
Tree topology	Correct topology estimated	Correct topology estimated	Correct topology estimated	Correct topology estimated
Migration parameters	Biased estimation compared to unbiased sampling conditions	Increased bias and uncertainty compared to unbiased sampling conditions	Increased bias and uncertainty compared to unbiased sampling conditions	Correct estimation
Alternative sampling strategy				
Tree topology	Correct topology estimated	Correct topology estimated	Correct topology estimated	Correct topology estimated
Migration parameters	Similar performance as in unbiased sampling conditions	Decreased bias and uncertainty compared to unbiased sampling conditions	Decreased bias and uncertainty compared to unbiased sampling conditions	Correct estimation
Conclusion	Sensitive to spatial sampling bias that can be reverted using alternative sampling strategies	Sensitive to equal deme size assumption and ghost demes, performs better on spatially balanced samples	Sensitive to equal deme size assumption and ghost demes, performs better on spatially balanced samples	Robust to bias when informed with exact case counts

4.4. Practical implications for the analysis of empirical data sets

Computation time is an important consideration in real-life situations. CTMC is a fast algorithm that can handle many sequences while facing little convergence issues, which made it the predominant approach. For example, CTMC and its extensions have been extensively used during the SARS-CoV-2 pandemic (Candido et al., 2020; Dellicour et al., 2021b; Lemey et al., 2020; Perez et al., 2022; Kaleta et al., 2022; Alteri et al., 2021; Dellicour et al., 2021c; Butera et al., 2021). In general, researchers analyzed large data sets whose composition was corrected or reflected case counts (Candido et al., 2020; Lemey et al., 2020) or the number of hospitalizations per geographical location (Dellicour et al., 2021b). According to our results, even though the pool of available sequences is not representative of the underlying transmission process, CTMC inference should be little impacted when using even subsamples of the available sequences. However, we did not test sampling strategies based on case counts in our simulations.

With BASTA and MASCOT, computation time can become rapidly cumbersome and even impractical - also as a result of poor mixing of structured coalescent model parameters - when the number of

sequences and locations increase. In such cases, these approaches are not able to discriminate which migration routes are the most important in the migration process leading to bimodal structured coalescent posterior densities, as observed for MASCOT on large samples of 500 sequences in the seven demes framework and for BASTA on the SARS-CoV-2 data set. Repeating these problematic analyses with different starting values did not redeem these issues. Other studies have reported similar issues (Richardson et al., 2018). However, these problematic inferences can potentially be overcome by informing structured coalescent models with additional covariate data on viral population size dynamics. Indeed, as a result of adding such data, MASCOT-GLM not only outperformed the other approaches at estimating spatiotemporal parameters but also displayed improved mixing as expected with GLM approaches which improves the computational burden. However, such improvements depend on the availability and informativeness of the case count data used, notably on the early viral population size dynamics. This is illustrated in our analysis of the SARS-CoV-2 data for which the addition of WHO data led to improved chain mixing and past migration inference compared to the Our World in Data data, knowing that the dynamics are rather similar in the two data sets but they go back to January, 4th 2020 for the WHO data and to January, 23rd 2020 for the Our World in Data data.

4.5. Limitations

We acknowledge several limitations of our study. First, BASTA and MASCOT are expected to perform better on even samples, a condition that we did not directly test. In the representative (uniform) samples, location frequencies inform CTMC and thus it would be expected to be favored over BASTA and MASCOT. Still, we show that MASCOT and BASTA perform better on even (region and region+year) samples in the three demes framework even if they are derived from biased large biobanks. This result suggests that BASTA and MASCOT perform better on even samples with no ghost demes. Second, our subsampling procedure in the simulation analysis could leave some locations unsampled, which can be considered as an extreme case of sampling bias. While this happened in only a few highly biased samples in the three demes framework, it is very common in the seven demes framework even in absence of sampling bias. It is difficult to determine whether the poor performance of MASCOT and BASTA in absence of bias in the seven demes framework compared to the three demes framework is due to ghost demes or is simply due to the higher number of locations. Additionally, we cannot rule out that the effects of sampling bias we observe are due to unsampled locations/unspecified ghost demes rather than unrepresentative sampling. We did not include unsampled locations as ghost demes in such conditions. However, this is unlikely to improve migration rate estimation (Ewing and Rodrigo, 2006). Third, the

impact of sampling bias certainly depends on the underlying overall migration rate as shown by (Liu et al., 2022), an impact that we did not investigate here.

Another limitation concerns the incorporation of epidemiological data in phylogeographic models. Here, deme sizes in MASCOT-GLM are informed by case count data but this kind of data may not be readily available (Grubaugh et al., 2019b) and is known to be often biased due to varying testing effort and strategy, as well as differential testing behaviors by age (Buckee et al., 2021). It is difficult to predict how MASCOT-GLM would perform if parameterized with biased case counts, a case that we did not address in our simulations. The comparison between the WHO and WID cases data, however, suggests that biased coverage of the true case load could bias such inference. If case count data are not reliable, one could use hospitalization data instead (Dellicour et al., 2021b). Further, a similar approach is available under the CTMC framework but we did not test it here. This framework consists in modeling the migration process with CTMC and the overall population dynamics with the GLM extension (Gill et al., 2016) of the skygrid coalescent model (Gill et al., 2013). In this extension, case count over all locations could be used as a predictor of the viral population size over time. Yet, such an approach assumes a panmictic population and remains rare.

Finally, it is difficult to generalize our results in regards to the number of demes. Our choice of the number of locations was influenced by the RABV scenario in Morocco. While a scenario with three demes was doable, the one with seven demes turned out to be difficult to analyze, notably due to computational burden (Supplementary Materials). More research and development is needed for data sets with a large number of locations (> 15) and it currently seems unlikely that such analyses are possible at all with BASTA and MASCOT.

4.6. Perspectives

In conclusion, sampling bias can be tackled at different levels of data generation and analysis in phylogeographic analyses: sample constitution, inference model choice, and data integration (e.g. through an integrated GLM). Other studies also assess the impact of sampling bias in post hoc analyses (Chaillon et al., 2020; Vrancken et al., 2020) or explicitly model sampling patterns (Guindon and De Maio, 2021). Although the exploration of the impact of sampling bias has increased over the recent years and more robust methodologies have been developed, many aspects remain unclear, among which the impact of unsampled locations, biased epidemiological data incorporation, or the relative performances on even versus representative samples. Whenever possible, we would advise to opt for an even sampling strategy

across geographical locations, compare the inferences of the different approaches or compare the inferences over multiple subsamples when analyzing real data sets. These considerations are all the more important in a world of ever-growing genome sequence generation and concern not only human viral diseases but also zoonoses and epizooties.

Chapter 5

Impact of contact heterogeneity on respiratory diseases transmission in households

Households are an ideal setting for the study of respiratory diseases transmission. Modeling studies of household transmission data have helped characterize the role of children for infections such as influenza and SARS-CoV-2. However, estimates obtained in these studies may be biased since they do not account for the heterogeneous nature of household contacts. Here, we quantified the impact of contact heterogeneity between household members on the estimation of the relative susceptibility and infectivity of children. We simulated epidemics of SARS-CoV-2-like and influenza-like infections in a synthetic database of 1,000 households assuming heterogeneous contact levels. Contacts were assumed more frequent in the father-mother pair, followed by the child-mother pair, then the child-child pair, and finally the child-father pair with the least contact frequency. Child susceptibility and infectivity were then estimated while accounting for heterogeneous contacts or not. We showed that the relative susceptibility of children was under-estimated by approximately 20% in the two disease scenarios. Concerning the relative infectivity of children, it is underestimated by 20% when children and adults had different infectivity levels. This study shows how in small communities, heterogeneous contact patterns should be evaluated and accounted for. New household studies collecting both disease and contact data are needed to cast light on the role of complex contacts on disease transmission and improve the estimation of key biological parameters.

1. Introduction

Households constitute an ideal setting for the study of respiratory diseases transmission. These diseases generally transmit through aerosols, a transmission route that is favored in closed indoor spaces like households, or through droplets during close contacts that are characteristic of contacts between household members (Tsang et al., 2016; Wang et al., 2021). Household transmission represents a non-negligible part of respiratory disease transmission, and, in extreme cases like influenza infections, living with an infected individual is the most important risk factor of infection (Longini et al., 1982). In addition, the study of respiratory diseases transmission is simplified in households because case contacts are well-defined which facilitates their follow-up after exposure and the estimation of the secondary attack rate (SAR), defined as the proportion of susceptible household contacts that are infected after the index case is detected.

Mathematical models of disease transmission in households have helped characterize the role of children (Cauchemez et al., 2004; Cauchemez et al., 2009; Tsang et al., 2019; Endo et al., 2019; Dattner et al., 2021) and vaccination (Tsang et al., 2019; Prunas et al., 2022) in the dynamics of transmission. This is generally done by estimating their relative susceptibility and infectivity compared to adults. For example, child susceptibility was shown to be half adult susceptibility for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections (Davies et al., 2020a; Viner et al., 2021; Dattner et al., 2021; Franco et al., 2022), and twice as susceptible to influenza infections (Cauchemez et al., 2004; Cauchemez et al., 2009). The relative infectivity and susceptibility estimated in these studies can be caused by biological factors (e.g. different levels of viral shedding when infected or different propensity to get infected when exposed), but also by the level of physical contacts in the household (Lordan et al., 2021). However, so far, household transmission models have always ignored the second source of heterogeneity, implicitly assuming estimated values were indicative of different biological parameters between children and adults. To date, only one study has tested the hypothesis of homogeneous mixing in the household environment that is underlying in all these studies (Goeyvaerts et al., 2018). It concluded that, (*i*) on average, children have less contacts with their father than with other siblings, (*ii*) the overall rate of physical contacts between children decreases with age, and (*iii*) the magnitude of contacts decreases with household size. The study shows that the assumption of homogeneous mixing does not hold in the household environment. As a result, part of the estimated differences between children and adults in households is expected to be due to different mixing patterns in the household. It is important to determine by how much the complex mixing patterns in the households may bias estimates of biological

susceptibility/infectivity that are derived in household studies.

Here, we aim at investigating how heterogeneous contact patterns in households could bias estimates of respiratory diseases transmission, notably the force of infection between household members, the relative susceptibility of children compared to adults, and their relative infectivity. We simulate epidemics in households using realistic contact patterns (Goeyvaerts et al., 2018) and we estimate key transmission parameters, accounting or not for the heterogeneous nature of the contact patterns.

2. Methods

2.1. Household composition in the simulated data set

We constituted a synthetic database by randomly sampling with replacement 1,000 households from a subset of the households (n = 225) of the multi-center household study RECOVER (Verberk et al., 2022). From the RECOVER study, we retained households with two to five household members that correspond either to heterosexual couples, or to single-parent or hetero-parental two-generation families. We excluded same-sex couples and homo-parental families because of the lack of estimates in the study by Goeyvaerts et al. (2018) on contact levels between partners of same-sex couples, and more specifically, between same-sex parents and their children. From the original household study RECOVER, we kept two types of information for each household member: (*i*) whether the individual is the index case, and (*ii*) the category of the individual (i.e., mother, father, or child).

2.2. Simulation of household epidemics

2.2.1. *In silico* follow-up protocol

We assumed that the 1,000 households from the synthetic database were recruited and followed up starting from the symptom onset of the index case, and for up to 20 days. Since our aim is to ascertain how the misspecification of contact intensity may impact the estimation of transmission rates, we decided to consider a simple inference context: in the simulated data set we analyze, all cases exhibit symptoms and testing is perfect.

2.2.2. Force of infection within households

In the simulations, the probability that an individual k in household h gets infected between time t and time $t + dt$ is:

$$\Lambda_k(t,t+dt) = 1 - exp\left(\alpha \times dt + \sum_{l \in I_h\{\xi_l < t\}} \frac{\beta}{n/2}\, \kappa_{k,l}\, \mu_{s,k}\, \mu_{i,l} \int_t^{t+dt} f(u-\xi_l|s_l)du\right) \tag{5.1}$$

where:

- α is the instantaneous hazard of infection in the community.
- $l \in I_h\{\xi_l < t\}$ correspond to the infected individuals in household h that were infected before time t, with ξ_l their infection date.
- $\frac{\beta}{n/2}$ models the dependency between the baseline transmission rate β and the household size n. Here, the baseline transmission rate β corresponds to the transmission rate in heterosexual couples, when $n = 2$, $\kappa_{k,l} = \kappa_{mother,father} = 1$, $\mu s,k = \mu_{s,adult} = 1$, and $\mu i,l = \mu_{i,adult} = 1$. See below for the meaning of the parameters, and the reference categories.
- $\kappa_{k,l}$ is the relative contact rate between recipient k and infector l according to the type of the pair. We used the mother-father pair as a reference, which means that for this type of pair $\kappa_{k,l} = 1$. For the other pairs, we used the odds-ratios estimated by Goeyvaerts et al. (2018) during weekdays. Mother-child pairs were assumed to be 11% less in contact than mother-father pairs, father-child pairs 58% less in contact, and pairs of children 24% less in contact.
- $\mu_{s,k}$ is the relative susceptibility of recipient k according to their age. For adults, $\mu_{s,adult} = 1$.
- $\mu_{i,l}$ is the relative infectivity of infector l according to their age. For adults, $\mu_{i,adult} = 1$.
- $f(t-\xi_l|s_l)$ is the density probability function of the generation time conditioned on the incubation period s_l of infector l. Here, the generation time is defined as the distribution of the interval between the infection time of the infector and the infection time of the recipient. We used the distribution estimated by Ferretti et al. (2020) for SARS-CoV-2 infections.

If k gets infected between t and $t+dt$, its exact time of infection ξ_k is drawn uniformly between t and $t+dt$, and its incubation period s_k is drawn from a log-normal distribution with log-mean=1.63 and log-standard deviation=0.5, previously estimated by McAloon et al. (2020) for SARS-CoV-2 infections. If symptom onset occurs after the end of the follow-up, the individual is not detected. We simulate continuous times of infection and symptom onset. For realistic reasons, we truncated the time of symptom onset and kept only the day of symptom onset to perform the inference.

We tested two scenarios. The first one corresponds to a SARS-CoV-2-like scenario, with child sus-

ceptibility being half adult susceptibility, and 20% less infectious than adults (Table 8). The second corresponds to an influenza-like scenario with children being twice as susceptible but as infectious as than adults. For each scenario, the value of the baseline transmission rate in heterosexual couples β was chosen so that the overall SAR is approximately 37% (Table 8). Finally, we simulated epidemics in the synthetic household database 100 times for each scenario.

Table 8: Parameter values used in the simulations.

Parameter	SARS-CoV-2	Influenza virus
Force of infection in the community α	0.001	0.001
Transmission rate in heterosexual couples β	0.46	0.99
Relative susceptibility of children $\mu_{s,child}$	0.5	2
Relative infectivity of children $\mu_{i,child}$	0.8	1

2.3. Statistical inference

Statistical inference was performed in a Bayesian framework with data augmentation (Cauchemez et al., 2004). In the section above, we detailed the model by using adults as the reference. In the inference model, we used children as a reference because pairs of children were more numerous than pairs of adults which improved inference. We estimated the hazard of infection in the community α, the force of infection between two children in a household of size four $\frac{\beta}{4/2}\kappa_{child,child}\mu_{s,child}\mu_{i,child}$, the relative susceptibility $1/\mu_{s,child}$ and relative infectivity $1/\mu_{i,child}$ of adults compared to children using a simple Metropolis-Hastings algorithm. For α, we assumed an exponential prior distribution with parameter equal to 500 which means that the instantaneous incidence rate is 200/100,000 inhabitants in the population, and for $\frac{\beta}{4/2}\kappa_{child,child}\mu_{s,child}\mu_{i,child}$, we assumed a uniform prior distribution between 0 and 10. We used a log-normal distribution with log-mean = 0 and log-standard deviation = 1 for $1/\mu_{s,child}$ and $1/\mu_{i,child}$.

Infection dates and symptom onset dates were augmented after each parameter iteration. Infection dates were sampled from the incubation period distribution estimated by McAloon et al. (2020), and the exact time of symptom onset was sampled uniformly over the observed day of symptom onset.

For each simulation, we launched one Markov chain Monte Carlo (MCMC) chain for 50,000 iterations. We discarded a burn-in of 5,000 steps and applied a thinning of 20 for the estimation of the posterior distributions. Convergence was assessed visually and by calculating the effective sample size (ESS) using the "effectiveSize" function in the "coda" R package for every parameter of every MCMC chain. ESS values exceeded 500 for all parameters in all chains. The comparison of the prior and posterior

distributions of model parameters is available in Fig. E1 for the SARS-CoV-2 infection scenario and in Fig. E2 for the influenza virus infection scenario in Appendix E.

2.4. Comparison of simulated and estimated parameters

The estimates of β, $\mu_{s,child}$, and $\mu_{i,child}$ were compared to the values used in the simulations using two metrics:

- the mean relative bias defined as $MRB = \frac{1}{100}\sum_{i=1}^{100}\frac{1}{\theta_i}(\hat{\theta}_i - \theta_i)$;
- and the 95% coverage defined as $coverage_{95\%} = \frac{1}{100}\sum_{i=1}^{100} 1_{\{\theta_i \in CrI_{95\%}(D_i)\}}$.

We denote θ_i the true value of the parameter, D_i the parameter posterior distribution, $\hat{\theta}_i$ the median estimate, and $CrI_{95\%}$ the 95% credible interval.

3. Results

In the COVID-19 scenario depicted in Fig. 27, the three parameters of within household transmission are well estimated when the inference model accounts for heterogeneous contact patterns between household members ("correct" inference model in Fig. 27). The transmission rate in heterosexual couples is relatively well estimated with a mean relative bias lower than 3% (Fig. 27D) and a 95% coverage of 96% (Fig. 27G). The estimation of child relative susceptibility is also satisfying with a mean relative bias around -5% (Fig. 27E) and a 95% coverage of 91% (Fig. 27H). Finally, the 20% reduction of child infectivity is well estimated with a mean relative bias of about 4% (Fig. 27F) and a high 95% coverage of 98% (Fig. 27I). The slight overestimation of the transmission rate in heterosexual couples mirrors the slight underestimation of child relative susceptibility as the two parameters are negatively correlated. When the inference model does not account for contact heterogeneity ("incorrect" inference model in Fig. 27), the estimation of the parameters of within household transmission is largely biased. The transmission rate is overestimated by 29% and the 95% CrI never contains the true value (Fig. 27D and G). Child relative susceptibility and child relative infectivity are underestimated by more than 20% (Fig. 27E-F) and their 95% coverage does not exceed 20% (Fig. 27H-I). Given that heterosexual couples have the strongest level of contact in the simulations, their net transmission rate is higher than the net transmission rate in pairs of children or between parents and children. When the inference model assumes that all household members have the same level of contact patterns, it has to compensate for the higher transmission rate in pairs of adults and the lower transmission rates between children and in parent-child pairs by increasing the transmission rate in heterosexual couples and reducing the susceptibility and infectivity of children.

The extent of the bias that we observe results from the values used to model contact heterogeneity in the simulations.

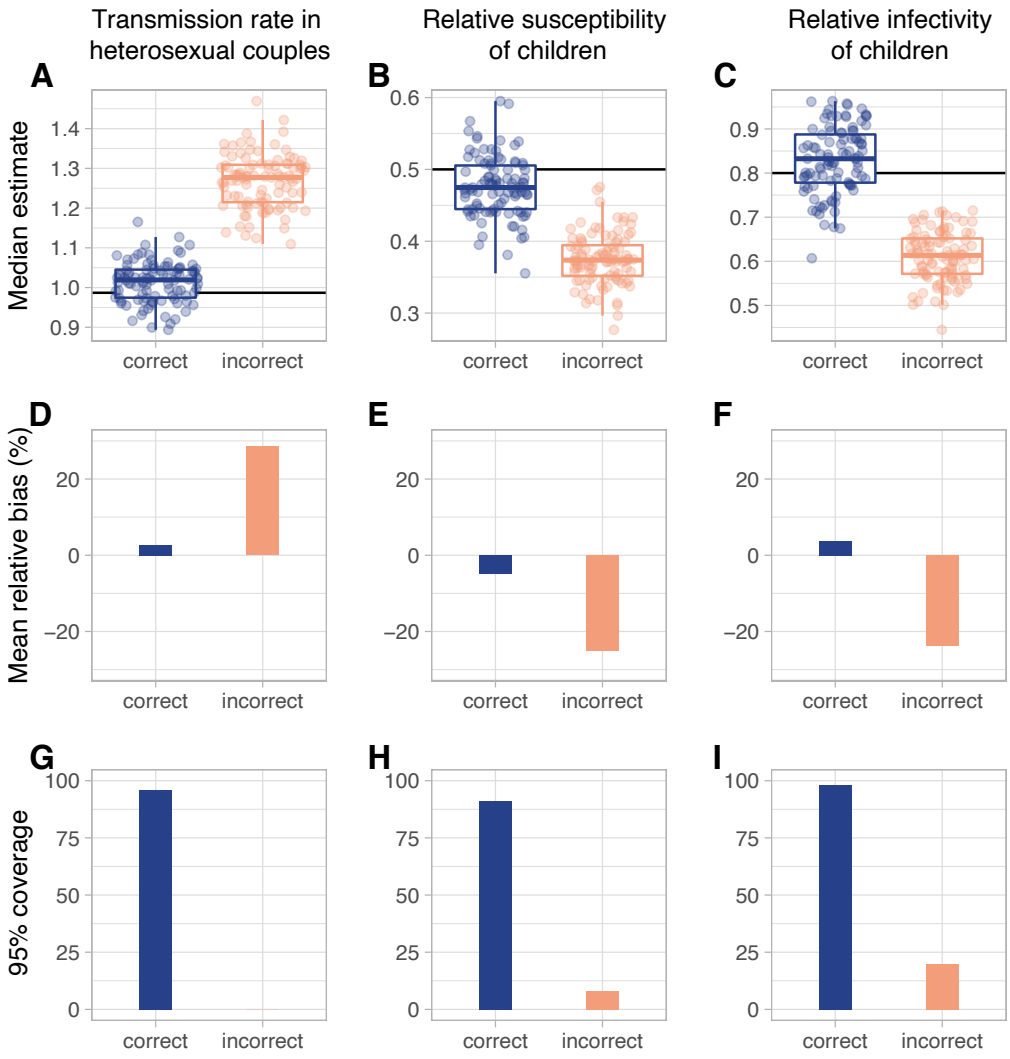


Figure 27: Impact of contact patterns on the estimation of within household transmission and children infectivity and susceptibility in severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections. (**A-C**) Posterior median estimates of the transmission rate in heterosexual couples, child relative susceptibility, and child relative infectivity for the correct (heterogeneous mixing) inference model in dark blue (n=100), and the incorrect (homogeneous) inference model in light orange (n=100). The black horizontal line corresponds to the true value used in the simulations. (**D-F**) Relative bias between the posterior median estimate and the true value for the transmission rate in heterosexual couples, child relative susceptibility, and child relative infectivity. Positive values indicate overestimation and negative values underestimation. Relative bias is expressed in percentage. (**G-I**) 95% coverage of the transmission rate in heterosexual couples, child relative susceptibility, and child relative infectivity.

We obtain very similar results for the flu scenario presented in Fig. 28. When contact heterogeneity is accounted for, the transmission rate in heterosexual couple is slightly overestimated by around 3% with a 95% coverage of 91% (Fig. 28D and G). Child relative susceptibility is underestimated by about 5% with a 95% coverage of 91% (Fig. 28E and H). In contrast, when homogeneous mixing between household

members is assumed, the transmission rate in heterosexual couples is overestimated by 19% (Fig. 28D) and child relative susceptibility is underestimated by 18% (Fig. 28E) given that the 95% coverage does not exceed 40% for both parameters (Fig. 28G-H). Just like in the COVID-19 scenario, estimation bias in the incorrect inference model results from the compensation of contact heterogeneity in the simulations. The results for child relative infectivity are less clear in the flu scenario in which adults and children have the same infectivity levels. Indeed, the parameter is overestimated by 8% with a 95% coverage of 85% (Fig. 28F and I) with the correct model and it is underestimated by 8% with a 95% coverage of 82% with the incorrect model (Fig. 28F and I).

4. Discussion

In this study, we showed that the estimates of the child relative susceptibility and child relative infectivity can be biased when heterogeneous contact patterns between household members are not accounted for in the inference. When considering the transmission of SARS-CoV-2 or influenza viruses in households with heterogeneous contacts derived from Goeyvaerts et al. (2018), the incorrect assumption of homogeneous mixing in the inference model leads to the underestimation of child relative susceptibility and child relative infectivity by around 20%. This underestimation compensates the lower contact rate between children and other household members compared to the contact rate in heterosexual couples in the simulated epidemics (Goeyvaerts et al., 2018). Biased estimates of child relative susceptibility and child relative infectivity may mislead the parameterization of disease transmission models that are often used to design intervention measures.

To circumvent this problem, it seems important to integrate information about contact patterns in household transmission models. Using the results of a household contact survey such as Goeyvaerts et al. (2018) to inform an observational study in a different country seems problematic since household contact patterns are expected to vary across socioeconomic levels and cultural practices. Ideally, the study design of household transmission studies should integrate the collection of epidemiological data on transmission and contact data between household members. The behavior of household members may change when one or multiple members develop symptoms, and it is therefore important to monitor variations in contact patterns during the study period. In addition, behavioral change upon symptomatic infection may depend on socioeconomic factors. For instance, physical distancing and self-isolation are not possible in crowded households (VoPham et al., 2021). Finally, the way contact data are collected may be influential. Contact surveys may be subject to reporting bias (selective revealing or suppression of information) because participants may under-report undesirable behaviors like not implementing physical distancing.

Alternatively, wearable electronic devices that measure close-proximity interactions are highly valuable in contexts with complex networks and for the study of infectious disease transmission (Starnini et al., 2017). However, records typically do not exceed a few days due to the limited autonomy of these devices.

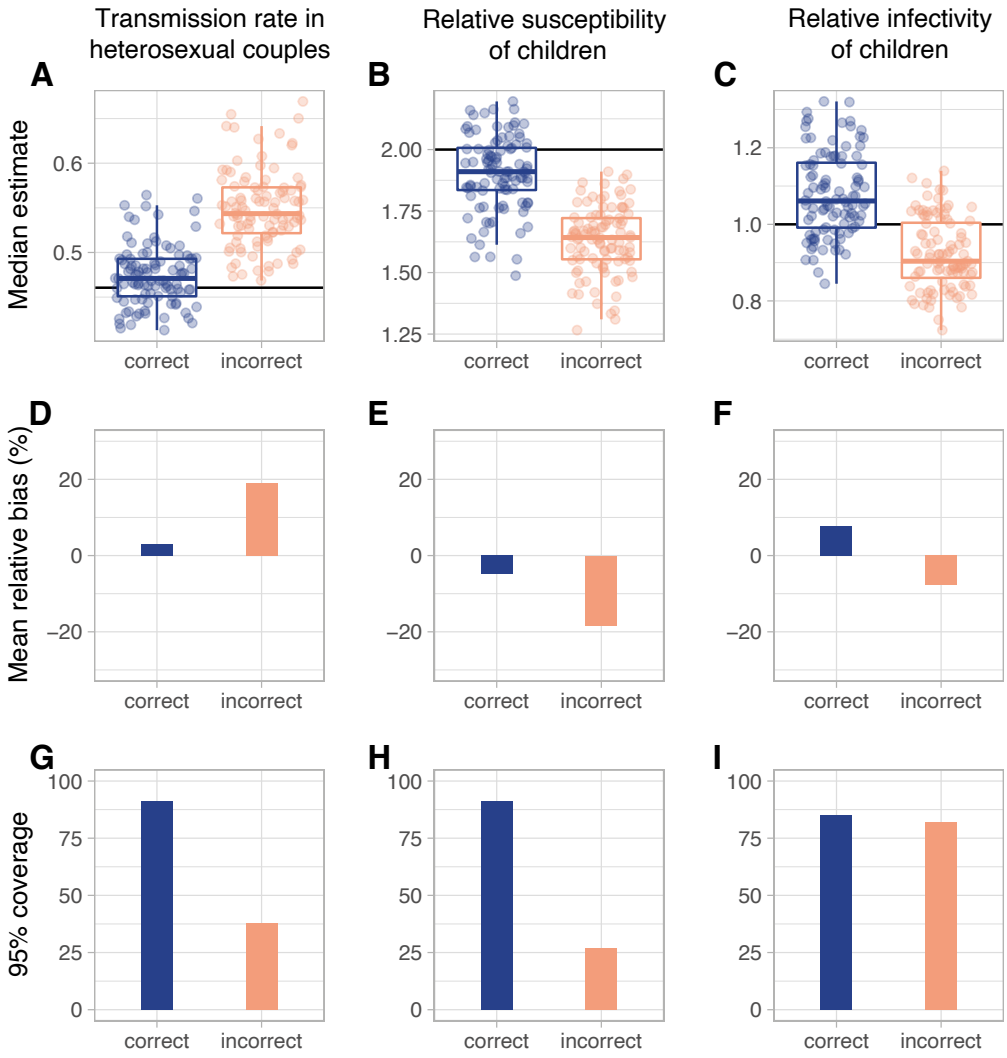


Figure 28: Impact of contact patterns on the estimation of within household transmission, child infectivity, and child susceptibility in influenza virus infections. (**A-C**) Posterior median estimates of the transmission rate in heterosexual couples, child relative susceptibility, and child relative infectivity for the correct (heterogeneous mixing) inference model in dark blue (n=100), and the incorrect (homogeneous) inference model in light orange (n=100). The black horizontal line corresponds to the true value used in the simulations. (**D-F**) Relative bias between the posterior median estimate and the true value for the transmission rate in heterosexual couples, child relative susceptibility, and child relative infectivity. Positive values indicate overestimation and negative values underestimation. Relative bias is expressed in percentage. (**G-I**) 95% coverage of the transmission rate in heterosexual couples, child relative susceptibility, and child relative infectivity.

Here, we simulated epidemics in households so that around 37% of household contacts get infected. The choice of this value for the SAR is relatively arbitrary given that estimates from empirical data vary from a few percents to 45% for the historical variant of SARS-CoV-2 (Madewell et al., 2020), and from 4% to 45% for influenza viruses (Cauchemez et al., 2009; Lau et al., 2012; Glatman-Freedman et al., 2012;

Azman et al., 2013). Simulating epidemics with a lower SAR would reduce the number of infected pairs, and thus, the statistical power to estimate child relative susceptibility and child relative infectivity. Besides, we made simplistic assumptions in the simulation model assuming that all infected individuals eventually develop symptoms and testing is perfect which would not have been adapted for the analysis of empirical data. We also assumed that the risk of infection in the community is the same for adults and children, although it may vary with age (Cauchemez et al., 2011; Tsang et al., 2019). However, household studies are not designed to evaluate the risk of infection in the community, notably because participants are followed up over short time periods, and such studies are generally deployed for the study of diseases with intense transmission in households. For these reasons, the interpretation of the hazard of infection in the community should subject to caution, as its estimation is loosely reliable.

In conclusion, the heterogeneous nature of contacts in households is expected to bias estimates of key parameters that are estimated from household studies, such as the relative susceptibility and infectivity of children. It is therefore important that these complex household contact patterns are accounted for in future household studies. Data are scarce and many knowledge gaps remain concerning the changes of household contact patterns that may occur following infections. Future household transmission studies should collect that on both disease and contact patterns, raising new challenges related to the study design, and model development.

General discussion

1. Synthesis of the results

1.1. Dog rabies circulation and its control

In Chapter 1, I highlight the main mechanisms underlying rabies circulation in dog populations: small R_0 leading to short transmission chains, role of imports from neighboring endemic areas, co-circulation of viral lineages, and high inter-individual heterogeneity in behavior and transmission. A recent modeling study analyzing an extensive contact tracing data set has confirmed the intricate role of all these mechanisms in the maintenance of rabies circulation in an African and endemic setting, the Serengeti district in Tanzania (Mancy et al., 2022). In Chapter 1, I also underline that there is a high variability across settings related to the role of wildlife, importations, and dog ecology. For example, the Iranian epidemiological context is underlaid by the co-circulation of independent viral lineages in dogs and wildlife that belong to very different rabies biogeographical clades (Dellicour et al., 2019). On the other hand, I show in Chapter 2 that a unique viral clade circulates in dogs, the most important reservoir of rabies in Cambodia (Ly et al., 2009), with presumably no or few cross-country transmission events. More intriguingly, the epidemiological situation in Cambodia contrasts with its neighbors where multiple clades, either related to Chinese lineages (Nguyen et al., 2011), or other Southeast Asian lineages (Ahmed et al., 2015; Benjathummarak et al., 2016) circulate. Besides, (illegal) dog trade, not only for dog meat consumption, concerns the entire Southeast Asian area (Chevalier et al., 2021) which means that dogs do cross borders and external introductions very likely occur but do not lead to sustained transmission chains. Parts of rabies transmission dynamics in Cambodia thus remain unclear, notably the role of humans, long-distance transmission events, and spatial heterogeneity.

The comparison of the contributions of epidemiological and phylodynamic modeling in Chapter 1 points to the complementarity of the two methodological approaches. Phylodynamics helped unravel large-scale mechanisms of transmission, while epidemiological modeling focused more on the design and evaluation of interventions. Nevertheless, using simulations, several epidemiological modeling studies illustrated how population structure heterogeneity may participate to the maintenance of rabies despite its low transmissibility (Leung and Davis, 2017; Hudson et al., 2019b; Kadowaki et al., 2018). Recent efforts to better characterize dog populations in terms of population density (Chevalier et al., 2021;

Mancy et al., 2022; Thanapongtharm et al., 2021), roaming behavior (Bombara et al., 2017; Wilson-Aggarwal et al., 2021; Kittisiam et al., 2021), serological status (Velander et al., 2022), and ecological interactions with humans (Chevalier et al., 2021; Thanapongtharm et al., 2021) constitute a first glimpse at the individual-level heterogeneity and setting-specific features of dogs. For example, Chevalier et al. (2021) evaluated in two contrasted Cambodian provinces dog density, age-specific survival rates, dog ownership determinants, and dog management behavior. Their results are highly valuable for the development of more realistic models. They also indicate that transmission dynamics may not be solely governed by dog behavior in Cambodia but also by the ecological interactions between dogs and humans. Indeed, the vast majority of dogs are owned and owners do trade their dogs on a regular basis but in a very opportunistic way. Although, this trade mostly occurs at small spatial scales, long-distance trade can also happen which implies that humans might have mediated the long-distance transmission events unraveled in Cambodia in Chapter 2, similar to the North African context (Dellicour et al., 2017; Talbi et al., 2010). More extensive field studies are necessary to fill the gap about dog population structure and ecology, and set up more realistic models of rabies transmission in dogs

Dog vaccination, a control measure that has been recommended by the WHO for several decades (World Health Organization (WHO) et al., 2018), is the most effective way of controlling rabies circulation. However, parenteral vaccination of stray and free-roaming dog populations requires large and skilled dog-catching teams that are very costly (Yale et al., 2022), thus it is rarely implemented by local authorities despite the importance of these dog populations in RABV circulation (Leung and Davis, 2017; Hudson et al., 2019b; Kadowaki et al., 2018). The complementary use of oral rabies vaccination (ORV) in dogs represents a promising avenue to achieve high vaccination coverage in low-resource settings where the majority of dogs are free-roaming such as in India (Yale et al., 2022) and Cambodia (Chevalier et al., 2021). Unfortunately, the absence of political engagement and resource mobilization in Cambodia constitute the primary barrier to the implementation of ambitious vaccination policies. As long as rabies is not a notifiable disease in Cambodia, there will be no wide access to dog vaccination nor frequent vaccination campaigns and, the advancement towards rabies elimination will remain compromised. One way to promote political engagement is to show that elimination is feasible or at least that the costs to reduce the number of cases are limited. Our results in Chapter 2 are encouraging in this sense. The relatively strict endemicity of rabies in Cambodia implies that any well-designed vaccination campaign could effectively reduce rabies burden in dogs and human on the short term. However, as local transmission is controlled, re-importations will play a greater role threatening long-term control or elimination. Additional measures such as legislation on vaccination status at importation could help prevent potentially

harmful introductions but they pose numerous technical challenges.

In brief, data are necessary at the individual and country levels to better understand setting-specific dynamics of RABV in dogs and design adapted control strategies in the most cost-efficient way.

1.2. SARS-CoV-2 household dynamics and interventions to control its transmission

Despite being a close relative of SARS-CoV, SARS-CoV-2 transmits in a very different way (Abdelrahman et al., 2020; Cevik et al., 2021) which led to many unknowns and knowledge gaps at the start of the pandemic at the beginning of 2020. At the time, estimating key transmission parameters at the individual-level was crucial to assess variations in infectiousness through time by age and symptom status. Household studies rapidly proved to be valuable and have continued to provide insightful knowledge over the entire pandemic period (Pitzer and Cohen, 2020; Yang and Kenah, 2022). In parallel, the advent of vaccination gave to public health authorities a means to control circulation while relaxing social distancing measures. Randomized controlled trials were carried out to evaluate vaccine efficacy against severe outcome (Baden et al., 2021), and in a lesser extent, against infection (Polack et al., 2020). However, these vaccine efficacy estimates do not reflect the protective power of vaccination in real life conditions that are not controlled. Besides, randomized controlled trials are not adapted to evaluate the effect of vaccines against transmission in breakthrough infections (Lipsitch et al., 2022). Vaccine effectiveness measures protection in real world conditions and can be estimated in observational studies such as cohort studies, household studies, case control studies, or with the screening method. At the time of the study presented in Chapter 3, vaccine effectiveness against infection was already well quantified but evidence about vaccine effectiveness against transmission in breakthrough infections remained limited. Regev-Yochay et al. (2021) showed that individuals with breakthrough infections display lower viral loads compared to infected individuals with no prior vaccination. Shah et al. (2021) quantified the indirect effect of vaccination on the risk of infection in individuals who live with a vaccinated individual. Finally, Harris et al. (2021) estimated a reduction of the odds of SARS-CoV-2 infection in household contacts when the index case is vaccinated. In Chapter 3, I estimate that vaccine is effective against transmission, but a high uncertainty around the point estimate remains. Prunas et al. (2022) estimated a similar effect with the same order of magnitude, but they also showed that this effect is transient. Indeed, vaccine effectiveness against transmission in breakthrough infections decreases with time from vaccination. More than 90 days after the second dose, there is no more protective effect against transmission. Vaccine effectiveness against infection also decreases, but fortunately, vaccine effectiveness

against severe outcome remains high (Feikin et al., 2022). Vaccine effectiveness against infection and transmission is transient due to the waning of vaccine-induced immunity, and vaccine effectiveness has been challenged during the pandemic by the emergence of new SARS-CoV-2 variants. For example, our results in Chapter 3 are applicable to the Alpha variant that represented more than 95% of the cases in Israel at the time of the study (Ritchie et al., 2020), but it got rapidly replaced by the Delta variant in June 2021, itself replaced by Omicron in January 2022. Vaccine effectiveness against Delta and Omicron is lower compared to Alpha (Andrews et al., 2022; Bernal et al., 2021). This reduction is due to vaccine variant-mismatch because the more recent variants have accumulated mutations in the spike protein, the main target of vaccine-induced neutralizing antibodies, compared to the ancestral D614G strain used for vaccine production (Hewins et al., 2022). The adaptive immune evasion of SARS-CoV-2 is one of the main challenges for the control of disease dynamics on the long-term.

Nonpharmaceutical interventions like social and physical distancing proved to be highly valuable to control SARS-CoV-2 transmission at the population-level, notably during the pre-vaccine period (Flaxman et al., 2020; Zhang et al., 2020). The impact of nonpharmaceutical interventions was mainly assessed on aggregated data and rarely at the individual level, but it is possible that nonpharmaceutical interventions prevent transmission in the population when implemented by a critical mass of the population without conveying protection at the individual-level. In addition, the magnitude of the reduction of the infection risk when an individual isolates has been rarely quantified (Fazio et al., 2021). In Chapter 3, I account for physical distancing because household contacts were encouraged to isolate from the index case at enrollment, and not accounting for it would bias our estimates. Besides, social behavior, notably physical contacts within households, is expected to change upon infection by both the case and its contacts. Here, I estimate a strong effect of isolation behavior in Israeli households whose effectiveness against infection is similar to vaccines. Although these results were very promising, they need confirmation. Unfortunately, Omicron variants have completely replaced the Alpha variant by now, and since they are much more transmissible than Alpha, physical distancing in closed settings like households will likely be less effective.

Early on in the pandemic, age disparities in cases were observed. Most cases, especially severe cases, were adults suggesting that children are less susceptible to SARS-CoV-2 infection. Many studies have tried to quantify this potential reduction (Viner et al., 2021; Davies et al., 2020a; Zhu et al., 2021; Chung et al., 2021) by analyzing case notification or ascertainment data. However, lower notifications in children may result from multiple factors: (*i*) lower biological susceptibility, (*ii*) less social mixing,

and (*iii*) low case ascertainment of asymptomatic and paucisymptomatic cases that are more prevalent in children (Lordan et al., 2021). The last two factors may have confounding effects on the estimation of child susceptibility. Besides, the role of children in the transmission at the population-level has changed over the pandemic, in particular when adults were vaccinated. After the first vaccination campaign, a large part of the adult population was vaccinated, thus the proportion of transmission occurring in schools increased and living with a child constituted a risk factor for infection (Lessler et al., 2021). In our study presented in Chapter 3, I introduce a parameter for the relative susceptibility of children. Consistent with the literature, I estimate a lower susceptibility for children. To obtain more robust estimates of this parameter, it will be important to integrate to these analyses data describing contact rates between household members. Indeed, I show in Chapter 5 that ignoring contact heterogeneity in households could bias estimates of child relative susceptibility and infectivity. Some studies have also investigated the relative infectivity of children compared to adults (Dattner et al., 2021). In Chapter 3, I do not integrate child relative infectivity because there was not enough statistical power and I wanted to avoid model overparameterization.

Importantly, most results on SARS-CoV-2 transmission during the pandemic were published as preprints or peer-reviewed papers and were rapidly outdated due to the emergence of new variants and changes in priorities. Our estimates covered the Alpha wave in Israel and could not be generalized to the Delta variant that emerged a few months later at a time when vaccine-induced immunity had already decayed at the population level. A follow-up study covering the Delta wave in Israel is undergoing.

1.3. RABV and SARS-CoV-2: epidemiological research on different time-scales

The differing epidemiology of COVID-19 and rabies dictates the time-scale of their research. As underlined in Chapter 1, the transmission potential of rabies in dogs is low with an R_0 close to one. In contrast, the high transmissibility of COVID-19 leads to explosive dynamics, known as epidemic waves, which requires rapid evaluation within a few days or weeks. The short time-scale of SARS-CoV-2 research also depends on the recurrent emergence of variants with varying transmissibility. New variants should be identified and tracked in real-time, their transmission potential assessed in a timely manner, and their level of vaccine escape quantified when they replace the other variants. All of these steps may occur within a few months as exemplified by the Alpha variant studied in Chapter 3 that circulated in Israel for eight months only, from December 2020 to July 2021. RABV dynamics in relation to its genetic diversity are fairly different. As mentioned in Chapter 1, the co-circulation of RABV variants is com-

mon in rabies-affected countries and, so far, no difference of transmission potential between them has been shown. In this sense, rabies research spans more on the long-term. Another aspect of rabies and COVID-19 epidemiology that influences the time-scale of their research is the role of introductions in their maintenance or geographic expansion. While a few introductions of SARS-CoV-2 are sufficient to establish self-sustaining local transmission chains (Lemieux et al., 2021), transmission chains resulting from RABV introductions are short and should be numerous to allow RABV maintenance (Mancy et al., 2022). Rabies endemicity in Cambodia that I explore in Chapter 2 is a good example. Introductions from neighboring countries are very likely, but not sufficient in number to compete against the dominant Cambodian clade. Nevertheless, single introductions of rabies may lead in some cases like Bali to sustained transmission (Townsend et al., 2013b).

Beyond their contrasting epidemiology, COVID-19 and rabies have different research agendas and financial resources which also determines the type of data that are collected, the methodological choices, and the amount of scientific production. Rabies exemplifies the lack of detailed epidemiological data which is characteristic of neglected tropical diseases, epizooties in developing countries, and zoonoses. For such diseases, phylodynamics is a powerful approach as it requires little sampling efforts limiting the costs of data collection and generation. On the other side of the spectrum, studies on SARS-CoV-2 transmission using individual-level data are numerous. They take advantage of national health repositories or highly detailed follow-up in observational studies. Despite the large number of individual-level studies, uncertainties remain on the determinants of transmission, notably concerning partial immunity levels in populations following multiple waves of SARS-CoV-2 variants and more or less targeted vaccination campaigns.

Epidemiology, data availability, financial resources, and scientific opportunities are different for SARS-CoV-2 and RABV, but, as vaccine-preventable diseases, they share the same need for the design of effective vaccination campaigns. In the case of RABV, there is still room for improvement. Most endemic countries that have implemented regular vaccination campaigns are far from elimination and the design of national control strategies is restricted by financial constraints. For SARS-CoV-2, long-term transmission and recurrent emergence of new variants raises new challenges in terms of vaccination production, target population, and vaccination campaign frequency.

1.4. Limitations of quantitative tools for the study of infectious diseases spread

Although some of the quantitative tools have been developed a long time ago and are largely used, their limitations are not always well-characterized neither known by modelers. In this thesis, I also investigate the impact of sampling bias and model misspecification on parameter estimation in Chapters 4 and 5.

In Chapter 4, I focus on the impact of sampling bias on discrete phylogeography reconstruction and our conclusions are less clear-cut than previously thought (De Maio et al., 2015) because none of the three most popular algorithms outperforms the others nor is robust to sampling bias at the same time. CTMC modeling is a valid approach, even when spatial sampling is very biased, but it requires careful sensitivity analyses based on subsamples. Importantly, a recent study by Gascuel and Steel (2020) investigated identifiability issues in CTMC modeling, suggesting that further work is needed to precisely assess the limits of parameter identifiability in CTMC. The approximations of the structured coalescent on the other hand were previously thought to be more robust to spatial sampling bias (De Maio et al., 2015). However, in our study, they appear to have limited estimation capacities on real-world data. Interestingly, the addition of unbiased case count data largely improves the performances of one of the approximations of the structured coalescent model (MASCOT-GLM) that even outperforms CTMC. This work has many practical implications that go from study design to sensitivity analyses. Indeed, one can either organize genetic data collection in a way that is representative of the transmission process, prioritize sample size over representativeness to collect as many sequences as possible, perform sensitivity analyses on subsamples that maximize the spatiotemporal coverage, or analyze genetic data collected in an opportunistic way along with case count data that were collected in the least biased way. In the end, our work highlights the importance of testing and discussing the impact of sampling bias, notably on data sets obtained by opportunistic sampling.

In Chapter 5, I explore how assumptions on mixing patterns in household transmission studies impact the estimation of age-varying susceptibility and infectivity parameters. This study is particularly relevant for airborne diseases such as flu and COVID-19 whose transmission in confined spaces is primarily driven by droplets (Lei et al., 2018; Jayaweera et al., 2020). In such settings, successful transmission depends on case infectivity, contact susceptibility, and contact rate between the infector and the infectee. I show that not accounting for heterogeneous contact rates between household members leads to biased infectivity and susceptibility estimates. Our simulation study highlights the importance of collecting transmission-related data along with household contact data which, to the best of our knowledge, has

not been implemented yet. Designing such a study raises many practical challenges: should household follow-up rely rather on detailed testing and symptom onset data, or seroconversion data? Seroconversion data collection being less constraining for household members and less costly, but more challenging in terms of modeling. Should all household members be surveyed to collect their contacts or only a fraction of them to increase participant adherence? When should household members detail their contacts: at inclusion or during follow-up? How frequently should household members provide information on their contacts? etc. Importantly, SARS-CoV-2 is known to be transmitted not only by droplets during close contacts but also by aerosols (very small pathogen-laden particles in air that can stagnate when ventilation is low) that allow long-term transmission (Jayaweera et al., 2020). For the time being, models of SARS-CoV-2 transmission in households assume that droplets are the primary route of transmission and overlook the potential role of aerosols.

1.5. Comparison of epidemiological and phylodynamic modeling

In this thesis, I use epidemiological modeling on individual-level data of partially observed transmission chains and phylodynamics on genetic sequences from cases whose link between one another is unknown. The two types of approaches allow the study of different phenomena: the former models between-host transmission while the latter models evolutionary processes that shape pathogen genetic diversity. Depending on the rate of within-host evolution relative to the timing of transmission, a phylogenetic tree does not exactly represent a transmission chain, and the pathogen effective population size might not be simply proportional to case counts. Consequently, the estimates of pathogen dispersal by the two approaches may not be easily comparable. For example, the comparison of simulated transmission chains to estimated phylogenetic trees in Chapter 4 is not trivial. Since migration rates estimated by discrete phylogeography do not correspond to the ones of the mobility matrix used in the simulations, I compare counts of migration and introduction events. Actually, model benchmarking in phylodynamics is generally done on phylogenetic trees simulated with a coalescent or birth-death model (Gill et al., 2013; Boskova et al., 2014; De Maio et al., 2015; Müller et al., 2018), rather than alignments of sequences generated from the simulation of viral genetic sequences along transmission chains themselves simulated by an epidemiological model like in Chapter 4. This relatively new approach opens a new avenue for model benchmarking in phylodynamics and allows to test algorithms in more complex situations that are closer to the analysis of real-world data. Besides, the simulated epidemics of RABV decribed in Chapter 4 could be used to assess the performances of algorithms of maximum likelihood-based ancestral character reconstruction such as the CTMC models implemented in PastML (Ishikawa et al., 2019) or

TreeTime (Sagulenko et al., 2018).

In the field of epidemiological modeling, models are often very refined and adapted to a specific context. Until recently, source codes were not systematically in open access but a culture shift is currently operating, notably under the requirements of scientific journals to foster open science and analysis reproducibility. Besides, new tools such as Odin (FitzJohn and Fischer, 2022) or RStan (Stan Development Team, 2020) now facilitate inference even for complex epidemiological models. In contrast, phylodynamics relies on more complex and less flexible models and inference is generally performed using packages like BEAST. In addition to accelerating data analysis and result generation, the use of packages allows non-experts to apply complex methodologies to their own data sets but this comes at the expense of potential misuse or misinterpretation. Although the core hypotheses in phylodynamics are quite rigid, models of population dynamics (Gill et al., 2016) and phylogeographic models (Lemey et al., 2014; Müller et al., 2019) have been extended to integrate covariates related to demography, mobility, case counts, climate etc. Inversely, the use of genetic sequences to inform epidemiological models remains challenging. When it comes to individual-level data, epidemiological modeling is an ideal framework related because of its flexibility while accounting for individual heterogeneities in phylodynamics requires the development of new analytical tools.

Still, epidemiological modeling and phylodynamics share common features among which the inference framework. Indeed, Bayesian statistics are generally preferred over maximum likelihood inference, and the Metropolis-Hastings MCMC algorithm as exemplified in this thesis is the most widely used inference algorithm. Furthermore, both approaches can be affected by bias that can occur at the different stages of the analysis, from data collection and preparation, to model choice and specification, or prior choice. Simulation studies provide important insights to assess the impact of such biases and identify the limits of parameter identifiability. In this thesis, simulation studies have helped better characterizing the impact of sampling bias in discrete phylogeography (Chapter 4) and of incorrect mixing assumption (Chapter 5). The development of digital tools to collect, store, and share epidemiological and genetic data have paved the way to big data in epidemiology. This raises new technical challenges related to data storage, optimization of computational times, and inference approach, as Bayesian inference might be computationally impractical on very large data sets.

2. Perspectives

2.1. On SARS-CoV-2 and RABV

Many questions remain regarding RABV spread in dog populations. The role of dog demography, dog movement, dog social structure, dog ecology, and interactions with humans very likely depends on the cultural practices that vary across countries, but also within countries. For example, the epidemiology of rabies in the Northeastern part of Cambodia, close to the Lao and Vietnamese borders, is expected to be very different from that in the rest of the country because hunting practices, dog density, and relationship to dogs are extremely different (Chapter 2). Field studies to collect dog behavior and dog ecology data are key to better understand rabies dynamics and epidemiological data are still needed in most endemic countries. In parallel, RABV genetic data that are already available on GenBank could be exploited in large-scale phylogenetic studies to explore dog rabies transmission across countries in a cost-effective way. This would provide a first overview of the worldwide spread of the virus that should be completed by the above-mentioned finer-scale studies.

SARS-CoV-2 transmission is now well-characterized although there is room left for studies deepening knowledge on the impact of the mosaicism of population immunity on SARS-CoV-2 transmission. The immune profile of individuals against SARS-CoV-2 is now highly heterogeneous within and across countries because it depends on the biological features of the individuals, their personal infection history, and the types and number of vaccine doses they received that greatly varies across countries. On top of that, immunity against SARS-CoV-2 wanes with time from infection or vaccination. Consequently, there is now no clear reference group to estimate vaccination efficacy and vaccine effectiveness. Besides, epidemiological studies that are generally used to evaluate herd immunity at the population level such as cross-sectional serosurveys are not sufficient to assess the mosaicism of population immunity. The level of protection at the individual-level should not focus on individual infection and vaccination histories but on biological correlates like antibody titers. Assessing the quantitative relationship between protection and biological correlates and using this information to interprete serosurvey results will help assess herd immunity.

2.2. On methodological approaches

As demonstrated in this thesis, epidemiological modeling is particularly adapted to the integration of host-related determinants, like susceptibility, infectivity, and behavior. Nevertheless, all modeling frameworks are not suitable to account for complex host behaviors that vary in space and time. Network mod-

els constitute a good candidate and have already been used to design infection prevention measures in hospital settings using a simulation approach (Smith et al., 2020), using contact data recorded by using proximity sensors. This approach could be extended to households to measure more precisely heterogeneous contact patterns.

Contrary to epidemiological modeling, phylodynamics is still a relatively new field in which model limitations are still to be defined. While phylogeography is highly valuable, the impact of sampling bias cannot be overlooked and methodological advances that incorporate sampling procedure or uncertainty concerning sample representativeness are to be developed. Besides, theoretical work has shown that CTMC fails at estimating state change counts and root location at the same time (Gascuel and Steel, 2020). Similar questions related to parameter identifiability concern the structured coalescent in discrete phylogeography and the RRW model in continuous phylogeography. Additional data such as case counts for the structured coalescent model (Chapter 4), or sequence-free cases for the RRW model (Kalkauskas et al., 2021) were shown to improve phylogeography inference, but they lead to other sampling issues that also need attention. Another prospect for phylodynamics relates to its predictive power and its use to evaluate control measure which has been only little explored so far (Dellicour et al., 2020).

As previously underlined, epidemiological modeling and phylodynamics provide complementary insights on the transmission process. That is why, interdisciplinary studies that exploit the potential of both approaches have emerged, notably for RABV and SARS-CoV-2. We have identified so far two types of interdisciplinary studies that either propose a new unified modeling framework that integrates epidemiological and genetic data (Salje et al., 2021) - however, only few approaches currently exist and they require extensive model testing and validation -, or that combine multiple modeling approaches and sources of data. In the latter case, investigators benefit from epidemiological modeling, phylodynamics, and basic epidemiological investigation, but tremendous efforts of coordination and communication between the different fields of expertise might complexify scientific production.

3. Conclusion

In this thesis, I show that quantitative studies in epidemiology encompass a wide range of concepts, techniques, and data. At both ends of the spectrum, we find epidemiological modeling that can incorporate refined information on disease transmission at the individual level and phylodynamics that makes use of genetic sequences embedding coarse information on transmission at the population level. The complementary insights of these two approaches help understanding disease transmission and often guide

decision making, as shown in the context of SARS-CoV-2 and RABV. Ultimately, the development of modeling frameworks that unify epidemiological and genetic data is an exciting and promising prospect for epidemiology.

APPENDICES

Appendix A

Supplementary information on the scoping review

All supplementary information presented in this Appendix are also available online at https://mlayan.github.io/RabiesScopingReview/ and archived on the open-access repository Zenodo (DOI: 10.5281/zenodo.4743553).

1. Supplementary Tables

Table A1: General characteristics of the included studies. Studies 1 to 22 correspond to phylodynamic studies, studies 23 to 52 correspond to mathematical modeling studies, and studies 53 to 59 correspond to interdisciplinary studies.

#	Study context	New methodology	Study area	Temporal scale	Additional host species	Reference
1	Global dynamics	NO	World	1969 - 2004	Domestic animals, wildlife	Bourhy et al. 2008
2	Dynamics in an endemic area; Maintenance of viral epidemic cycles	YES	West Africa	1986 - 2007	-	Lemey et al. 2009a
3	Dynamics in an endemic area	NO	West Africa	1986 - 2007	-	Talbi et al. 2009
4	Dynamics in an endemic area	NO	China	1931 - 2009	-	Meng et al. 2011
5	Dynamics in an endemic area	NO	Ghana	1979 - 2009	Domestic animals, humans, wildlife	Hayman et al. 2011
6	Dynamics in an endemic area	NO	Brazil	1985 - 2006	-	Carnieli et al. 2011
7	Dynamics in an endemic area; Contribution of wildlife	NO	China	2003 - 2008	Domestic animals, wildlife	Yu et al. 2012
8	Incursion in a rabies-free area	NO	KwaZulu Natal Province, South Africa	1980 - 2011	Humans	Mollentze et al. 2013
9	Role of lineage incursions	NO	China	2003 - 2010	Domestic animals, humans, wildlife	Guo et al. 2013
10	Lineage dynamics in dogs and wildlife	NO	Brazil	2002 - 2005	Domestic animals, humans, wildlife	Carnieli et al. 2013
11	Dynamics in an endemic area	NO	Philippines	2004 - 2010	Domestic animals, wildlife	Tohma et al. 2014
12	Role of lineage incursions; Inference of clade reservoirs	NO	Middle East	1972 - 2014	Wildlife	Horton et al. 2015
13	Dynamics in an endemic area; Identification of factors associated with spread	NO	Tanzania	2003 - 2012	-	Brunker et al. 2015

Table A1 continued from previous page

14	Dynamics after incursion	NO	Indonesia	1997 - 2010	Domestic animals	Dibia et al. 2015
15	Dynamics in an endemic area; Identification of factors associated with spread	NO	China	1989 - 2012	Domestic animals, humans, wildlife	Yao et al. 2015
16	Global dynamics of dogs and wildlife lineages; Host shifting	NO	World	1950 - 2015	Domestic animals, humans, wildlife	Troupin et al. 2016
17	Role of lineage incursions; Causes of lineage displacement	NO	Yunnan	1963 - 2013	Domestic animals, humans, wildlife	Zhang et al. 2017
18	Dynamics after incursion	NO	Shaanxi province, China	2009 - 2012	Humans	Ma et al. 2017
19	Dynamics in an endemic area; Identification of factors associated with velocity	NO	North Africa	2001 - 2008	Domestic animals, wildlife	Dellicour et al. 2017
20	Identification of factors associated with velocity	NO	Tanzania	2004 - 2013	Domestic animals, wildlife	Brunker et al. 2018b
21	Dynamics in an endemic area	NO	China	1983 - 2016	Domestic animals, humans, wildlife	Wang et al. 2019
22	Dynamics in an endemic area; Identification of factors associated with dispersal velocity and direction	YES	Iran	2008 - 2015	Wildlife	Dellicour et al. 2019
23	Critical dog vaccination coverage	NO	Memhis & Shelby County, USA; Hermosillo, Mexico; Central Java, Indonesia; Kuala Lumpur, Malaysia	1948; 1987 - 1988; 1985 - 1986; 1946 - 1953	Dogs	Coleman and Dye 1996
24	New methodology	YES	São Paulo, Brazil	11 weeks	-	Ortega et al. 2000
25	Role of dog density; Efficacy of control strategies	NO	Machakos district, Kenya	2 years	-	Kitala et al. 2002
26	Synchrony of rabies epidemics	NO	East Africa	1971 - 2000	-	Hampson et al. 2007
27	Dog-human transmission; Cost-effectiveness study	NO	N'Djaména, Chad	2001 - 2006	Humans	Zinsstag et al. 2009
28	Efficacy of control strategies	NO	None[a]	2 years	-	Carroll et al. 2010
29	Dynamics in an endemic area; Efficacy of control strategies	NO	China	1996 - 2010	Humans	Zhang et al. 2011
30	Rabies control with multiple introduction sources	NO	Serengeti district, Tanzania	2002 - 2007	-	Beyer et al. 2011
31	Dynamics in an endemic area; Efficacy of control strategies	NO	Guangdong, China	2006 - 2014	Humans	Hou et al. 2012
32	Seasonal dynamics	NO	China	2004 - 2010	Humans	Zhang et al. 2012

Table A1 continued from previous page

33	Dynamics in dogs and wildlife; Efficacy of control strategies	NO	Serengeti and Ngorongoro districts, Tanzania	2002 - 2006	Wildlife	Fitzpatrick et al. 2012
34	Vaccination allocation strategies	NO	Serengeti district, Tanzania	132 months	-	Beyer et al. 2012
35	Role of introduction; Efficacy of control strategies	NO	Bali, Indonesia	2008 - 2010	-	Townsend et al. 2013b
36	Disease detection; Efficacy of control strategies	NO	None[a]	-	-	Townsend et al. 2013a
37	Risk assessment of introduction	NO	North Australia	End of outbreak	Humans	Dürr and Ward 2015
38	Dynamics; Efficacy of control strategies	NO	Region IV, Philippines	2010 - 2012	-	Ferguson et al. 2015
39	Dynamics in an endemic area; Dog-human transmission	NO	Hebei and Fujian provinces, China	2004 - 2012	Humans	Chen et al. 2015
40	Population structure; Efficacy of control strategies	NO	Northern Australia and New South Wales regions, Australia	300 or 800 days	-	Sparkes et al. 2016
41	Population structure; Efficacy of control strategies	NO	None[a]	125 years	-	Leung and Davis 2017
42	Individual heterogeneity; Efficacy of control strategies	NO	N'Djaména, Chad	-	-	Laager et al. 2018
43	Risk assessment of introduction; Efficacy of control strategies	NO	Ibaraki and Hokkaido prefectures, Japan	End of outbreak	-	Kadowaki et al. 2018
44	Drivers of resurgence	NO	N'Djaména, Chad	2012 - 2016	Humans	Laager et al. 2019
45	Contact heterogeneity	NO	Mayo-Kebbi Est region, Chad	2016	-	Wilson-Aggarwal et al. 2019
46	Cost-effectiveness study of human and cattle rabies	NO	Lemuna-Bilbilo and Bishoftu districts, Ethiopia	2013 - 2014	Cattle, humans	Beyene et al. 2019
47	Dynamics in an endemic area	NO	Sarawak state, Malaysia	2017 - 2019	Humans	Abdul Taib et al. 2019
48	Dynamics in dogs and wildlife; Dog-human transmission	NO	Zheijang province, China	2004 - 2017	Chinese ferret badger, humans	Huang et al. 2019
49	Contact heterogeneity; Efficacy of control strategies	NO	Northern Peninsula Area, Australia	End of outbreak	-	Hudson et al. 2019a
50	Dynamics in a disease-free area; Efficacy of control strategies	NO	Torres Strait islands, Australia	3 years	-	Brookes et al. 2019
51	Risk assessment of introduction; Efficacy of control strategies	YES	Northern Peninsula Area, Australia	-	Humans	Hudson et al. 2019b
52	Role of human movement	NO	Central African Republic	300 years	-	Colombi et al. 2020

Table A1 continued from previous page

53	Dynamics in an endemic area; Identification of factors associated with spread	NO	North Africa	2001 - 2008	Domestic animals, wildlife	Talbi et al. 2010
54	Dynamics in an endemic area; Unexhaustive sampling	NO	KwaZulu Natal Province, South Africa	2010 - 2011	-	Mollentze et al. 2014
55	Disease detection; Dynamics in an archipelago	NO	Tablas and Luzon Islands, Philippines	2004 - 2013	-	Tohma et al. 2016
56	Dynamics in an endemic area; Introduction	NO	Bangui, Central African Republic	1986 - 2012	-	Bourhy et al. 2016
57	Introduction; Efficacy of control strategies	NO	N'Djaména, Chad	2012 - 2015	Humans	Zinsstag et al. 2017
58	Dynamics in an endemic area; New methodology	YES	Bangui, Central African Republic	2003 - 2012	-	Cori et al. 2018
59	Dynamics in an endemic area; Identification of factors associated with spread	NO	Yunnan province, China	2008 - 2015	Humans, non-flying mammals	Tian et al. 2018

[a] Simulation study not grounded in a specific geographic area.

Table A2: Description of the phylogeographic models with an emphasis on data source and potential sources of bias.

Phylodynamics category	Model description	Software	Genetic sequences	Sequence length (pb)	Data source	Parameters inferred	Reference
Parsimony discrete phylogeography	Maximum parsimony	PAUP*	N (n = 130)	1335	Not informed	Nucleotide substitution rate, Migration rates, tMRCA	Bourhy et al. 2008
Bayesian discrete phylogeography	Symmetric DTA + BSSVS	BEAST 1	N (n = 101)	1335	Bourhy et al. 2008	Migration rates	Lemey et al. 2009a
Parsimony discrete phylogeography	Maximum parsimony	PAUP*	N (n = 231) and G (n = 34)	1335 and 1572	Passive	Nucleotide substitution rate, Migration rates, tMRCA	Talbi et al. 2009
Parsimony discrete phylogeography	DELTRAN and ACCTRAN optimization, and Monte Carlo simulations	MigraPhyla	N (n = 200)	1353	Passive and GenBank	Nucleotide substitution rate, Migration rates, tMRCA	Meng et al. 2011
Bayesian discrete phylogeography	DTA + BSSVS[a]	BEAST 1	N (n = 303)	405	Passive and GenBank	Migration rates, Root state, tMRCA	Hayman et al. 2011
Bayesian discrete phylogeography	Symmetric DTA + BSSVS	BEAST 1	N (n = 71)	1332	GenBank	Nucleotide substitution rate, Migration rates, tMRCA	Carnieli et al. 2011
Parsimony discrete phylogeography	DELTRAN optimization and Monte Carlo simulations	PAUP* and MigraPhyla	N (n = 210)	720	Active and Passive	Nucleotide substitution rate, Migration rates, tMRCA	Yu et al. 2012
Bayesian discrete phylogeography	Asymmetric DTA + BSSVS	BEAST 1	intergenic G-L (n = 635)	590	GenBank	Migration rates, Root state, tMRCA	Mollentze et al. 2013
Bayesian discrete phylogeography	Symmetric DTA + BSSVS	BEAST 1	N (n = 232)	1353	Active and Passive	Nucleotide substitution rate, Migration rates, tMRCA	Guo et al. 2013
Bayesian continuous phylogeography	Relaxed random walk (lognormal prior)	BEAST 1	N (n = 53)	1388	Passive and GenBank	Nucleotide substitution rate, Lineage velocity, tMRCA	Carnieli et al. 2013
Bayesian discrete phylogeography	Symmetric DTA + BSSVS	BEAST 1	G (n = 263)	1572	Passive	Nucleotide substitution rate, Markov jumps, Migration rates, tMRCA	Tohma et al. 2014
Bayesian discrete phylogeography	Symmetric DTA + BSSVS	BEAST 1	N (n = 139)	400	Passive and Systematic	Markov jumps, Migration rates, Root state, tMRCA	Horton et al. 2015

Table A2 continued from previous page

Bayesian discrete phylogeography	Asymmetric DTA + BSSVS	BEAST 1	partial N gene (n = 430), N (n = 100), and whole genome (n = 58)	405, 1350, and -	Active and Passive	Nucleotide substitution rate, Markov jumps, Migration rates, Population dynamics	Brunker et al. 2015
Bayesian discrete phylogeography	Symmetric DTA + BSSVS	BEAST 1	N (n = 62)	534	Passive	Migration rates, Root state	Dibia et al. 2015
Bayesian discrete phylogeography	Symmetric DTA + BSSVS	BEAST 1	N (n = 203)	1353	GenBank	Migration rates	Yao et al. 2015
Bayesian phylodynamics	Coalescent models	BEAST 1	Whole genome (n = 248)	-	GenBank; WHO Collaborative Center for Reference and Research on Rabies; French National Reference Centre for Rabies	Nucleotide substitution rate, tMRCA	Troupin et al. 2016
Bayesian discrete phylogeography	Asymmetric DTA + BSSVS	BEAST 1	N (n = 401)	1350	Active and Passive	Nucleotide substitution rate, Markov jumps, Migration rates	Zhang et al. 2017
Bayesian discrete phylogeography	Asymmetric DTA + BSSVS	BEAST 1	N (n = 20) and G (n = 16)	1353 and 1575	Passive	Nucleotide substitution rate, Migration rates, tMRCA	Ma et al. 2017
Bayesian continuous phylogeography	Relaxed random walk (lognormal prior)	BEAST 1	N, P, intergenic G-L (n = 250)	3080	Talbi et al. 2009	Velocity, distance and duration of movement; Velocity-associated landscape features	Dellicour et al. 2017
Bayesian discrete phylogeography and Bayesian continuous phylogeography	Asymmetric GLM + BSSVS and Relaxed random walk (gamma prior)	BEAST 1	Whole genome (n = 152)	-	Brunker et al. 2015	Nucleotide substitution rate, Markov jumps, Migration rates; Velocity, distance and duration of movement; Velocity-associated landscape features	Brunker et al. 2018b
Bayesian discrete phylogeography	Symmetric DTA + BSSVS	BEAST 1	N (n = 112)	1353	GenBank	Nucleotide substitution rate, Migration rates, Population dynamics, tMRCA	Wang et al. 2019

Table A2 continued from previous page

Bayesian discrete phylogeography and Bayesian continuous phylogeography	Symmetric DTA + BSSVS and Relaxed random walk (not indicated)	BEAST 1	Whole genome (n = 101)	-	Passive	Nucleotide substitution rate; Direction of spread; Velocity, distance and duration of movement; Velocity- and direction-associated landscape features	Dellicour et al. 2019

Abbrevations: ACCTRAN, Accelerated Transformation; BSSVS, Bayesian Stochastic Search Variable Selection; DELTRAN, Delayed Transformation; DTA, Discrete Trait Analysis; GLM, Generalized Linear Model; tMRCA, time to the Most Recent Common Ancestor; WHO World Health Organization.

[a] Symmetry not indicated in the original publication.

Table A3: Description of the mathematical models with their key quantitative results.

Model category	Original model reference	Spatially explicit	Inference method	Dog population structure	Quantitative results	Type of sensitivity analysis	Outcome in sensitivity analysis	Key parameters in sensitivity analysis	Optimal control strategy	Reference
Deterministic compartmental	Anderson et al., 1991	NO	Regression	-	R, critical vaccination coverage	-	-	-	-	Coleman and Dye 1996
Fuzzy compartmental	-	NO	Simulation	-	-	-	-	-	-	Ortega et al. 2000
Deterministic compartmental	Anderson et al., 1980	NO	Simulation	-	R	-	-	-	Annual dog vaccination with 70% coverage or biannual dog vaccination with a 60% coverage	Kitala et al. 2002
Deterministic compartmental	-	NO	Simulation	-	Characteristic period and damping time of oscillations	Simulation under a range model parameter value with a specific focus on the uncertainty in the variance in rates of vaccine delivery	Rabies dynamics	Robust	-	Hampson et al. 2007
Deterministic compartmental	Hampson et al., 2017; Kitala et al., 2002; Coleman et al. 2004	NO	Maximum likelihood	-	Re, weekly numbers of rabid-dog cases for the whole city, human exposures, dog-to-dog transmission rate, dog-to-human transmission rate, number of exposed dogs	Simulation under a range model parameter value	Rabies dynamics	Rabies-related mortality rate, dog-to-dog transmission rate, probability of clinical outcome ++, carrying capacity, initial number of dogs per km^2 +	70% dog vaccination coverage	Zinsstag et al. 2009

Table A3 continued from previous page

Deterministic compartmental	Anderson et al., 1981; Smith & Cheeseman, 2002; Hampson et al., 2007	NO	Simulation	-	Time to eradication	Simulation under a range model parameter value	Rabies eradication	Robust when vaccination and fertility control are implemented	≥50% dog vaccination and fertility control coverage	Carroll et al. 2010
Deterministic compartmental	-	NO	Simulation	-	R	Simulation under a range of R and the initial number of infected dogs	R	Dog-to-dog-transmission rate, birth rate, vaccination coverage	Dog population management and dog vaccination	Zhang et al. 2011
Stochastic metapopulation	-	YES	Simulation	-	-	-	-	-	-	Beyer et al. 2011
Deterministic compartmental	-	NO	Simulation	Domestic, stray	R	Simulation under a range model parameter value	R	Rate at which rabid domestic dogs become rabid stray dogs	75% dog vaccination coverage and stray dog management	Hou et al. 2012
Deterministic compartmental	Zhang et al., 2011	NO	Least-square fitting	-	R	Simulation under a range model parameter value	R	Vaccination coverage, birth rate, baseline contact rate	Dog population management, dog vaccination and public awareness	Zhang et al. 2012
Stochastic compartmental	-	NO	MCMC	-	R, transmission rates, critical level of vaccine coverage	-	-	-	70% dog vaccination coverage	Fitzpatrick et al. 2012
Stochastic metapopulation	Beyer et al., 2011	YES	Simulation	-	Allocation strategies efficacy	Simulation when considering no heterogeneity in transmission between adjacent villages	Rabies dynamics	Robust to distances among villages	Frequent dog vaccination campaigns targeting the reduction in metapopulation risk	Beyer et al. 2012
Stochastic agent-based	-	NO	Maximum likelihood	-	R, vaccination coverage	Simulation under a range model parameter value	Rabies eradication	R, vaccination coverage	Dog vaccination coverage targeting even and 70% coverage	Townsend et al. 2013b

Table A3 continued from previous page

Stochastic branching process	-	NO	Simulation	-	-	-	-	-	Reactive dog vaccination with rapid surveillance system followed by a 2-year monitoring period	Townsend et al. 2013a
Stochastic agent-based	-	YES	Simulation	-	R	Simulations under different control strategies	Outbreak duration, number of rabid dogs	Incubation period, transmission probability given a bite, distance jernel, bite probability given a contact, vaccine efficacy, index community, delay in starting the control strategy of movement restrictions between communities	Reactive dog vaccination until all targeted dogs are vaccinated	Dürr and Ward 2015
Stochastic agent-based	-	NO	Simulation	-	Rc, time to elimination	Simulations under different distributions of the number of offspring cases resulting from each rabies case and the generation interval	Rabies elimination	R, number of offspring cases, mean generation interval	Dog vaccination targeting even coverage	Ferguson et al. 2015
Deterministic metapopulation	Zhang et al., 2011	YES	Simulation	-	Two-patch R, isolated R	Simulation under a range model parameter value	R	Mobility rate	-	Chen et al. 2015
Deterministic compartmental	-	NO	Simulation	Island and peri-urban free-roaming, wild	Time to detection	Simulation under a range model parameter value	Rabies dynamics	Contact rate, incubation period, transmission rate	90% dog vaccination coverage	Sparkes et al. 2016

Table A3 continued from previous page

Deterministic compartmental	-	NO	Simulation	Stray, owned free-roaming, owned confined	-	Theoretical critical proportion of stray dog vaccination coverage required to prevent outbreaks as a function of the proportion of vaccinated owned dogs	Rabies elimination	Dog population size and composition, vaccination coverage per dog subpopulation	Stray dog vaccination coverage should be based on dog population composition	Leung and Davis 2017
Stochastic agent-based	-	YES	Simulation	-	-	Simulation under a range of detection probability	Outbreak size, duration and probability	Network construction parameter, infectious period, incubation period, vaccination coverage	Targeted dog vaccination based on their social and roaming behaviors, public awareness, locally reactive interventions and reporting of 60% of cases by the surveillance system	Laager et al. 2018
Stochastic agent-based	Townsend et al., 2013	YES	Simulation	Stray, confined	Parameter of the influence of dog population density on secondary case locations, mean outbreak sizes under 50% vaccination coverage, mean epidemic durations	Simulation under a range model parameter value for R, the vaccination coverage, the unintended release of a rabid dog by their owner, and the preferential contact by rabid stray dogs	Outbreak size	Probability of unintentional release of rabid dogs by their owner, vaccination coverage, probability that a rabid stray dog selects a stray dog to bite	Dog owner awareness besides dog registration, capture of free-roaming dogs, mandatory dog vaccination and quarantine of imported animals	Kadowaki et al. 2018

Table A3 continued from previous page

Deterministic metapopulation	Zinsstag et al., 2009; Zinsstag et al., 2017	YES	Least-square fitting	-	Time to elimination, averted cases	Estimation of the transmission rate for different model specification (homogeneous mixing, metapopulation model, with importation, without importation)	Outbreak size	Rabies-related mortality rate, incubation period	-	Laager et al. 2019
Stochastic network	Epimodel	YES	Simulation	-	Network metrics	-	-	-	-	Wilson-Aggarwal et al. 2019
Stochastic compartmental	Fitzpatrick et al., 2014	NO	MCMC	-	Carrying capacities, dog-to-dog transmission rates, dog-to-human bite rate, incubation period, infectious period, probability of human developing rabies following exposure and without treatment, post-exposure prophylaxis coverage, annual number of rabid dogs, vaccination coverage	-	Cost-effectiveness sensitivity analysis	-	90% dog vaccination coverage	Beyene et al. 2019
Deterministic compartmental	Zhang et al., 2011	NO	Simulation	-	-	Simulation under a range model parameter value	Rabies dynamics	Vaccination coverage, birth rate	$\geq$50% dog vaccination coverage	Abdul Taib et al. 2019
Deterministic compartmental	-	NO	Least-square fitting	-	R	-	-	-	-	Huang et al. 2019

Table A3 continued from previous page

Stochastic agent-based	Dürr and Ward, 2015; Hudson et al., 2019	YES	Simulation	Explorer, roamer, stay-at-home	-	-	-	-	Dog vaccination targeting explorer and roamer dogs	Hudson et al. 2019a
Stochastic agent-based	-	NO	Simulation	-	R in the first month	Simulation under a range model parameter value	Outbreak duration and size, Re	Dog population size, degree of connectivity, incubation period, clinical periods	90% dog vaccination coverage	Brookes et al. 2019
Stochastic agent-based	Dürr and Ward 2015	YES	Simulation	-	Outbreak duration, number of rabid dogs	-	-	-	-	Hudson et al. 2019b
Stochastic metapopulation	-	YES	Simulation	-	Persistence probability, endemic prevalence per patch, R, birth rate	-	-	-	Dog movement bans and massive dog vaccination campaigns in urban areas	Colombi et al. 2020

Abbreviations: MCMC, Markov Chain Monte Carlo; R, reproduction ratio; Rc, constrained reproduction number.

Table A4: Description of the interdisciplinary studies combining phylodynamics and mathematical modeling or integrating epidemiological and genetic data. The first part of the table describes the genetic data and phylodynamic approach of the interdisciplinary studies, while the second part of the table describes the epidemiological data and the epidemiological model.

Phylodynamics category	**Model description**	**Software**	**Genetic sequences**	**Sequence length (pb)**	**Data source**	**Parameters inferred**	**Reference**
Bayesian discrete phylogeography	Asymmetric DTA + BSSVS	BEAST 1	N, P, intergenic G-L (n = 250)	3080	Passive	Markov jumps, migration rates, reward-associated distances, spread-associated landscape features, tMRCA	Talbi et al. 2010
Bayesian phylodynamics	Exponential growth coalescent	BEAST 1	intergenic G-L (n = 176)	760	Passive	tMRCA	Mollentze et al. 2014
Bayesian discrete phylogeography	Symmetric DTA + BSSVS	BEAST 1	G, P (n = 39)	2463	Passive	Migration rates, tMRCA	Tohma et al. 2016
Bayesian continuous phylogeography	Relaxed random walk (lognormal prior)	BEAST 1	N, P, M, G, intergenic G-L (n = 88)	5061	Passive	Nucleotide substitution rate, number of introductions, tMRCA, velocity	Bourhy et al. 2016
Bayesian phylodynamics	Birth-Death	BEAST 2	N (n = 29)	1350	Passive	Re	Zinsstag et al. 2017
-	-	-	N, P, M, G, intergenic G-L (n = 151)	5061	Passive	-	Cori et al. 2018
Bayesian discrete phylogeography	DTA[a]and GLM on the effective population sizes	BEAST 1	N (n = 543)	1350	Passive	tMRCA, velocity, diffusion coefficients, velocity-associated landscape features	Tian et al. 2018
Bayesian continuous phylogeography	Relaxed random walk (lognormal prior)		G (n = 491)	1575			

Table A4 continued from previous page

Model category	**Original model reference**	**Spatially explicit**	**Inference method**	**Dog population structure**	**Quantitative results**	**Type of sensitivity analysis**	**Outcome in sensitivity analysis**	**Key parameters in sensitivity analysis**	**Optimal control strategy**	**Reference**
Spatial spread along phylogenetic trees	-	YES	Simulation	-	Spatial distribution of cases	-	-	-	-	Talbi et al. 2010
Directed Acyclic Graph	Morelli et al., 2012	YES	MCMC	-	Distance between direct and indirect cases, number of cases, incubation period, infectious period, strength of exogeneous and individual sources,	Parameter estimation from simulated outbreaks under a range of detection probabilities	Transmission events	Robust to when detection probability is above 0.5	-	Mollentze et al. 2014
Deterministic compartmental	Hampson et al., 2007; Kitala et al., 2002; Panjeti and Real, 2011	NO	Simulation	-	Simulations under different values of the birth and mortality rates, the dog detection probability, different patterns of vaccine immunity, and under a strategy of pulse vaccination	Time to detection	Time to detection	Reporting rate	-	Tohma et al. 2016
Re	-	NO	particle MCMC	-	Number and rate of introductions, outbreak periodicity, Re	Estimation of the introduction rate of rabid cases in Bangui and R(t) under a range of detection probability (10%-50%)	Temporal variation of Re	Robust to detection probability	-	Bourhy et al. 2016

Table A4 continued from previous page

Deterministic compartmental and stochastic compartmental	Zinsstag et al., 2009	NO	Least-square fitting	-	Vaccination rate, dog-to-dog transmission rate, dog-to-human transmission rate, R_0	Simulation of rabies epidemics under a range of detection probability and other model parameters	Rabies dynamics	Robust to detection probability	-	Zinsstag et al. 2017
Graph model	-	YES	Graph pruning and intersection	-	R, rate of importation, rate of unobserved importation	Cluster reconstruction of the rabies outbreak under a range of detection probability and cutoff value at the graph pruning step	Clusters and R	Reporting rate, graph pruning cutoff	-	Cori et al. 2018
Deterministic compartmental	Anderson et al., 1981; Hampson et al., 2007; Zinsstag et al., 2009	NO	MCMC	-	Carrying capacity, dog- to-dog transmission rate, dog-to-human transmission rate, R_0	Simulation of rabies epidemics under 1,000 parameter sets drawn from prior distributions	Rabies dynamics	Transmission rates, initial conditions	-	Tian et al. 2018

Abbreviations: BSSVS, Bayesian Stochastic Search Variable Selection; DTA, Discrete Trait Analysis; GLM, Generalized Linear Model; MCMC, Markov Chain Monte Carlo; R0, basic reproduction ratio; Re, effective reproduction ratio; tMRCA, time to the Most Recent Common Ancestor.

[a] Symmetry not indicated in the original publication.

Table A6: Detailed list of the estimated parameters in phylodynamic models.

Viral lineages	Sequence	Nucleotide substitution rate (subs.site^{-1}.year^{-1}, 95% HPD)	Migration rate (migrations. year^{-1}, 95% HPD)	Velocity[a] (km.year^{-1}, 95% HPD)	Diffusion coefficient (km^2.year^{-1}, 95% HPD)	Factors facilitating spread	Factors preventing spread	Reference
Global	N G	2.3e-4 (1.1e-4 - 3.6e-4) 3.9e-4 (1.2e-4 - 6.5e-4)	-	-	-	-	-	Bourhy et al. 2008
Africa 2	N	3.82e-4 (2.62e-4 - 5.02e-4)	-	-	-	-	-	Talbi et al. 2009
Africa 1	G N, P, intergenic G-L	3.25e-4 (2.22e-4 - 4.32e-4) -	-	Algeria $v_{Great\ circle\ distances}$ = 26 (18 - 34) $v_{Road\ distances}$ = 33 (23 - 43) Morocco $v_{Great\ circle\ distances}$ = 42 (26 - 58) $v_{Road\ distances}$ = 51 (34 - 72)	-	Roads	-	Talbi et al. 2010
Arctic, Cosmopolitan, Indian, SEA-1, SEA-2, SEA-3	N	4.316e-4 (3.11e-4 - 5.632e-4)	-	-	-	-	-	Meng et al. 2011
-	N	1.519e-3[c]	-	-	-	-	-	Carnieli et al. 2011
Asian	N	clade I: 1.274e-3 (8.3705e-4 - 1.2515e-3) clade II: 9.629e-4 (3.519e-4 - 1.628e-3)	-	-	-	-	-	Yu et al. 2012
Cosmopolitan, SEA-1, SEA-2, SEA-3	N	5.23e-4 (3.94e-4 - 6.68e-4)	-	-	-	-	-	Guo et al. 2013

Table A6 continued from previous page

-	N	6.81e-4 (6.66e-4 - 6.96e-4)	-	$v_{Overall}$ = 12.88[b] v_{Dogs} = 30.5[b] $v_{Cerdocyon\,thous}$ = 9.0[b]	-	-	-	Carnieli et al. 2013
Asian 2	G	5.81e-4 (4.47e-4 - 7.27e-4)	-	-	-	-	-	Tohma et al. 2014
Africa 1[b]	whole-genome	1.44e-4 (5.78e-7 - 3.19e-4)	-	-	-	-	-	Brunker et al. 2015
Global	N G	2.18e-4 (1.68 - 2.71)[d] 3.20e-4 (2.60 - 3.80)[d]	-	-	-	-	-	Troupin et al. 2016
Africa 1, Africa 2	N, P, M, G, intergenic G-L	5.9e-4 (4.4e-4 - 7.5e-4)	-	v = 0.9 (0.65 - 1.2)	-	-	-	Bourhy et al. 2016
Arctic, Cosmopolitan, Indian, SEA-1, SEA-2, SEA-3	N	1.88e-4 (1.37e-4 - 2.41e-4)	-	-	-	-	-	Zhang et al. 2017
-	N G	2.4e-4 (1.5e-4 - 3.2e-4) 3.4e-4 (2.7e-4 - 4.1e-4)	-	-	-	-	-	Ma et al. 2017
Africa 1	N, P, intergenic G-L	-	-	$v_{wavefront}$ 15 - 22[b]	D = 2874 (1900 - 5420) $D_{weighted}$ = 1305 (1086 - 1574)	Grasslands, Urban areas, Human population density	Inacessibility, Elevation	Dellicour et al. 2017
Africa 1[b]	whole-genome	2.67e-4[b]	-	v = 4.46 (3.22 - 5.88) Coefficient of variation M = 3.10	-	Dog presence	Elevation Rivers	Brunker et al. 2018b
SEA-1, SEA-2, SEA-3	N, G	-	-	v = 57.5 (39.2 - 85.1)	D = 1733 (1082 - 2928)	Forest coverage	-	Tian et al. 2018

Table A6 continued from previous page

				$v_{weighted}$ = 23.4 (2.4 - 32.6)	$D_{weighted}$ = 1064 (116 - 1638)	Spread towards forest coverage, croplands, accessible areas		
Arctic-like 2, Central Asian 1, SEA-1, SEA-2, SEA-3, SEA-5	N	2.848e-4 (2.292e-4 - 3.449e-4)	5.81e-3 (3.92e-3 - 7.77e-3)	-	-	-	-	Wang et al. 2019
-	whole-genome	2.13e-4 (1.88e-4 - 2.35e-4)	-	v = 55.5 (38.9 - 142.4) $v_{weighted}$ = 18.1 (16.3 - 20.8)	D = 2676 (1935 - 5066) $D_{weighted}$ = 1643 (1356 - 2325)	Accessible areas Spread towards accessible areas associated with relatively high human population density	Barren vegetation areas Spread towards grasslands	Dellicour et al. 2019

[a] Depending on the study, estimates of RABV velocity were obtained with different methodologies.
[b] 95% HPD intervals are not specified in the original papers
[c] 95% HPD upper value
[d] Median with the 95% HPD interval
Abbreviations: SEA-1, South-East Asia 1; SEA-2, South-East Asia 2; SEA-3, South-East Asia 3.

Table A7: Detailed list of the estimated parameters in mathematical models.

Study category	**Location**	**Reproduction ratio**	**Transmission rates (95% CI)**	**Rabies-related parameters (95% CI)**	**Vaccination coverage (95% upper CL or 95% CI)**	**Dog ecology parameters - Carrying capacity K (dogs.km^{-2}, 95% CI) - Birth rate b (year^{-1}, 95% CI)**	**Reference**
Mathematical modeling	Memphis and Shelby County, USA	R = 2.334 (1.778 - 2.89, S.E.)	-	-	Critical vaccination coverage: 57.1% (71%)	-	Coleman and Dye 1996
Mathematical modeling	Hermosillo, Mexico	R = 1.981 (1.563 - 2.399, 95% S.E.)	-	-	Critical vaccination coverage: 49.5% (64.5%)	-	Coleman and Dye 1996
Mathematical modeling	Central Java, Indonesia	R = 1.789 (1.338 - 2.24, 95% S.E.)	-	-	Critical vaccination coverage: 44.1% (62.8%)	-	Coleman and Dye 1996
Mathematical modeling	Kuala Lumpur, Malaysia	R = 1.627 (1.325 - 2.091, 95% S.E.)	-	-	Critical vaccination coverage: 38.5% (55.2%)	-	Coleman and Dye 1996
Mathematical modeling	Machakos, Kenya	R = 2.44 (1.52 - 3.36, 95% CI)	-	-	-	-	Kitala et al. 2002
Mathematical modeling	N'Djaména, Chad	R = 1.01	dog-to-dog transmission contact-rate (km^2.dogs^{-1}.week^{-1}, 95% CI): 0.0807 (0.0804 - 0.0809) dog-to-human transmission contact-rate (km^2.dogs^{-1}.week^{-1}, 95% CI): 0.0002 (0.00017 - 0.00024)	-	-	-	Zinsstag et al. 2009
Mathematical modeling	China	R = 2[a]	-	-	-	-	Zhang et al. 2011
Mathematical modeling	Guangdong, China	R = 1.65[a]	-	-	-	-	Hou et al. 2012
Mathematical modeling	China	R = 1.03[a]	-	-	-	-	Zhang et al. 2012

Table A7 continued from previous page

Mathematical modeling	Ngorongoro, Tanzania	R = 1.24	Dog-to-dog transmission rate (expected number of secondary cases, 95% CI): 1.16 (0.85 - 1.54)	-	Annual dog vaccination coverage required: 39% (67%)	-	Fitzpatrick et al. 2012
Mathematical modeling	Seregenti, Tanzania	R = 1.18	Dog-to-dog transmission rate (expected number of secondary cases, 95% CI): 1.09 (0.98 - 1.21)	-	Annual dog vaccination coverage required: 30% (42%)	-	Fitzpatrick et al. 2012
Mathematical modeling	Australia	R from 0 to 6.1 (median 1.8) for reactive vaccination[b] R from 0 to 6.1 (median 1.7) for reactive culling[b] R from 0 to 5.7 (median 1.7) for movement bans between communities[b]	-	-	-	-	Dürr and Ward 2015
Mathematical modeling	Hebei and Fujian provinces, China	R = 1.0319 (Hebei: 0.5477, Fujian: 0.8197)[a]	-	-	-	-	Chen et al. 2015
Mathematical modeling	Guizhou and Guangxi provinces, China	R = 4.9211 (Guizhou: 1.5998, Guangxi: 6.1905)[a]	-	-	-	-	Chen et al. 2015
Mathematical modeling	Sichuan and Shaanxi provinces, China	R = 1.5085 (Sichuan: 1.3414, Shaanxi: 1.0061)[a]	-	-	-	-	Chen et al. 2015
Interdisciplinary	Bangui, Central African Republic	R median monthly point estimates in the range 0.8 to 1.3	-	Introduction rate (importation.week^{-1}, median, 95% CI): 0.13 (0.07 - 0.27)	-	-	Bourhy et al. 2016
Interdisciplinary	N'Djaména, Chad	R = 1.14	Dog-to-dog transmission rate (km^2.dogs^{-1}.week^{-1}): 0.0292	-	Background dog vaccination rate (week^{-1}): 2.96e-3	-	Zinsstag et al. 2017

Table A7 continued from previous page

			Dog-to-human transmission rate (transmission. $dogs^{-1}.week^{-1}$): 2.34e-5				
Interdisciplinary	Bangui, Central African Republic	R = 0.92 (95% CI, 0.85 - 1.01)	-	Rate of importation (importation.$year^{-1}$, 95% CI): 4.91 (3.60 - 6.65) Rate of unobserved importation (importation.$year^{-1}$, 95% CI): 2.40 (1.09 - 4.14)	-	-	Cori et al. 2018
Interdisciplinary	Yunnan, China	R = 1.05	Dog-to-dog transmission rate (transmission.$year^{-1}$, 95% CI): 6.62 (5.60 - 7.64) Dog-to-human transmission rate (transmission.$year^{-1}$, 95% CI): 0.0004 (0.0002 - 0.0006)	-	-	Carrying capacity (dogs.km^{-2}, 95% CI): 13.59 (7.80 - 19.37)	Tian et al. 2018
Mathematical modeling	Bishoftu, Ethiopia	-	Dog-to-dog transmission (transmission. $dogs^{-1}.day^{-1}$, 95% CI): 0.46 (0.39-0.47)	Incubation period (days, 95% CI): 22.5 (19.9 - 24.6) Infectious period (days, 95% CI): 3.13 (2.85 - 3.4) Rabid dog-human bite rate (transmission.$dogs^{-1}$, 95% CI): 0.5 (0.45 - 0.56) Probability of developing rabies (95% CI): 0.16 (0.15 - 0.17)	PEP coverage: 0.8 (0.74-0.84)	Carrying capacity (dogs.km^{-2}, 95% CI): 361 (346 - 367)	Beyene et al. 2019

Table A7 continued from previous page

Mathematical modeling	Lemuna-bilbilo, Ethiopia	-	Dog-to-dog transmission (transmission.dogs^{-1}.day^{-1}, 95% CI): 0.40 (0.37-0.44)	Incubation period (days, 95% CI): 21.6 (19.8 - 24.8)	PEP coverage: 0.59 (0.45-0.59)	Carrying capacity (dogs.km^{-2}, 95% CI): 2.6 (41.8 - 45.2)	Beyene et al. 2019
Mathematical modeling	Zheijiang, China	R = 1.0114[a]	-	Infectious period (days, 95% CI): 3.08 (2.88 - 3.35) Rabid dog-human bite rate (transmission.dogs^{-1}, 95% CI): 0.56 (0.46 - 0.57) Probability of developing rabies (95% CI): 0.16 (0.15 - 0.17)	-	-	Huang et al. 2019
Mathematical modeling	Kubin community, Australia	R (1st month) = 2.50 (95% range 1.0 - 7.0) in Kubin community[b]	-	-	-	-	Brookes et al. 2019
Mathematical modeling	Saibai community, Australia	R (1st month) = 1.73 (95% range 0 - 6.0) in Saibai community[b]	-	-	-	-	Brookes et al. 2019
Mathematical modeling	Warraber community, Australia	R (1st month) = 3.23 (95% range 1.0 - 8.0) in Warraber community[b]	-	-	-	-	Brookes et al. 2019
Mathematical modeling	Central African Republic	R = 1.03 (95% CI: 1.02 - 1.04)	-	-	-	Annual birth rate (95% CI): 1.59 (1.19 - 1.99)	Colombi et al. 2020

Abbreviations: CI, Confidence Interval; CL, Confidence Limit; R, reproduction ratio; PEP, Post-Exposure Prophylaxis.

[a] Based on the choice of parameter values

[b] Simulation study

2. Rabies epidemiological situation and methodologies implemented to study rabies dispersal and control at the continent level

The situation of North America is not detailed since the study of Coleman and Dye (1996) only estimated the critical vaccination coverage and R from an epidemic in the Tennessee in the 1940s and an epidemic in Mexico in the 1980s.

2.1. Africa

Current situation

Endemic

Models

- Large variety with development of multi-host, metapopulation and network models (Laager et al., 2019; Wilson-Aggarwal et al., 2019; Zinsstag et al., 2009; Kitala et al., 2002; Beyene et al., 2019; Colombi et al., 2020; Beyer et al., 2011; Fitzpatrick et al., 2012; Townsend et al., 2013a; Hampson et al., 2007; Beyer et al., 2012; Laager et al., 2018).
- Parsimony, Bayesian discrete and continuous phylogeography (Lemey et al., 2009a; Mollentze et al., 2013; Brunker et al., 2015; Talbi et al., 2009; Hayman et al., 2011; Brunker et al., 2018b; Dellicour et al., 2017).
- Interdisciplinary studies (Mollentze et al., 2014; Cori et al., 2018; Talbi et al., 2010; Bourhy et al., 2016; Zinsstag et al., 2017).

Data

Bite incidence in humans, dog, and human rabies incidence, contact tracing, dog mobility data, dog census, RABV genetic sequences from dogs, wildlife, and humans.

Modelling aims

- Better understanding of the spatial and temporal dynamics of rabies spread (Zinsstag et al., 2009; Hampson et al., 2007; Brunker et al., 2015; Talbi et al., 2009; Hayman et al., 2011; Talbi et al., 2010; Bourhy et al., 2016).
- Identification of environmental factors impacting RABV spread and the main patterns of dispersal (Colombi et al., 2020; Mollentze et al., 2013; Brunker et al., 2015; Brunker et al., 2018b; Dellicour et al., 2017; Talbi et al., 2010).
- Role of introductions (Mollentze et al., 2013; Bourhy et al., 2016; Zinsstag et al., 2017), spatial heterogeneity (Laager et al., 2019; Colombi et al., 2020; Beyer et al., 2011), dog population structure (Wilson-Aggarwal et al., 2019; Laager et al., 2018) and wildlife (Beyer et al., 2011; Fitzpatrick et al., 2012) in dog rabies maintenance.
- Feasible and effective control strategies (Zinsstag et al., 2009; Kitala et al., 2002; Beyene et al., 2019; Beyer et al., 2011; Fitzpatrick et al., 2012; Townsend et al., 2013a; Beyer et al., 2012; Laager et al., 2018; Zinsstag et al., 2017).

- Development of new methodologies (Mollentze et al., 2014; Cori et al., 2018). Impact of underreporting (Mollentze et al., 2014; Zinsstag et al., 2017; Laager et al., 2019).

Mains findings

Rabies spread was studied at multiple geographical scales (transborder area, country, district, city, neighborhood). Studies at small spatial scales supported that local scale elimination is achievable on the short term (Laager et al., 2019; Zinsstag et al., 2017; Bourhy et al., 2016), but introduction events participate in rabies maintenance (Laager et al., 2019; Mollentze et al., 2014; Cori et al., 2018; Beyer et al., 2011) and impede control efforts (Bourhy et al., 2016; Zinsstag et al., 2017). Rabies was shown to circulate at low intensity within two cities, Bangui (Bourhy et al., 2016) and N'Djaména (Laager et al., 2019; Zinsstag et al., 2017; Zinsstag et al., 2009), but connections between urban areas are expected to accentuate rabies spread (Colombi et al., 2020). Indeed, human-mediated movements strongly impact rabies dispersal within countries in North Africa (Dellicour et al., 2017; Talbi et al., 2010) and the Central African Republic (Colombi et al., 2020). According to the setting, they may counteract the effects of control measures.

Spatial and individual heterogeneity were not sufficient to explain rabies maintenance in settings with low circulation (Laager et al., 2019; Wilson-Aggarwal et al., 2019) and there is no current evidence of the role of wildlife in the maintenance of rabies (Beyer et al., 2011; Fitzpatrick et al., 2012). At the continental scale, there are both co-circulating RABV lineages (Lemey et al., 2009a; Brunker et al., 2015; Talbi et al., 2009; Hayman et al., 2011; Brunker et al., 2018b) and spatial clustering of RABV lineages (Lemey et al., 2009a; Mollentze et al., 2013; Brunker et al., 2015; Talbi et al., 2009; Brunker et al., 2018b) which points at the role of human-mediated movements.

The dog vaccination coverage recommended by the WHO (70%) has been generally shown to be sufficient to reach dog rabies elimination (Zinsstag et al., 2009; Kitala et al., 2002; Fitzpatrick et al., 2012), except in Ethiopia where a 90% vaccination coverage was recommended (Beyene et al., 2019). Vaccination strategies targeting at-risk dog populations are more effective (Beyer et al., 2012). The role of underreporting is not clear (Laager et al., 2019; Zinsstag et al., 2017; Mollentze et al., 2014) but heterogeneous vaccination coverage is shown to disrupt vaccination (Townsend et al., 2013b).

2.2. Asia

Current situation

- Disease-free in Japan.
- Recent introductions in Indonesia and Philippines.
- Endemic in China and continental South-East Asia.

Models

- Mostly deterministic models implemented to analyze dog rabies in China (Zhang et al., 2012; Hou et al., 2012; Zhang et al., 2011; Chen et al., 2015; Huang et al., 2019), Malaysia (Coleman and Dye, 1996; Abdul Taib et al., 2019) and Indonesia (Coleman and Dye, 1996).
- Agent-based models (Townsend et al., 2013b; Kadowaki et al., 2018; Ferguson et al., 2015).

- Phylogeography (Zhang et al., 2017; Meng et al., 2011; Guo et al., 2013; Wang et al., 2019; Tohma et al., 2014; Dibia et al., 2015; Ma et al., 2017; Yu et al., 2012; Yao et al., 2015).
- Interdisciplinary studies (Tohma et al., 2016; Tian et al., 2018).

Data

Human rabies cases from passive surveillance, dog rabies cases from active (China) or passive surveillance, contact tracing, dog vaccination data, dog density from surveys, dog movements from household surveys, historical records of dog rabies epidemics in Osaka and of dog and human censuses, RABV genetic sequences from dogs, wildlife and humans.

Modelling aims

- Better understanding of the spatial and temporal dynamics of rabies spread (Zhang et al., 2011; Zhang et al., 2012; Chen et al., 2015; Abdul Taib et al., 2019; Zhang et al., 2017; Meng et al., 2011; Guo et al., 2013; Wang et al., 2019; Tohma et al., 2014; Dibia et al., 2015; Ma et al., 2017; Yu et al., 2012; Yao et al., 2015; Tohma et al., 2016; Tian et al., 2018).
- Identification of circulating lineages (Wang et al., 2019; Dibia et al., 2015; Tian et al., 2018) and environmental factors impacting rabies spread (Yao et al., 2015; Tian et al., 2018).
- Spatiotemporal dynamics and interactions of canine and wildlife RABV lineages (Huang et al., 2019; Yu et al., 2012).
- Feasible and effective control strategies (Coleman and Dye, 1996; Huang et al., 2019; Ferguson et al., 2015; Townsend et al., 2013b; Zhang et al., 2011; Hou et al., 2012; Zhang et al., 2012).
- Impact of human-mediated movement (Townsend et al., 2013b; Ferguson et al., 2015) and vaccination coverage (Townsend et al., 2013b) on the efficacy of control strategies.
- Modelling dynamics following an introduction and assessment of the efficacy of current contingency plans (Kadowaki et al., 2018).
- Estimation of the time from introduction to detection according to the value of R (Tohma et al., 2016).

Main findings

Rabies introductions in the disease-free islands of the Philippines result from single introductions from neighboring rabies-endemic islands followed by local transmission (Tohma et al., 2014; Tohma et al., 2016).

At the continental scale, RABV lineages are spatially clustered (Guo et al., 2013; Wang et al., 2019) but transboundary movements markedly influence rabies spread (Guo et al., 2013). China is endemic for rabies and multiple RABV lineages co-circulate across the country, notably Asian, Arctic-like and Cosmopolitan lineages (Meng et al., 2011; Yu et al., 2012; Yao et al., 2015). It is thought to be one of the main sources of RABV lineages in Asia (Meng et al., 2011; Guo et al., 2013).

A decade after achieving rabies elimination, it resurged in Yunnan and is currently circulating uncontrolled. This Chinese province corresponds to a crossroads area where multiple RABV lineages circu-

late, probably resulting from multiple transboundary movements (Zhang et al., 2017; Tian et al., 2018). Moreover, rabies dispersal velocity is weakly associated with forest coverage, croplands and accessible areas (Tian et al., 2018). Whereas human-mediated movement is not statistically associated with rabies velocity in the Yunnan province (Tian et al., 2018), it is suspected to have played a role in rabies dispersal in the Shaanxi province (Ma et al., 2017). More studies are needed to unravel the interactions between RABV, reservoir ecology and humans in Asia.

In general, rabies is estimated to spread at low grade with an R lower than two (Townsend et al., 2013b; Zhang et al., 2011; Hou et al., 2012; Zhang et al., 2012; Chen et al., 2015; Huang et al., 2019; Abdul Taib et al., 2019). Occasional long distance migrations which were documented in the Philippines (Tohma et al., 2016), Indonesia (Dibia et al., 2015) and China (Chen et al., 2015; Yao et al., 2015; Guo et al., 2013) might contribute to disease persistence.

The role of wildlife has been poorly studied and remains unclear in endemic areas (Huang et al., 2019; Yu et al., 2012).

Dog vaccination is the most effective strategy (Zhang et al., 2011; Hou et al., 2012; Huang et al., 2019; Abdul Taib et al., 2019) and may be improved by complementary measures such as domestic and stray dog management (Zhang et al., 2011; Hou et al., 2012; Zhang et al., 2012), dog confinement (Ferguson et al., 2015), or increasing public awareness (Zhang et al., 2012; Huang et al., 2019; Kadowaki et al., 2018). Homogeneous vaccination coverage was shown to yield better elimination prospects (Townsend et al., 2013a; Chen et al., 2015; Ferguson et al., 2015) which might be due to its robustness to human-mediated movements (Townsend et al., 2013a). In Japan, Kadowaki et al. (2018) showed that the current contingency plan is adapted to the rapid detection, control and elimination of rabies after an introduction. The authors emphasized the benefits of dog owner awareness and the control of stray dogs in the improvement of the plan (Kadowaki et al., 2018).

The time to detection is also a crucial factor in the success of rabies elimination after introduction. The faster the disease is detected, the higher the odds of eradicating it (Townsend et al., 2013a). For example, it's estimated that the surveillance system detected rabies circulation one year after its introduction in the Luzon island group in the Philippines (Tohma et al., 2016). This delay would have been greater with a lower reporting capacity (Tohma et al., 2016).

2.3. Middle East

Current situation

Endemic

Models

Phylogeography (Dellicour et al., 2019; Horton et al., 2015)

Data

RABV genetic sequences from dogs, wildlife, and humans.

Modelling aims

- Spatiotemporal dynamics and interactions of canine and wildlife RABV lineages (Dellicour et al., 2019; Horton et al., 2015).
- Identification of circulating lineages and environmental factors impacting rabies spread (Dellicour et al., 2019).

Main findings

Many lineages circulate that are phylogenetically related to Asian, Arctic/Artic-like, or Cosmopolitan lineages resulting from sustained circulation in dogs and wildlife after introduction (Dellicour et al., 2019; Horton et al., 2015). There is a strong spatial segregation of RABV lineages circulating in Iran. Overall, their spread is not driven by road connectivity, but humans presumably play a role since lineages tend to disperse towards and remain in highly populated areas. Lineages were less likely to spread towards grasslands and to occur in areas with barren vegetation. These results may be influenced by biased sampling towards populated areas however (Dellicour et al., 2019).

Wildlife seems to play a role in rabies maintenance in dog populations (Dellicour et al., 2019; Horton et al., 2015) but data are not sufficiently available to study host shift and dynamics between reservoirs.

2.4. South America

Current situation

Endemic for bat rabies and localized resurgences of rabies in dogs.

Models

- Fuzzy compartmental model (Ortega et al., 2000).
- Phylogeography (Carnieli et al., 2013; Carnieli et al., 2011).

Data

Serological data and RABV genetic sequences from dogs and wildlife.

Modelling aims

- Implementation of a fuzzy logic approach to model rabies spread (Ortega et al., 2000).
- Spatiotemporal dynamics of wild fox (Carnieli et al., 2013) and dog (Carnieli et al., 2013; Carnieli et al., 2011) RABV lineages.

Main findings

Despite extensive dog vaccination campaigns, multiple dog-related RABV lineages circulate in Brazil with a relatively recent common ancestor estimated in the 1950s (Carnieli et al., 2013; Carnieli et al., 2011). Dog lineages are generally spatially clustered (Carnieli et al., 2013; Carnieli et al., 2011) and lineages circulating in wild foxes and dogs are phylogenetically and dynamically independent (Carnieli et al., 2013).

2.5. Oceania

Current situation

Rabies-free.

Models

- Agent-based models (Dürr and Ward, 2015; Hudson et al., 2019a; Hudson et al., 2019b; Brookes et al., 2019).
- Compartmental models (Sparkes et al., 2016).

Data

Dog population structure, dog roaming behavior (GPS data, questionnaires/interviews of dog owners), dog contacts, census data.

Modelling aims

Modelling dynamics following an introduction (Dürr and Ward, 2015; Hudson et al., 2019a; Hudson et al., 2019b; Brookes et al., 2019; Sparkes et al., 2016) and assessment of the most effective control strategies (Dürr and Ward, 2015; Hudson et al., 2019a; Brookes et al., 2019; Sparkes et al., 2016).

Main findings

Australian studies focused on rabies spread in remote rural and peri-urban locations where rabies is expected to be introduced and where surveillance systems might be weakened by the remoteness.

Rabies dynamics are expected to differ between dog categories, such as explorer dogs, roaming dogs or domestic dogs (Hudson et al., 2019b; Brookes et al., 2019; Sparkes et al., 2016), and consequently between rural and peri-urban areas (Sparkes et al., 2016).

Reactive vaccination after the detection of rabies introduction is the only beneficial strategy (Dürr and Ward, 2015; Hudson et al., 2019a; Brookes et al., 2019; Sparkes et al., 2016). A 90% dog vaccination coverage is recommended to break down rabies spread (Brookes et al., 2019; Sparkes et al., 2016) and targeting at-risk dogs should enhance vaccination campaigns efficacy (Hudson et al., 2019a; Sparkes et al., 2016).

3. PRISMA-ScR Checklist

Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) Checklist

SECTION	ITEM	PRISMA-ScR CHECKLIST ITEM	REPORTED ON PAGE #
TITLE			
Title	1	Identify the report as a scoping review.	Title
ABSTRACT			
Structured summary	2	Provide a structured summary that includes (as applicable): background, objectives, eligibility criteria, sources of evidence, charting methods, results, and conclusions that relate to the review questions and objectives.	Abstract
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known. Explain why the review questions/objectives lend themselves to a scoping review approach.	Introduction (Background)
Objectives	4	Provide an explicit statement of the questions and objectives being addressed with reference to their key elements (e.g., population or participants, concepts, and context) or other relevant key elements used to conceptualize the review questions and/or objectives.	Introduction (Objectives)
METHODS			
Protocol and registration	5	Indicate whether a review protocol exists; state if and where it can be accessed (e.g., a Web address); and if available, provide registration information, including the registration number.	Methods (Search strategy)
Eligibility criteria	6	Specify characteristics of the sources of evidence used as eligibility criteria (e.g., years considered, language, and publication status), and provide a rationale.	Methods (Search strategy, Selection of studies)
Information sources*	7	Describe all information sources in the search (e.g., databases with dates of coverage and contact with authors to identify additional sources), as well as the date the most recent search was executed.	Methods (Search Strategy)
Search	8	Present the full electronic search strategy for at least 1 database, including any limits used, such that it could be repeated.	Methods (Search strategy)
Selection of sources of evidence†	9	State the process for selecting sources of evidence (i.e., screening and eligibility) included in the scoping review.	Methods (Selection of studies)
Data charting process‡	10	Describe the methods of charting data from the included sources of evidence (e.g., calibrated forms or forms that have been tested by the team before their use, and whether data charting was done independently or in duplicate) and any processes for obtaining and confirming data from investigators.	Methods (Data extraction and analysis)
Data items	11	List and define all variables for which data were sought and any assumptions and simplifications made.	Methods (Data extraction and analysis)
Critical appraisal of individual sources of evidence§	12	If done, provide a rationale for conducting a critical appraisal of included sources of evidence; describe the methods used and how this information was used in any data synthesis (if appropriate).	Methods (Data extraction and analysis)
Synthesis of results	13	Describe the methods of handling and summarizing the	Methods (Data

SECTION	ITEM	PRISMA-ScR CHECKLIST ITEM	REPORTED ON PAGE #
		data that were charted.	extraction and analysis)
RESULTS			
Selection of sources of evidence	14	Give numbers of sources of evidence screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally using a flow diagram.	Results (General characteristics of selected studies) Fig 1
Characteristics of sources of evidence	15	For each source of evidence, present characteristics for which data were charted and provide the citations.	Results (Topics addressed by the studies) S1-6 Tables
Critical appraisal within sources of evidence	16	If done, present data on critical appraisal of included sources of evidence (see item 12).	Results (Potential sources of bias in the data, Sensitivity analyses)
Results of individual sources of evidence	17	For each included source of evidence, present the relevant data that were charted that relate to the review questions and objectives.	Results Supplementary materials
Synthesis of results	18	Summarize and/or present the charting results as they relate to the review questions and objectives.	Fig 3, Tables 1-3
DISCUSSION			
Summary of evidence	19	Summarize the main results (including an overview of concepts, themes, and types of evidence available), link to the review questions and objectives, and consider the relevance to key groups.	Discussion (Insights on rabies epidemiology and control) Fig 4
Limitations	20	Discuss the limitations of the scoping review process.	Discussion (Insights on rabies epidemiology and control)
Conclusions	21	Provide a general interpretation of the results with respect to the review questions and objectives, as well as potential implications and/or next steps.	Discussion (Conclusions)
FUNDING			
Funding	22	Describe sources of funding for the included sources of evidence, as well as sources of funding for the scoping review. Describe the role of the funders of the scoping review.	Funding section

JBI = Joanna Briggs Institute; PRISMA-ScR = Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews.
* Where *sources of evidence* (see second footnote) are compiled from, such as bibliographic databases, social media platforms, and Web sites.
† A more inclusive/heterogeneous term used to account for the different types of evidence or data sources (e.g., quantitative and/or qualitative research, expert opinion, and policy documents) that may be eligible in a scoping review as opposed to only studies. This is not to be confused with *information sources* (see first footnote).
‡ The frameworks by Arksey and O'Malley (6) and Levac and colleagues (7) and the JBI guidance (4, 5) refer to the process of data extraction in a scoping review as data charting.
§ The process of systematically examining research evidence to assess its validity, results, and relevance before using it to inform a decision. This term is used for items 12 and 19 instead of "risk of bias" (which is more applicable to systematic reviews of interventions) to include and acknowledge the various sources of evidence that may be used in a scoping review (e.g., quantitative and/or qualitative research, expert opinion, and policy document).

From: Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMAScR): Checklist and Explanation. Ann Intern Med. 2018;169:467–473. doi: 10.7326/M18-0850.

Appendix B

Supplementary information on dog rabies spread in Cambodia

R scripts and related files needed to run all the landscape phylogeographic analyses, as well as BEAST XML files, are all available at https://github.com/sdellicour/rabv_cambodia.

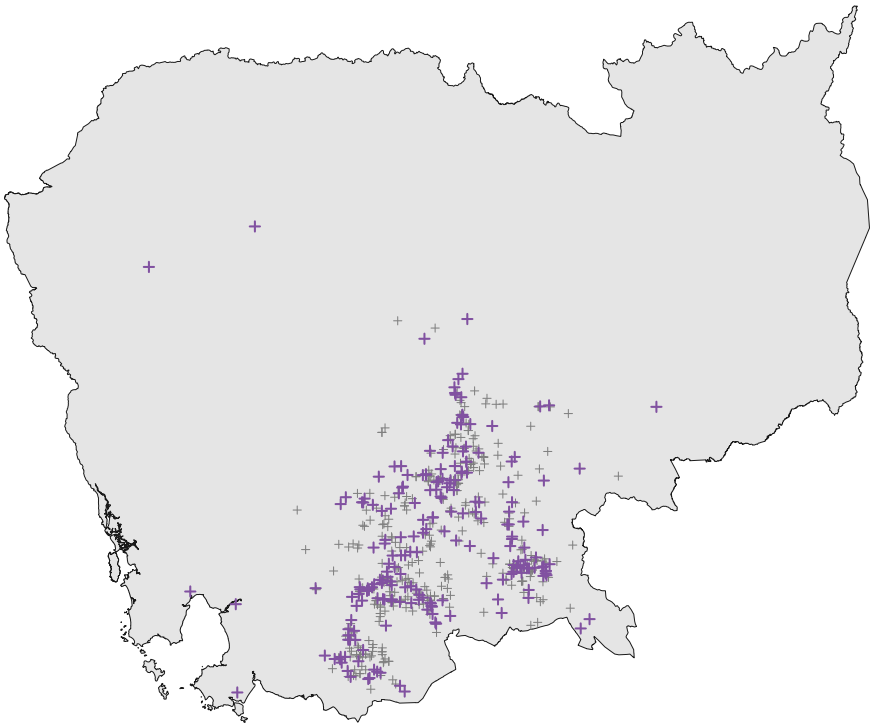


Figure B1: Selection of the samples that were sequenced in the context of the present study. Selected samples and non-selected samples are displayed in purple and grey, respectively. See the Material and Methods section for the detailed procedure implemented to select the samples to sequence while maximising the spatio-temporal coverage of available sequences included in continuous phylogeographic inference.

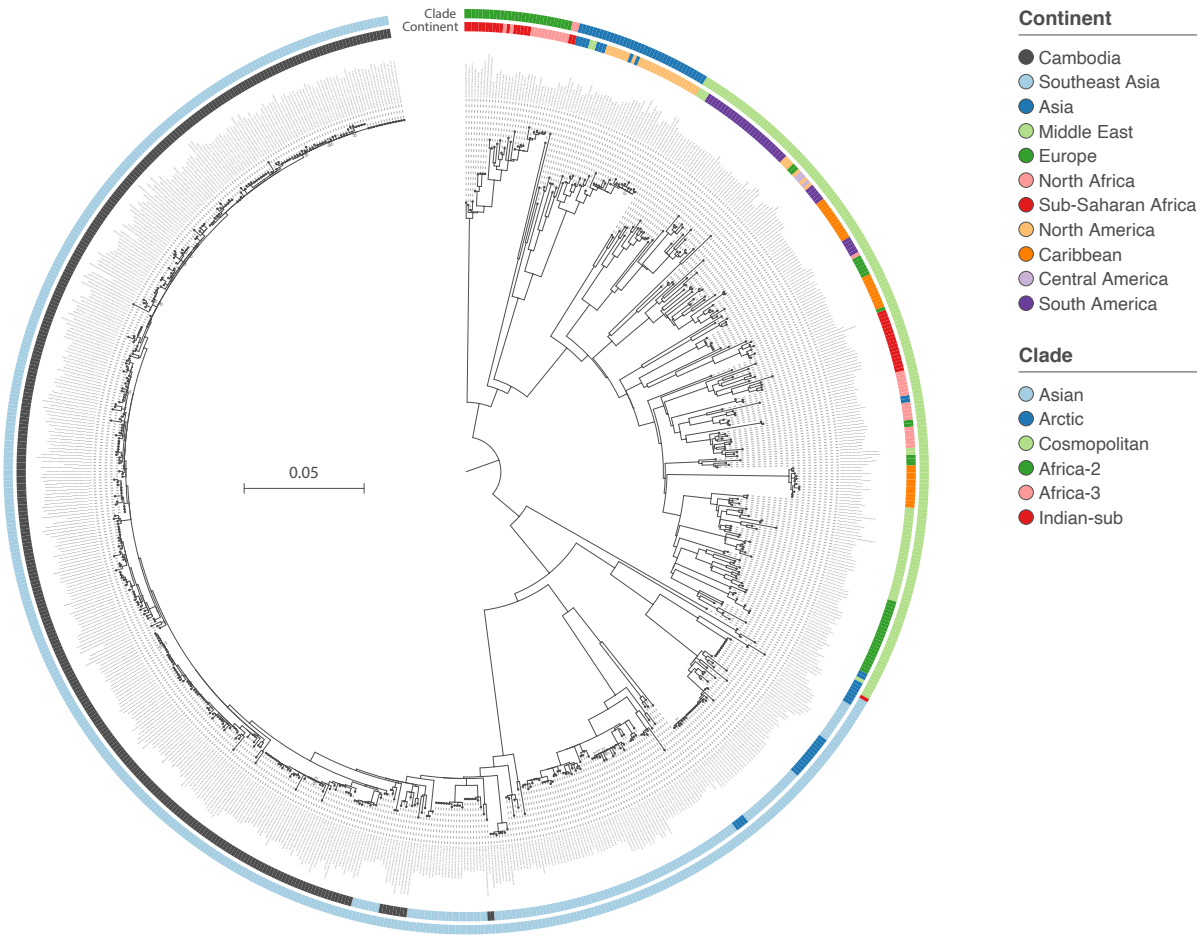


Figure B2: Maximum likelihood tree of worldwide representatives and Cambodian RABV N genes. Tips are colored by continent, except Cambodian sequences that are colored in dark grey, and by RABV clade that is determined for each tip with the online RABV genotyping tool RABV-GLUE (http://rabv-glue.cvr.gla.ac.uk/#/home).

Table B1: Source of data for each environmental raster.

Original raster	**Source**	**URL**
Elevation raster	SRTM (Shuttle Radar Topography Mission)	webmap.ornl.gov
Land cover raster	IGBP (International Geosphere Biosphere Programme)	www.igbp.net
Annual mean temperature	WorldClim database, version 2.0 (bioclimatic variable 'bio1')	worldclim.org
Annual precipitation	WorldClim database, version 2.0 (bioclimatic variable 'bio12')	worldclim.org
Human population density	GRUMP (Global Rural-Urban Mapping Project)	www.map.ox.ac.uk

Table B2: Investigating the impact of several environmental factors on the dispersal location of RABV lineages in Cambodia. We report approximated Bayes factor (BF) supports for the association between environmental values and tree node locations. The results are based on 1,000 posterior trees obtained by spatially-explicit phylogeographic inference. Following Kass and Raftery (1995), we consider a BF value >20 as strong support (in bold).

	Tendency of viral lineages to avoid circulating within specific environmental conditions		**Tendency of viral lineages to preferentially circulate within specific environmental conditions**	
Environmental factor	**Full genomes**	**N genes**	**Full genomes**	**N genes**
Forest areas	**>99**	**>99**	0.0	0.0
Savannas	**>99**	**>99**	0.0	0.0
Grasslands	3.8	3.5	0.3	0.3
Croplands	0.0	0.0	**99**	**>99**
Water areas	0.3	0.5	3.7	2.0
Human population density (log_{10})	0.0	0.0	**23.3**	**>99**
Elevation	6.6	**44.0**	0.2	0.0
Annual mean temperature	0.0	0.0	**>99**	**>99**
Annual precipitation	6.0	**68.2**	0.2	0.0

Table B3: Investigating the impact of several environmental factors on the dispersal velocity of RABV lineages in Cambodia, based on the analysis of the full genomes data set. The results are based on 1,000 posterior trees obtained by spatially-explicit phylogeographic inference. *R* and *C* indicate if the considered environmental raster was considered as a resistance (*R*) or conductance (*C*) factor, and *k* is the rescaling parameter used to transform the initial raster (see the text for further detail). For regression coefficients and Q values we report both the median estimate and the 95% HPD interval. The Bayes factor (BF) supports are only reported when $p(Q > 0)$ is at least 90%. Following Kass and Raftery (1995), we consider a BF value >20 as strong support for a significant correlation between the environmental distances and dispersal durations.

Environmental factor	*k*	Regression coefficient	*Q statistic*	$p(Q > 0)$	BF
Forest areas (C)	10	0.095 [0.051, 0.132]	-0.054 [-0.096, -0.024]	0.00	-
	100	0.051 [0.018, 0.091]	-0.098 [-0.142, -0.052]	0.00	-
	1000	0.027 [0.004, 0.064]	-0.122 [-0.164, -0.072]	0.00	-
Forest areas (R)	10	0.118 [0.066, 0.261]	-0.031 [-0.083, 0.108]	0.17	-
	100	0.029 [0.012, 0.206]	-0.118 [-0.161, 0.058]	0.06	-
	1000	0.017 [0.001, 0.192]	-0.131 [-0.174, 0.041]	0.05	-
Savannas (C)	10	0.122 [0.074, 0.167]	-0.025 [-0.065, -0.008]	0.01	-
	100	0.071 [0.030, 0.113]	-0.078 [-0.126, -0.043]	0.00	-
	1000	0.035 [0.007, 0.072]	-0.113 [-0.156, -0.069]	0.00	-
Savannas (R)	10	0.107 [0.069, 0.187]	-0.043 [-0.066, 0.038]	0.06	-
	100	0.009 [0.004, 0.106]	-0.138 [-0.179, -0.034]	0.01	-
	1000	0.000 [0.000, 0.076]	-0.147 [-0.188, -0.060]	0.01	-
Grasslands (C)	10	0.140 [0.093, 0.184]	-0.008 [-0.032, 0.003]	0.08	-
	100	0.108 [0.063, 0.153]	-0.040 [-0.077, -0.012]	0.00	-
	1000	0.076 [0.037, 0.118]	-0.072 [-0.115, -0.034]	0.00	-
Grasslands (R)	10	0.131 [0.076, 0.189]	-0.016 [-0.069, 0.027]	0.10	-
	100	0.026 [0.002, 0.106]	-0.121 [-0.175, -0.039]	0.01	-
	1000	0.001 [0.000, 0.058]	-0.146 [-0.189, -0.074]	0.01	-
Croplands (C)	10	0.131 [0.073, 0.288]	-0.018 [-0.069, 0.136]	0.27	-
	100	0.098 [0.013, 0.350]	-0.050 [-0.149, 0.192]	0.13	-
	1000	0.088 [0.002, 0.331]	-0.061 [-0.163, 0.182]	0.11	-
Croplands (R)	10	0.091 [0.052, 0.134]	-0.057 [-0.100, -0.022]	0.00	-
	100	0.078 [0.040, 0.121]	-0.070 [-0.117, -0.029]	0.00	-
	1000	0.076 [0.038, 0.119]	-0.072 [-0.121, -0.031]	0.00	-
Water areas (C)	10	0.096 [0.055, 0.145]	-0.051 [-0.093, -0.014]	0.00	-
	100	0.043 [0.015, 0.088]	-0.106 [-0.149, -0.054]	0.00	-
	1000	0.028 [0.006, 0.067]	-0.122 [-0.165, -0.067]	0.00	-
Water areas (R)	10	0.044 [0.015, 0.104]	-0.103 [-0.151, -0.042]	0.00	-
	100	0.002 [0.000, 0.032]	-0.144 [-0.186, -0.093]	0.00	-
	1000	0.002 [0.000, 0.023]	-0.146 [-0.186, -0.095]	0.00	-
Human population density (log10, C)	10	0.152 [0.104, 0.195]	0.003 [-0.010, 0.011]	0.74	-
	100	0.153 [0.101, 0.205]	0.004 [-0.027, 0.029]	0.61	-
	1000	0.137 [0.087, 0.231]	-0.013 [-0.056, 0.074]	0.30	-
Human population density (log10, R)	10	0.144 [0.096, 0.188]	-0.006 [-0.017, 0.013]	0.17	-
	100	0.054 [0.024, 0.099]	-0.095 [-0.135, -0.052]	0.00	-
	1000	0.011 [0.003, 0.043]	-0.137 [-0.177, -0.086]	0.00	-
Elevation (C)	10	0.137 [0.090, 0.178]	-0.012 [-0.036, 0.001]	0.03	-
	100	0.087 [0.044, 0.136]	-0.061 [-0.107, -0.024]	0.00	-
	1000	0.046 [0.018, 0.101]	-0.101 [-0.148, -0.051]	0.00	-
Elevation (R)	10	0.150 [0.098, 0.207]	-0.001 [-0.017, 0.042]	0.46	-
	100	0.086 [0.047, 0.201]	-0.062 [-0.109, 0.060]	0.06	-
	1000	0.046 [0.022, 0.177]	-0.101 [-0.147, 0.042]	0.04	-
Annual mean temperature (C)	10	0.153 [0.102, 0.196]	0.002 [-0.001, 0.014]	0.96	3.6

Table B3 continued from previous page

	100	0.153 [0.102, 0.196]	0.003 [0.000, 0.014]	0.96	3.7
	1000	0.153 [0.102, 0.196]	0.003 [0.000, 0.014]	0.96	3.7
Annual mean temperature (R)	10	0.150 [0.101, 0.191]	-0.001 [-0.005, 0.012]	0.28	-
	100	0.150 [0.100, 0.191]	-0.001 [-0.005, 0.011]	0.25	-
	1000	0.150 [0.100, 0.191]	-0.001 [-0.005, 0.011]	0.25	-
Annual precipitation (C)	10	0.133 [0.089, 0.177]	-0.016 [-0.037, 0.001]	0.04	-
	100	0.127 [0.083, 0.170]	-0.022 [-0.046, -0.002]	0.02	-
	1000	0.126 [0.082, 0.169]	-0.023 [-0.047, -0.002]	0.01	-
Annual precipitation (R)	10	0.145 [0.095, 0.187]	-0.006 [-0.021, 0.015]	0.23	-
	100	0.139 [0.090, 0.185]	-0.012 [-0.031, 0.014]	0.14	-
	1000	0.138 [0.089, 0.185]	-0.012 [-0.032, 0.013]	0.13	-

Table B4: Investigating the impact of several environmental factors on the dispersal velocity of RABV lineages in Cambodia, based on the analysis of the full N genes data set. The results are based on 1,000 posterior trees obtained by spatially-explicit phylogeographic inference. *R* and *C* indicate if the considered environmental raster was considered as a resistance (*R*) or conductance (*C*) factor, and *k* is the rescaling parameter used to transform the initial raster (see the text for further detail). For regression coefficients and Q values we report both the median estimate and the 95% HPD interval. The Bayes factor (BF) supports are only reported when $p(Q > 0)$ is at least 90%, which was not the case in the context of the analyses reported in the present table.

Environmental factor	*k*	**Regression coefficient**	*Q statistic*	$p(Q > 0)$	**BF**
Forest areas (C)	10	0.129 [0.092, 0.173]	-0.014 [-0.039, 0.006]	0.10	-
	100	0.099 [0.063, 0.142]	-0.044 [-0.080, -0.010]	0.00	-
	1000	0.069 [0.035, 0.110]	-0.075 [-0.115, -0.035]	0.00	-
Forest areas (R)	10	0.092 [0.054, 0.145]	-0.050 [-0.089, -0.010]	0.01	-
	100	0.023 [0.005, 0.066]	-0.118 [-0.163, -0.073]	0.00	-
	1000	0.012 [0.001, 0.048]	-0.128 [-0.174, -0.084]	0.00	-
Savannas (C)	10	0.152 [0.112, 0.199]	0.009 [-0.012, 0.022]	0.84	-
	100	0.120 [0.083, 0.162]	-0.023 [-0.057, 0.004]	0.05	-
	1000	0.085 [0.051, 0.121]	-0.059 [-0.102, -0.022]	0.00	-
Savannas (R)	10	0.049 [0.032, 0.079]	-0.094 [-0.124, -0.068]	0.00	-
	100	0.004 [0.001, 0.020]	-0.138 [-0.182, -0.100]	0.00	-
	1000	0.001 [0.000, 0.012]	-0.142 [-0.187, -0.103]	0.00	-
Grasslands (C)	10	0.136 [0.099, 0.181]	-0.007 [-0.017, 0.001]	0.03	-
	100	0.109 [0.072, 0.149]	-0.036 [-0.059, -0.014]	0.00	-
	1000	0.079 [0.046, 0.119]	-0.064 [-0.095, -0.034]	0.00	-
Grasslands (R)	10	0.134 [0.090, 0.182]	-0.007 [-0.041, 0.008]	0.16	-
	100	0.042 [0.006, 0.096]	-0.102 [-0.159, -0.047]	0.00	-
	1000	0.005 [0.000, 0.029]	-0.137 [-0.182, -0.095]	0.00	-
Croplands (C)	10	0.074 [0.044, 0.115]	-0.069 [-0.101, -0.038]	0.00	-
	100	0.013 [0.004, 0.046]	-0.128 [-0.170, -0.087]	0.00	-
	1000	0.001 [0.000, 0.027]	-0.140 [-0.187, -0.099]	0.00	-
Croplands (R)	10	0.116 [0.078, 0.163]	-0.028 [-0.056, -0.001]	0.02	-
	100	0.105 [0.068, 0.150]	-0.039 [-0.072, -0.009]	0.01	-
	1000	0.103 [0.066, 0.148]	-0.042 [-0.075, -0.011]	0.00	-
Water areas (C)	10	0.100 [0.066, 0.141]	-0.043 [-0.073, -0.014]	0.00	-
	100	0.049 [0.023, 0.082]	-0.094 [-0.134, -0.056]	0.00	-
	1000	0.030 [0.010, 0.061]	-0.112 [-0.153, -0.074]	0.00	-
Water areas (R)	10	0.059 [0.030, 0.102]	-0.085 [-0.129, -0.036]	0.00	-
	100	0.002 [0.000, 0.026]	-0.140 [-0.184, -0.097]	0.00	-
	1000	0.001 [0.000, 0.016]	-0.142 [-0.187, -0.100]	0.00	-
Human population density (log10, C)	10	0.123 [0.086, 0.167]	-0.021 [-0.033, -0.007]	0.00	-
	100	0.094 [0.061, 0.140]	-0.049 [-0.071, -0.026]	0.00	-
	1000	0.070 [0.040, 0.115]	-0.073 [-0.104, -0.040]	0.00	-
Human population density (log10, R)	10	0.144 [0.108, 0.190]	0.000 [-0.025, 0.027]	0.51	-
	100	0.041 [0.018, 0.083]	-0.101 [-0.146, -0.053]	0.00	-
	1000	0.018 [0.004, 0.048]	-0.124 [-0.171, -0.079]	0.00	-
Elevation (C)	10	0.148 [0.110, 0.193]	0.005 [-0.007, 0.014]	0.81	-
	100	0.113 [0.075, 0.153]	-0.030 [-0.067, 0.000]	0.03	-
	1000	0.063 [0.034, 0.105]	-0.080 [-0.123, -0.034]	0.00	-
Elevation (R)	10	0.121 [0.086, 0.162]	-0.023 [-0.036, -0.009]	0.00	-
	100	0.047 [0.025, 0.081]	-0.095 [-0.131, -0.064]	0.00	-
	1000	0.022 [0.009, 0.049]	-0.120 [-0.158, -0.084]	0.00	-
Annual mean temperature (C)	10	0.142 [0.102, 0.189]	-0.001 [-0.006, 0.008]	0.16	-
	100	0.142 [0.102, 0.189]	-0.002 [-0.006, 0.008]	0.15	-

Table B4 continued from previous page

	1000	0.142 [0.102, 0.189]	-0.002 [-0.006, 0.008]	0.15	-
Annual mean temperature (R)	10	0.145 [0.105, 0.192]	0.001 [-0.003, 0.011]	0.75	-
	100	0.145 [0.105, 0.192]	0.001 [-0.003, 0.011]	0.76	-
	1000	0.145 [0.105, 0.192]	0.001 [-0.003, 0.011]	0.76	-
Annual precipitation (C)	10	0.146 [0.107, 0.192]	0.003 [-0.010, 0.016]	0.69	-
	100	0.146 [0.106, 0.190]	0.002 [-0.014, 0.016]	0.61	-
	1000	0.146 [0.106, 0.190]	0.002 [-0.014, 0.016]	0.60	-
Annual precipitation (R)	10	0.124 [0.084, 0.167]	-0.020 [-0.033, -0.006]	0.01	-
	100	0.117 [0.078, 0.159]	-0.027 [-0.043, -0.010]	0.01	-
	1000	0.116 [0.077, 0.159]	-0.028 [-0.044, -0.011]	0.01	-

Table B5: Accession number and epidemiological and phylogenetic information for the 776 rabies virus samples analyzed in Chapter 2.

ID	date	continent	country	province	district	municipality	village	latitude	longitude	host	type	clade	dataset
KM366202	30/10/2008	Southeast Asia	Cambodia	Takeo	Baty	Lumpong	TRAPAING KRALANH	11.318	104.765	Dog	N gene	Asian	Mey et al. 2016
KM366203	02/01/2006	Southeast Asia	Cambodia	Battambang	Battambang	O Cha	ANGDONG CHENH	13.104	103.174	Dog	N gene	Asian	Mey et al. 2016
KM366204	01/09/2004	Southeast Asia	Cambodia	Siem Reap	Puok	Doung Keo	KOUK PH-NAO	13.473	103.756	Dog	N gene	Asian	Mey et al. 2016
KM366205	17/07/2010	Southeast Asia	Cambodia	Prey Veng	Pearaing	Kampong Popil	KHSAM TBONG	11.643	105.159	Dog	N gene	Asian	Mey et al. 2016
KM366206	04/05/2004	Southeast Asia	Cambodia	Kg Chh-nang	Toek Phus	Kraing Ske	PRA SHET	12.22	104.464	Dog	N gene	Asian	Mey et al. 2016
KM366207	19/09/2007	Southeast Asia	Cambodia	Kandal	Ponhieleu	Chhvang	TA AUK	11.688	104.789	Dog	N gene	Asian	Mey et al. 2016
KM366208	18/05/2004	Southeast Asia	Cambodia	Prey Veng	Mesang	Svay Chrum	PO TA-MOM	11.375	105.562	Dog	N gene	Asian	Mey et al. 2016
KM366209	16/05/2002	Southeast Asia	Cambodia	Udoor Meanchey	Anloung Veng	Anloung Veng	TUOL KANDAL	14.229	104.104	Dog	N gene	Asian	Mey et al. 2016
KM366210	06/08/2000	Southeast Asia	Cambodia	Kampong Speu	Bosed	Por Muoy Real	TA NUON	11.138	104.55	Dog	N gene	Asian	Mey et al. 2016
KM366211	03/09/2010	Southeast Asia	Cambodia	Kg Chh-nang	Boribou	Po Peal	KRA CHY	12.407	104.447	Dog	N gene	Asian	Mey et al. 2016
KM366212	14/06/2008	Southeast Asia	Cambodia	Kandal	Khsach Kandal	Prek Ampil	PREK DAUN-HEM	11.649	104.962	Dog	N gene	Asian	Mey et al. 2016
KM366213	12/03/1998	Southeast Asia	Cambodia	Kampong Chhnang	Kampong Tralach	Lonvaek	ANLONG TNOUT	11.891	104.712	Dog	N gene	Asian	Mey et al. 2016
KM366214	16/06/2003	Southeast Asia	Cambodia	Phnom Penh	Meanchey	Chbar Am-pov Ii	NA	11.537	104.939	Dog	N gene	Asian	Mey et al. 2016
KM366215	10/02/2002	Southeast Asia	Cambodia	Takeo	Prey Kam-bass	Prek Pdov	SAYVAR	11.162	104.88	Dog	N gene	Asian	Mey et al. 2016
KM366216	20/05/2010	Southeast Asia	Cambodia	Kg Speu	Kong Pisey	Sdok	TRAPAING-LEUK	11.216	104.616	Dog	N gene	Asian	Mey et al. 2016
KM366217	25/03/2009	Southeast Asia	Cambodia	Koh Kong	Botumsakor	Angdong Toek	ANGDONG TOEK	11.191	103.478	Dog	N gene	Asian	Mey et al. 2016
KM366218	07/11/2003	Southeast Asia	Cambodia	Kampong Speu	Samroung Tong	Vorsar	CHAMBOK	11.472	104.566	Dog	N gene	Asian	Mey et al. 2016
KM366219	16/07/2004	Southeast Asia	Cambodia	Kg Speu	Phnom Sruoch	Treng Tray-oeng	PHUM 3	11.273	104.213	Dog	N gene	Asian	Mey et al. 2016

Table B5 continued from previous page

KM366220	04/11/2011	Southeast Asia	Cambodia	Kg Cham	Choeung Prey	Sdoeung Chey	KDOY	12.195	105.107	Dog	N gene	Asian	Mey et al. 2016
KM366221	08/08/2011	Southeast Asia	Cambodia	Kratie	Chhong	Prak Tamann	PREAK DACH	12.277	106.027	Dog	N gene	Asian	Mey et al. 2016
KM366222	27/06/2011	Southeast Asia	Cambodia	Kg Thom	Baray	Chrolorng	HANTHVEA	12.267	105.126	Dog	N gene	Asian	Mey et al. 2016
KM366223	24/02/2010	Southeast Asia	Cambodia	Kg Cham	Kang Meas	Sour Kong	PREK KRUOS	11.982	105.18	Dog	N gene	Asian	Mey et al. 2016
KM366244	22/12/2000	Southeast Asia	Cambodia	Takeo	Baty	Cham Bak	TANOP	11.091	104.436	Dog	N gene	Asian	Mey et al. 2016
KM366224	02/08/2000	Southeast Asia	Cambodia	Takeo	Tram Kak	Trapaing Kragnog	KDOUCH	11.236	104.794	Dog	N gene	Asian	Mey et al. 2016
KM366225	18/10/2011	Southeast Asia	Cambodia	Prey Veng	Ba Phnom	Cheu Kach	CHEU KACH	11.256	105.406	Dog	N gene	Asian	Mey et al. 2016
KM366226	24/03/2001	Southeast Asia	Cambodia	Phnom Penh	Mean Chey	Niroude	TA NGOV	11.556	104.851	Dog	N gene	Asian	Mey et al. 2016
KM366227	26/07/2010	Southeast Asia	Cambodia	Svay Rieng	Svay Chrum	Tasok	BOENG ANDENG	11.528	104.939	Dog	N gene	Asian	Mey et al. 2016
KM366277	13/12/2002	Southeast Asia	Cambodia	Phnom Penh	Mean Chey	Chbar Empao I	NA	11.087	105.683	Dog	N gene	Asian	Mey et al. 2016
KM366228	04/07/2008	Southeast Asia	Cambodia	Phnom Penh	Reussey Keo	Chroy Changva	2	11.576	104.936	Dog	N gene	Asian	Mey et al. 2016
KM366229	20/05/2011	Southeast Asia	Cambodia	Phnom Penh	Dangkor	Cham Chao	PHUM 4	11.539	104.824	Dog	N gene	Asian	Mey et al. 2016
KM366230	02/08/2011	Southeast Asia	Cambodia	Banteay Meanchey	Poipet	Poipet	PALELAY	13.649	102.57	Dog	N gene	Asian	Mey et al. 2016
KM366231	21/08/2008	Southeast Asia	Cambodia	Kg Speu	Thpong	Rong Roeurng	THMEY DONG TUNG	11.72	104.524	Dog	N gene	Asian	Mey et al. 2016
KM366232	02/12/2003	Southeast Asia	Cambodia	Kampong Cham	Prey Chhor	Lvea	LVEA	11.999	105.019	Dog	N gene	Asian	Mey et al. 2016
KM366233	30/08/2002	Southeast Asia	Cambodia	Kandal	Shang	Treuy Sla	PORLEU	11.352	105.021	Dog	N gene	Asian	Mey et al. 2016
KM366234	01/04/1999	Southeast Asia	Cambodia	Kandal	Ksach Kandal	Prek Takov	PREK LOVEAR	11.603	104.956	Dog	N gene	Asian	Mey et al. 2016
KM366235	06/02/2009	Southeast Asia	Cambodia	Takeo	Prey Kabbas	Prey Phdao	CHHUM ROM	11.142	104.874	Dog	N gene	Asian	Mey et al. 2016
KM366236	17/07/2008	Southeast Asia	Cambodia	Kandal	Ang Snuol	Makak	CHONG BOENG	11.644	104.714	Dog	N gene	Asian	Mey et al. 2016

Table B5 continued from previous page

KM366237	11/03/2008	Southeast Asia	Cambodia	Koh Kong	Kapong Sila	Kampong Sila	KRAING AT	11.103	103.922	Dog	N gene	Asian	Mey et al. 2016
KM366238	23/02/2009	Southeast Asia	Cambodia	Kg Cham	Srey Santhor	Reussey Srok	REUSSEY SROK	11.927	105.207	Dog	N gene	Asian	Mey et al. 2016
KM366239	07/01/1999	Southeast Asia	Cambodia	Phnom Penh	Chamcar Morn	Olimpic	NA	11.568	104.933	Dog	N gene	Asian	Mey et al. 2016
KM366240	13/12/2001	Southeast Asia	Cambodia	Kampong Chhnang	Samaki Mean Chey	Svay	PHSAR TRACH	11.815	104.674	Dog	N gene	Asian	Mey et al. 2016
KM366241	27/01/1999	Southeast Asia	Cambodia	Kandal	Leuk Dek	Prek Talorp	KAMPONG CHAM-LONG	11.256	105.278	Dog	N gene	Asian	Mey et al. 2016
KM366242	20/09/2001	Southeast Asia	Cambodia	Kampong Thom	Kampong Svay	Achar Leak	ACHAR LEAK	12.725	104.893	Dog	N gene	Asian	Mey et al. 2016
KM366243	26/03/2007	Southeast Asia	Cambodia	Takeo	Samrorng	Rovieng	THMEY	11.196	104.788	Dog	N gene	Asian	Mey et al. 2016
KM366245	24/04/2000	Southeast Asia	Cambodia	Kampong Speu	Kang Pisey	Angpopeal	SAMRONG REASMEI	11.354	104.72	Dog	N gene	Asian	Mey et al. 2016
KM366246	29/11/2006	Southeast Asia	Cambodia	Koh Kong	Sre Ambil	Sre Ambil	SRE AM-BIL	11.122	103.748	Dog	N gene	Asian	Mey et al. 2016
KM366247	11/04/2002	Southeast Asia	Cambodia	Kampong Thom	Santuk	Kampong Thmar	KAMPONG THMAR	12.499	105.127	Dog	N gene	Asian	Mey et al. 2016
KM366248	27/04/2006	Southeast Asia	Cambodia	Kandal	Sa Ang	Kraing Yov	SAMRONG	11.299	104.957	Dog	N gene	Asian	Mey et al. 2016
KM366249	04/07/2005	Southeast Asia	Cambodia	Siem Reap	Siem Reap	Svay Dangkom	KRUOS	13.371	103.84	Dog	N gene	Asian	Mey et al. 2016
KM366250	16/05/2003	Southeast Asia	Cambodia	Kandal	Muk Kampoul	Rokar Korng 1	PEAM	11.85	105	Dog	N gene	Asian	Mey et al. 2016
KM366251	15/05/1998	Southeast Asia	Cambodia	Kandal	Angnaul	Kambol	AMPIL	11.543	104.757	Dog	N gene	Asian	Mey et al. 2016
KM366252	18/09/1998	Southeast Asia	Cambodia	Phnom Penh	Tuol Kork	Psar Daum Kor	NA	11.553	104.906	Dog	N gene	Asian	Mey et al. 2016
KM366253	28/05/2007	Southeast Asia	Cambodia	Prey Veng	Baphnom	Cheu Chach	TREA	11.26	105.424	Dog	N gene	Asian	Mey et al. 2016
KM366254	16/07/2009	Southeast Asia	Cambodia	Phnom Penh	Mean Chey	Boeng Tompun	SANSAM-KOSAL	11.531	104.911	Dog	N gene	Asian	Mey et al. 2016
KM366255	26/09/2011	Southeast Asia	Cambodia	Kandal	Sa Arng	Prek Ombil	ANLONG TASEK LEU	11.184	105.021	Dog	N gene	Asian	Mey et al. 2016

Table B5 continued from previous page

KM366256	18/02/2002	Southeast Asia	Cambodia	Kratie	Ang Snuol	Snuol	SNUOL	12.097	106.46	Dog	N gene	Asian	Mey et al. 2016
KM366257	08/10/2006	Southeast Asia	Cambodia	Prey Veng	Mesang	Chhi Puch	TRORK	11.309	105.548	Dog	N gene	Asian	Mey et al. 2016
KM366258	20/03/2006	Southeast Asia	Cambodia	Kg Cham	Srey Santhor	Prek Rumdeng	SVAY TANUN KOR	11.9	105.257	Dog	N gene	Asian	Mey et al. 2016
KM366259	25/10/2004	Southeast Asia	Cambodia	Phnom Penh	Chamkar Mon	Boeng Kengkang1	WILDAID ORGANIZATION	11.553	104.929	Dog	N gene	Asian	Mey et al. 2016
KM366260	05/06/2002	Southeast Asia	Cambodia	Kampot	Angkor Chey	Tani	REUSSEI	10.772	104.684	Dog	N gene	Asian	Mey et al. 2016
KM366261	11/10/2005	Southeast Asia	Cambodia	Kratie	Chhloung	Han Chey	HAN CHEY 3	12.255	105.941	Dog	N gene	Asian	Mey et al. 2016
KM366262	10/02/2009	Southeast Asia	Cambodia	Kandal	Kandal Stoeng	Spean Thmor	MEUN TRA	11.459	104.872	Dog	N gene	Asian	Mey et al. 2016
KM366263	23/12/2011	Southeast Asia	Cambodia	Takeo	Bati	Dong	YUTHKA	11.234	104.908	Dog	N gene	Asian	Mey et al. 2016
KM366264	04/11/2001	Southeast Asia	Cambodia	Prey Veng	Mesang	Chi Phoch	SAMRONG VEAL	11.36	105.559	Dog	N gene	Asian	Mey et al. 2016
KM366265	08/02/2000	Southeast Asia	Cambodia	Phnom Penh	Mean Chey	Steung Mean Chey	NA	11.527	104.892	Dog	N gene	Asian	Mey et al. 2016
KM366266	30/12/2010	Southeast Asia	Cambodia	Rattanakiri	Ban Long	Ta Veng	TA VENG	13.738	107.019	Dog	N gene	Asian	Mey et al. 2016
KM366267	21/08/2010	Southeast Asia	Cambodia	Phnom Penh	Dangkor	Chomchao	NA	11.527	104.846	Dog	N gene	Asian	Mey et al. 2016
KM366268	07/09/2005	Southeast Asia	Cambodia	Phnom Penh	Mean Chey	Prek Pra	OU ANDAUNG1	11.492	104.952	Dog	N gene	Asian	Mey et al. 2016
KM366269	22/01/1999	Southeast Asia	Cambodia	Kampot	Banteay Meas	Samrong Loeu	TRAM SAR SAR	10.728	104.597	Dog	N gene	Asian	Mey et al. 2016
KM366270	10/11/2003	Southeast Asia	Cambodia	Kampot	Angkor Chey	Tany	PRAL	10.783	104.665	Dog	N gene	Asian	Mey et al. 2016
KM366271	25/09/2001	Southeast Asia	Cambodia	Siem Reap	Angloung Veng	Trapaing Prey	PRAMBEY RUOY	14.226	103.942	Dog	N gene	Asian	Mey et al. 2016
KM366272	29/12/2003	Southeast Asia	Cambodia	Kg Chhnang	Samaki Meanchey	Svay	PORITHIKRAI	11.833	104.718	Dog	N gene	Asian	Mey et al. 2016
KM366273	31/08/2004	Southeast Asia	Cambodia	Kapot	Angkor Chey	Tany	BREAL	10.783	104.665	Dog	N gene	Asian	Mey et al. 2016

Table B5 continued from previous page

KM366274	08/04/2008	Southeast Asia	Cambodia	Prey Veng	Baphnom	Beong Preah	ANG-KROUNG	11.323	105.453	Dog	N gene	Asian	Mey et al. 2016
KM366275	06/09/2000	Southeast Asia	Cambodia	Prey Veng	Baphnom	Choeung-phnom	RONG-DAMREY	11.257	105.386	Dog	N gene	Asian	Mey et al. 2016
KM366276	14/09/2006	Southeast Asia	Cambodia	Kapot	Chhouk	Krang Svay	DAMNAK TRAB KHANG-CHOEURNG	10.862	104.478	Dog	N gene	Asian	Mey et al. 2016
KM366278	24/05/2001	Southeast Asia	Cambodia	Kampong Speu	Samrong	Trapaing Kang	TRAPAING KHSANG	11.485	104.643	Dog	N gene	Asian	Mey et al. 2016
KM366279	22/12/2008	Southeast Asia	Cambodia	Takeo	Samrorng	Boeng Tranh Tboung	BOENG TRANH	11.088	104.749	Dog	N gene	Asian	Mey et al. 2016
KM366280	12/01/2008	Southeast Asia	Cambodia	Pursat	Sampov Meas	Phteas Prey	PEAL NHEK 2	12.536	103.921	Dog	N gene	Asian	Mey et al. 2016
KM366281	17/07/2008	Southeast Asia	Cambodia	Kandal	Khsach Kandal	Svaychrom	BARACHA-NARAM	11.578	104.959	Dog	N gene	Asian	Mey et al. 2016
KM366282	30/08/2000	Southeast Asia	Cambodia	Kampong Cham	Kang Meas	Kchao	VAR RIN	11.903	105.135	Dog	N gene	Asian	Mey et al. 2016
KM366283	22/06/2009	Southeast Asia	Cambodia	Battambang	Rukhakiri	Prek Chik	KAM RENG	13.045	102.507	Dog	N gene	Asian	Mey et al. 2016
KM366284	26/07/2011	Southeast Asia	Cambodia	Kapot	Dangtung	Torteung	TOCH	10.712	104.526	Dog	N gene	Asian	Mey et al. 2016
KM366285	02/07/1998	Southeast Asia	Cambodia	Kampong Cham	Srey San-thor	Prek Damboque	TAKOY	11.897	105.167	Dog	N gene	Asian	Mey et al. 2016
KM366286	14/11/2011	Southeast Asia	Cambodia	Kg Speu	Borseth	Svay Rompea	TRAM SAR SAR	11.195	104.566	Dog	N gene	Asian	Mey et al. 2016
KM366287	12/09/1998	Southeast Asia	Cambodia	Kampong Cham	Ta Bong Kahmom	Kor	KBAL O	11.957	105.719	Dog	N gene	Asian	Mey et al. 2016
KM366288	16/02/2009	Southeast Asia	Cambodia	Kg Thom	Baray	Sralao	TUOL AMPIL	12.325	105.107	Dog	N gene	Asian	Mey et al. 2016
KM366289	08/02/2007	Southeast Asia	Cambodia	Phnom Penh	Meanchey	Chbar Am-pov2	NA	11.533	104.939	Dog	N gene	Asian	Mey et al. 2016
KM366290	26/12/2001	Southeast Asia	Cambodia	Kampong Cham	Bateay	Bateay	SVAY POK	11.995	104.954	Dog	N gene	Asian	Mey et al. 2016
KM366291	22/12/2003	Southeast Asia	Cambodia	Banteay Meanchey	Thmor Puok	Thmor Puok	THMOR PUOK	13.939	103.058	Dog	N gene	Asian	Mey et al. 2016

Table B5 continued from previous page

KM366292	01/05/2006	Southeast Asia	Cambodia	Kg Thom	Baray	Svay Phloeung	SVAY PHLOE-UNG	12.228	105.157	Dog	N gene	Asian	Mey et al. 2016
KM366293	22/02/2007	Southeast Asia	Cambodia	Banteay Meanchey	Pranet Preah	Toek Chor	TOEK CHOR	13.603	103.401	Dog	N gene	Asian	Mey et al. 2016
KM366294	25/12/2006	Southeast Asia	Cambodia	Phnom Penh	7Makara	Mithpheap	NA	11.559	104.899	Dog	N gene	Asian	Mey et al. 2016
KM366295	07/08/1999	Southeast Asia	Cambodia	Kampong Speu	Cbar Morn	Cbar Morn	BEK CHAN	11.485	104.491	Dog	N gene	Asian	Mey et al. 2016
KM366296	01/10/2007	Southeast Asia	Cambodia	Kg Speu	Boset	Po Angkrong	PREY TAPHEM	11.138	104.662	Dog	N gene	Asian	Mey et al. 2016
KM366297	25/06/2004	Southeast Asia	Cambodia	Kandal	Kien Svay	Kbal Koh	PREK THOM	11.507	105.026	Dog	N gene	Asian	Mey et al. 2016
KM366298	16/03/2002	Southeast Asia	Cambodia	Banteay Meanchey	Serey Sophorn	Toeuk Thla	KAMPONG SVAY	13.592	102.976	Dog	N gene	Asian	Mey et al. 2016
KM366299	26/01/2009	Southeast Asia	Cambodia	Preah Vi-hear	Khcham San	Khcham San	TORK SRALAO	14.214	104.936	Dog	N gene	Asian	Mey et al. 2016
KM366300	09/06/2002	Southeast Asia	Cambodia	Kampong Speu	Borseth	Svayrompear	SLAPLENG	11.212	104.585	Dog	N gene	Asian	Mey et al. 2016
KM366301	19/08/2009	Southeast Asia	Cambodia	Kg Chh-nang	Rolea Pha Air	Rolea Pha Air	PREY KHMER	12.165	104.665	Dog	N gene	Asian	Mey et al. 2016
KM366302	28/01/2004	Southeast Asia	Cambodia	Svay Rieng	Svay Chrum	Krorl Kour	THLORK	11.137	105.632	Dog	N gene	Asian	Mey et al. 2016
KM366303	12/11/2007	Southeast Asia	Cambodia	Svay Rieng	Svay Thom	Svay Thom	KRANG-LEAV	11.002	105.599	Dog	N gene	Asian	Mey et al. 2016
KM366304	07/05/2001	Southeast Asia	Cambodia	Kandal	Angsnoul	Prey Pouch	PREY POUCH	11.466	104.718	Dog	N gene	Asian	Mey et al. 2016
KM366305	08/05/2002	Southeast Asia	Cambodia	Prey Veng	Sdech	Rom Chak	CHUNG RUK	11.163	105.367	Dog	N gene	Asian	Mey et al. 2016
KM366306	26/10/1998	Southeast Asia	Cambodia	Kampong Speu	Kangpisey	Roka Kaoh	ROKA KAOH	11.43	104.692	Dog	N gene	Asian	Mey et al. 2016
KM366307	15/07/1998	Southeast Asia	Cambodia	Kandal	Kien Svay	Phum Thom	RETEING	11.495	105.046	Dog	N gene	Asian	Mey et al. 2016
KM366308	28/07/2003	Southeast Asia	Cambodia	Prey Veng	Pmesang	Chiphuch	KROSAING	11.313	105.513	Dog	N gene	Asian	Mey et al. 2016
KM366309	25/01/1999	Southeast Asia	Cambodia	Kampong Cham	Kung Meas	Rokaar	CHROY KRA BAO	11.865	105.128	Dog	N gene	Asian	Mey et al. 2016
KM366310	29/05/2008	Southeast Asia	Cambodia	Kg Cham	Siem Reap	Sala Kam-roeuk	SLOR KRAM	13.341	103.865	Dog	N gene	Asian	Mey et al. 2016

Table B5 continued from previous page

KM366311	18/04/2007	Southeast Asia	Cambodia	Kapot	Dang Tong	Angkor Meas	SNOR TOCH	10.761	104.399	Dog	N gene	Asian	Mey et al. 2016
KM366312	14/11/2005	Southeast Asia	Cambodia	Kg Thom	Baray	Andoung Po	CHIVIPHEAP	12.292	105.152	Dog	N gene	Asian	Mey et al. 2016
KM366313	17/02/2004	Southeast Asia	Cambodia	Kg Cham	Kang Meas	Roka A	SVAY SRANEAH 1	11.855	105.113	Dog	N gene	Asian	Mey et al. 2016
KM366314	09/05/2009	Southeast Asia	Cambodia	Banteay Meanchey	Poi Pet	Nimit	NIMIT	13.615	102.739	Dog	N gene	Asian	Mey et al. 2016
KM366315	20/04/2006	Southeast Asia	Cambodia	Takeo	Prey Kab-bas	Prey Phdao	PREY KHNHEY	11.159	104.873	Dog	N gene	Asian	Mey et al. 2016
KM366316	16/04/2005	Southeast Asia	Cambodia	Kg Chh-nang	Kampong Tralach	Long Vek	PHSAR TRACH	11.846	104.728	Dog	N gene	Asian	Mey et al. 2016
KM366317	20/01/2003	Southeast Asia	Cambodia	Takeo	Samrounrng	Khvav	ANGKONH	11.159	104.708	Dog	N gene	Asian	Mey et al. 2016
KM366318	28/06/2010	Southeast Asia	Cambodia	Kg Thom	Baray	So Young	SO YOUNG	12.237	105.126	Dog	N gene	Asian	Mey et al. 2016
KM366319	14/10/2008	Southeast Asia	Cambodia	Svay Rieng	Chan Trea	Prey Angkunh	CHREY THOM	11.035	106.012	Dog	N gene	Asian	Mey et al. 2016
KM366320	11/05/2007	Southeast Asia	Cambodia	Kg Thom	Siem Reap	Sala Kam-roeurn	TAVEAN	13.355	103.878	Dog	N gene	Asian	Mey et al. 2016
KM366321	25/10/2001	Southeast Asia	Cambodia	Takeo	Sarong	Samrong	KRAING ROAUT	11.128	104.798	Dog	N gene	Asian	Mey et al. 2016
KM366322	04/08/2011	Southeast Asia	Cambodia	Kandal	Koh Thom	Prek Thmey	PREK THMEY	11.125	105.06	Dog	N gene	Asian	Mey et al. 2016
KM366323	28/10/2006	Southeast Asia	Cambodia	Kg Speu	Thpong	Monorum	THNAL	11.723	104.544	Dog	N gene	Asian	Mey et al. 2016
KM366324	15/09/2003	Southeast Asia	Cambodia	Kampong Thom	Santok	Kampong Thnor	KAMPONG THMOR	12.499	105.127	Dog	N gene	Asian	Mey et al. 2016
KM366325	16/01/2001	Southeast Asia	Cambodia	Kampot	Kampong Trach	Damnak Kantoude	DAMNAK KAN-TOUDE	10.577	104.466	Dog	N gene	Asian	Mey et al. 2016
KM366326	29/03/2000	Southeast Asia	Cambodia	Kampong Cham	Bathay	Taingkraing	PRASATH	11.868	104.933	Dog	N gene	Asian	Mey et al. 2016
KM366327	06/09/1999	Southeast Asia	Cambodia	Takeo	Prey Kam-bas	Prey Phdov	PREY PH-DOV	11.137	104.872	Dog	N gene	Asian	Mey et al. 2016
KM366328	29/01/1999	Southeast Asia	Cambodia	Phnom Penh	Mean Chey	Cbar Am-pao Ii	NA	11.537	104.939	Dog	N gene	Asian	Mey et al. 2016

Table B5 continued from previous page

KM366329	28/08/1998	Southeast Asia	Cambodia	Kandal	Saang	Troy Slar	N1L	11.352	105.021	Dog	N gene	Asian	Mey et al. 2016
KM366330	19/01/1998	Southeast Asia	Cambodia	Phnom Penh	Roussey Keo	NA	NA	11.626	104.908	Dog	N gene	Asian	Mey et al. 2016
KM366331	17/06/2009	Southeast Asia	Cambodia	Kg Som	Stoeng Hav	Tomnub Rolok	PHUM 2	10.738	103.632	Dog	N gene	Asian	Mey et al. 2016
KM366332	20/10/1998	Southeast Asia	Cambodia	Kandal	Kien Svay	Kampong Svay	KAMPONG SVAY	11.418	105.055	Dog	N gene	Asian	Mey et al. 2016
KM366333	21/12/2009	Southeast Asia	Cambodia	Kapot	Angkor Chey	Bra Phnom	DAMNAK KRASIANG	10.841	104.655	Dog	N gene	Asian	Mey et al. 2016
KM366334	08/06/1999	Southeast Asia	Cambodia	Kampot	Chhuok	Chhuok	KRASAING	10.834	104.443	Dog	N gene	Asian	Mey et al. 2016
KM366335	16/08/2000	Southeast Asia	Cambodia	Kandal	Angsnoule	Samrong Leu	PREY PCHITH	11.558	104.707	Dog	N gene	Asian	Mey et al. 2016
KM366336	19/04/2010	Southeast Asia	Cambodia	Kapot	Angkor Chey	Phnom Kong	SKOUR TUNG	10.753	104.663	Dog	N gene	Asian	Mey et al. 2016
KM366337	05/09/2009	Southeast Asia	Cambodia	Kg Speu	Thpong	Veal Pon	PREY MICH	11.788	104.609	Dog	N gene	Asian	Mey et al. 2016
KM366338	16/01/2010	Southeast Asia	Cambodia	Kandal	Ang Snuol	Damnak Ampil	TRAPAING TRACH	11.53	104.685	Dog	N gene	Asian	Mey et al. 2016
KM366339	04/02/2004	Southeast Asia	Cambodia	Phnom Penh	Chamkar Mon	Tuol Tom-poung 2	1	11.564	104.924	Dog	N gene	Asian	Mey et al. 2016
KM366340	22/04/2004	Southeast Asia	Cambodia	Phnom Penh	Daun Penh	Sras Chak	NA	11.56	104.923	Dog	N gene	Asian	Mey et al. 2016
KM366341	27/05/2004	Southeast Asia	Cambodia	Phnom Penh	Russeykeo	Chroy Changva	2	11.576	104.936	Dog	N gene	Asian	Mey et al. 2016
KM366342	27/07/2004	Southeast Asia	Cambodia	Phnom Penh	Reussey Keo	Chroy Changva	DOEUM KOR	11.587	104.927	Dog	N gene	Asian	Mey et al. 2016
KM366343	09/06/2004	Southeast Asia	Cambodia	Phnom Penh	Russeykeo	Chroy Changva	4	11.591	104.94	Dog	N gene	Asian	Mey et al. 2016
KM366344	11/04/2004	Southeast Asia	Cambodia	Phnom Penh	Mean Chey	Boeng Tompun	THNOAT CHRUM	11.518	104.909	Dog	N gene	Asian	Mey et al. 2016
KM366345	11/06/2004	Southeast Asia	Cambodia	Phnom Penh	Russey Keo	Prek Leap	PREK LEAP	11.636	104.925	Dog	N gene	Asian	Mey et al. 2016
KM366346	24/03/2005	Southeast Asia	Cambodia	Phnom Penh	Chamkar Mon	Tuolsvayprey1	5	11.559	104.926	Dog	N gene	Asian	Mey et al. 2016
KM366347	22/06/2005	Southeast Asia	Cambodia	Phnom Penh	Mean Chey	Nirot	RUSSEY SROS	11.53	104.949	Dog	N gene	Asian	Mey et al. 2016

Table B5 continued from previous page

KM366348	23/10/2006	Southeast Asia	Cambodia	Phnom Penh	Russeykeo	Toekthla	NA	11.547	104.888	Dog	N gene	Asian	Mey et al. 2016
KM366349	16/12/2004	Southeast Asia	Cambodia	Phnom Penh	Russey Keo	Chroy Changva	2	11.576	104.936	Dog	N gene	Asian	Mey et al. 2016
KM366350	30/07/2004	Southeast Asia	Cambodia	Phnom Penh	Dangkor	Samrong Krorm	TRAPAING THNONG	11.569	104.812	Dog	N gene	Asian	Mey et al. 2016
KX148226	2002	Middle East	Afghanistan	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Arctic	Troupin et al. 2016
KX148197	2015	North Africa	Algeria	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Cosmopolitan	Troupin et al. 2016
KX148216	1986	South America	Brazil	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Cosmopolitan	Troupin et al. 2016
KX148109	1986	South America	Brazil	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Bats	Troupin et al. 2016
KX148217	1986	South America	Brazil	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Cosmopolitan	Troupin et al. 2016
KX148249	1998	Southeast Asia	Cambodia	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Asian	Troupin et al. 2016
KX148252	1999	Southeast Asia	Cambodia	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Asian	Troupin et al. 2016
KX148251	1998	Southeast Asia	Cambodia	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Asian	Troupin et al. 2016
KX148253	1998	Southeast Asia	Cambodia	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Asian	Troupin et al. 2016
KX148250	1999	Southeast Asia	Cambodia	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Asian	Troupin et al. 2016
KX148267	1992	Asia	China	NA	NA	NA	NA	NA	NA	Human Homo sapiens	Genome	Asian	Troupin et al. 2016

Table B5 continued from previous page

GU345747	1986	Asia	China	Ningxia	NA	NA	NA	NA	NA	Human Homo sapiens	Genome	Asian	Troupin et al. 2016
KX148101	1979	Middle East	Egypt	NA	NA	NA	NA	NA	NA	Human Homo sapiens	Genome	Cosmopolitan	Troupin et al. 2016
KX148200	1988	Sub-Saharan Africa	Ethiopia	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Cosmopolitan	Troupin et al. 2016
KX148128	1992	Europe	France	NA	NA	NA	NA	NA	NA	Red fox Vulpes vulpes	Genome	Cosmopolitan	Troupin et al. 2016
KX148268	1990	South America	French Guiana	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Bats	Troupin et al. 2016
KX148105	1980	North America	Greenland	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Arctic	Troupin et al. 2016
KX148160	1991	Europe	Hungary	NA	NA	NA	NA	NA	NA	Human Homo sapiens	Genome	Cosmopolitan	Troupin et al. 2016
KX148138	1993	Europe	Hungary	NA	NA	NA	NA	NA	NA	Red fox Vulpes vulpes	Genome	Cosmopolitan	Troupin et al. 2016
KX148246	1997	Asia	India	NA	NA	NA	NA	NA	NA	Human Homo sapiens	Genome	Indian-Sub	Troupin et al. 2016
AY956319	2004	Asia	India	NA	NA	NA	NA	NA	NA	Human Homo sapiens	Genome	Arctic	Troupin et al. 2016
KX148266	2003	Southeast Asia	Indonesia	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Asian	Troupin et al. 2016
KX148189	1985	Middle East	Iran	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Cosmopolitan	Troupin et al. 2016
KX148159	1974	Middle East	Iran	NA	NA	NA	NA	NA	NA	Sheep Ovis aries	Genome	Cosmopolitan	Troupin et al. 2016

Table B5 continued from previous page

KX148186	1984	Middle East	Iran	NA	NA	NA	NA	NA	NA	Wolf Canis pallipes	Genome	Cosmopolitan	Troupin et al. 2016
KX148190	1976	Middle East	Iran	NA	NA	NA	NA	NA	NA	Jackal Canis aureus	Genome	Cosmopolitan	Troupin et al. 2016
KX148188	1991	Middle East	Iran	NA	NA	NA	NA	NA	NA	Wolf Canis pallipes	Genome	Cosmopolitan	Troupin et al. 2016
KX148185	1996	Middle East	Iran	NA	NA	NA	NA	NA	NA	Wolf Canis pallipes	Genome	Cosmopolitan	Troupin et al. 2016
KX148258	2002	Southeast Asia	Laos	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Asian	Troupin et al. 2016
KX148256	2002	Southeast Asia	Laos	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Asian	Troupin et al. 2016
KX148257	2002	Southeast Asia	Laos	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Asian	Troupin et al. 2016
KX148255	1999	Southeast Asia	Laos	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Asian	Troupin et al. 2016
AB981663	2011	Southeast Asia	Laos	Champasak	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Asian	Troupin et al. 2016
AB981664	2011	Southeast Asia	Laos	Vientiane Capital	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Asian	Troupin et al. 2016
KX148236	1993	North Africa	Mauritania	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Africa-2	Troupin et al. 2016
KX148237	1993	North Africa	Mauritania	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Africa-2	Troupin et al. 2016
KX148112	1991	North America	Mexico	NA	NA	NA	NA	NA	NA	Human Homo sapiens	Genome	Cosmopolitan	Troupin et al. 2016

Table B5 continued from previous page

KX148195	2004	North Africa	Morocco	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Cosmopolitan	Troupin et al. 2016
KX148193	2008	North Africa	Morocco	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Cosmopolitan	Troupin et al. 2016
KX148194	1989	North Africa	Morocco	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Cosmopolitan	Troupin et al. 2016
KX148196	1990	North Africa	Morocco	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Cosmopolitan	Troupin et al. 2016
KF155001	2009	North Africa	Morocco	NA	NA	NA	NA	NA	NA	Cow Bos taurus	Genome	Cosmopolitan	Troupin et al. 2016
KX148247	1999	Southeast Asia	Myanmar	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Asian	Troupin et al. 2016
KX148248	1999	Southeast Asia	Myanmar	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Asian	Troupin et al. 2016
KX148228	1998	Asia	Nepal	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Arctic	Troupin et al. 2016
KX148201	1986	Sub-Saharan Africa	Nigeria	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Cosmopolitan	Troupin et al. 2016
KX148170	1990	Middle East	Oman	NA	NA	NA	NA	NA	NA	Fox Und.	Genome	Cosmopolitan	Troupin et al. 2016
HE802675	2010	Asia	Pakistan	NA	NA	NA	NA	NA	NA	Cow Bos taurus	Genome	Arctic	Troupin et al. 2016
KX148260	2004	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Asian	Troupin et al. 2016
KX148259	1994	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Asian	Troupin et al. 2016
KX148263	1994	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Asian	Troupin et al. 2016

Table B5 continued from previous page

KX148262	1994	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Asian	Troupin et al. 2016
KX148261	1994	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Asian	Troupin et al. 2016
KX148114	1995	Europe	Poland	NA	NA	NA	NA	NA	NA	Red fox Vulpes vulpes	Genome	Cosmopolitan	Troupin et al. 2016
KX148168	1987	Middle East	Saudi Arabia	NA	NA	NA	NA	NA	NA	Fox Und.	Genome	Cosmopolitan	Troupin et al. 2016
KX148198	1993	Sub-Saharan Africa	Somalia	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Cosmopolitan	Troupin et al. 2016
KX148103	1981	Sub-Saharan Africa	South Africa	NA	NA	NA	NA	NA	NA	Human Homo sapiens	Genome	Cosmopolitan	Troupin et al. 2016
EU293111	1983	Southeast Asia	Thailand	NA	NA	NA	NA	NA	NA	Human Homo sapiens	Genome	Asian	Troupin et al. 2016
KX148166	1993	Middle East	Turkey	NA	NA	NA	NA	NA	NA	Dog Canis familiaris	Genome	Cosmopolitan	Troupin et al. 2016
AY705373	1983	North America	USA	NA	NA	NA	NA	NA	NA	Human Homo sapiens	Genome	Bats	Troupin et al. 2016
FJ392392	1980	Sub-Saharan Africa	Namibia	NA	NA	NA	NA	NA	NA	Mongoose	N gene	unknown	Troupin et al. 2016
FJ392382	2002	Sub-Saharan Africa	South Africa	NA	NA	NA	NA	NA	NA	Mongoose	N gene	unknown	Troupin et al. 2016
FJ392378	1999	Sub-Saharan Africa	South Africa	NA	NA	NA	NA	NA	NA	Mongoose	N gene	Africa-3	Troupin et al. 2016
FJ392376	2001	Sub-Saharan Africa	South Africa	NA	NA	NA	NA	NA	NA	Mongoose	N gene	Africa-3	Troupin et al. 2016

Table B5 continued from previous page

AY854576	2000	Caribbean	Cuba	NA	NA	NA	NA	NA	NA	Mongoose	N gene	Cosmopolitan	Troupin et al. 2016
AY854575	2000	Caribbean	Cuba	NA	NA	NA	NA	NA	NA	Mongoose	N gene	Cosmopolitan	Troupin et al. 2016
AY854567	2000	Caribbean	Cuba	NA	NA	NA	NA	NA	NA	Mongoose	N gene	Cosmopolitan	Troupin et al. 2016
AY854574	2001	Caribbean	Cuba	NA	NA	NA	NA	NA	NA	Mongoose	N gene	Cosmopolitan	Troupin et al. 2016
AY854561	2002	Caribbean	Cuba	NA	NA	NA	NA	NA	NA	Mongoose	N gene	Cosmopolitan	Troupin et al. 2016
AY854578	2002	Caribbean	Cuba	NA	NA	NA	NA	NA	NA	Mongoose	N gene	Cosmopolitan	Troupin et al. 2016
AY854552	2000	Caribbean	Cuba	NA	NA	NA	NA	NA	NA	Mongoose	N gene	Cosmopolitan	Troupin et al. 2016
AY854559	2002	Caribbean	Cuba	NA	NA	NA	NA	NA	NA	Mongoose	N gene	Cosmopolitan	Troupin et al. 2016
AY854545	2000	Caribbean	Cuba	NA	NA	NA	NA	NA	NA	Mongoose	N gene	Cosmopolitan	Troupin et al. 2016
AY854530	2000	Caribbean	Cuba	NA	NA	NA	NA	NA	NA	Mongoose	N gene	Cosmopolitan	Troupin et al. 2016
AY854531	2002	Caribbean	Cuba	NA	NA	NA	NA	NA	NA	Mongoose	N gene	Cosmopolitan	Troupin et al. 2016
AY854525	2001	Caribbean	Cuba	NA	NA	NA	NA	NA	NA	Mongoose	N gene	Cosmopolitan	Troupin et al. 2016
AY854506	2001	Caribbean	Cuba	NA	NA	NA	NA	NA	NA	Mongoose	N gene	Cosmopolitan	Troupin et al. 2016
AY854515	2001	Caribbean	Cuba	NA	NA	NA	NA	NA	NA	Mongoose	N gene	Cosmopolitan	Troupin et al. 2016
KJ957437	2011	Caribbean	Grenada	NA	NA	NA	NA	NA	NA	Mongoose	N gene	Cosmopolitan	Troupin et al. 2016
KJ957434	2011	Caribbean	Grenada	NA	NA	NA	NA	NA	NA	Mongoose	N gene	Cosmopolitan	Troupin et al. 2016
KM067274	2012	Caribbean	Grenada	NA	NA	NA	NA	NA	NA	Mongoose	N gene	Cosmopolitan	Troupin et al. 2016
KJ957444	2012	Caribbean	Grenada	NA	NA	NA	NA	NA	NA	Mongoose	N gene	Cosmopolitan	Troupin et al. 2016
KJ957449	2013	Caribbean	Grenada	NA	NA	NA	NA	NA	NA	Mongoose	N gene	Cosmopolitan	Troupin et al. 2016

Table B5 continued from previous page

KJ957450	2013	Caribbean	Grenada	NA	NA	NA	NA	NA	NA	Mongoose	N gene	Cosmopolitan	Troupin et al. 2016
KJ957440	2011	Caribbean	Grenada	NA	NA	NA	NA	NA	NA	Mongoose	N gene	Cosmopolitan	Troupin et al. 2016
FJ228497	2006	Caribbean	Puerto Rico	NA	NA	NA	NA	NA	NA	Mongoose	N gene	Cosmopolitan	Troupin et al. 2016
JQ513538	1997	Caribbean	Puerto Rico	NA	NA	NA	NA	NA	NA	Mongoose	N gene	Cosmopolitan	Troupin et al. 2016
FJ228495	2004	Caribbean	Puerto Rico	NA	NA	NA	NA	NA	NA	Mongoose	N gene	Cosmopolitan	Troupin et al. 2016
FJ228496	2006	Caribbean	Puerto Rico	NA	NA	NA	NA	NA	NA	Mongoose	N gene	Cosmopolitan	Troupin et al. 2016
JQ513537	1997	Caribbean	Puerto Rico	NA	NA	NA	NA	NA	NA	Mongoose	N gene	Cosmopolitan	Troupin et al. 2016
JQ513528	1996	Caribbean	Puerto Rico	NA	NA	NA	NA	NA	NA	Mongoose	N gene	Cosmopolitan	Troupin et al. 2016
JQ513536	1997	Caribbean	Puerto Rico	NA	NA	NA	NA	NA	NA	Mongoose	N gene	Cosmopolitan	Troupin et al. 2016
KP860152	01/08/2013	Asia	Taiwan	Nantou	Nantou	NA	NA	NA	NA	Ferret Badger	N gene	Asian	Troupin et al. 2016
A20	23/02/2016	Southeast Asia	Cambodia	NA	NA	Prey Kabbas Ko	NA	11.095	104.952	Dog	Genome	Asian	Cambodian project
a31	28/03/2016	Southeast Asia	Cambodia	NA	NA	Trapeang Thma	NA	10.803	104.296	Dog	Genome	Asian	Cambodian project
A35	04/04/2016	Southeast Asia	Cambodia	NA	NA	Prey Rumduol Khang Cheung	NA	11.132	104.524	Dog	Genome	Asian	Cambodian project
A39	26/04/2016	Southeast Asia	Cambodia	NA	NA	Dot Kambaor	NA	11.402	104.763	Dog	Genome	Asian	Cambodian project
A41	03/05/2016	Southeast Asia	Cambodia	NA	NA	Ta Ei	NA	11.052	104.956	Dog	Genome	Asian	Cambodian project
A51	19/05/2016	Southeast Asia	Cambodia	NA	NA	Prey Preah Andoung	NA	11.39	105.589	Dog	Genome	Asian	Cambodian project
A56	26/05/2016	Southeast Asia	Cambodia	NA	NA	Preaek Ta Meak	NA	11.746	105.01	Dog	Genome	Asian	Cambodian project
A57	27/05/2016	Southeast Asia	Cambodia	NA	NA	Tram Sasar	NA	11.195	104.566	Dog	Genome	Asian	Cambodian project
A75	06/07/2016	Southeast Asia	Cambodia	NA	NA	Prey Phdau Tboung	NA	11.137	104.872	Dog	Genome	Asian	Cambodian project

Table B5 continued from previous page

A77	08/07/2016	Southeast Asia	Cambodia	NA	NA	Thnal Totueng	NA	11.495	104.667	Dog	Genome	Asian	Cambodian project
A83	11/08/2016	Southeast Asia	Cambodia	NA	NA	Pratheat	NA	11.04	105.064	Dog	Genome	Asian	Cambodian project
A85	13/08/2016	Southeast Asia	Cambodia	NA	NA	Trapeang Skon	NA	11.325	105.533	Dog	Genome	Asian	Cambodian project
A90	20/08/2016	Southeast Asia	Cambodia	NA	NA	Prohut	NA	11.719	105.241	Dog	Genome	Asian	Cambodian project
a091	22/08/2016	Southeast Asia	Cambodia	NA	NA	Borei Kam-meakkar	NA	11.511	104.762	Dog	Genome	Asian	Cambodian project
a093	23/08/2016	Southeast Asia	Cambodia	NA	NA	Tuol Kh-long	NA	11.1	104.491	Dog	Genome	Asian	Cambodian project
A96	29/08/2016	Southeast Asia	Cambodia	NA	NA	Prey Ba Srei	NA	11.139	105.356	Dog	Genome	Asian	Cambodian project
A100	09/09/2016	Southeast Asia	Cambodia	NA	NA	Snao Touch	NA	10.761	104.399	Dog	Genome	Asian	Cambodian project
A101	13/09/2016	Southeast Asia	Cambodia	NA	NA	Khnar Ta Nong	NA	11.205	104.559	Dog	Genome	Asian	Cambodian project
A103	19/09/2016	Southeast Asia	Cambodia	NA	NA	Trapeang Prei	NA	11.212	104.51	Dog	Genome	Asian	Cambodian project
A104	19/09/2016	Southeast Asia	Cambodia	NA	NA	Prey Dang Tuek	NA	11.401	104.712	Dog	Genome	Asian	Cambodian project
a106	21/09/2016	Southeast Asia	Cambodia	NA	NA	Tumnob	NA	12.816	105.169	Dog	Genome	Asian	Cambodian project
a109	22/09/2016	Southeast Asia	Cambodia	NA	NA	Mream Khang Tboung	NA	11.35	105.672	Dog	Genome	Asian	Cambodian project
A111	29/09/2016	Southeast Asia	Cambodia	NA	NA	Khnach Khang Tboung	NA	11.13	104.704	Dog	Genome	Asian	Cambodian project
A114	06/10/2016	Southeast Asia	Cambodia	NA	NA	Srangae Cheung	NA	12.053	105.161	Dog	Genome	Asian	Cambodian project
A116	08/10/2016	Southeast Asia	Cambodia	NA	NA	Prasat	NA	11.148	105.543	Dog	Genome	Asian	Cambodian project
A121	18/10/2016	Southeast Asia	Cambodia	NA	NA	Ballangk Khang Lech	NA	12.698	104.906	Dog	Genome	Asian	Cambodian project
a131	18/11/2016	Southeast Asia	Cambodia	NA	NA	Krasang Chi Meae	NA	11.333	104.724	Dog	Genome	Asian	Cambodian project

Table B5 continued from previous page

A133	18/11/2016	Southeast Asia	Cambodia	NA	NA	Prey Ta Mean	NA	11.305	104.677	Dog	Genome	Asian	Cambodian project
A137	25/11/2016	Southeast Asia	Cambodia	NA	NA	Veang	NA	11.325	105.546	Dog	Genome	Asian	Cambodian project
A148	14/12/2016	Southeast Asia	Cambodia	NA	NA	Kakruos	NA	10.965	105.863	Dog	Genome	Asian	Cambodian project
A149	15/12/2016	Southeast Asia	Cambodia	NA	NA	Trapeang Tuk	NA	11.146	104.619	Dog	Genome	Asian	Cambodian project
S70	12/01/2017	Southeast Asia	Cambodia	NA	NA	Phum Prey Tear	NA	11.515	104.841	Dog	Genome	Asian	Cambodian project
S18	16/01/2017	Southeast Asia	Cambodia	NA	NA	Angk Svay	NA	11.021	105.917	Dog	Genome	Asian	Cambodian project
S06	16/01/2017	Southeast Asia	Cambodia	NA	NA	Doun Ov	NA	10.925	104.441	Dog	Genome	Asian	Cambodian project
S43	24/01/2017	Southeast Asia	Cambodia	NA	NA	Chong Ampil	NA	11.664	105.424	Dog	Genome	Asian	Cambodian project
S58	26/01/2017	Southeast Asia	Cambodia	NA	NA	Phnom Penh 3	NA	11.615	104.898	Dog	Genome	Asian	Cambodian project
S28	30/01/2017	Southeast Asia	Cambodia	NA	NA	Prey Sniet	NA	11.623	105.253	Dog	Genome	Asian	Cambodian project
S40	31/01/2017	Southeast Asia	Cambodia	NA	NA	Buor	NA	10.623	104.757	Dog	Genome	Asian	Cambodian project
S758	01/02/2017	Southeast Asia	Cambodia	NA	NA	Ta Kaev Ti Pir	NA	11.664	105.07	Dog	Genome	Asian	Cambodian project
S24	01/03/2017	Southeast Asia	Cambodia	NA	NA	Trapenang Antong	NA	11.668	104.645	Dog	Genome	Asian	Cambodian project
S29	01/03/2017	Southeast Asia	Cambodia	NA	NA	Pra Boeng	NA	12.091	105.051	Dog	Genome	Asian	Cambodian project
S50	06/03/2017	Southeast Asia	Cambodia	NA	NA	Boeng Trav	NA	11.962	105.163	Dog	Genome	Asian	Cambodian project
S47	29/03/2017	Southeast Asia	Cambodia	NA	NA	Toap Sdach	NA	10.707	104.617	Dog	Genome	Asian	Cambodian project
S35	17/04/2017	Southeast Asia	Cambodia	NA	NA	Meanok Kaeut	NA	11.873	104.627	Dog	Genome	Asian	Cambodian project
S81	19/04/2017	Southeast Asia	Cambodia	NA	NA	Prey Andoung	NA	11.195	105.545	Dog	Genome	Asian	Cambodian project
S02	11/05/2017	Southeast Asia	Cambodia	NA	NA	Phum Ti Prammuoy	NA	11.991	105.459	Dog	Genome	Asian	Cambodian project

Table B5 continued from previous page

S03	08/05/2017	Southeast Asia	Cambodia	NA	NA	Samraong	NA	12.407	105.09	Dog	Genome	Asian	Cambodian project
S64	24/07/2017	Southeast Asia	Cambodia	NA	NA	Prey Khmaer	NA	11.352	104.688	Dog	Genome	Asian	Cambodian project
S806	31/07/2017	Southeast Asia	Cambodia	NA	NA	Kampong Luong	NA	10.993	104.978	Dog	Genome	Asian	Cambodian project
S09	07/08/2017	Southeast Asia	Cambodia	NA	NA	Pou	NA	11.793	105.086	Dog	Genome	Asian	Cambodian project
S2716	22/08/2017	Southeast Asia	Cambodia	NA	NA	Khlouk	NA	11.2	104.514	Dog	Genome	Asian	Cambodian project
S48	24/08/2017	Southeast Asia	Cambodia	NA	NA	Tranh Veaeng	NA	11.75	104.425	Dog	Genome	Asian	Cambodian project
S13	28/08/2017	Southeast Asia	Cambodia	NA	NA	Oknha Tep	NA	11.473	104.663	Dog	Genome	Asian	Cambodian project
S818	11/09/2017	Southeast Asia	Cambodia	NA	NA	Phsar Daek Kraom	NA	11.807	104.771	Dog	Genome	Asian	Cambodian project
S716	12/09/2017	Southeast Asia	Cambodia	NA	NA	Prey Cham-bak	NA	11.199	104.856	Dog	Genome	Asian	Cambodian project
S745	27/09/2017	Southeast Asia	Cambodia	NA	NA	Boeng Trav	NA	11.962	105.163	Dog	Genome	Asian	Cambodian project
S6616	06/10/2017	Southeast Asia	Cambodia	NA	NA	Phnom Penh	NA	11.559	104.918	Dog	Genome	Asian	Cambodian project
S718	13/11/2017	Southeast Asia	Cambodia	NA	NA	Prasaeur	NA	12.176	105.321	Dog	Genome	Asian	Cambodian project
S616	21/11/2017	Southeast Asia	Cambodia	NA	NA	Prey Chek	NA	11.309	105.649	Dog	Genome	Asian	Cambodian project
S681	05/12/2017	Southeast Asia	Cambodia	NA	NA	Tean Phle-ung	NA	11.554	105.63	Dog	Genome	Asian	Cambodian project
S82	12/12/2017	Southeast Asia	Cambodia	NA	NA	Trapeang Lbaeuk	NA	10.783	104.357	Dog	Genome	Asian	Cambodian project
S664	15/12/2017	Southeast Asia	Cambodia	NA	NA	Chantum	NA	11.85	105.637	Dog	Genome	Asian	Cambodian project
S53	20/12/2017	Southeast Asia	Cambodia	NA	NA	Prey Chek	NA	11.309	105.649	Dog	Genome	Asian	Cambodian project
W06	11/01/2012	Southeast Asia	Cambodia	NA	NA	Kbal Dam-rei	NA	11.291	105.44	Dog	Genome	Asian	Cambodian project
W008	12/01/2012	Southeast Asia	Cambodia	NA	NA	Boeng	NA	10.795	104.42	Dog	Genome	Asian	Cambodian project

Table B5 continued from previous page

W15	20/01/2012	Southeast Asia	Cambodia	NA	NA	Trapeang Khnar	NA	10.982	104.671	Dog	Genome	Asian	Cambodian project
W18	27/01/2012	Southeast Asia	Cambodia	NA	NA	Noreay	NA	11.141	104.653	Dog	Genome	Asian	Cambodian project
W19	30/01/2012	Southeast Asia	Cambodia	NA	NA	Veal Ampil	NA	12.348	105.132	Dog	Genome	Asian	Cambodian project
W27	06/02/2012	Southeast Asia	Cambodia	NA	NA	Preaek Pok	NA	10.676	104.454	Dog	Genome	Asian	Cambodian project
W34	15/02/2012	Southeast Asia	Cambodia	NA	NA	Prey Totueng	NA	11.403	104.788	Dog	Genome	Asian	Cambodian project
W36	20/02/2012	Southeast Asia	Cambodia	NA	NA	Traeuy Trak	NA	11.274	104.681	Dog	Genome	Asian	Cambodian project
W37	20/02/2012	Southeast Asia	Cambodia	NA	NA	Reaksmei Rumdaoh	NA	12.243	105.138	Dog	Genome	Asian	Cambodian project
W51	06/03/2012	Southeast Asia	Cambodia	NA	NA	Svay Pear	NA	11.078	104.924	Dog	Genome	Asian	Cambodian project
W53	07/03/2012	Southeast Asia	Cambodia	NA	NA	Daeum Chrey	NA	11.845	104.991	Dog	Genome	Asian	Cambodian project
W54	09/03/2012	Southeast Asia	Cambodia	NA	NA	Kaev Chamraeun	NA	11.166	104.874	Dog	Genome	Asian	Cambodian project
W58	15/03/2012	Southeast Asia	Cambodia	NA	NA	Krang Ta Char	NA	11.718	104.531	Dog	Genome	Asian	Cambodian project
W65	25/03/2012	Southeast Asia	Cambodia	NA	NA	Thmei	NA	11.72	104.524	Dog	Genome	Asian	Cambodian project
W76	09/04/2012	Southeast Asia	Cambodia	NA	NA	Khloy Kaeut	NA	12.241	105.138	Dog	Genome	Asian	Cambodian project
W098	30/04/2012	Southeast Asia	Cambodia	NA	NA	Trapeang Reang	NA	10.781	104.39	Dog	Genome	Asian	Cambodian project
W106	08/05/2012	Southeast Asia	Cambodia	NA	NA	Dambouk Khpos	NA	11.206	104.24	Dog	Genome	Asian	Cambodian project
W107	10/05/2012	Southeast Asia	Cambodia	NA	NA	Kampong Popil	NA	11.653	105.141	Dog	Genome	Asian	Cambodian project
W154	26/06/2012	Southeast Asia	Cambodia	NA	NA	Angkor Phdiek	NA	11.124	104.756	Dog	Genome	Asian	Cambodian project
W162	04/07/2012	Southeast Asia	Cambodia	NA	NA	Klaeng Poar	NA	12.016	105.237	Dog	Genome	Asian	Cambodian project
W190	06/08/2012	Southeast Asia	Cambodia	NA	NA	Ba Baong	NA	11.386	105.328	Dog	Genome	Asian	Cambodian project

Table B5 continued from previous page

W196	20/08/2012	Southeast Asia	Cambodia	NA	NA	Kok Ta Rie	NA	11.211	104.788	Dog	Genome	Asian	Cambodian project
W202	28/08/2012	Southeast Asia	Cambodia	NA	NA	Kampong Reab	NA	11.134	105.017	Dog	Genome	Asian	Cambodian project
W225	26/09/2012	Southeast Asia	Cambodia	NA	NA	Prey Sralet	NA	11.723	105.218	Dog	Genome	Asian	Cambodian project
W233	05/10/2012	Southeast Asia	Cambodia	NA	NA	Svay Damnak	NA	11.81	105.044	Dog	Genome	Asian	Cambodian project
W236	08/10/2012	Southeast Asia	Cambodia	NA	NA	Ta Teaen	NA	10.897	104.483	Dog	Genome	Asian	Cambodian project
W256	07/11/2012	Southeast Asia	Cambodia	NA	NA	Samraong	NA	11.15	104.896	Dog	Genome	Asian	Cambodian project
W261	12/11/2012	Southeast Asia	Cambodia	NA	NA	Chhuk Kieb	NA	11.154	104.463	Dog	Genome	Asian	Cambodian project
W262	13/11/2012	Southeast Asia	Cambodia	NA	NA	Veal	NA	11.834	105.06	Dog	Genome	Asian	Cambodian project
W268	16/11/2012	Southeast Asia	Cambodia	NA	NA	Prey Ta Phem	NA	11.138	104.662	Dog	Genome	Asian	Cambodian project
W277	26/11/2012	Southeast Asia	Cambodia	NA	NA	Ta Kay	NA	11.897	105.167	Dog	Genome	Asian	Cambodian project
W278	26/11/2012	Southeast Asia	Cambodia	NA	NA	Kdei Ta Hok	NA	11.12	104.942	Dog	Genome	Asian	Cambodian project
W290	12/12/2012	Southeast Asia	Cambodia	NA	NA	Doun Tao	NA	12.195	105.107	Dog	Genome	Asian	Cambodian project
W295	20/12/2012	Southeast Asia	Cambodia	NA	NA	Sala Lekh Pram	NA	11.934	104.723	Dog	Genome	Asian	Cambodian project
W304	31/12/2012	Southeast Asia	Cambodia	NA	NA	Prey Khley	NA	11.016	104.458	Dog	Genome	Asian	Cambodian project
X01	03/01/2013	Southeast Asia	Cambodia	NA	NA	Preaek Ta Tep	NA	11.63	104.959	Dog	Genome	Asian	Cambodian project
X02	03/01/2013	Southeast Asia	Cambodia	NA	NA	Boeng Kdol	NA	11.716	104.847	Dog	Genome	Asian	Cambodian project
X04	08/01/2013	Southeast Asia	Cambodia	NA	NA	Prey Khcheay	NA	10.961	104.441	Dog	Genome	Asian	Cambodian project
X13	12/01/2013	Southeast Asia	Cambodia	NA	NA	Trapeang Sya	NA	11.23	104.599	Dog	Genome	Asian	Cambodian project
X19	16/01/2013	Southeast Asia	Cambodia	NA	NA	Phsar Slab Leaeng	NA	11.212	104.585	Dog	Genome	Asian	Cambodian project

Table B5 continued from previous page

X23	21/01/2013	Southeast Asia	Cambodia	NA	NA	Lvea Sa Leu	NA	11.458	105.184	Dog	Genome	Asian	Cambodian project
X24	23/01/2013	Southeast Asia	Cambodia	NA	NA	Damnak Pring Lech	NA	11.84	105.42	Dog	Genome	Asian	Cambodian project
X38	11/02/2013	Southeast Asia	Cambodia	NA	NA	Thmei	NA	10.898	104.44	Dog	Genome	Asian	Cambodian project
X40	12/02/2013	Southeast Asia	Cambodia	NA	NA	Thma Sa	NA	11.291	104.76	Dog	Genome	Asian	Cambodian project
X54	26/02/2013	Southeast Asia	Cambodia	NA	NA	Stueng Chveng	NA	11.918	105.006	Dog	Genome	Asian	Cambodian project
X67	07/03/2013	Southeast Asia	Cambodia	NA	NA	Krang Doung	NA	11.742	104.54	Dog	Genome	Asian	Cambodian project
X73	13/03/2013	Southeast Asia	Cambodia	NA	NA	Speam Thmei	NA	11.841	104.98	Dog	Genome	Asian	Cambodian project
X75	14/03/2013	Southeast Asia	Cambodia	NA	NA	Chakto Tis	NA	11.813	104.775	Dog	Genome	Asian	Cambodian project
X81	18/03/2013	Southeast Asia	Cambodia	NA	NA	Angk Sangkae	NA	11.114	104.946	Dog	Genome	Asian	Cambodian project
X89	25/03/2013	Southeast Asia	Cambodia	NA	NA	Thum Thmei	NA	10.711	104.436	Dog	Genome	Asian	Cambodian project
X91	28/03/2013	Southeast Asia	Cambodia	NA	NA	Boeng	NA	11.336	105.636	Dog	Genome	Asian	Cambodian project
X96	08/04/2013	Southeast Asia	Cambodia	NA	NA	Sangkaeub	NA	11.882	104.896	Dog	Genome	Asian	Cambodian project
X104	23/04/2013	Southeast Asia	Cambodia	NA	NA	Andoung Tuek	NA	11.186	103.473	Dog	Genome	Asian	Cambodian project
X108	26/04/2013	Southeast Asia	Cambodia	NA	NA	Phum Damnak Thum	NA	11.537	104.897	Dog	Genome	Asian	Cambodian project
X112	29/04/2013	Southeast Asia	Cambodia	NA	NA	Kandal	NA	11.336	105.473	Dog	Genome	Asian	Cambodian project
X125	13/05/2013	Southeast Asia	Cambodia	NA	NA	Varint Ti Bei	NA	11.895	105.133	Dog	Genome	Asian	Cambodian project
X146	21/06/2013	Southeast Asia	Cambodia	NA	NA	Angk Roung	NA	11.247	104.698	Dog	Genome	Asian	Cambodian project
X153	02/07/2013	Southeast Asia	Cambodia	NA	NA	Prey Kab-bas Ka	NA	11.1	104.949	Dog	Genome	Asian	Cambodian project
X186	16/08/2013	Southeast Asia	Cambodia	NA	NA	Svay Antor Ti Muoy	NA	11.587	105.422	Dog	Genome	Asian	Cambodian project

Table B5 continued from previous page

X199	29/08/2013	Southeast Asia	Cambodia	NA	NA	Mondol Bei	NA	13.369	103.868	Dog	Genome	Asian	Cambodian project
X200	29/08/2013	Southeast Asia	Cambodia	NA	NA	Bat Doeng	NA	11.681	104.701	Dog	Genome	Asian	Cambodian project
X218	23/09/2013	Southeast Asia	Cambodia	NA	NA	Lvea Kraom	NA	11.86	105.017	Dog	Genome	Asian	Cambodian project
X246	30/10/2013	Southeast Asia	Cambodia	NA	NA	Tras	NA	12.488	105.139	Dog	Genome	Asian	Cambodian project
X250	01/11/2013	Southeast Asia	Cambodia	NA	NA	Roka Ar	NA	11.852	105.094	Dog	Genome	Asian	Cambodian project
X259	05/11/2013	Southeast Asia	Cambodia	NA	NA	Sampong Chey	NA	12.185	105.13	Dog	Genome	Asian	Cambodian project
X266	12/11/2013	Southeast Asia	Cambodia	NA	NA	Chong Doung	NA	12.456	105.114	Dog	Genome	Asian	Cambodian project
X273	29/11/2013	Southeast Asia	Cambodia	NA	NA	Preaek Thaong	NA	11.628	104.956	Dog	Genome	Asian	Cambodian project
X274	29/11/2013	Southeast Asia	Cambodia	NA	NA	Thma Da	NA	12.187	105.187	Dog	Genome	Asian	Cambodian project
X279	09/12/2013	Southeast Asia	Cambodia	NA	NA	Ampil Rung	NA	11.773	104.75	Dog	Genome	Asian	Cambodian project
X295	28/12/2013	Southeast Asia	Cambodia	NA	NA	Ou Kandaol Tboung	NA	11.5	105.442	Dog	Genome	Asian	Cambodian project
Y11	14/01/2014	Southeast Asia	Cambodia	NA	NA	Thnal	NA	12.291	106.326	Dog	Genome	Asian	Cambodian project
Y17	22/01/2014	Southeast Asia	Cambodia	NA	NA	Khnar	NA	11.19	104.555	Dog	Genome	Asian	Cambodian project
Y19	22/01/2014	Southeast Asia	Cambodia	NA	NA	Khsach Prachheh Leu	NA	12.293	105.615	Dog	Genome	Asian	Cambodian project
Y025	28/01/2014	Southeast Asia	Cambodia	NA	NA	Trapeang Chak	NA	11.237	104.646	Dog	Genome	Asian	Cambodian project
Y030	06/02/2014	Southeast Asia	Cambodia	NA	NA	Trapeang Teab	NA	11.252	104.651	Dog	Genome	Asian	Cambodian project
Y40	12/02/2014	Southeast Asia	Cambodia	NA	NA	Sdok	NA	11.704	104.59	Dog	Genome	Asian	Cambodian project
Y52	24/02/2014	Southeast Asia	Cambodia	NA	NA	Ba Reach	NA	11.367	105.481	Dog	Genome	Asian	Cambodian project

Table B5 continued from previous page

Y65	14/03/2014	Southeast Asia	Cambodia	NA	NA	Keatha Vong Kraom	NA	10.719	104.598	Dog	Genome	Asian	Cambodian project
Y79	08/04/2014	Southeast Asia	Cambodia	NA	NA	Khnach Khang Cheung	NA	11.139	104.707	Dog	Genome	Asian	Cambodian project
Y82	17/04/2014	Southeast Asia	Cambodia	NA	NA	Kaoh Roka	NA	11.754	105.005	Dog	Genome	Asian	Cambodian project
Y086	25/04/2014	Southeast Asia	Cambodia	NA	NA	Sramaoch Haer	NA	11.218	104.824	Dog	Genome	Asian	Cambodian project
Y98	27/05/2014	Southeast Asia	Cambodia	NA	NA	Chong	NA	11.884	104.917	Dog	Genome	Asian	Cambodian project
Y106	09/06/2014	Southeast Asia	Cambodia	NA	NA	Svay Teab	NA	11.72	105.442	Dog	Genome	Asian	Cambodian project
Y110	11/06/2014	Southeast Asia	Cambodia	NA	NA	Trach Koh	NA	11.224	104.706	Dog	Genome	Asian	Cambodian project
Y113	25/06/2014	Southeast Asia	Cambodia	NA	NA	Trapeang Ta Sek	NA	10.769	104.502	Dog	Genome	Asian	Cambodian project
Y125	15/07/2014	Southeast Asia	Cambodia	NA	NA	Peam Ta Aek	NA	11.546	105.029	Dog	Genome	Asian	Cambodian project
Y127	16/07/2014	Southeast Asia	Cambodia	NA	NA	Kampong Sambuor	NA	13.127	103.22	Dog	Genome	Asian	Cambodian project
Y137	29/07/2014	Southeast Asia	Cambodia	NA	NA	Prey Tbaeng	NA	11.604	105.511	Dog	Genome	Asian	Cambodian project
Y172	08/09/2014	Southeast Asia	Cambodia	NA	NA	Pnov	NA	12.363	105.097	Dog	Genome	Asian	Cambodian project
Y173	08/09/2014	Southeast Asia	Cambodia	NA	NA	Thmei	NA	11.792	104.941	Dog	Genome	Asian	Cambodian project
Y175	11/09/2014	Southeast Asia	Cambodia	NA	NA	Sameakki	NA	11.795	104.979	Dog	Genome	Asian	Cambodian project
Y189	06/10/2014	Southeast Asia	Cambodia	NA	NA	Snay Poal	NA	11.661	105.221	Dog	Genome	Asian	Cambodian project
Y198	22/10/2014	Southeast Asia	Cambodia	NA	NA	Thmei	NA	12.024	105.142	Dog	Genome	Asian	Cambodian project
Y204	31/10/2014	Southeast Asia	Cambodia	NA	NA	Trapeang Traeunh	NA	11.709	104.394	Dog	Genome	Asian	Cambodian project
Y209	14/11/2014	Southeast Asia	Cambodia	NA	NA	Chheu Teal Chrum	NA	10.836	104.531	Dog	Genome	Asian	Cambodian project

Table B5 continued from previous page

Y210	17/11/2014	Southeast Asia	Cambodia	NA	NA	Tang Thlaeung	NA	12.027	104.94	Dog	Genome	Asian	Cambodian project
Y214	19/11/2014	Southeast Asia	Cambodia	NA	NA	Skon	NA	12.054	105.079	Dog	Genome	Asian	Cambodian project
Y226	06/12/2014	Southeast Asia	Cambodia	NA	NA	Reay Pay Leu	NA	11.936	105.092	Dog	Genome	Asian	Cambodian project
Y228	11/12/2014	Southeast Asia	Cambodia	NA	NA	Roung Damrei	NA	11.257	105.386	Dog	Genome	Asian	Cambodian project
Y229	11/12/2014	Southeast Asia	Cambodia	NA	NA	Ta Pov	NA	11.883	104.803	Dog	Genome	Asian	Cambodian project
Y236	18/12/2014	Southeast Asia	Cambodia	NA	NA	Sameakki	NA	11.001	104.973	Dog	Genome	Asian	Cambodian project
Y247	30/12/2014	Southeast Asia	Cambodia	NA	NA	Ta Poy	NA	11.888	104.915	Dog	Genome	Asian	Cambodian project
Z04	12/01/2015	Southeast Asia	Cambodia	NA	NA	Krouch	NA	12.23	105.144	Dog	Genome	Asian	Cambodian project
Z18	02/02/2015	Southeast Asia	Cambodia	NA	NA	Sanlung	NA	11.344	105.513	Dog	Genome	Asian	Cambodian project
Z27	25/02/2015	Southeast Asia	Cambodia	NA	NA	Thmei	NA	11.458	105.431	Dog	Genome	Asian	Cambodian project
Z43	16/03/2015	Southeast Asia	Cambodia	NA	NA	Thkov	NA	11.117	104.826	Dog	Genome	Asian	Cambodian project
Z44	17/03/2015	Southeast Asia	Cambodia	NA	NA	Snae Rean	NA	11.345	105.452	Dog	Genome	Asian	Cambodian project
Z46	23/03/2015	Southeast Asia	Cambodia	NA	NA	Prey Lieb	NA	10.588	104.785	Dog	Genome	Asian	Cambodian project
Z47	24/03/2015	Southeast Asia	Cambodia	NA	NA	Khley	NA	10.898	104.451	Dog	Genome	Asian	Cambodian project
Z52	08/04/2015	Southeast Asia	Cambodia	NA	NA	Svay Ming	NA	11.424	104.82	Dog	Genome	Asian	Cambodian project
Z58	28/04/2015	Southeast Asia	Cambodia	NA	NA	Srah Ta Oem	NA	11.284	105.654	Dog	Genome	Asian	Cambodian project
Z59	29/04/2015	Southeast Asia	Cambodia	NA	NA	Chek	NA	11.449	105.518	Dog	Genome	Asian	Cambodian project
Z63	07/05/2015	Southeast Asia	Cambodia	NA	NA	Prey Khmau	NA	10.951	104.476	Dog	Genome	Asian	Cambodian project
Z65	08/05/2015	Southeast Asia	Cambodia	NA	NA	Krang Thum	NA	11.449	104.595	Dog	Genome	Asian	Cambodian project

Table B5 continued from previous page

Z66	08/05/2015	Southeast Asia	Cambodia	NA	NA	Bang Kokir	NA	10.582	103.761	Dog	Genome	Asian	Cambodian project
Z70	18/05/2015	Southeast Asia	Cambodia	NA	NA	Knaor Dambang	NA	12.016	105.018	Dog	Genome	Asian	Cambodian project
Z75	22/05/2015	Southeast Asia	Cambodia	NA	NA	Preaek Chik	NA	11.964	105.44	Dog	Genome	Asian	Cambodian project
Z76	28/05/2015	Southeast Asia	Cambodia	NA	NA	Ruessei Thlok	NA	11.058	105.379	Dog	Genome	Asian	Cambodian project
Z81	04/06/2015	Southeast Asia	Cambodia	NA	NA	Trapeang Dar	NA	11.256	104.634	Dog	Genome	Asian	Cambodian project
Z96	06/07/2015	Southeast Asia	Cambodia	NA	NA	Prey Ba Krong	NA	11.172	104.504	Dog	Genome	Asian	Cambodian project
Z101	24/07/2015	Southeast Asia	Cambodia	NA	NA	Kang Ta Noeng Ti Pram	NA	11.851	105.066	Dog	Genome	Asian	Cambodian project
Z102	24/07/2015	Southeast Asia	Cambodia	NA	NA	Cheach Thum	NA	11.921	105.856	Dog	Genome	Asian	Cambodian project
Z104	07/08/2015	Southeast Asia	Cambodia	NA	NA	Varint Ti Muoy	NA	11.903	105.135	Dog	Genome	Asian	Cambodian project
Z112	12/08/2015	Southeast Asia	Cambodia	NA	NA	Pou Poat	NA	11.376	105.524	Dog	Genome	Asian	Cambodian project
Z119	27/08/2015	Southeast Asia	Cambodia	NA	NA	Dei Leu	NA	12.299	105.669	Dog	Genome	Asian	Cambodian project
Z121	31/08/2015	Southeast Asia	Cambodia	NA	NA	Boeng Veaeng	NA	11.514	105.443	Dog	Genome	Asian	Cambodian project
Z139	29/09/2015	Southeast Asia	Cambodia	NA	NA	Krang Trea	NA	11.421	104.86	Dog	Genome	Asian	Cambodian project
Z141	29/09/2015	Southeast Asia	Cambodia	NA	NA	Popeal Khae	NA	11.491	105.1	Dog	Genome	Asian	Cambodian project
Z143	07/10/2015	Southeast Asia	Cambodia	NA	NA	Ruessei Chuor	NA	11.304	105.452	Dog	Genome	Asian	Cambodian project
Z151	22/10/2015	Southeast Asia	Cambodia	NA	NA	Tuk Meas	NA	10.663	104.562	Dog	Genome	Asian	Cambodian project
Z153	26/10/2015	Southeast Asia	Cambodia	NA	NA	Preaek Ta Sa	NA	11.236	105.286	Dog	Genome	Asian	Cambodian project
Z159	31/10/2015	Southeast Asia	Cambodia	NA	NA	Trapeang	NA	11.11	103.751	Dog	Genome	Asian	Cambodian project
Z163	18/11/2015	Southeast Asia	Cambodia	NA	NA	Trapeang Pnov	NA	11.936	104.764	Dog	Genome	Asian	Cambodian project

Table B5 continued from previous page

Z174	11/12/2015	Southeast Asia	Cambodia	NA	NA	Pou Doh	NA	10.795	104.393	Dog	Genome	Asian	Cambodian project
Z176	16/12/2015	Southeast Asia	Cambodia	NA	NA	Chi Aok	NA	12.374	105.093	Dog	Genome	Asian	Cambodian project
Z179	21/12/2015	Southeast Asia	Cambodia	NA	NA	Chambak Chrum	NA	11.331	105.627	Dog	Genome	Asian	Cambodian project
MW690139	24/08/2015	Southeast Asia	Malaysia	Pekan Kaki Bukit	Perlis	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
MW690140	02/09/2015	Southeast Asia	Malaysia	Kangar	Perlis	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
MW690141	31/03/2016	Southeast Asia	Malaysia	Chuping	Perlis	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
MW690142	14/02/2017	Southeast Asia	Malaysia	Padang Terap	Kedah	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
MW690143	07/07/2017	Southeast Asia	Malaysia	Serian	Sarawak	NA	NA	NA	NA	Felis catus	N gene	Asian	GenBank
MW690144	11/07/2017	Southeast Asia	Malaysia	Serian	Sarawak	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
MW690145	13/07/2017	Southeast Asia	Malaysia	Serian	Sarawak	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
MW690146	13/07/2017	Southeast Asia	Malaysia	Serian	Sarawak	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
MW690147	08/09/2017	Southeast Asia	Malaysia	Kuching	Sarawak	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
MW690148	28/12/2017	Southeast Asia	Malaysia	Bau	Sarawak	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
MW690149	17/01/2018	Southeast Asia	Malaysia	Kota Samarahan	Sarawak	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank

Table B5 continued from previous page

MW690150	02/02/2018	Southeast Asia	Malaysia	Kuching	Sarawak	NA	NA	NA	NA	Felis catus	N gene	Asian	GenBank
MW690151	06/04/2018	Southeast Asia	Malaysia	Kota Samarahan	Sarawak	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
MW690152	11/06/2018	Southeast Asia	Malaysia	Kangar	Perlis	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
MW690153	25/09/2018	Southeast Asia	Malaysia	Sri Aman	Sarawak	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
MW690154	06/11/2018	Southeast Asia	Malaysia	Kubang Pasu	Kedah	NA	NA	NA	NA	Felis catus	N gene	Asian	GenBank
MW690155	12/11/2018	Southeast Asia	Malaysia	Jitra	Kedah	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
MW055103	2018	Central America	Guatemala	San Luis	Peten	NA	NA	NA	NA	Bos taurus	N gene	Bats	GenBank
MW055104	2018	Central America	Guatemala	Mataquescuintla	Jalapa	NA	NA	NA	NA	Bos taurus	N gene	Bats	GenBank
MW055105	2018	Central America	Guatemala	La Libertad	Peten	NA	NA	NA	NA	Bos taurus	N gene	Bats	GenBank
MW055106	2018	Central America	Guatemala	Huehuetenango	NA	NA	NA	NA	NA	Bos taurus	N gene	Bats	GenBank
MW055107	2018	Central America	Guatemala	Totonicapan	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Cosmopolitan	GenBank
MW055128	12/01/2018	North America	USA	KY	NA	NA	NA	NA	NA	Canidae	N gene	Cosmopolitan	GenBank
MW055129	2018	Caribbean	Haiti	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Cosmopolitan	GenBank
MW055131	22/12/2017	Middle East	Egypt	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Cosmopolitan	GenBank
MW055163	06/04/2017	Asia	India	Goa	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Arctic	GenBank

Table B5 continued from previous page

MW055214	2017	Southeast Asia	Vietnam	Phu Tho Province	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
MW055215	2017	Southeast Asia	Vietnam	Phu Tho Province	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
MW055216	2017	Southeast Asia	Vietnam	Phu Tho Province	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
MW055217	2017	Southeast Asia	Vietnam	Phu Tho Province	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
MW055218	2017	Southeast Asia	Vietnam	Phu Tho Province	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
MW055219	2017	Southeast Asia	Vietnam	Phu Tho Province	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
MW055220	2018	Southeast Asia	Vietnam	Phu Tho Province	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
MW055221	2018	Southeast Asia	Vietnam	Phu Tho Province	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
MW055222	2018	Southeast Asia	Vietnam	Phu Tho Province	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
MW055223	2018	Southeast Asia	Vietnam	Phu Tho Province	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
MW055225	2017	Southeast Asia	Vietnam	Phu Tho Province	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
MW055226	2017	Southeast Asia	Vietnam	Phu Tho Province	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
MW055227	2017	Southeast Asia	Vietnam	Phu Tho Province	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank

Table B5 continued from previous page

MW055228	2017	Southeast Asia	Vietnam	Phu Tho Province	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
MW055229	2018	Southeast Asia	Vietnam	Phu Tho Province	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
MW055230	2018	Southeast Asia	Vietnam	Phu Tho Province	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
MW055231	2018	Southeast Asia	Vietnam	Phu Tho Province	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
MW055232	2017	Southeast Asia	Vietnam	Hoa Binh Province	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
MW055233	2017	Southeast Asia	Vietnam	Phu Tho Province	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
MN510456	2015	Asia	China	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
LC550018	24/10/2019	Southeast Asia	Philippines	Region III	Zambales	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
LC550019	18/10/2019	Southeast Asia	Philippines	Region III	Bulacan	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
LC550020	18/10/2019	Southeast Asia	Philippines	Region III	Bulacan	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
LC550021	18/10/2019	Southeast Asia	Philippines	Region III	Bulacan	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
LC550022	03/10/2019	Southeast Asia	Philippines	Region III	Bulacan	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
LC550023	03/10/2019	Southeast Asia	Philippines	Region III	Nueva Ecija	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank

Table B5 continued from previous page

LC550024	13/08/2019	Southeast Asia	Philippines	Region III	Pampanga	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
LC550025	05/08/2019	Southeast Asia	Philippines	Region III	Pampanga	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
LC550026	05/07/2019	Southeast Asia	Philippines	Region III	Pampanga	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
LC550027	21/06/2019	Southeast Asia	Philippines	Region III	Pampanga	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
MT079890	01/01/2015	Europe	Georgia	Kvemo Kartli	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Cosmopolitan	GenBank
MT241241	2019-05	Asia	China	Shanghai	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
MK124742	06/06/2018	Asia	China	Jiangsu	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
MT454639	2016	Sub-Saharan Africa	South Africa	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Cosmopolitan	GenBank
MK598338	2006	Europe	Hungary	NA	NA	NA	NA	NA	NA	Vulpes vulpes	Genome	Cosmopolitan	GenBank
MK598340	2007	Europe	Hungary	NA	NA	NA	NA	NA	NA	Vulpes vulpes	Genome	Cosmopolitan	GenBank
MK598344	2009	Europe	Hungary	NA	NA	NA	NA	NA	NA	Vulpes vulpes	Genome	Cosmopolitan	GenBank
MK598346	2010	Europe	Hungary	NA	NA	NA	NA	NA	NA	Vulpes vulpes	Genome	Cosmopolitan	GenBank
MK598347	2016	Europe	Hungary	NA	NA	NA	NA	NA	NA	Vulpes vulpes	Genome	Cosmopolitan	GenBank
MK598348	2017	Europe	Hungary	NA	NA	NA	NA	NA	NA	Vulpes vulpes	Genome	Cosmopolitan	GenBank
MK598361	2013	Europe	Hungary	NA	NA	NA	NA	NA	NA	Vulpes vulpes	Genome	Cosmopolitan	GenBank

Table B5 continued from previous page

MK598388	2013	Europe	Hungary	NA	NA	NA	NA	NA	NA	Vulpes vulpes	Genome	Cosmopolitan	GenBank
MK598394	2014	Europe	Hungary	NA	NA	NA	NA	NA	NA	Vulpes vulpes	Genome	Cosmopolitan	GenBank
MK598398	1996	Europe	Hungary	NA	NA	NA	NA	NA	NA	Felis catus	Genome	Cosmopolitan	GenBank
MN534897	2018	Middle East	Qatar	NA	NA	NA	NA	NA	NA	Canidae	Genome	Cosmopolitan	GenBank
MN233900	2007	North America	USA	Alaska	NA	NA	NA	NA	NA	Vulpes lagopus	Genome	Arctic	GenBank
MN233903	2005	North America	Canada	Manitoba	NA	NA	NA	NA	NA	Vulpes vulpes	Genome	Arctic	GenBank
MN233906	1996	North America	Canada	Newfoundland and Labrador	NA	NA	NA	NA	NA	Vulpes vulpes	Genome	Arctic	GenBank
MN233915	2002	North America	Canada	Newfoundland and Labrador	NA	NA	NA	NA	NA	Vulpes vulpes	Genome	Arctic	GenBank
MN233930	2012	North America	Canada	Newfoundland and Labrador	NA	NA	NA	NA	NA	Vulpes vulpes	Genome	Arctic	GenBank
MN233951	1993	North America	Canada	Northern Territories	NA	NA	NA	NA	NA	Vulpes lagopus	Genome	Arctic	GenBank
MN233976	2002	North America	Canada	Nunavut	NA	NA	NA	NA	NA	Vulpes vulpes	Genome	Arctic	GenBank
MN233979	2005	North America	Canada	Nunavut	NA	NA	NA	NA	NA	Vulpes vulpes	Genome	Arctic	GenBank
MN233998	2014	North America	Canada	Nunavut	NA	NA	NA	NA	NA	Vulpes lagopus	Genome	Arctic	GenBank
MN234022	2001	North America	Canada	Quebec	NA	NA	NA	NA	NA	Vulpes vulpes	Genome	Arctic	GenBank
MN234026	2003	North America	Canada	Quebec	NA	NA	NA	NA	NA	Vulpes vulpes	Genome	Arctic	GenBank
MN857167	26/06/2019	Southeast Asia	Philippines	Caloocan City	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank

Table B5 continued from previous page

MN857168	28/05/2019	Southeast Asia	Philippines	Quezon City	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MN857169	31/05/2019	Southeast Asia	Philippines	Pasay City	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MN857170	27/05/2019	Southeast Asia	Philippines	Angono	Rizal	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MN857171	06/05/2019	Southeast Asia	Philippines	San Miguel	Bulacan	NA	NA	NA	NA	Feliformia	Genome	Asian	GenBank
MN075931	1999	Southeast Asia	Thailand	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MF467498	2011	South America	Ecuador	Morona Santiago	NA	NA	NA	NA	NA	Bos taurus	N gene	Bats	GenBank
MF467499	2011	South America	Ecuador	Morona Santiago	NA	NA	NA	NA	NA	Homo sapiens	N gene	Bats	GenBank
MF467501	2014	South America	Ecuador	Loja	NA	NA	NA	NA	NA	Equus caballus	N gene	Bats	GenBank
MF467502	2014	South America	Ecuador	Loja	NA	NA	NA	NA	NA	Equus caballus	N gene	Bats	GenBank
MF467504	12/02/2016	South America	Ecuador	Loja	NA	NA	NA	NA	NA	Bos taurus	N gene	Bats	GenBank
MN726833	2017	Sub-Saharan Africa	Tanzania	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Cosmopolitan	GenBank
MN726836	2013	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MN726837	2013	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MN726838	2013	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MN726839	2015	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank

Table B5 continued from previous page

MN726841	2014	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MN726842	2019	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MN726846	2015	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MN726847	2017	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MN726848	2016	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MN726851	2014	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MN726852	2018	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MN726853	2012	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MN726854	2015	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MN726855	2016	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MN726856	2012	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MN726858	2016	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MN726859	2015	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank

Table B5 continued from previous page

MN726861	2015	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MN726862	2013	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MN726863	2015	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MN726864	2016	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MN726866	2014	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MN726867	2012	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MN726868	2016	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MN726869	2013	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Felis catus	Genome	Asian	GenBank
MN726870	2014	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MN726871	2013	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MN726872	2017	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MN726873	2014	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MN726874	2012	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank

Table B5 continued from previous page

MN726876	2018	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MN726877	2015	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MN726878	2018	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MN726879	2012	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MN726880	2019	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MN726881	2015	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MN726882	2019	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
MK956097	2015	South America	Brazil	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Cosmopolitan	GenBank
MK956098	2015	South America	Brazil	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Cosmopolitan	GenBank
MK956099	2015	South America	Brazil	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Cosmopolitan	GenBank
MK956101	2015	South America	Brazil	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Cosmopolitan	GenBank
LC422729	25/09/2017	Asia	Mongolia	Ulaanbaatar	Bayanzurkh district	NA	NA	NA	NA	Bos taurus	N gene	Cosmopolitan	GenBank
MK760679	23/04/2008	Middle East	Iran	Hamadan	Asadabad	NA	NA	NA	NA	Canis lupus familiaris	Genome	Cosmopolitan	GenBank

Table B5 continued from previous page

MK760687	11/04/2009	Middle East	Iran	Kurdistan	Gharveh	NA	NA	NA	NA	Canis lupus familiaris	Genome	Cosmopolitan	GenBank
MK760695	08/01/2011	Middle East	Iran	Gilan	Masal	NA	NA	NA	NA	Canidae	Genome	Cosmopolitan	GenBank
MK760709	19/10/2012	Middle East	Iran	Mazandaran	Savadkooh	NA	NA	NA	NA	Canis lupus familiaris	Genome	Cosmopolitan	GenBank
MK760727	29/12/2013	Middle East	Iran	Fars	Eqlid	NA	NA	NA	NA	Canidae	Genome	Cosmopolitan	GenBank
MK760766	04/10/2015	Middle East	Iran	Mazandaran	Behshahr	NA	NA	NA	NA	Canis lupus familiaris	Genome	Cosmopolitan	GenBank
MK540895	1991	North America	Canada	Ontario	NA	NA	NA	NA	NA	Vulpes vulpes	Genome	Arctic	GenBank
MK540903	1993	North America	Canada	Ontario	NA	NA	NA	NA	NA	Vulpes vulpes	Genome	Arctic	GenBank
MK981888	1993	North Africa	Tunisia	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Cosmopolitan	GenBank
MH481711	2018	Sub-Saharan Africa	Liberia	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Africa-2	GenBank
MF197743	2010-08	Europe	Poland	Lesser Poland	NA	NA	NA	NA	NA	Canidae	Genome	Cosmopolitan	GenBank
KY860587	2001	Middle East	Turkey	NA	NA	NA	NA	NA	NA	Canidae	Genome	Cosmopolitan	GenBank
KY860589	2006	Middle East	Turkey	NA	NA	NA	NA	NA	NA	Canidae	Genome	Cosmopolitan	GenBank
KY860590	2007	Middle East	Turkey	NA	NA	NA	NA	NA	NA	Canidae	Genome	Cosmopolitan	GenBank
KY860593	2010	Middle East	Turkey	NA	NA	NA	NA	NA	NA	Canidae	Genome	Cosmopolitan	GenBank
KY860601	2014	Middle East	Turkey	NA	NA	NA	NA	NA	NA	Canidae	Genome	Cosmopolitan	GenBank
KY860606	2001	Middle East	Turkey	NA	NA	NA	NA	NA	NA	Canidae	Genome	Cosmopolitan	GenBank

Table B5 continued from previous page

KY860612	1989	Middle East	Turkey	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Cosmopolitan	GenBank
MG458313	1996	Asia	Russia	NA	NA	NA	NA	NA	NA	Vulpes lagopus	Genome	Arctic	GenBank
MG458319	2012	Caribbean	Grenada	NA	NA	NA	NA	NA	NA	Herpestidae	Genome	Cosmopolitan	GenBank
KY553261	1994	Sub-Saharan Africa	South Africa	NA	NA	NA	NA	NA	NA	Civettictis civetta	N gene	Cosmopolitan	GenBank
KY553266	1991	Sub-Saharan Africa	Zimbabwe	NA	NA	NA	NA	NA	NA	Civettictis civetta	N gene	Africa-3	GenBank
MF538630	11/10/2014	Sub-Saharan Africa	Chad	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Africa-2	GenBank
MF538631	26/01/2015	Sub-Saharan Africa	Chad	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Africa-2	GenBank
KT932673	2014-06	Asia	China	Yunnan Province	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
KY214283	2016	Asia	China	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
MF143207	31/07/2013	North America	USA	South Portland	Cumberland, ME	NA	NA	NA	NA	Urocyon cinereoargenteus	Genome	RAC-SK	GenBank
KX708500	2014	North America	Mexico	Yucatan	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Cosmopolitan	GenBank
KX708504	2008	North America	Mexico	Yucatan	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Cosmopolitan	GenBank
KY002901	2001	Asia	Russia	Omsk Oblast	NA	NA	NA	NA	NA	Vulpes corsac	N gene	Cosmopolitan	GenBank
KX148130	1994	Europe	Slovenia	NA	NA	NA	NA	NA	NA	Canidae	Genome	Cosmopolitan	GenBank

Table B5 continued from previous page

KX148254	2001	Southeast Asia	Vietnam	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Asian	GenBank
KR781547	2012	South America	Brazil	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Cosmopolitan	GenBank
KR781552	2012	South America	Brazil	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Cosmopolitan	GenBank
KR781586	2009	South America	Brazil	NA	NA	NA	NA	NA	NA	Cerdocyon thous	N gene	Cosmopolitan	GenBank
KR781600	2010	South America	Brazil	NA	NA	NA	NA	NA	NA	Cerdocyon thous	N gene	Cosmopolitan	GenBank
KR781601	2011	South America	Brazil	NA	NA	NA	NA	NA	NA	Cerdocyon thous	N gene	Cosmopolitan	GenBank
LM645029	2007-04	North America	Greenland	NA	NA	NA	NA	NA	NA	Vulpes lagopus	N gene	Arctic	GenBank
LM645046	2006-03	North America	Greenland	NA	NA	NA	NA	NA	NA	Felis catus	N gene	Arctic	GenBank
LM645052	2006-02	North America	Greenland	NA	NA	NA	NA	NA	NA	Vulpes lagopus	N gene	Arctic	GenBank
KU564993	09/01/2013	Sub-Saharan Africa	Chad	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Africa-2	GenBank
KX036362	01/12/2001	North America	Greenland	NA	NA	NA	NA	NA	NA	Vulpes lagopus	Genome	Arctic	GenBank
KU198462	1990	North America	USA	NA	NA	NA	NA	NA	NA	Vulpes lagopus	Genome	Arctic	GenBank
KU198465	1992	North America	Canada	NA	NA	NA	NA	NA	NA	Vulpes vulpes	Genome	Arctic	GenBank
KR906774	22/09/2011	Sub-Saharan Africa	Tanzania	Serengeti District	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Cosmopolitan	GenBank
KC522613	2012	Europe	Slovenia	NA	NA	NA	NA	NA	NA	Canidae	Genome	Cosmopolitan	GenBank
KU550097	15/07/2014	Central America	Costa Rica	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Bats	GenBank

Table B5 continued from previous page

KU550099	28/07/2014	Central America	Costa Rica	NA	NA	NA	NA	NA	NA	Bos taurus	N gene	Bats	GenBank
KU550100	05/09/2014	Central America	Costa Rica	NA	NA	NA	NA	NA	NA	Bos taurus	N gene	Bats	GenBank
KU550101	18/09/2014	Central America	Costa Rica	NA	NA	NA	NA	NA	NA	Bos taurus	N gene	Bats	GenBank
KU550102	23/09/2014	Central America	Costa Rica	NA	NA	NA	NA	NA	NA	Bos taurus	N gene	Bats	GenBank
KU550103	18/08/2014	Central America	Costa Rica	NA	NA	NA	NA	NA	NA	Homo sapiens	N gene	Cosmopolitan	GenBank
KT119658	27/01/2005	Sub-Saharan Africa	Central African Republic	Bangui	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Africa-2	GenBank
KT119761	14/07/2008	Sub-Saharan Africa	Central African Republic	Bangui	NA	NA	NA	NA	NA	Capra hircus	Genome	Africa-2	GenBank
KT119766	07/04/2009	Sub-Saharan Africa	Central African Republic	Kabo Ouham	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Africa-2	GenBank
KT119776	17/05/1989	Sub-Saharan Africa	Democratic Republic of Congo	Kinshasa	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Cosmopolitan	GenBank
KT119781	16/07/1992	Sub-Saharan Africa	Central African Republic	Bangui	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Cosmopolitan	GenBank
KT276361	1989	South America	French Guiana	Iracoubo	NA	NA	NA	NA	NA	Bos taurus	N gene	Bats	GenBank
KT276362	1991	South America	French Guiana	Matoury	NA	NA	NA	NA	NA	Bos taurus	N gene	Bats	GenBank
KT276365	1997	South America	French Guiana	Mana	NA	NA	NA	NA	NA	Bos taurus	N gene	Bats	GenBank
KR534229	2010	Sub-Saharan Africa	Tanzania	Serengeti District	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Cosmopolitan	GenBank
GQ412744	1931	Asia	China	Beijing	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Cosmopolitan	GenBank
AB981665	01/11/2011	Southeast Asia	Laos	Vientiane Capital	NA	NA	NA	NA	NA	Canis lupus	N gene	Asian	GenBank

Table B5 continued from previous page

AB981666	07/11/2011	Southeast Asia	Laos	Vientiane Capital	NA	NA	NA	NA	NA	Canis lupus	N gene	Asian	GenBank
AB981667	01/09/2011	Southeast Asia	Laos	Vientiane Capital	NA	NA	NA	NA	NA	Canis lupus	N gene	Asian	GenBank
AB981668	02/10/2011	Southeast Asia	Laos	Vientiane Capital	NA	NA	NA	NA	NA	Canis lupus	N gene	Asian	GenBank
AB981669	13/06/2011	Southeast Asia	Laos	Vientiane Capital	NA	NA	NA	NA	NA	Canis lupus	N gene	Asian	GenBank
AB981670	21/02/2011	Southeast Asia	Laos	Champasak	NA	NA	NA	NA	NA	Canis lupus	N gene	Asian	GenBank
AB981671	12/03/2012	Southeast Asia	Laos	Vientiane Capital	NA	NA	NA	NA	NA	Canis lupus	N gene	Asian	GenBank
AB981672	07/03/2012	Southeast Asia	Laos	Vientiane Capital	NA	NA	NA	NA	NA	Canis lupus	N gene	Asian	GenBank
AB981673	02/02/2012	Southeast Asia	Laos	Vientiane Capital	NA	NA	NA	NA	NA	Canis lupus	N gene	Asian	GenBank
AB981674	17/02/2012	Southeast Asia	Laos	Champasak	NA	NA	NA	NA	NA	Canis lupus	N gene	Asian	GenBank
AB981675	04/03/2012	Southeast Asia	Laos	Champasak	NA	NA	NA	NA	NA	Canis lupus	N gene	Asian	GenBank
AB981676	07/03/2012	Southeast Asia	Laos	Vientiane Capital	NA	NA	NA	NA	NA	Canis lupus	N gene	Asian	GenBank
AB981677	02/03/2012	Southeast Asia	Laos	Champasak	NA	NA	NA	NA	NA	Canis lupus	N gene	Asian	GenBank
LC029897	03/09/2012	Sub-Saharan Africa	Uganda	NA	NA	NA	NA	NA	NA	Bos taurus	N gene	Cosmopolitan	GenBank
KJ957446	09/01/2013	Caribbean	Grenada	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Cosmopolitan	GenBank
KJ957447	17/01/2013	Caribbean	Grenada	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Cosmopolitan	GenBank
KJ957448	14/02/2013	Caribbean	Grenada	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Cosmopolitan	GenBank
JX276418	15/05/2012	Asia	China	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank

Table B5 continued from previous page

KF864235	1996	South America	Argentina	NA	NA	NA	NA	NA	NA	Bos taurus	N gene	Bats	GenBank
KF864238	1999	South America	Argentina	NA	NA	NA	NA	NA	NA	Bos taurus	N gene	Bats	GenBank
KF864239	1999	South America	Argentina	NA	NA	NA	NA	NA	NA	Bubalus bubalis	N gene	Bats	GenBank
KF864242	1999	South America	Argentina	NA	NA	NA	NA	NA	NA	Bos taurus	N gene	Bats	GenBank
KF864254	2001	South America	Argentina	NA	NA	NA	NA	NA	NA	Bos taurus	N gene	Bats	GenBank
KF864300	2005	South America	Argentina	NA	NA	NA	NA	NA	NA	Bos taurus	N gene	Bats	GenBank
KF864307	2000	South America	Argentina	NA	NA	NA	NA	NA	NA	Bos taurus	N gene	Bats	GenBank
KF864348	2006	South America	Argentina	NA	NA	NA	NA	NA	NA	Bos taurus	N gene	Bats	GenBank
KF864400	2008	South America	Argentina	NA	NA	NA	NA	NA	NA	Bos taurus	N gene	Bats	GenBank
JX987737	11/07/2003	Asia	Nepal	Pokhara region	NA	NA	NA	NA	NA	Bubalus bubalis	N gene	Indian-Sub	GenBank
JX648416	2003	South America	Peru	NA	NA	NA	NA	NA	NA	NA	N gene	Bats	GenBank
JX648424	2002	South America	Peru	NA	NA	NA	NA	NA	NA	NA	N gene	Bats	GenBank
JX648439	2007	South America	Peru	NA	NA	NA	NA	NA	NA	NA	N gene	Bats	GenBank
JX648463	2002	South America	Peru	NA	NA	NA	NA	NA	NA	NA	N gene	Bats	GenBank
JX648495	2004	South America	Peru	NA	NA	NA	NA	NA	NA	NA	N gene	Bats	GenBank
JQ513529	1997	Caribbean	Puerto Rico	NA	NA	NA	NA	NA	NA	Felis catus	N gene	Cosmopolitan	GenBank
JQ513531	1997	Caribbean	Puerto Rico	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Cosmopolitan	GenBank
JQ686013	2000	North America	USA	NA	NA	NA	NA	NA	NA	Urocyon cinereoargenteus	N gene	Bats	GenBank

Table B5 continued from previous page

JQ798959	2011	Asia	China	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
JF693454	1995	South America	Colombia	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Cosmopolitan	GenBank
EU851125	2002	Europe	Spain	Melilla	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Cosmopolitan	GenBank
EU851127	1990	Europe	Spain	Melilla	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Cosmopolitan	GenBank
EU851129	1985	Europe	Spain	Melilla	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Cosmopolitan	GenBank
JN008438	1986	Europe	Bosnia and Herzegovina	NA	NA	NA	NA	NA	NA	Canidae	N gene	Cosmopolitan	GenBank
JF973774	1972	Europe	Serbia	NA	NA	NA	NA	NA	NA	Canidae	N gene	Cosmopolitan	GenBank
JF973779	1977	Europe	Serbia	NA	NA	NA	NA	NA	NA	Feliformia	N gene	Cosmopolitan	GenBank
JF973781	1978	Europe	Montenegro	NA	NA	NA	NA	NA	NA	Bos taurus	N gene	Cosmopolitan	GenBank
JF973791	1998	Europe	Serbia	NA	NA	NA	NA	NA	NA	Canidae	N gene	Cosmopolitan	GenBank
JF973793	1997	Europe	Serbia	NA	NA	NA	NA	NA	NA	Feliformia	N gene	Cosmopolitan	GenBank
JF973796	1999	Europe	Bosnia and Herzegovina	NA	NA	NA	NA	NA	NA	Canidae	N gene	Cosmopolitan	GenBank
JF973804	2000	Europe	Serbia	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Cosmopolitan	GenBank
HQ317918	1956	Asia	China	NA	NA	NA	NA	NA	NA	Homo sapiens	Genome	Asian	GenBank

Table B5 continued from previous page

GQ472473	2007-04	Asia	China	Guangxi	Nanning	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
GU937042	2009	Asia	South Korea	NA	NA	NA	NA	NA	NA	Bos taurus	N gene	Arctic	GenBank
GU992304	2004	Middle East	Afghanistan	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Arctic	GenBank
GU992307	1983	Southeast Asia	Thailand	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
GU992321	1988	Europe	France	NA	NA	NA	NA	NA	NA	NA	N gene	Cosmopolitan	GenBank
GU992322	1993	North Africa	Morocco	NA	NA	NA	NA	NA	NA	NA	N gene	Cosmopolitan	GenBank
GU992323	1976	Europe	France	NA	NA	NA	NA	NA	NA	NA	N gene	Cosmopolitan	GenBank
GQ303555	05/02/2009	Southeast Asia	Thailand	NA	NA	NA	NA	NA	NA	Homo sapiens	N gene	Asian	GenBank
GQ303556	12/01/2009	Southeast Asia	Thailand	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
GQ406343	2005-07	Asia	China	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Cosmopolitan	GenBank
EU159387	1985	Asia	China	NA	NA	NA	NA	NA	NA	Homo sapiens	N gene	Asian	GenBank
EU159394	2006	Asia	China	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
EU159399	1994	Asia	China	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
EU853566	1986	North Africa	Tunisia	NA	NA	NA	NA	NA	NA	Homo sapiens	N gene	Cosmopolitan	GenBank
EU853567	1996	North Africa	Algeria	NA	NA	NA	NA	NA	NA	Homo sapiens	N gene	Cosmopolitan	GenBank
EU853568	1996	North Africa	Algeria	NA	NA	NA	NA	NA	NA	Homo sapiens	N gene	Cosmopolitan	GenBank

Table B5 continued from previous page

EU853570	1991	North Africa	Morocco	NA	NA	NA	NA	NA	NA	Homo sapiens	N gene	Cosmopo-litan	GenBank
EU853572	1986	North Africa	Morocco	NA	NA	NA	NA	NA	NA	Homo sapiens	N gene	Cosmopo-litan	GenBank
EU853573	1986	North Africa	Tunisia	NA	NA	NA	NA	NA	NA	Homo sapiens	N gene	Cosmopo-litan	GenBank
EU853574	1986	North Africa	Tunisia	NA	NA	NA	NA	NA	NA	Homo sapiens	N gene	Cosmopo-litan	GenBank
EU853576	1986	North Africa	Tunisia	NA	NA	NA	NA	NA	NA	Homo sapiens	N gene	Cosmopo-litan	GenBank
EU853577	1986	North Africa	Tunisia	NA	NA	NA	NA	NA	NA	Homo sapiens	N gene	Cosmopo-litan	GenBank
EU853578	1986	North Africa	Tunisia	NA	NA	NA	NA	NA	NA	Homo sapiens	N gene	Cosmopo-litan	GenBank
EU853579	1986	North Africa	Tunisia	NA	NA	NA	NA	NA	NA	Homo sapiens	N gene	Cosmopo-litan	GenBank
EU853586	2003	Sub-Saharan Africa	Central African Republic	NA	NA	NA	NA	NA	NA	Canis lupus fa-miliaris	N gene	Cosmopo-litan	GenBank
EU853608	1991	North Africa	Mauritania	NA	NA	NA	NA	NA	NA	Canis lupus fa-miliaris	N gene	Africa-2	GenBank
EU853611	1994	North Africa	Mauritania	NA	NA	NA	NA	NA	NA	Canis lupus fa-miliaris	N gene	Africa-2	GenBank
EU853613	1992	North Africa	Mauritania	NA	NA	NA	NA	NA	NA	Equus asinus	N gene	Africa-2	GenBank
EU853624	1991	North Africa	Mauritania	NA	NA	NA	NA	NA	NA	Camelus dromedar-ius	N gene	Africa-2	GenBank
EU853638	1995	Sub-Saharan Africa	Senegal	NA	NA	NA	NA	NA	NA	Homo sapiens	N gene	Africa-2	GenBank
EU853642	1991	North Africa	Mauritania	NA	NA	NA	NA	NA	NA	Capra hircus	N gene	Africa-2	GenBank
EU853646	1990	Sub-Saharan Africa	Niger	NA	NA	NA	NA	NA	NA	Canis lupus fa-miliaris	N gene	Africa-2	GenBank

Table B5 continued from previous page

EU853651	2004	Sub-Saharan Africa	Central African Republic	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Africa-2	GenBank
EU853656	1987	Sub-Saharan Africa	Chad	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Africa-2	GenBank
FJ719760	2008-10	Asia	China	Zhejiang	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
FJ228485	1999	North America	Mexico	Chihuahua	NA	NA	NA	NA	NA	Bos taurus	N gene	RAC-SK	GenBank
FJ228492	2002	Central America	El Salvador	NA	NA	NA	NA	NA	NA	Homo sapiens	N gene	Bats	GenBank
FJ228517	2001	Central America	Honduras	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Cosmopolitan	GenBank
FJ228521	1991	North America	Mexico	Durango	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Cosmopolitan	GenBank
FJ228523	1998	North America	Mexico	Yucatan	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Cosmopolitan	GenBank
FJ228524	1995	North America	Mexico	Yucatan	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Cosmopolitan	GenBank
FJ228528	1994	North America	USA	Florida	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Cosmopolitan	GenBank
EU514572	2007	Sub-Saharan Africa	Niger	NA	NA	NA	NA	NA	NA	NA	N gene	Africa-2	GenBank
EU514575	2005	North Africa	Mauritania	NA	NA	NA	NA	NA	NA	NA	N gene	Africa-2	GenBank
EU514576	2005	North Africa	Mauritania	NA	NA	NA	NA	NA	NA	NA	N gene	Africa-2	GenBank
EU514577	2005	North Africa	Mauritania	NA	NA	NA	NA	NA	NA	NA	N gene	Africa-2	GenBank
EU514578	2006	North Africa	Mauritania	NA	NA	NA	NA	NA	NA	NA	N gene	Africa-2	GenBank

Table B5 continued from previous page

EU514580	2006	North Africa	Mauritania	NA	NA	NA	NA	NA	NA	NA	N gene	Africa-2	GenBank
EU514581	2007	North Africa	Mauritania	NA	NA	NA	NA	NA	NA	NA	N gene	Africa-2	GenBank
FJ424484	2008-10	Europe	Italy	NA	NA	NA	NA	NA	NA	Vulpes vulpes	N gene	Cosmopolitan	GenBank
EU981929	2008-07	South America	Uruguay	NA	NA	NA	NA	NA	NA	Bos taurus	N gene	Bats	GenBank
EU086164	1999	Southeast Asia	Myanmar	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
EU086165	1999	Southeast Asia	Myanmar	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
EU086166	1999	Southeast Asia	Myanmar	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
EU086170	1997	Southeast Asia	Cambodia	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
EU086177	1987	Asia	China	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
EU086178	1969	Asia	China	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
EU086180	1989	Asia	China	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
EU086184	1993	Asia	China	NA	NA	NA	NA	NA	NA	Cervus nippon	N gene	Asian	GenBank
EU086192	2003	Southeast Asia	Indonesia	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
EU086193	1999	Southeast Asia	Laos	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank

Table B5 continued from previous page

EU086194	2002	Southeast Asia	Laos	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
EU086195	2002	Southeast Asia	Laos	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
EU086200	1994	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
EU086201	1994	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
EU086202	1994	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
EU086203	2000	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Homo sapiens	N gene	Asian	GenBank
EU086204	2001	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Homo sapiens	N gene	Asian	GenBank
EU086205	2004	Southeast Asia	Philippines	NA	NA	NA	NA	NA	NA	Homo sapiens	N gene	Asian	GenBank
EU086206	1983	Southeast Asia	Thailand	NA	NA	NA	NA	NA	NA	Homo sapiens	N gene	Asian	GenBank
EU086207	1983	Southeast Asia	Thailand	NA	NA	NA	NA	NA	NA	Homo sapiens	N gene	Asian	GenBank
EU086208	1983	Southeast Asia	Thailand	NA	NA	NA	NA	NA	NA	Homo sapiens	N gene	Asian	GenBank
EU086209	2001	Southeast Asia	Vietnam	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
EU086210	2001	Southeast Asia	Vietnam	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Asian	GenBank
EU886631	2001	Europe	Germany	NA	NA	NA	NA	NA	NA	Vulpes vulpes	Genome	Cosmopolitan	GenBank
EU886632	2002	Europe	Germany	NA	NA	NA	NA	NA	NA	Vulpes vulpes	Genome	Cosmopolitan	GenBank
EU886633	2004	Europe	Austria	NA	NA	NA	NA	NA	NA	Vulpes vulpes	Genome	Cosmopolitan	GenBank

Table B5 continued from previous page

EU886635	2005	Europe	Germany	NA	NA	NA	NA	NA	NA	Vulpes vulpes	Genome	Cosmopolitan	GenBank
EU293113	1990	South America	Guyana	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Bats	GenBank
EU038103	2006	Sub-Saharan Africa	Nigeria	Plateau State	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Africa-2	GenBank
DQ900547	1996-11	Sub-Saharan Africa	Tanzania	Serengeti	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Cosmopolitan	GenBank
DQ900549	1997-02	Sub-Saharan Africa	Tanzania	Serengeti	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Cosmopolitan	GenBank
DQ900564	1998-11	Sub-Saharan Africa	Tanzania	Serengeti National Park	NA	NA	NA	NA	NA	Otocyon megalotis	N gene	Cosmopolitan	GenBank
DQ837393	2005	Middle East	Israel	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Cosmopolitan	GenBank
DQ837410	2003	Middle East	Israel	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Cosmopolitan	GenBank
DQ837425	1998	Middle East	Jordan	NA	NA	NA	NA	NA	NA	Bos taurus	N gene	Cosmopolitan	GenBank
DQ837431	1950	Middle East	Israel	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Cosmopolitan	GenBank
DQ837435	1999	Middle East	Israel	NA	NA	NA	NA	NA	NA	Canidae	N gene	Cosmopolitan	GenBank
DQ837441	1997	Middle East	Israel	NA	NA	NA	NA	NA	NA	Bos taurus	N gene	Cosmopolitan	GenBank
DQ837443	2000	Middle East	Israel	NA	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Cosmopolitan	GenBank
DQ837484	1995	Middle East	Israel	NA	NA	NA	NA	NA	NA	Canidae	N gene	Cosmopolitan	GenBank

Table B5 continued from previous page

AB285215	25/10/2000	Sub-Saharan Africa	Zambia	NA	NA	NA	NA	NA	Canis lupus familiaris	Genome	Cosmopolitan	GenBank
DQ076122	1999	Asia	South Korea	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Arctic	GenBank
DQ076125	2002	Asia	South Korea	NA	NA	NA	NA	NA	Canis lupus familiaris	N gene	Arctic	GenBank

Appendix C

Supplementary information on SARS-CoV-2 transmission in Israeli households

All data and codes used to perform the analyses presented in Chapter 3 are available online at https://github.com/mlayan/VaccineEffectivenessSheba.

1. Differential testing instructions between vaccinated and unvaccinated household contacts

Testing instructions were different between household contacts according to their vaccination status and HCW status. Contacts who had received two vaccine doses at least seven days before detecting the COVID-19 patient were considered protected and encouraged to perform at least two PCR tests in the ten days following the detection of the patient. Contacts who did not meet this criterion were required to perform at least two PCR tests in the ten days following patient detection. If contacts tested positive, they were not required to perform a second test. Unvaccinated HCW were isolated at home whereas vaccinated HCW could come to the hospital for work provided a negative PCR test each time they reported to work.

The proportion of adult/teenager contacts who had at least two PCR tests or one positive PCR test was 79% among unvaccinated contacts and 70% among vaccinated contacts (Table C1). The positivity rate with one PCR was higher in unvaccinated contacts (30%) compared to vaccinated contacts (3%) but the proportion of contacts who performed at least two PCR tests was lower among unvaccinated contacts (49%) compared to vaccinated contacts (67%). When we stratify by HCW status, the proportion of contacts who performed at least two PCR tests is higher in HCW compared to non-HCW among vaccinated (89% in HCW versus 33% in non-HCW) and unvaccinated (76% in HCW and 45% in non-HCW) contacts. In general, HCW were more tested than the other household contacts but their share was lower in unvaccinated contacts (12%) compared to vaccinated contacts (61%).

Testing instructions and positivity rates were different between vaccinated contacts, unvaccinated contacts, HCW and non-HCW which makes it difficult to anticipate how vaccine effectiveness would be impacted. In a sensitivity analysis, we restricted our analysis to households where all negative contacts performed at least one or two PCR tests. The description of the households is detailed in Sections 8.1 and 8.2.

Table C1: Number and result of PCR tests performed by adult/teenager household contacts according to their vaccination status and HCW status. In accordance with the contact categories analyzed in our study, we report here household members corresponding to the adult/teenager category, i.e., all individuals above 12 years old. Among the 494 adult/teenager household contacts, 16 were considered not susceptible to infection over follow-up due to past infection over the preceding year, 353 were considered unvaccinated when they received testing instructions since they had not received two vaccine doses ≥7 days before the detection of the COVID-19 patient, and 125 were considered vaccinated when they received the testing instructions.

	Infection in the preceding year			**Unvaccinated**			**Vaccinated (≥7 days after two doses)**		
Healthcare worker	**No**	**Yes**	**All**	**No**	**Yes**	**All**	**No**	**Yes**	**All**
No test - no (%)	4 (27)	0 (0)	4 (25)	37 (12)	0 (0)	37 (10)	14 (29)	2 (3)	16 (13)
One negative test - no (%)	8 (53)	1 (100)	9 (56)	38 (12)	0 (0)	38 (11)	17 (35)	4 (5)	21 (17)
One positive test - no (%)	0 (0)	0 (0)	0 (0)	96 (31)	10 (24)	106 (30)	2 (4)	2 (3)	4 (3)
At least two tests - no (%)	3 (20)	0 (0)	3 (19)	140 (45)	32 (76)	172 (49)	16 (33)	68 (89)	84 (67)
Total - no	15	1	16	311	42	353	49	76	125

2. Endpoint phone questionnaire to collect household data

At the end of the follow-up, sociodemographic data, household composition characteristics, SARS-CoV-2 infection-related data, and social distancing behaviors were collected over a phone interview. We present below the questions translated in English that were asked to the participants.

- Age of the participant
- Sex of the participant
- What is your SARS-CoV-2 immunization status?
- How many people do you share your household with?
    - How is each of them related to you?
    - What is their age, sex, and SARS-CoV-2 immunization status (unvaccinated, recovered from Covid-19, partially vaccinated)?
- Where do you live and how many rooms are there in your house?
    - How many bathrooms and toilets?
- When was the first case of SARS-CoV-2 in your household diagnosed and who was it?
    - When did their symptoms begin?
- In the week leading to the index case's diagnosis, did any household members share a bedroom or a bathroom with them?
    - Did any household members dine with them?
    - Did any household members ride in the car with them?

- During the index case's mandated isolation period, did the rest of the household members isolate from them completely?
    - Did you still share a bathroom/toilet with them?
- Did you and the rest of your household members undergo PCR and rapid-antigenic testing after your exposure to the index case?
    - And if so, how many tests were conducted and on which dates? What were the results?
- Were the rest of the household members symptomatic at any point during the index case's isolation period?

3. Secondary attack rates in households in which a single index case was identified

In the baseline scenario, we calculated the SAR according to the type of contact for all households regardless of the number of identified index cases. In Table C2, we restricted the calculation of the SAR to households where a single index case was identified (n = 206). There was barely an impact on the SAR.

Table C2: Univariate secondary attack rates according to the type of contact or vaccination status of the index case restricted to households where a single index case was identified.

	No. of infected contacts	No. of susceptible contacts	SAR	
			%	95% CI
Contacts				
Unisolated and unvaccinated adults/teenagers	77	100	77	66, 85
Isolated and unvaccinated adults/teenagers	71	243	29	24, 35
Unisolated and vaccinated adults/teenagers	11	40	28	15, 44
Isolated and vaccinated adults/teenagers	9	81	11	5, 20
Unisolated children	64	96	67	56, 76
Isolated children	29	86	34	24, 45
Index case				
Vaccinated	8	42	19	9, 34
Unvaccinated	256	610	42	38, 46

Abbreviations: CI, confidence interval; SAR, secondary attack rate.

4. Model of SARS-CoV-2 transmission dynamics in households

4.1. Overview

We developed a statistical model describing SARS-CoV-2 transmission within households that accounts for tertiary infections (household members infected by a household case who is not the index case),

infection events in the community, household size, and varying follow up periods between households. For an individual i in household k, data consist in a vector $(a_i, s_i, d_i, v_i, m_i, t_{end})$ where a_i indicates whether i is an adult/teenager above 12 years old or a child, s_i is the infection status of individual i (symptomatic infection, asymptomatic infection or not infected), d_i is the symptom onset date for symptomatic cases or the date of the first positive RT-qPCR test for asymptomatic cases, nu_i is the vaccination status of i, m_i indicates whether i isolated from the index case when applicable, and t_{end} is the end of the follow up period of household k. For each confirmed case, we augmented their observed data with their unobserved date of infection. Infection dates were defined as continuous time to ensure the ordering of infection events within households. Time 0 corresponds to the first infection time in each household. Within household k, I_k denotes the list of SARS-CoV-2 cases and S_k denotes the list of susceptible individuals. Susceptible individuals correspond to household contacts that did not report a SARS-CoV-2 infection in the year preceding the follow-up. Model parameters were estimated in a Bayesian Markov Chain Monte Carlo (MCMC) framework (Cauchemez et al., 2004).

4.2. Transmission within households

If we consider infector i and infectee j in household k of size n, the instantaneous hazard that i infects j at time t is:

$$h_{i\to j}(t) = \frac{\beta}{(n/4)^{\delta}} \mu_{sus}(a_j, m_j, v_j) \mu_{inf}(v_i) \mu_{asymp}(s_i) f(t - \xi_i | d_i, s_i) \tag{C.1}$$

where:

- $\frac{\beta}{(n/4)^{\delta}}$ models the dependency between the transmission rate and the household size n. Here, β corresponds to the transmission rate in households of size 4.
- $\mu_{sus}(a_j, m_j, v_j)$ is the relative susceptibility of recipient j according to their age, vaccination status and isolation status. We define 5 categories of contacts (isolated and unvaccinated adults/teenagers, unisolated and vaccinated adults/teenagers, isolated and vaccinated adults/teenagers, isolated children, unisolated children) that are compared to adults/teenagers who did not isolate and were unvaccinated. For the reference group, $\mu_{sus}(a_j, m_j, v_j) = 1$.
- $\mu_{inf}(v_i)$ is the relative infectivity of infector i according to its vaccination status.

$$\mu_{inf}(v_i) = \begin{cases} 1 & \text{if } v_i = 0 \\ \pi_{inf} & \text{if } v_i = 1 \end{cases}$$

- $\mu_{asymp}(s_i)$ is the relative infectivity of infector i whether i is symptomatic or asymptomatic.

$$\mu_{asymp}(s_i) = \begin{cases} 1 & \text{if } s_i = 1 \\ \pi_{asymp} & \text{if } s_i = 0 \end{cases}$$

In the baseline scenario, we assumed a 40% reduction of the infectivity of asymptomatic cases compared to symptomatic cases ($\pi_{asymp} = 0.6$) as estimated by Byambasuren et al. (2020). In the sensitivity anal-

ysis, we explored the impact of similar infectivity levels between symptomatic and asymptomatic cases ($\pi_{asymp} = 1.0$) on the estimation of the relative susceptibility and relative infectivity parameters.

$f(\delta_t|d,s)$ is the density of the generation time defined as the distribution of the interval between the infection time ξ_i of the infector i and the infection time ξ_j of the recipient j conditioned on the symptom onset or date of detection, and the presence of symptoms of i.

For symptomatic cases, f is derived from the corrected infectivity profile estimated by Ashcroft et al. (2020), a shifted Γ distribution with shape=97.2, rate=3.7 and shift=25.6. The shift of the distribution corresponds to the symptom onset of i. According to McAloon et al., 2020, less than 2% of the symptomatic cases develop symptoms over the three days following their infection (McAloon et al., 2020). We assumed that incubation periods are of 3 days minimum and that the infectious period starts 3 days before the symptom onset, independent of the duration of the incubation period. For asymptomatic infectors, f is derived from the same estimate of the infectivity profile so that infectors are infectious starting from 2 days after their infection time and their infectivity peaks approximately 5 days after their infection.

4.3. Instantaneous risk of infection of a household member

The risk of infection of individual j in household k at time ξ is the sum of the hazard of infection within the community and the hazards of infection by infected household members:

$$\lambda_{j,k}(\xi) = \alpha + \sum_{i \in I_k \{\xi_i < \xi\}} h_{i \to j,k}(\xi) \tag{C.2}$$

where α is the instantaneous risk of infection in the community. It is assumed constant over the follow-up of households and the entire period of the study.

4.4. Likelihood function

Denote θ the vector of the transmission model parameters. The likelihood of the transmission process within the household conditional on the first date of infection ξ_1 in the household is:

$$P(\xi|\theta) = \prod_{i \in I} f(d_i - \xi_i) \prod_{i \in I - \{1\}} \lambda_i(\xi_i) e^{-\int_{\xi_1}^{\xi_i} \lambda_i(u)du} \prod_{j \in S} e^{-\int_{\xi_1}^{t_{end}} \lambda_j(u)du} \tag{C.3}$$

where $f(d_i - \xi_i)$ is the density of the incubation period for symptomatic cases or the density of the RT-qPCR detection period after infection for asymptomatic cases. The distribution of the incubation period was defined as a truncated log-normal distribution with log-mean=1.63 and log-sd=0.25 as estimated by McAloon et al., 2020 (McAloon et al., 2020). As previously mentioned, less than 2% of the symptomatic cases develop symptoms over the three days following their infection according to this distribution. We assumed that the incubation period lasts at least 3 days and does not exceed 30 days. For the RT-qPCR detection period, we assumed a Uniform(0,10) distribution.

Household contacts that reported a SARS-CoV-2 infection in the year preceding follow-up ($n = 20$) were considered protected from re-infection, and thus, did not contribute to the likelihood of the transmission process.

For the incubation period and the infectivity profile, we used distributions that were estimated on data from the historical lineages. Over the study period, up to 80% of the infections were caused by the alpha variant in Israel (Benenson et al., 2021). Yet, there is little knowledge on the mechanisms underlying the rapid spread of the alpha variant at the individual level. Modelling and phylodynamic studies support the hypothesis of its higher transmissibility compared to the historical lineages. The results concerning potentially shorter generation time and longer infectious period are less clear (Volz et al., 2021; Kissler et al., 2021; Davies et al., 2021). Since participants were not screened for the variants, we assumed that the infectivity profile and the incubation period remained unchanged between the alpha variant and the historical lineages.

5. Inference framework

We used a data augmentation MCMC approach to explore the joint posterior distribution of model parameters and the augmented dates of infection.

5.1. Priors

We choose a Uniform(0,1) prior distribution for the hazard of infection within the community α and a Uniform(0,5) prior distribution for the per capita transmission rate within households β. For the dependency between the transmission rate and the household size, the prior distribution for δ was a Uniform(-3,3) distribution.

For the relative susceptibility of the different categories of contacts μ_{sus} and the relative infectivity of vaccinated cases μ_{inf}, a log-normal prior with log-mean=0 and log-sd=1 was used. We investigated the impact of the log-sd on parameter estimation in a sensitivity analysis (Section 8.3).

5.2. Algorithm

Parameters were updated using a Metropolis-Hastings algorithm.

Data augmentation was performed at each iteration of the MCMC chain. After the update of the parameters, the infection times of all COVID-19 cases were updated. For symptomatic cases, the incubation period was sampled from the distribution mentioned above and the infection date was obtained as the difference of the symptom onset and the incubation period. For asymptomatic cases, we chose a conservative scenario according to which asymptomatic cases could have been infected up to ten days prior to their first positive RT-qPCR test independent of the Ct value of the test.

5.3. Implementation

The data augmentation Markov Chain Monte Carlo sampling algorithm was implemented in C++. Chains were run for 100,000 iterations and one out of 10 iterations were recorded. Marginal posteriors were sampled from MCMC chains after discarding a burn-in of 10,000 steps. Convergence was inspected visually.

6. Model adequacy

To assess the adequation of the model to the data, we performed a simulation study. We simulated from our household transmission model 2,000 data sets with a structure identical to that of the observed data (household size, age, vaccination and isolation status of the household members, symptom status of the index case, proportion of secondary asymptomatic cases, and follow-up period) with parameters equal to samples from their joint posterior distribution. We compared the observed secondary attack rate (SAR) to the one expected under the model estimates. There was a good agreement between the observed and expected SAR for households of size 2 to 5.

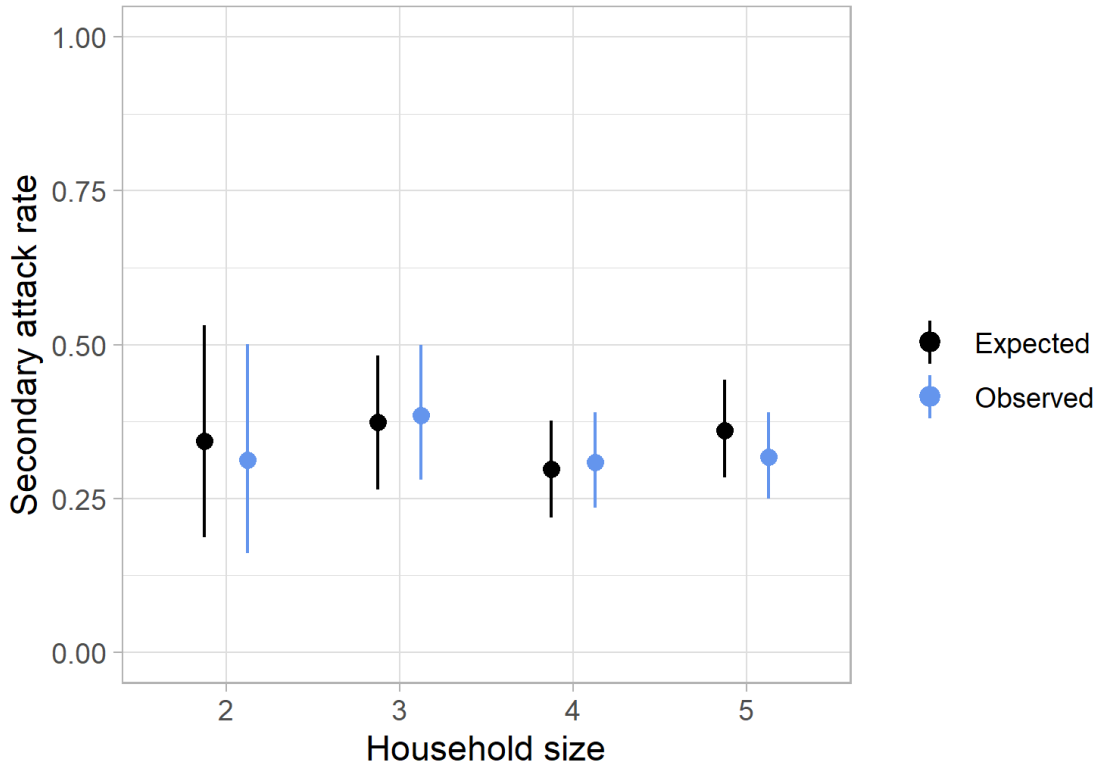


Figure C1: Observed and expected secondary attack rates under the model estimates for households of size 2 to 5. Black circles correspond to the observed SAR per household size. The mean and 95% credible interval are reported.

We also compared the observed and expected distributions of the number of infected individuals per household size. The observed values fall within the 95% credible interval of the expected values. There is a good agreement between the observed and expected distributions.

Table C3: Comparison of the observed and expected distributions of the number of infected individuals for households of size 2 to 5. The mean and the 95% credible interval are reported.

Household size	**Distribution**	**Number of infected individuals per household**				
		1	**2**	**3**	**4**	**5**
2	**Observed**	22	10			
2	**Expected**	20.5 (15-26)	11.5 (6-17)			
3	**Observed**	21	9	12		
3	**Expected**	19.2 (14-25)	13.4 (8-19)	9.4 (5-14)		
4	**Observed**	25	9	8	7	
4	**Expected**	21 (14-27)	15.3 (9-22)	8.8 (4-14)	3.9 (1-7)	
5	**Observed**	19	9	10	2	6
5	**Expected**	12.8 (7-19)	13.2 (8-19)	9.6 (4-15)	6.5 (2-11)	4 (1-8)

7. Parameter estimates

Table C4 compiles the median and 95% credible interval of the relative susceptibility and relative infectivity estimates that are presented in Figs. 17 and 19, Table C5 compiles the median and 95% credible interval of the person-to-person probability of transmission presented in Fig. 18, and Fig. C2 depicts the prior and posterior distributions of model parameters in the baseline analysis. We derive the person-to-person probability of transmission from the instantaneous risk of transmission from individual i to individual j presented in equation C.1:

$$P(i \rightarrow j) = 1 - exp\left(-\frac{\beta}{(n/4)^{\delta}}\ \mu_{sus}(a_j, m_j, v_j)\ \mu_{inf}(v_i)\ \mu_{asymp}(s_i)\right) \tag{C.4}$$

Table C4: Estimates of the relative susceptibility of household contacts and relative infectivity of cases in the baseline scenario and in the sensitivity analyses. The median and the 95% credible interval are reported. In the ≥15 days after 1st dose scenario, we assumed that vaccination is effective from 15 days after the 1st injection. In the 1 PCR test for all contacts and 2 PCR tests for all contacts, we restricted the analysis to the households where all household members performed at least one or two PCR tests, respectively, in the 10 days following the detection of the index case. In the 100% infectivity of asymptomatic cases, we assumed that symptomatic and asymptomatic cases have the same level of infectivity. In the relative susceptibility prior with log-sd=0.7 scenario, we used a log-sd=0.7 for the prior of the relative susceptibility parameters. In the relative infectivity prior with log-sd=0.7 scenario, we used a log-sd=0.7 for the prior of the relative infectivity of vaccinated cases compared to unvaccinated cases parameter. In the baseline scenario, we assumed that vaccination was effective from 7 days after the 2nd dose, the relative infectivity of asymptomatic cases compared to symptomatic cases was equal to 60% and the log-sd of the relative infectivity and relative susceptibility prior distributions was equal to 1. The posterior median and its associated 95% Bayesian credible interval are reported.

	Baseline scenario	**≥15 days after 1st dose**	**At least 1 PCR test for all negative contacts**	**At least 2 PCR tests for all negative contacts**	**100% infectivity of asymptomatic cases**	**Relative susceptibility prior and relative infectivity prior with log-sd=0.7**	**Relative susceptibility prior and relative infectivity prior with log-sd=2**
Relative susceptibility							
Isolated and unvaccinated adults/teenagers	0.12 (0.06-0.20)	0.12 (0.06-0.20)	0.10 (0.04-0.21)	0.11 (0.04-0.21)	0.12 (0.06-0.20)	0.16 (0.09-0.25)	0.09 (0.03-0.16)
Unisolated and vaccinated adults/teenagers	0.20 (0.08-0.41)	0.16 (0.06-0.35)	0.26 (0.08-0.61)	0.30 (0.09-0.76)	0.19 (0.07-0.37)	0.27 (0.13-0.50)	0.13 (0.02-0.31)
Isolated and vaccinated adults/teenagers	0.07 (0.02-0.15)	0.05 (0.02-0.12)	0.11 (0.04-0.25)	0.12 (0.04-0.27)	0.07 (0.02-0.15)	0.11 (0.05-0.21)	0.03 (0-0.08)
Unisolated children	0.48 (0.31-0.73)	0.44 (0.29-0.69)	0.50 (0.30-0.81)	0.45 (0.26-0.78)	0.45 (0.29-0.69)	0.56 (0.36-0.86)	0.41 (0.26-0.63)
Isolated children	0.16 (0.07-0.30)	0.15 (0.07-0.28)	0.24 (0.10-0.48)	0.23 (0.10-0.45)	0.16 (0.07-0.29)	0.21 (0.11-0.36)	0.11 (0.03-0.23)
Relative infectivity							
Vaccinated case	0.24 (0.06-0.69)	0.29 (0.07-0.77)	0.27 (0.06-0.82)	0.28 (0.07-0.85)	0.21 (0.05-0.62)	0.35 (0.13-0.85)	0.08 (0.01-0.47)

Table C5: Estimates of the person-to-person transmission probability from vaccinated adult/teenager cases and unvaccinated cases to the different categories of contacts within a household of size 4. Probabilities are reported in percentage with their 95% credible interval.

Contact	Vaccinated adult/teenager case	Unvaccinated case
Unisolated and unvaccinated adults/teenagers	22.2 (6.2-51.6)	65.5 (53.1-76.3)
Isolated and unvaccinated adults/teenagers	2.9 (0.7-9.2)	12.2 (6.3-18.4)
Isolated and vaccinated adults/teenagers	1.6 (0.3-6.2)	7 (2.5-13.9)
Unisolated and vaccinated adults/teenagers	4.7 (1-15.4)	19.1 (7.9-33.5)
Unisolated children	11.3 (2.8-30.2)	39.7 (28.9-50.4)
Isolated children	3.9 (0.8-12.1)	15.6 (7.5-26)

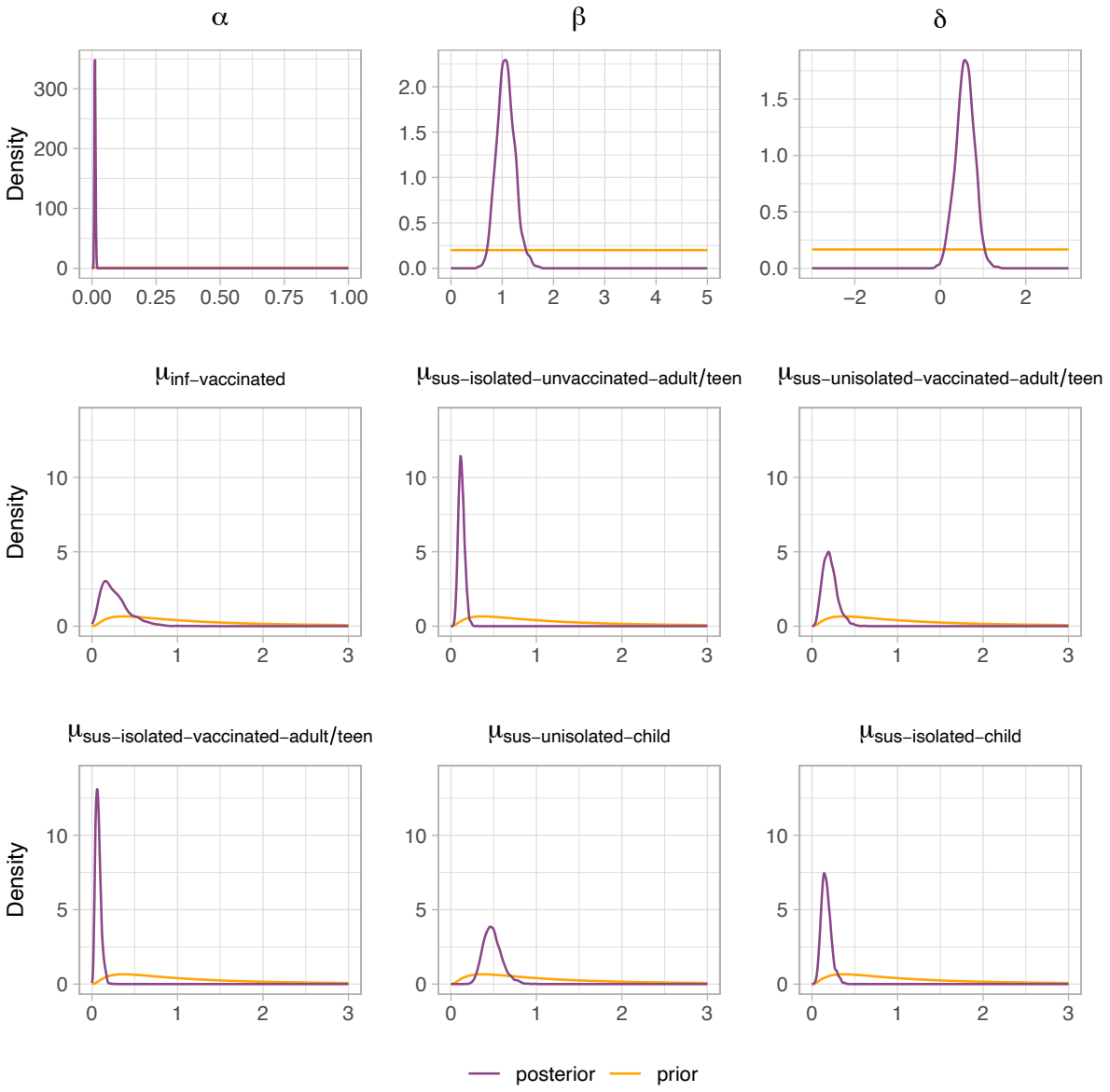


Figure C2: Prior and posterior distributions of model parameters in the baseline analysis.

8. Sensitivity analysis

8.1. Households where all contacts performed at least one PCR test in the ten days following the detection of the index case

In the baseline scenario, household contacts who did not report symptoms and did not perform at least one PCR test in the ten days following the detection of the index case were considered not infected over their follow-up (n = 125). In a sensitivity analysis, we verified the robustness of our estimates to this hypothesis by removing all the households with at least one contact whose outcome was not confirmed which restricted the analysis to 141 households (Fig. C3). There were 145 index cases, including 4 co-index cases, and 429 household contacts. 410 household contacts were susceptible to SARS-CoV-2 since they did not report any SARS-CoV-2 infection in the preceding year. Among susceptible contacts, 201 (49%) developed a SARS-CoV-2 infection (Table C6). This is slightly higher than the 40% in the baseline scenario. The characteristics of the index cases and household contacts are relatively similar to the baseline scenario, except the proportion of contact that is slightly higher in the sensitivity analysis. This is directly due to the more precise follow-up of household contacts by PCR testing.

Compared to the baseline scenario, the univariate SAR are higher in all contact categories due to the higher detection of asymptomatic cases (Table C7). The SAR in households with vaccinated index case(s) is equal to 8% in the sensitivity analysis but there are only 24 household contacts. There is low statistical power to precisely estimate the reduction of infectivity in vaccinated index cases compared to unvaccinated index cases which explains why the 95% credible interval is larger in the sensitivity analysis.

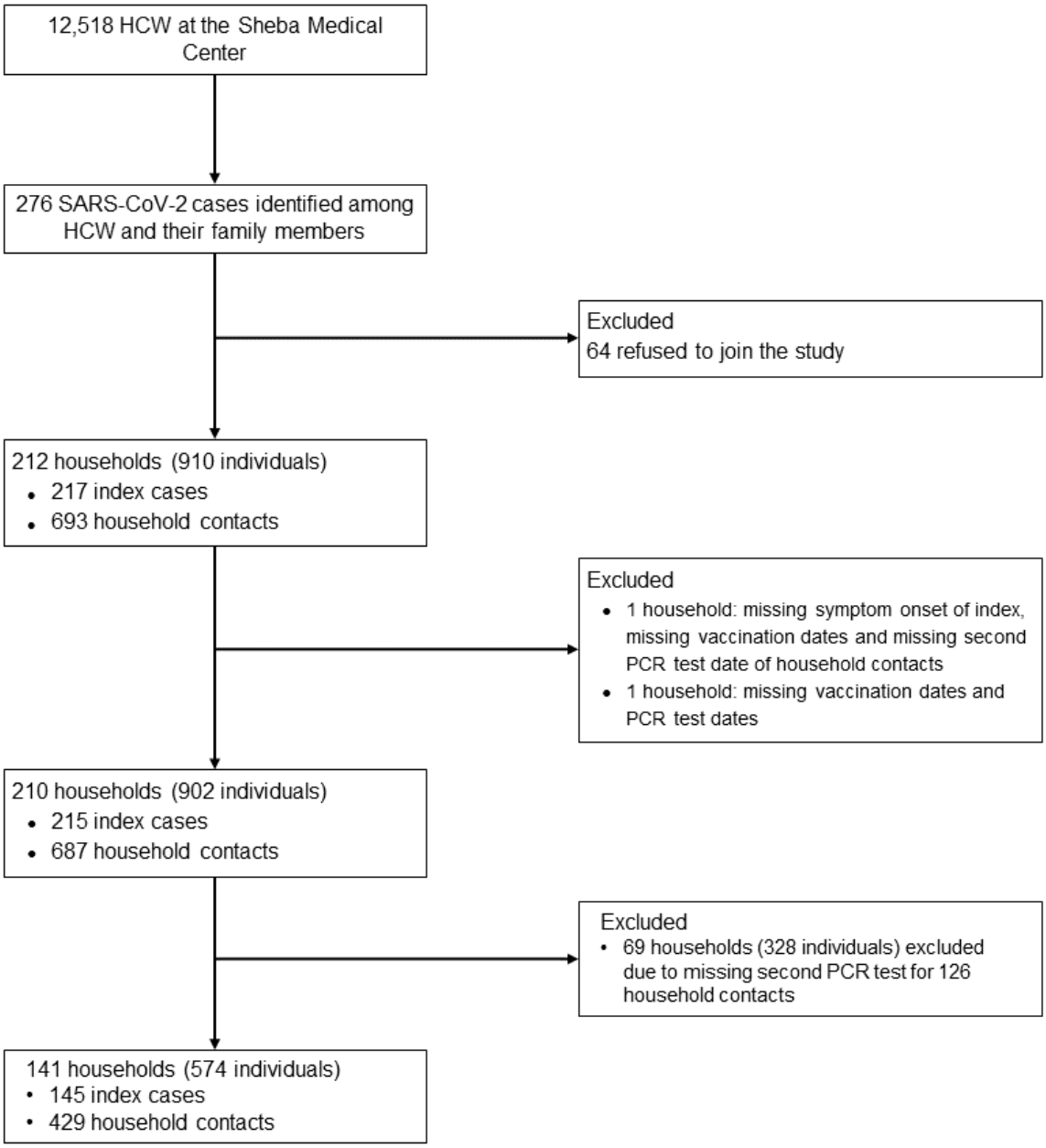


Figure C3: Flow chart of the households included in the sensitivity analysis where all contacts performed at least one PCR test in the ten days following the detection of the index case.

Table C6: Characteristics of the index cases and household contacts according to their age for households where all contacts had at least one PCR test in the ten days following the detection of the index case.

	Adult/teenager index cases (N = 130)	**Child index cases (N = 15)**	**All index cases (N = 145)**
Male sex - no. (%)	53 (41)	8 (53)	61 (42)
Age, years - mean (SD)	37 (15)	6 (4)	34 (17)
Cluster size - median (IQR)	2 (1 - 3)	2 (1 - 3.5)	2 (1 - 3)
Symptom status - no. (%)			
Symptomatic	117 (90)	6 (40)	123 (85)
Asymptomatic	13 (10)	9 (60)	22 (15)
Vaccination			
Vaccinated - no. (%)	10 (8)	-	10 (7)
Days from 2nd dose to detection - median (IQR)	45 (15 - 60)	-	45 (15 - 60)
	Adult/teenager household contacts (N = 304)	**Child household contacts (N = 125)**	**All household contacts (N = 429)**
Male sex - no. (%)	141 (46)	72 (58)	213 (50)
Age, years - mean (SD)	36 (17)	6 (4)	27 (20)
Infection and symptom status - no. (%)			
Past infection	15 (5)	4 (3)	19 (4)
Not infected	161 (53)	48 (38)	209 (49)
Symptomatic	94 (31)	32 (26)	126 (29)
Asymptomatic	33 (11)	41 (33)	74 (17)
Symptomatic (missing onset)	1 (0)	0 (0)	1 (0)
Vaccination			
Vaccinated - no. (%)	59 (19)	-	59 (14)
Days from 2nd dose to exposure - median (IQR)	23 (14 - 35)	-	23 (14 - 35)
Isolation - no. (%)			
Partial	60 (20)	15 (12)	75 (17)
Complete	140 (46)	37 (30)	177 (41)
Missing	2 (1)	0 (0)	2 (0)

Table C7: Univariate secondary attack rates according to the type of contact restricted to households where all contacts had at least one PCR test in the ten days following the detection of the index case.

	No. of infected contacts	No. of susceptible contacts	SAR - %	
			%	95% CI
Contacts				
Unisolated and unvaccinated adults/teenagers	66	81	81	71, 89
Isolated and unvaccinated adults/teenagers	45	147	31	23, 39
Unisolated and vaccinated adults/teenagers	7	19	37	16, 62
Isolated and vaccinated adults/teenagers	8	40	20	9, 36
Unisolated children	52	72	72	60, 82
Isolated children	21	49	43	29, 58
Index				
Vaccinated	2	24	8	1, 27
Unvaccinated	199	383	52	47, 57

8.2. Households where all contacts performed at least two PCR tests in the ten days following the detection of the index case

We further excluded households where at least one negative household contact did not perform at least 2 PCR tests in the 10 days following the detection of the index case. Here, the analysis is restricted to 130 households (Fig. C4). There were 134 index cases, including 4 co-index cases, and 388 household contacts, among whom 193 (50%) developed a SARS-CoV-2 infection (Table C8). This is slightly higher than in the 39% in the baseline scenario. The characteristics of the index cases and household contacts are relatively similar to the baseline scenario except for the median time from the 2nd dose to detection. There are less male child index cases and less symptomatic child index cases compared to the baseline scenario. There are more asymptomatic and symptomatic contact cases compared to the baseline scenario and the vaccination coverage is lower. Compared to the baseline scenario, the univariate SAR are higher in all contact categories (Table C9). The SAR in households with vaccinated index case(s) is equal to 10% in the sensitivity analysis but there are only 21 household contacts. There is low statistical power to precisely estimate the reduction of infectivity in vaccinated index cases compared to unvaccinated index cases which explains why the 95% credible interval is larger in the sensitivity analysis.

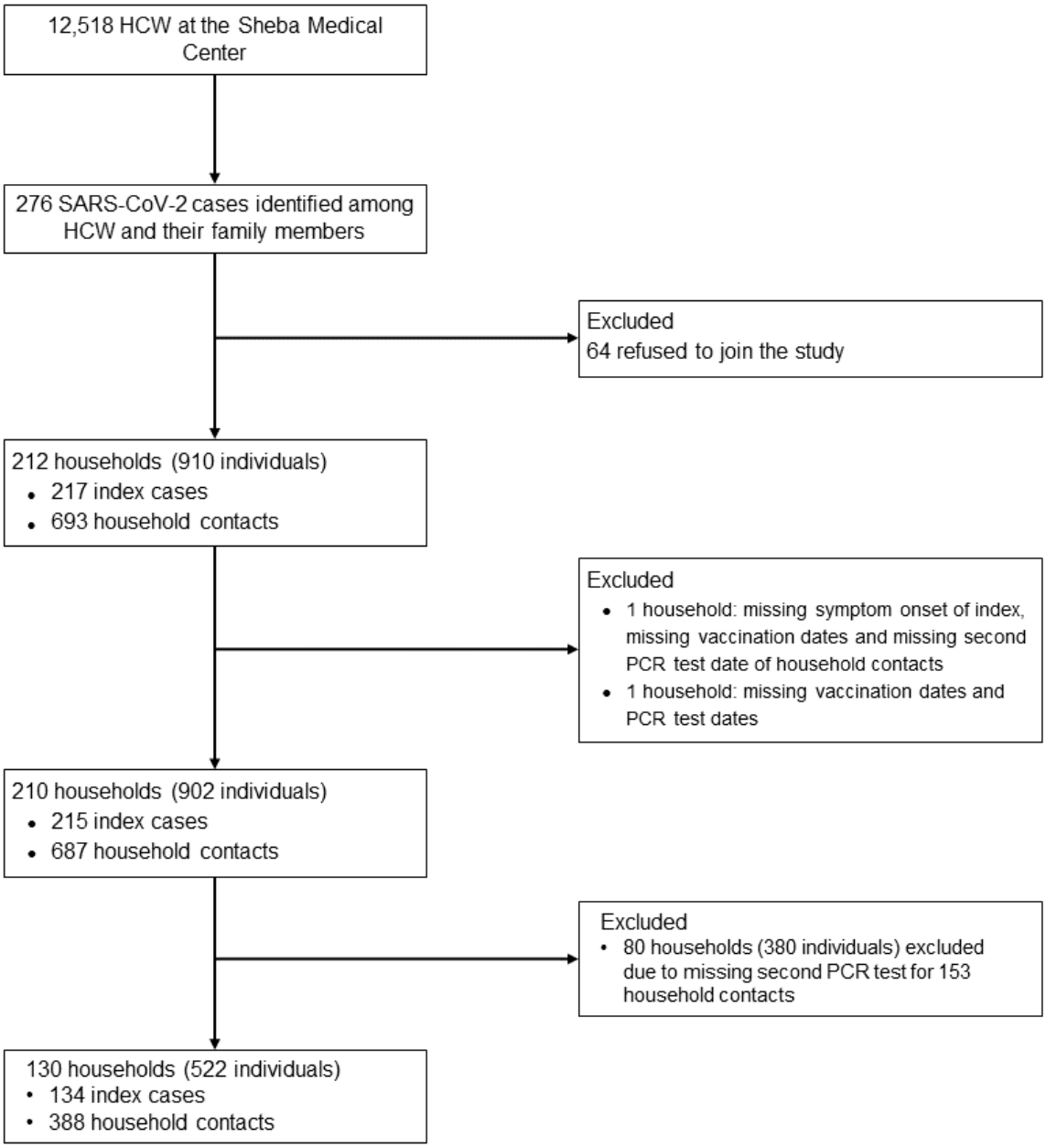


Figure C4: Flow chart of the households included in the sensitivity analysis where all contacts performed at least two PCR tests in the ten days following the detection of the index case.

Table C8: Characteristics of the index cases and household contacts according to their age for households where all contacts had two PCR tests in the ten days following the detection of the index case.

	Adult/teenager index cases (N = 120)	**Child index cases (N = 14)**	**All index cases (N = 134)**
Male sex - no. (%)	50 (42)	7 (50)	57 (43)
Age, years - mean (SD)	38 (15)	6 (4)	35 (17)
Cluster size - median (IQR)	2 (1 - 3)	2 (1.25 - 3.75)	2 (1 - 3)
Symptom status - no. (%)			
Symptomatic	108 (90)	5 (36)	113 (84)
Asymptomatic	12 (10)	9 (64)	21 (16)
Vaccination			
Vaccinated - no. (%)	9 (8)	-	9 (7)
Days from 2nd dose to detection - median (IQR)	46 (13 - 61)	-	46 (13 - 61)
	Adult/teenager household contacts (N = 273)	**Child household contacts (N = 115)**	**All household contacts (N = 388)**
Male sex - no. (%)	126 (46)	67 (58)	193 (50)
Age, years - mean (SD)	35 (17)	6 (4)	27 (19)
Infection and symptom status - no. (%)			
Past infection	8 (3)	2 (2)	10 (3)
Not infected	141 (52)	44 (38)	185 (48)
Symptomatic	91 (33)	28 (24)	119 (31)
Asymptomatic	32 (12)	41 (36)	73 (19)
Symptomatic (missing onset)	1 (0)	0 (0)	1 (0)
Vaccination			
Vaccinated - no. (%)	49 (18)	-	49 (13)
Days from 2nd dose to exposure - median (IQR)	24 (14 - 36)	-	24 (14 - 36)
Isolation - no. (%)			
Partial	53 (19)	11 (10)	64 (16)
Complete	120 (44)	35 (30)	155 (40)
Missing	2 (1)	0 (0)	2 (1)

Table C9: Univariate secondary attack rates according to the type of contact restricted to households where all contacts had two PCR tests in the ten days following the detection of the index case.

	No. of infected contacts	No. of susceptible contacts	SAR	
			%	95% CI
Contacts				
Unisolated and unvaccinated adults/teenagers	66	80	82	72, 90
Isolated and unvaccinated adults/teenagers	43	134	32	24, 41
Unisolated and vaccinated adults/teenagers	6	16	38	15, 65
Isolated and vaccinated adults/teenagers	7	33	21	9, 39
Unisolated children	49	68	72	60, 82
Isolated children	20	45	44	30, 60
Index				
Vaccinated	2	21	10	1, 30
Unvaccinated	191	354	54	49, 59

8.3. Prior distributions of the relative infectivity and relative susceptibility parameters

The log-sd of the relative susceptibility and relative infectivity parameters modifies the value range that is likely to be explored. In the baseline scenario, we used a log-sd=1.0 for both the relative infectivity and relative susceptibility parameters which corresponds to a relatively large 95% interval spanning from 0.14 to 7.1. With log-sd=0.7, the interval considerably shrinks around 1 and with log-sd=2 it spans from 0.02 to 50.4.

Table C10: 2.5% and 97.5% percentiles of the relative infectivity and relative susceptibility prior distributions.

Log-sd of the prior	2.5% percentile	97.5% percentile
0.7	0.25	3.9
1	0.14	7.1
2	0.02	50.4

8.4. Vaccination definition

8.4.1. Effective vaccination ≥15 days after the 1st dose

In a sensitivity analysis, we tested how a different definition of effective vaccination (≥15 days after the 1st dose or ≥7 days after the 2nd dose) affects parameter estimates. There are 3 additional vaccinated index cases and 16 additional vaccinated adult/teenager contacts in the 1 dose scenario (sensitivity analysis) compared to the 2 doses scenario (baseline). The 1 dose scenario slightly increases the share of vaccinated adult/teenager index cases and vaccinated adult/teenager contacts, but it did not impact parameter estimates (see Fig. 19).

8.4.2. Early vaccination

To determine the impact of early vaccinated cases on our estimates, we restricted the analysis to the households where the index case was either unvaccinated or infected ≥7days after the 2nd dose (n = 165), i.e., this analysis did not contain households where the index was vaccinated but infected before the vaccine was considered effective. In this analysis, we excluded 45 households compared to the baseline scenario. The estimations are represented in Fig. C5 and the values are gathered in Table C12. Compared to the baseline scenario, the relative susceptibility and relative infectivity parameters were not impacted.

Table C11: Vaccination status of the adult/teenager index cases and adult/teenager household contacts according to the definition of effective vaccination. Three individuals were exposed to the index case more than 15 days after they received the 1st vaccine dose but did not remember the exact date. Two index cases and three household contacts were detected or exposed about 15 days after they received the 1st vaccine dose. These individuals were considered as unvaccinated. The remaining 3 household contacts with a missing vaccination date were vaccinated less than 10 days before their exposure and thus do not verify the definition of effective vaccination.

	≥15 days after the 1st dose (sensitivity analysis)	**≥7 days after the 2nd dose (baseline)**
Adult/teenager index cases (N = 191)		
Vaccinated - no. (%)	26 (14)	15 (8)
Days from 1st dose to detection - median (IQR)	31 (21 - 67)	44 (13 - 59)
Missing vaccination date - no. (%)	2 (1)	0 (0)
Adult/teenager contacts (N = 494)		
Vaccinated - no. (%)	155 (31)	125 (25)
Days from 1st dose to exposure - median (IQR)	40 (29 - 55)	23 (14 - 36)
Missing vaccination date - no. (%)	6 (1)	0 (0)

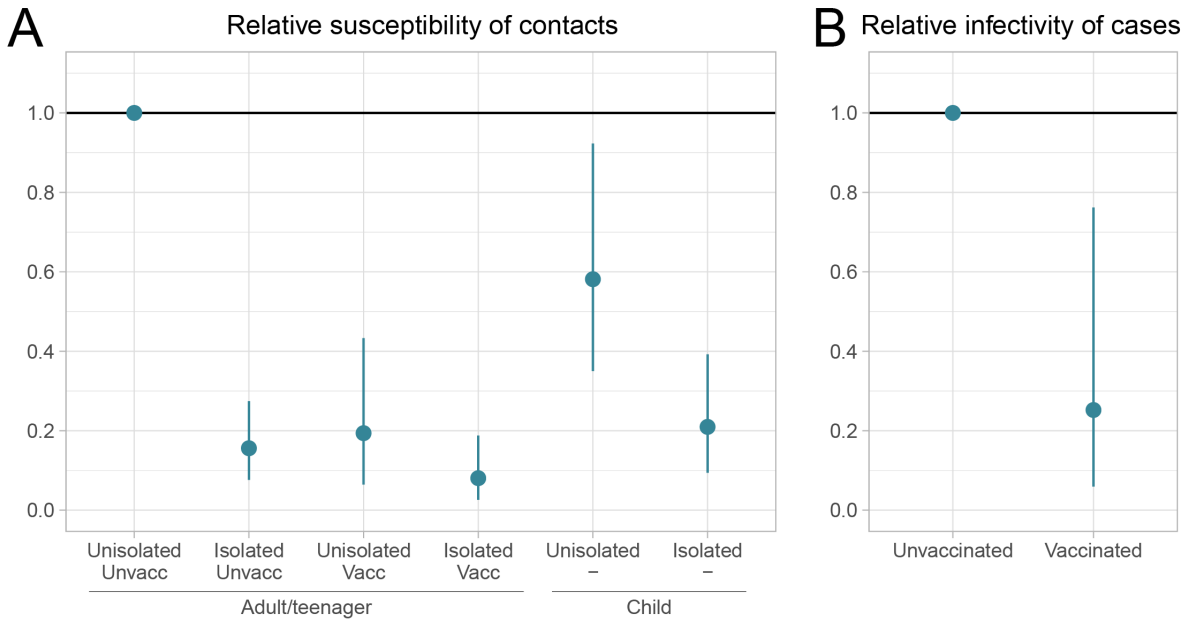


Figure C5: Estimates of the SARS-CoV-2 transmission parameters within households where the index case was not early vaccinated. (**A**) Estimated relative susceptibility of isolated and unvaccinated adults/teenagers, unisolated and vaccinated adults/teenagers, isolated and vaccinated adults/teenagers, unisolated children, and isolated children versus unisolated and unvaccinated adults/teenagers. (**B**) Estimated relative infectivity of vaccinated cases compared to unvaccinated cases. The median estimate and its associated 95% Bayesian credibility interval are reported.

Table C12: Estimates of the SARS-CoV-2 transmission parameters within households where the index case was not early vaccinated. The median and the 95% credible intervals are reported.

	Baseline scenario - median (95% credible interval)	**Household with no early vaccinated index case - median (95% credible interval)**
Relative susceptibility		
Isolated and unvaccinated adults/teenagers	0.12 (0.06-0.20)	0.16 (0.08-0.27)
Unisolated and vaccinated adults/teenagers	0.20 (0.08-0.41)	0.19 (0.06-0.43)
Isolated and vaccinated adults/teenagers	0.07 (0.02-0.15)	0.08 (0.03-0.19)
Unisolated children	0.48 (0.31-0.73)	0.58 (0.35-0.92)
Isolated children	0.16 (0.07-0.30)	0.21 (0.09-0.39)
Relative infectivity		
Vaccinated	0.24 (0.06-0.69)	0.25 (0.06-0.76)

Appendix D

Supplementary information on sampling bias in discrete phylogeography

All scripts used to simulate epidemics and perform the analyses presented in Chapter 4 are available at https://github.com/mlayan/Sampling_bias including examples of output files.

1. Impact and mitigation of bias in the seven demes framework

1.1. Estimation of genetic parameters

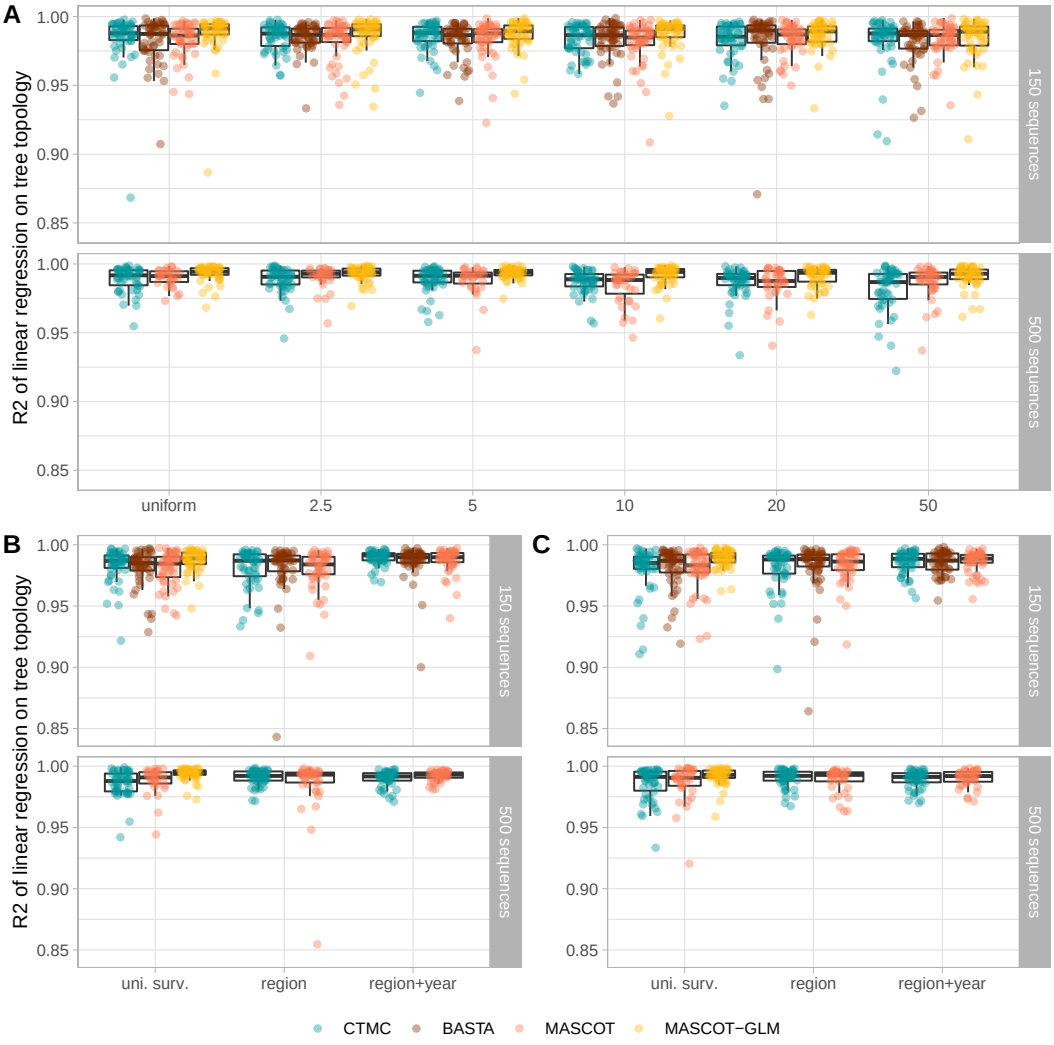


Figure D1: Comparison of the simulated and estimated tree topologies for all sampling conditions and the four algorithms. Pearson's determination coefficient of the pairwise divergence time between the simulated transmission chain and the MCC tree for the systematic bias (**A**), surveillance bias 10 (**B**), and surveillance bias 20 (**C**) sampling conditions.

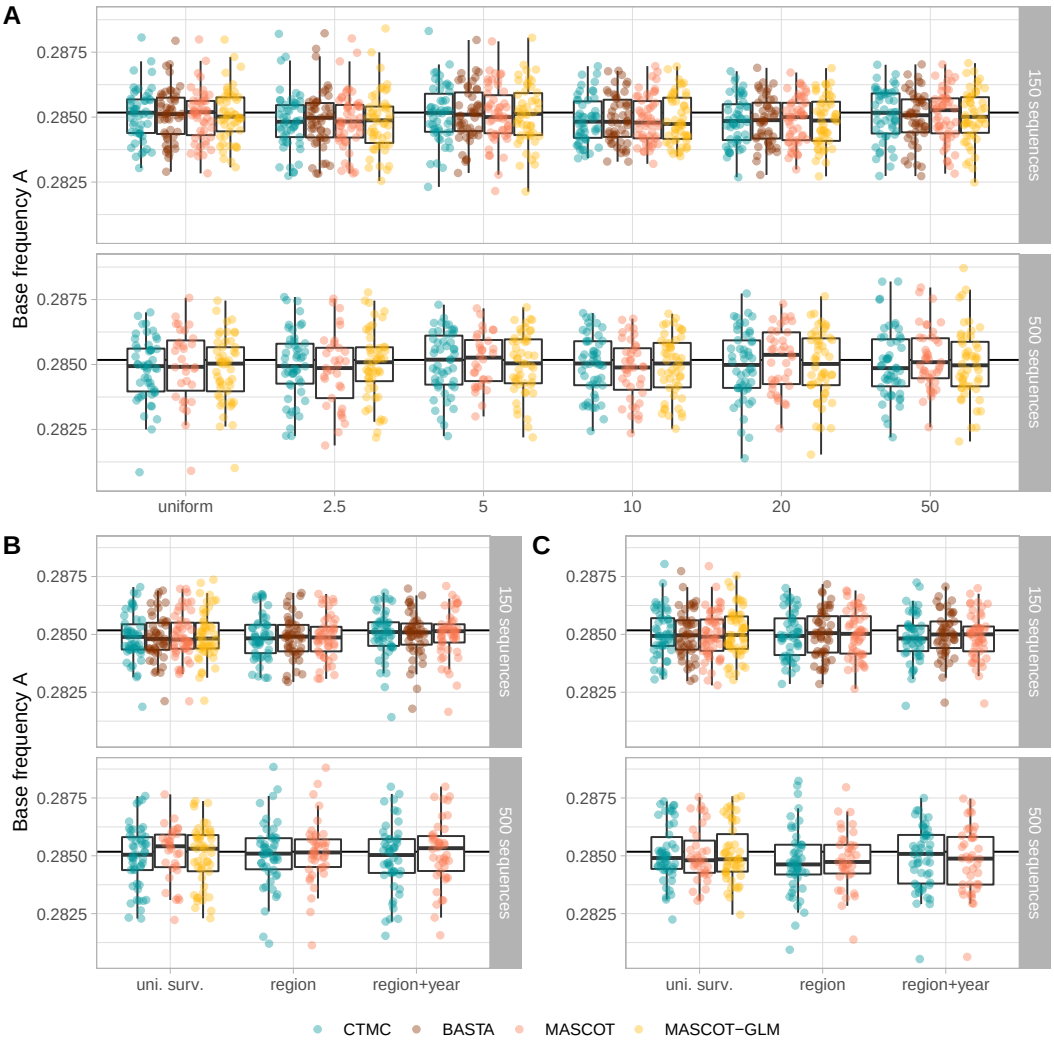


Figure D2: Median estimates of base frequency A for all sampling conditions and the four algorithms. Median estimate of the base frequency of A for the systematic bias (**A**), surveillance bias 10 (**B**), and surveillance bias 20 (**C**) sampling conditions. The true value of the parameter is represented as the horizontal black line.

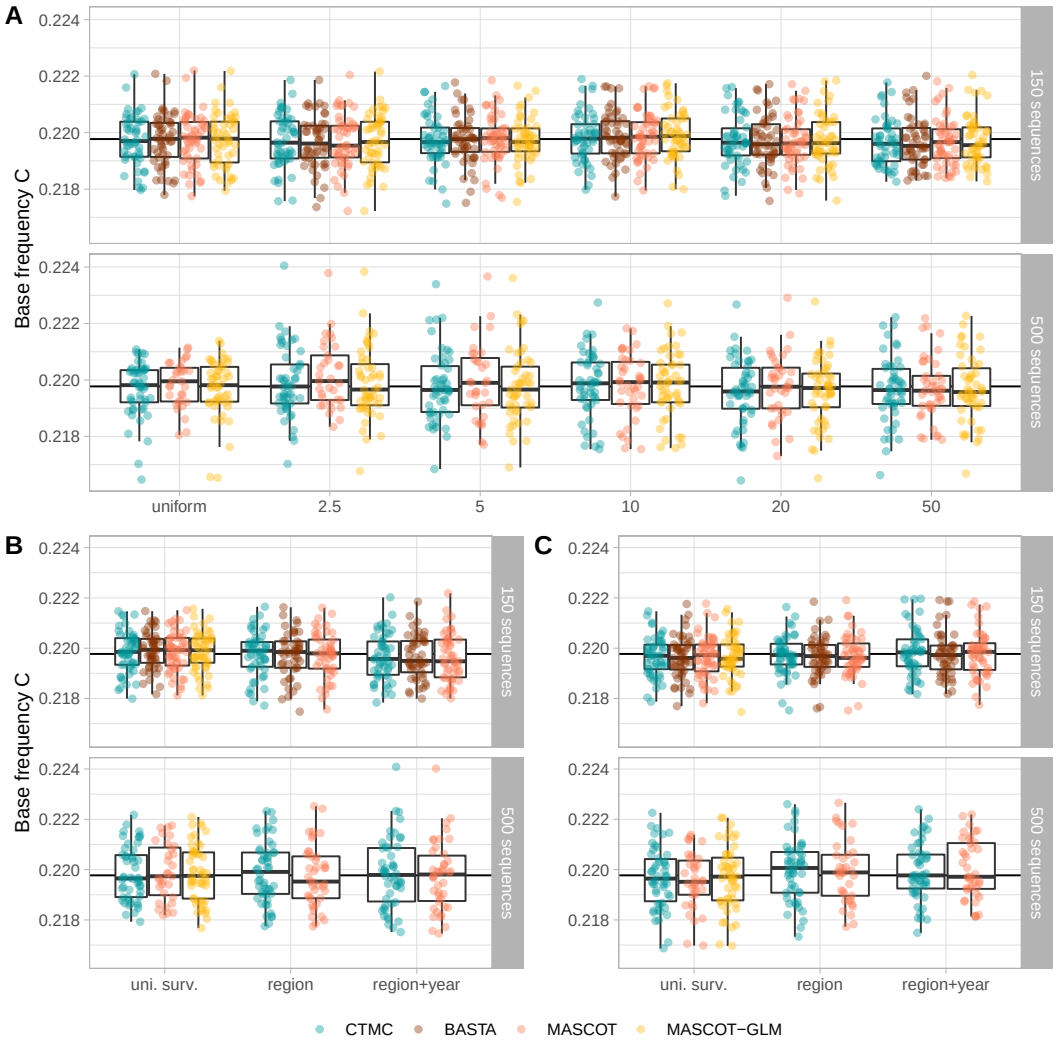


Figure D3: Median estimates of base frequency C for all sampling conditions and the four algorithms. Median estimate of the base frequency of C for the systematic bias (**A**), surveillance bias 10 (**B**), and surveillance bias 20 (**C**) sampling conditions. The true value of the parameter is represented as the horizontal black line.

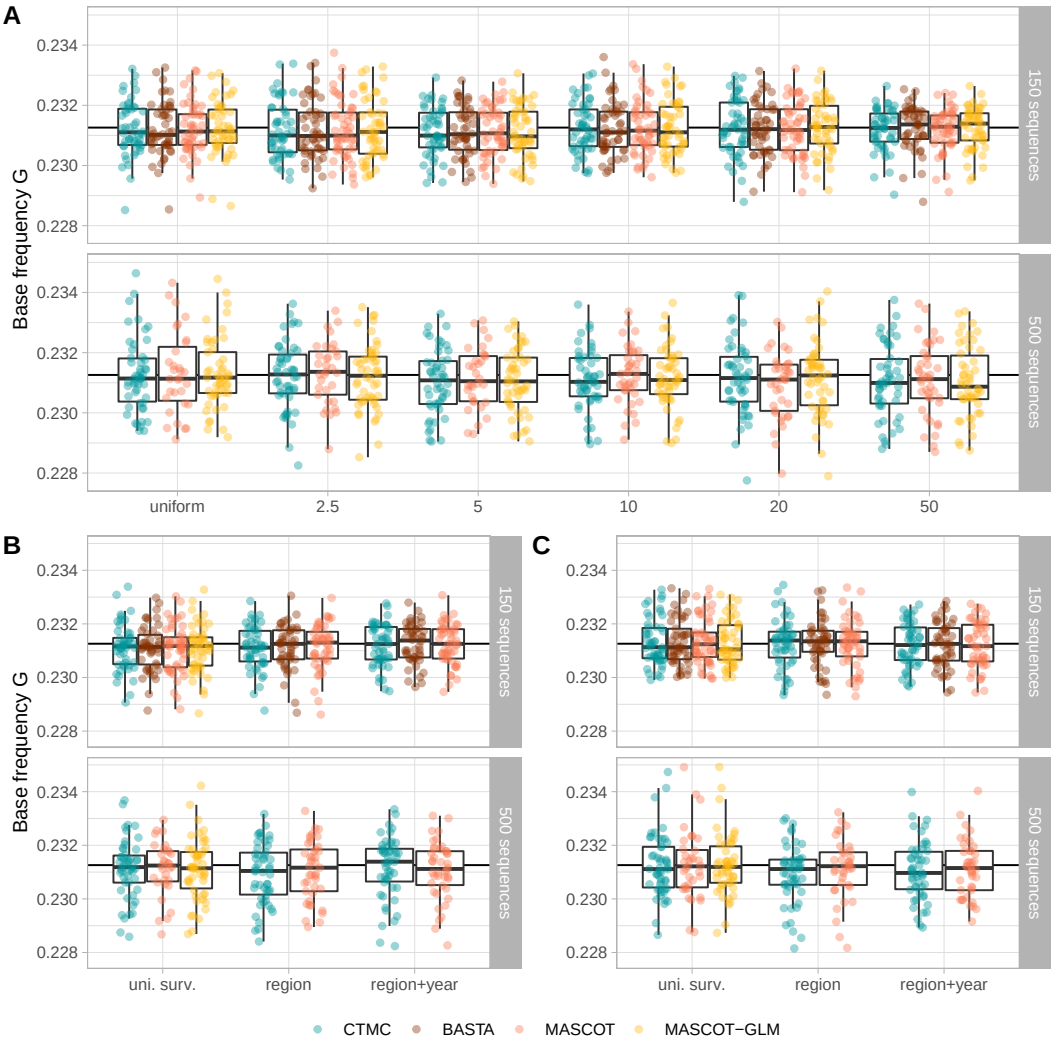


Figure D4: Median estimates of base frequency G for all sampling conditions and the four algorithms. Median estimate of the base frequency of G for the systematic bias (**A**), surveillance bias 10 (**B**), and surveillance bias 20 (**C**) sampling conditions. The true value of the parameter is represented as the horizontal black line.

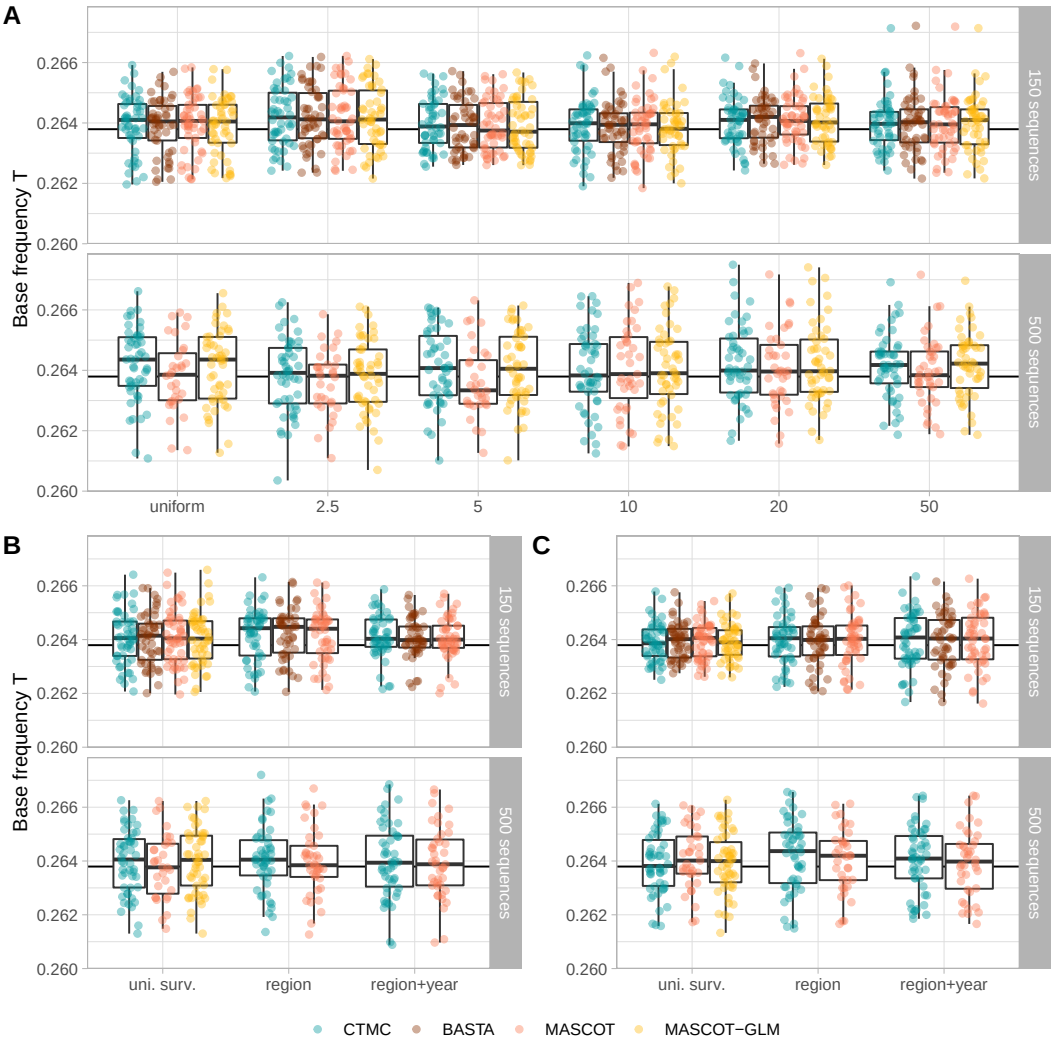


Figure D5: Median estimates of base frequency T for all sampling conditions and the four algorithms. Median estimate of the base frequency of T for the systematic bias (**A**), surveillance bias 10 (**B**), and surveillance bias 20 (**C**) sampling conditions. The true value of the parameter is represented as the horizontal black line.

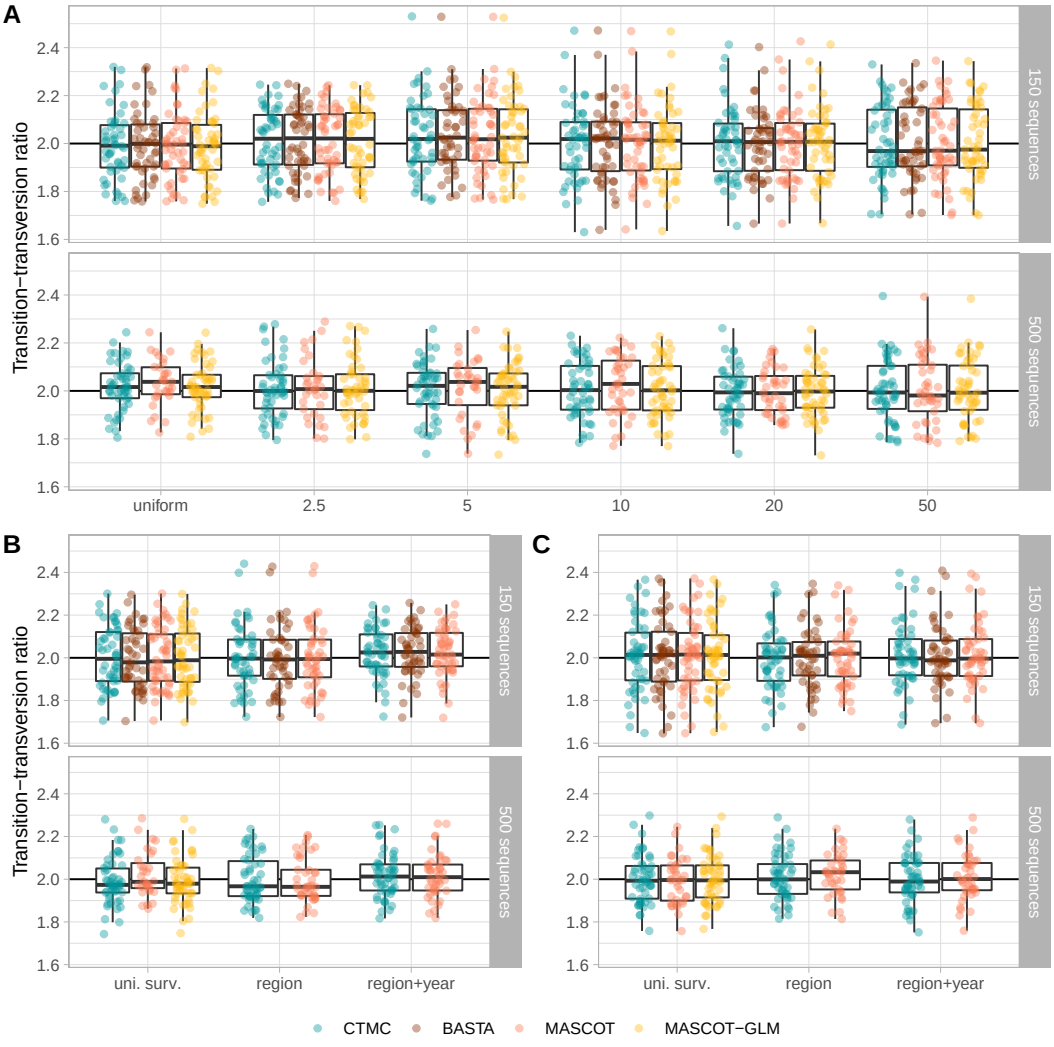


Figure D6: Median estimates of the transition-transversion ratio for all sampling conditions and the four algorithms. Median estimate of the transition-transversion ratio κ for the systematic bias (**A**), surveillance bias 10 (**B**), and surveillance bias 20 (**C**) sampling conditions. The true value of the parameter is represented as the horizontal black line.

1.2. Total migration counts

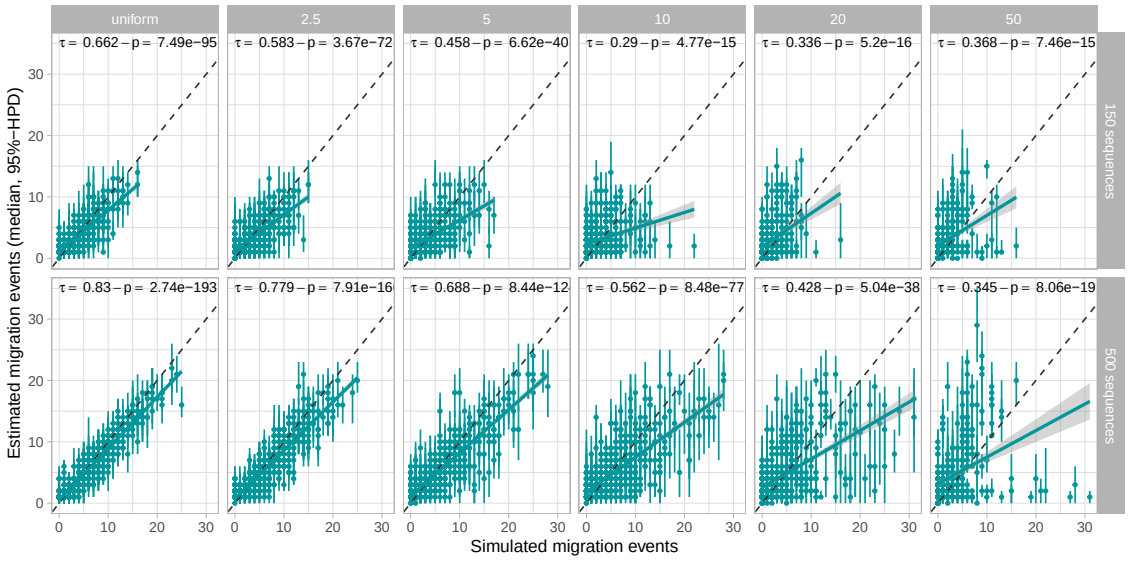


Figure D7: Impact of bias on the estimation of the total migration counts by CTMC. Median estimate of the base frequency of A for the systematic bias (**A**), surveillance bias 10 (**B**), and surveillance bias 20 (**C**) sampling conditions. The true value of the parameter is represented as the horizontal black line.

1.3. Lineage migration counts

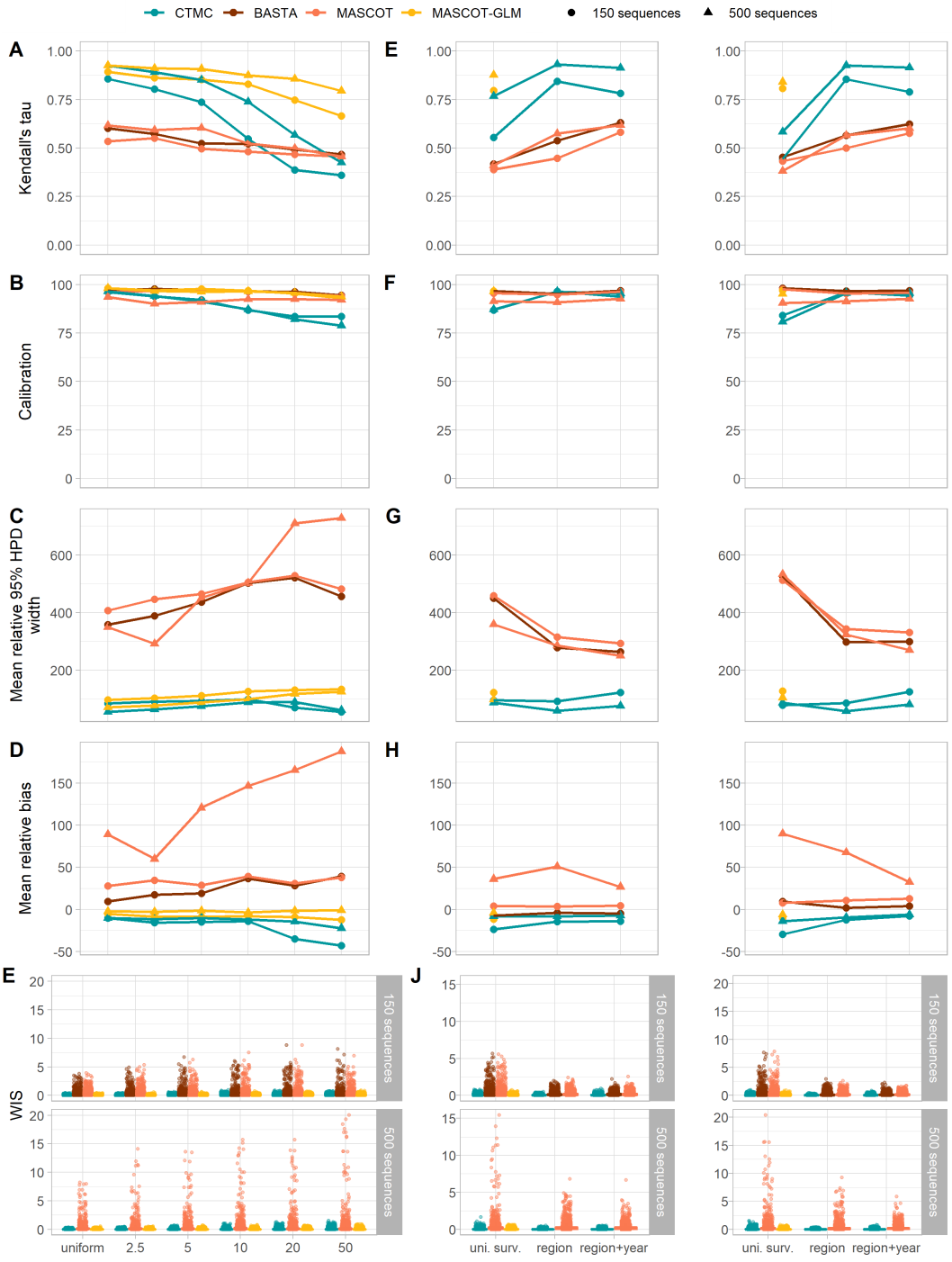


Figure D8: Impact and mitigation of spatial bias on the estimation of the lineage migration counts. (**A-E**) Impact of the increasing levels of spatial bias on the correlation, the calibration, the mean relative 95% HPD width, the mean relative bias, and the WIS between the simulated and the estimated lineage migration counts. (**F-J**) Mitigation of the impact of spatial bias on the correlation, the calibration, the mean relative 95% HPD width, the mean relative bias, and the WIS between the simulated and estimated lineage migration counts by using alternative sampling strategies. The mean relative bias and the mean relative 95% HPD width are not defined when the true value is null. We removed 65,399 out of 91,434 and 27,629 out of 41,916 simulated migration events in the small and large samples, respectively, due to true null values.

2. Impact and mitigation of bias in the three demes framework

2.1. Estimation of genetic parameters

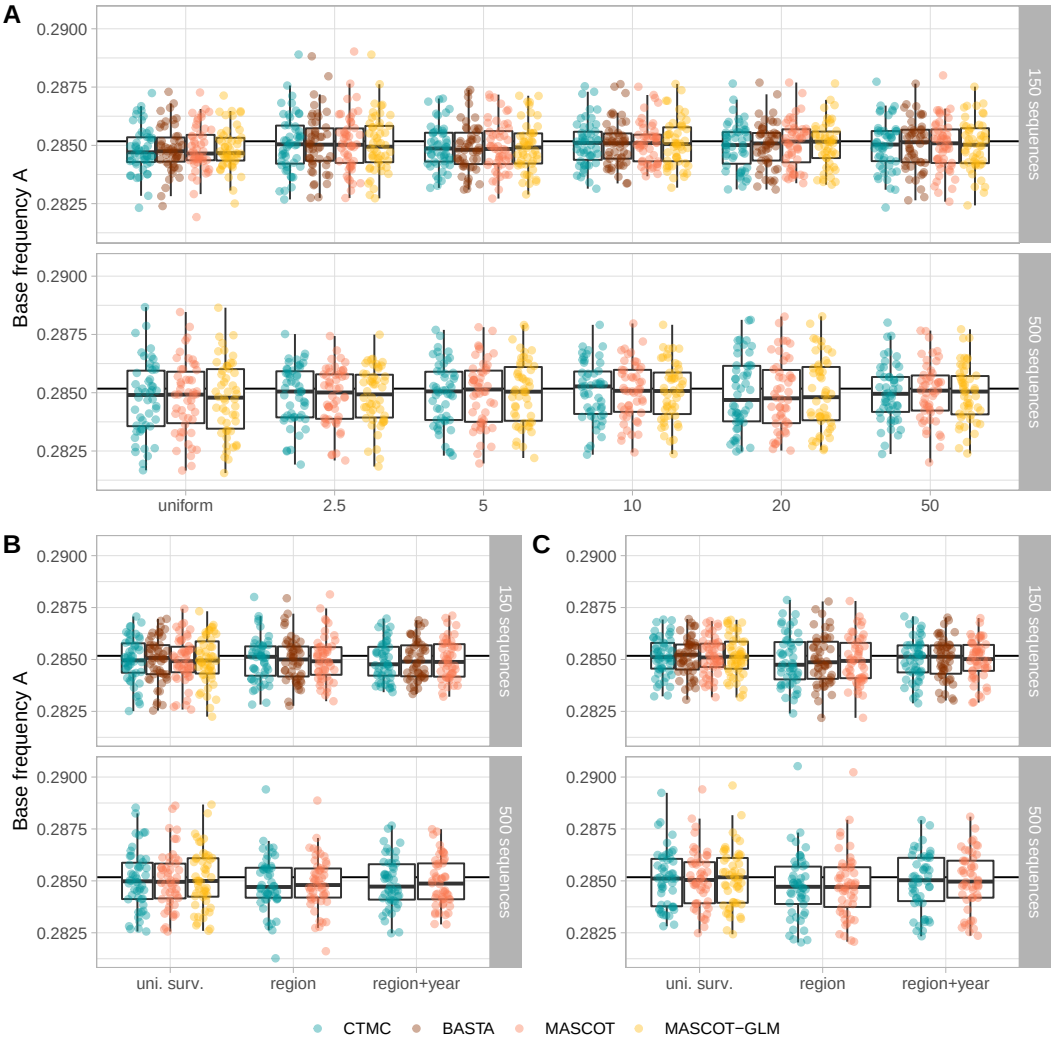


Figure D9: Median estimates of base frequency A for all sampling conditions and the four algorithms. Median estimate of the base frequency of A for the systematic bias (**A**), surveillance bias 10 (**B**), and surveillance bias 20 (**C**) sampling conditions. The true value of the parameter is represented as the horizontal black line.

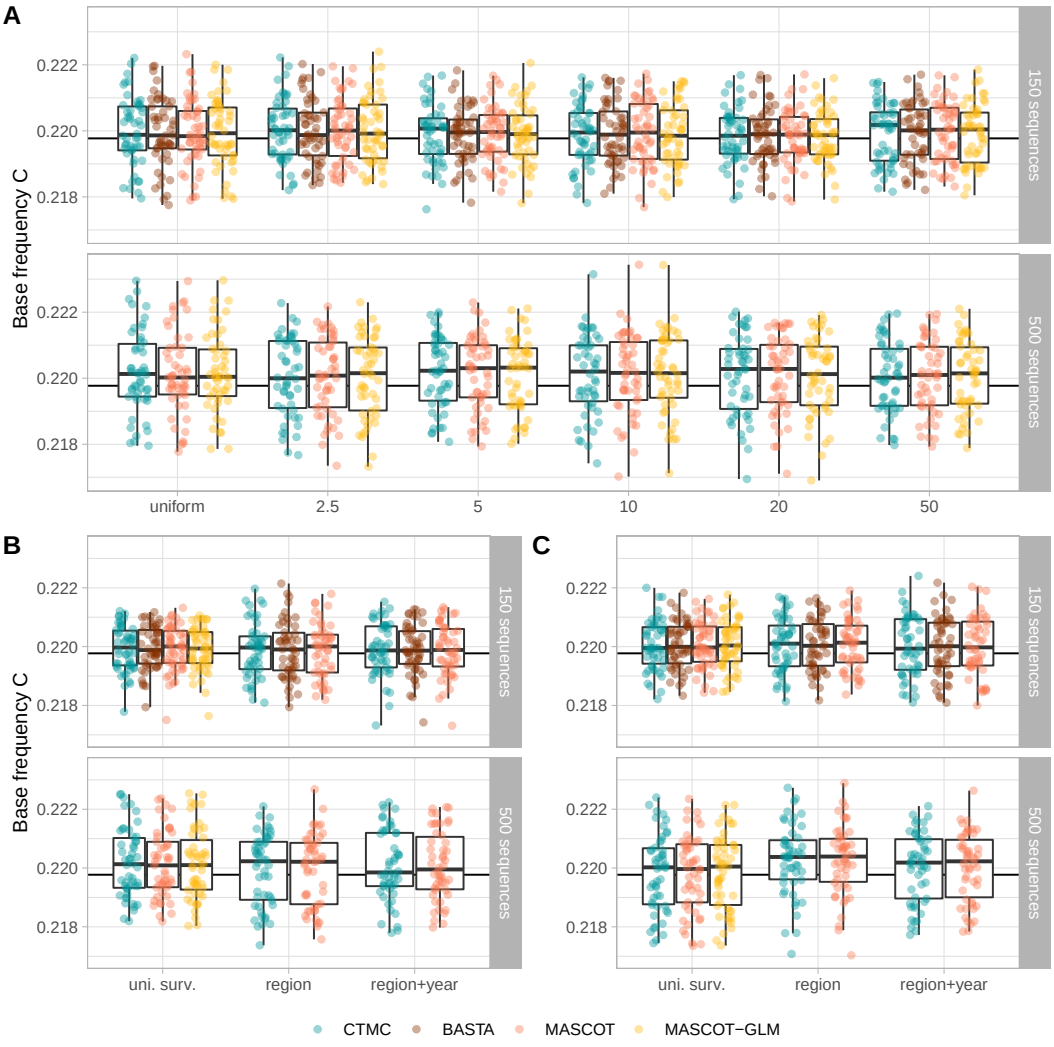


Figure D10: Median estimates of base frequency C for all sampling conditions and the four algorithms. Median estimate of the base frequency of C for the systematic bias (**A**), surveillance bias 10 (**B**), and surveillance bias 20 (**C**) sampling conditions. The true value of the parameter is represented as the horizontal black line.

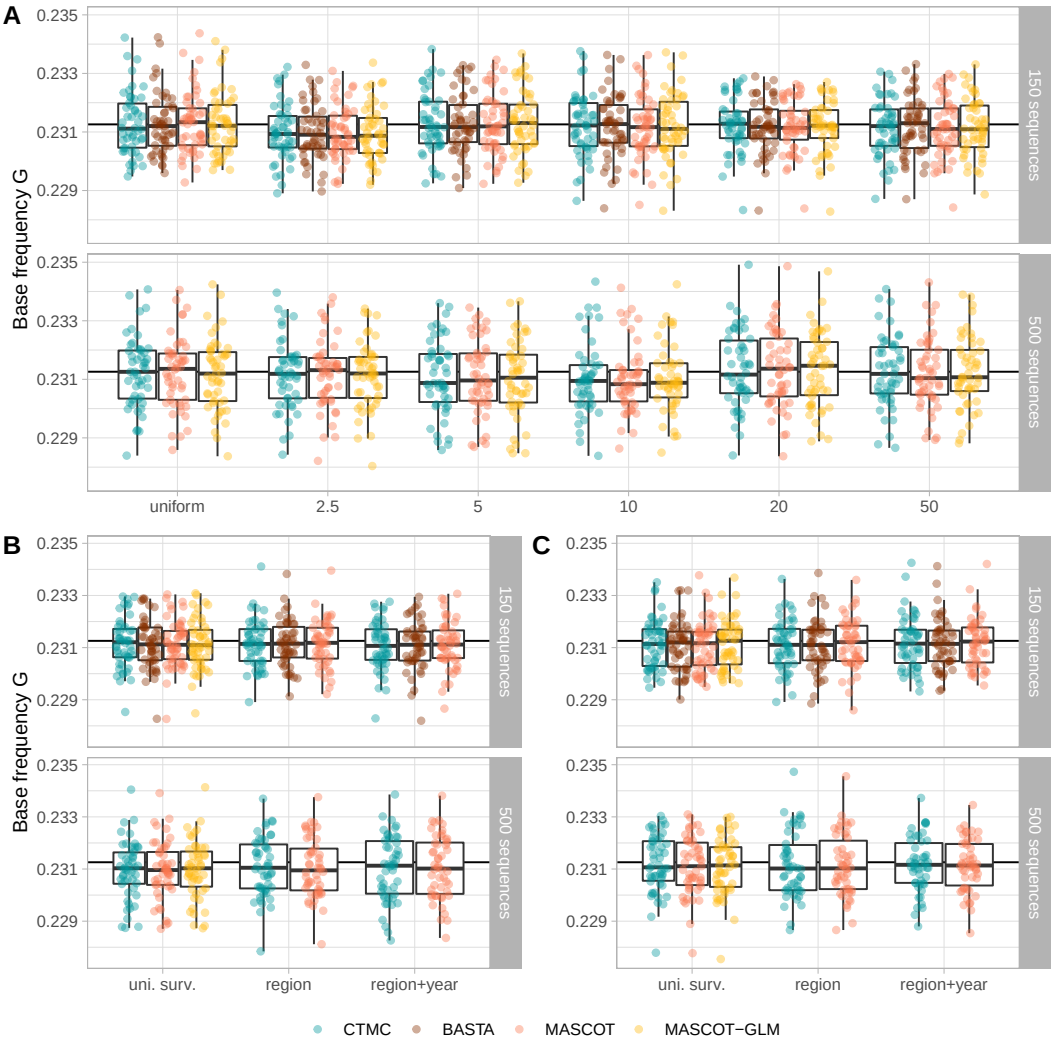


Figure D11: Median estimates of base frequency G for all sampling conditions and the four algorithms. Median estimate of the base frequency of G for the systematic bias (**A**), surveillance bias 10 (**B**), and surveillance bias 20 (**C**) sampling conditions. The true value of the parameter is represented as the horizontal black line.

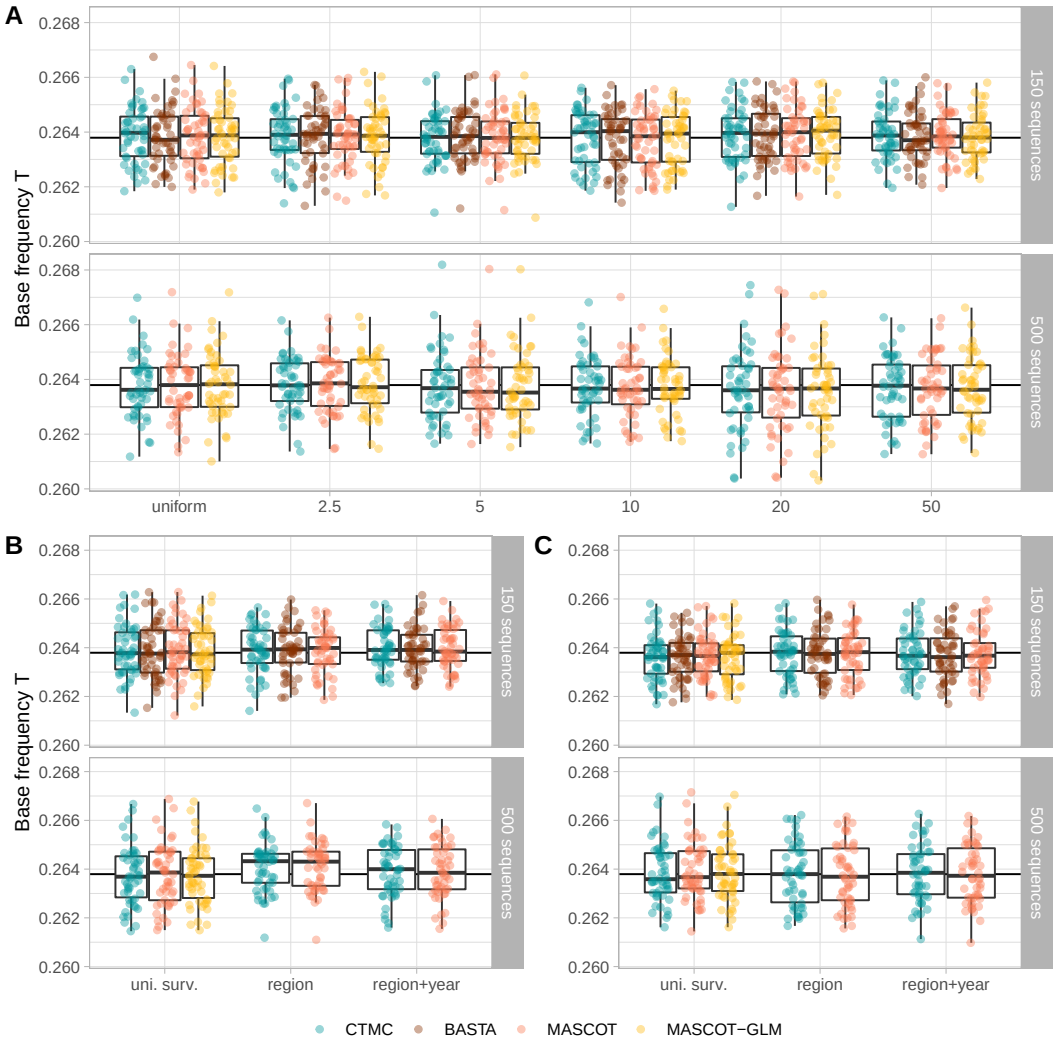


Figure D12: Median estimates of base frequency T for all sampling conditions and the four algorithms. Median estimate of the base frequency of T for the systematic bias (**A**), surveillance bias 10 (**B**), and surveillance bias 20 (**C**) sampling conditions. The true value of the parameter is represented as the horizontal black line.

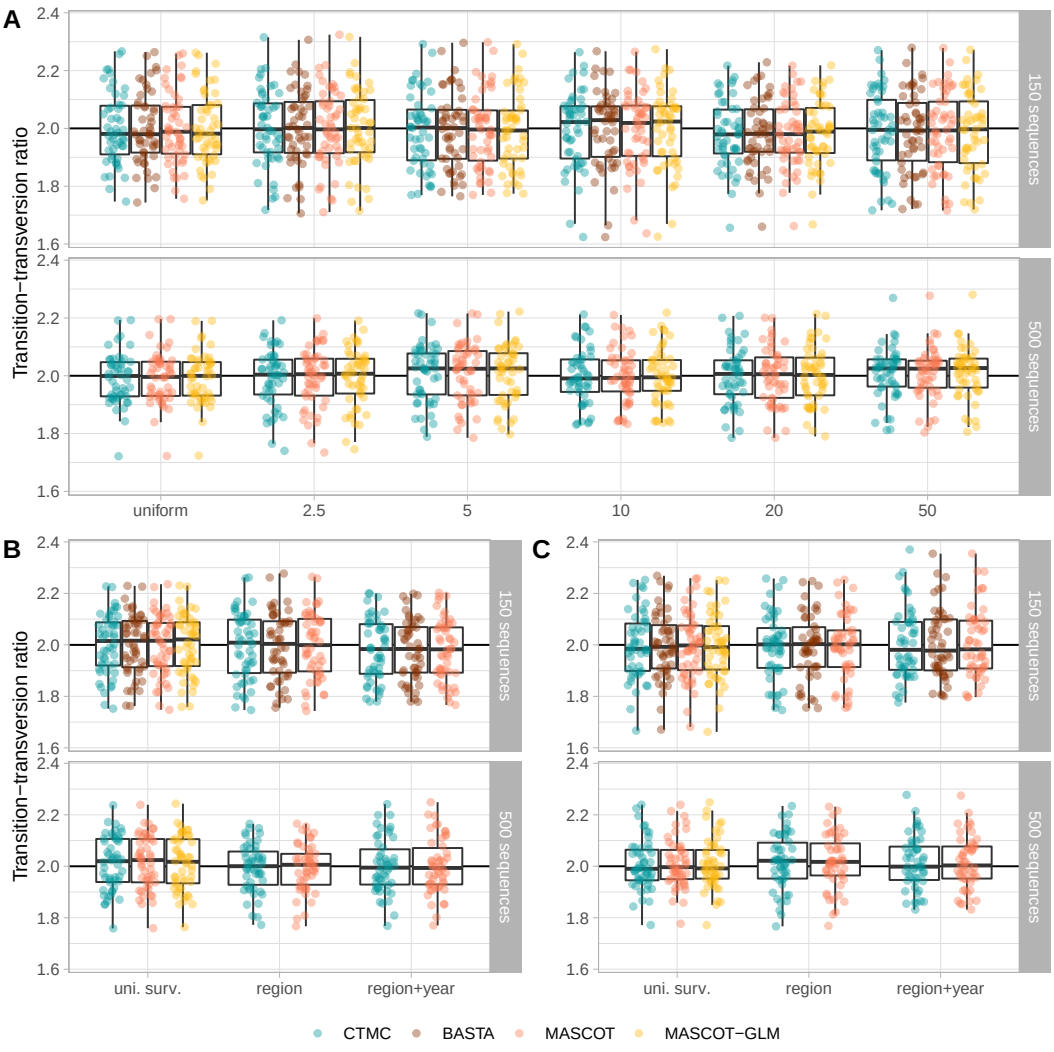


Figure D13: Median estimates of the transition-transversion ratio for all sampling conditions and the four algorithms. Median estimate of the transition-transversion ratio κ for the systematic bias (**A**), surveillance bias 10 (**B**), and surveillance bias 20 (**C**) sampling conditions. The true value of the parameter is represented as the horizontal black line.

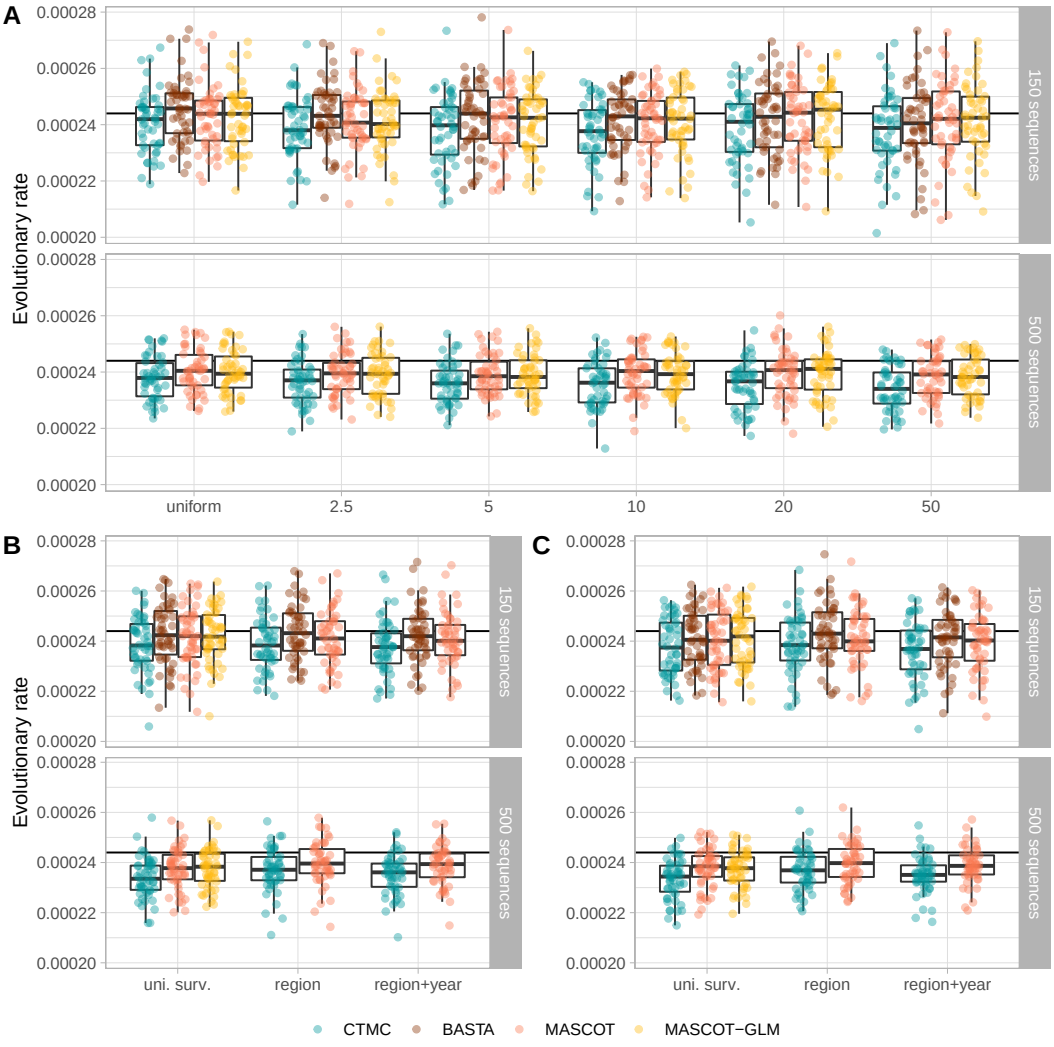


Figure D14: Median estimates of the evolutionary rate for all sampling conditions and the four algorithms. Median estimate of the evolutionary rate for the systematic bias (**A**), surveillance bias 10 (**B**), and surveillance bias 20 (**C**) sampling conditions. The true value of the parameter is represented as the horizontal black line.

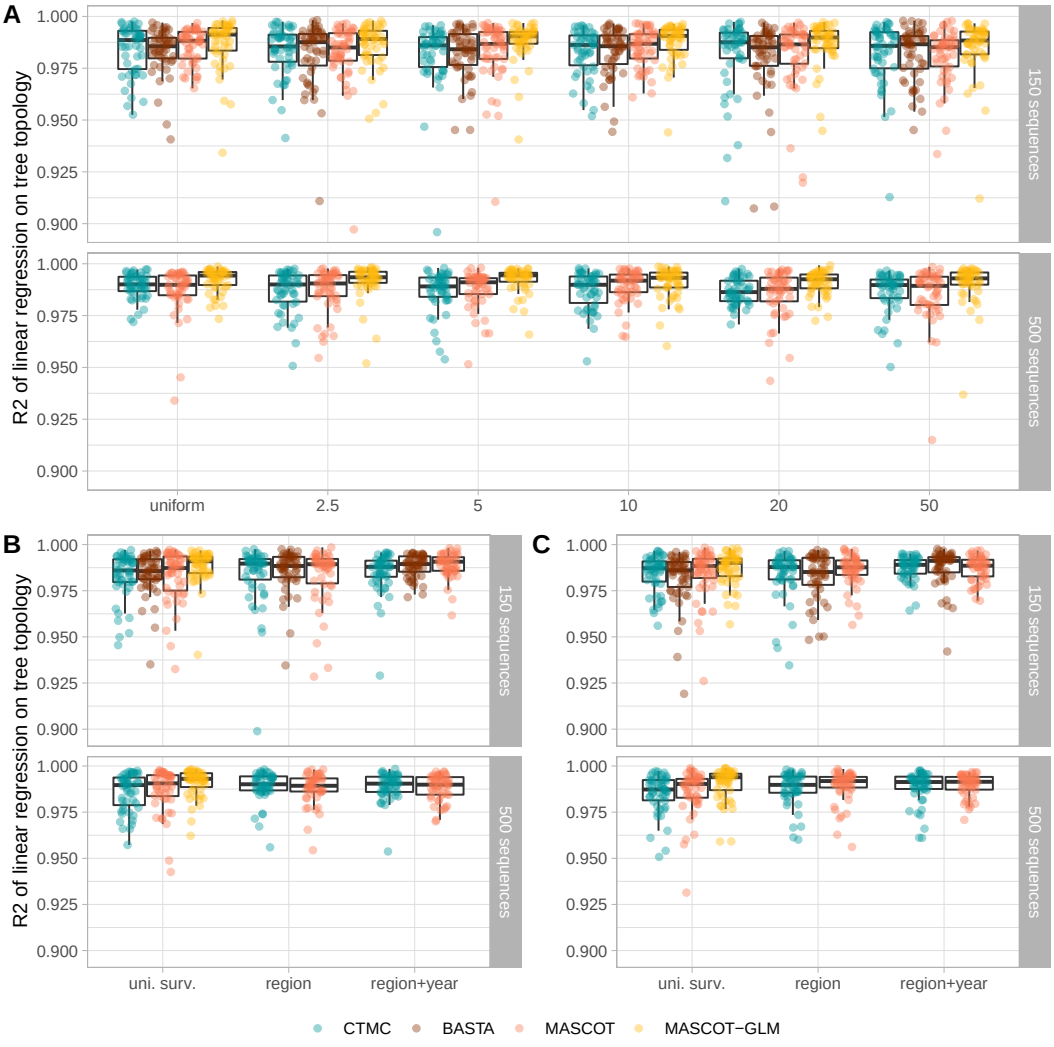


Figure D15: Comparison of the simulated and estimated tree topologies for all sampling conditions and the four algorithms. Pearson's determination coefficient of the pairwise divergence time between the simulated transmission chain and the MCC tree for the systematic bias (**A**), surveillance bias 10 (**B**), and surveillance bias 20 (**C**) sampling conditions.

2.2. Lineage migration counts

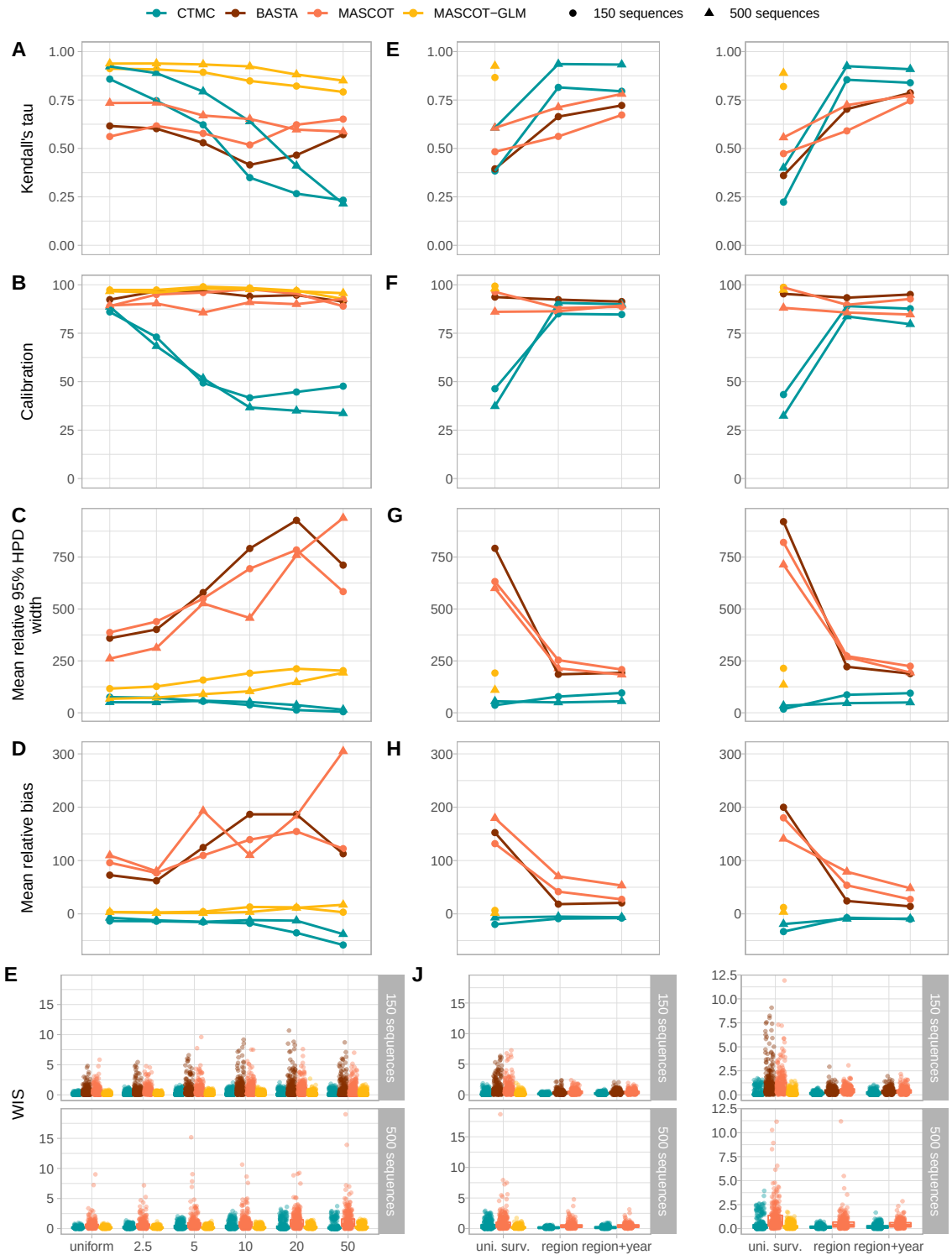


Figure D16: Impact and mitigation of spatial bias on the estimation of the lineage migration counts. (**A-E**) Impact of the increasing levels of spatial bias on the correlation, the calibration, the mean relative 95% HPD width, the mean relative bias, and the WIS between the simulated and the estimated lineage migration counts. (**F-J**) Mitigation of the impact of spatial bias on the correlation, the calibration, the mean relative 95% HPD width, the mean relative bias, and the WIS between the simulated and estimated lineage migration counts by using alternative sampling strategies. The mean relative bias and the mean relative 95% HPD width are not defined when the true value is null. We removed 3,126 out of 13,200 and 1,410 out of 9,588 simulated migration events in the small and large samples, respectively, due to null true values.

2.3. Total migration counts

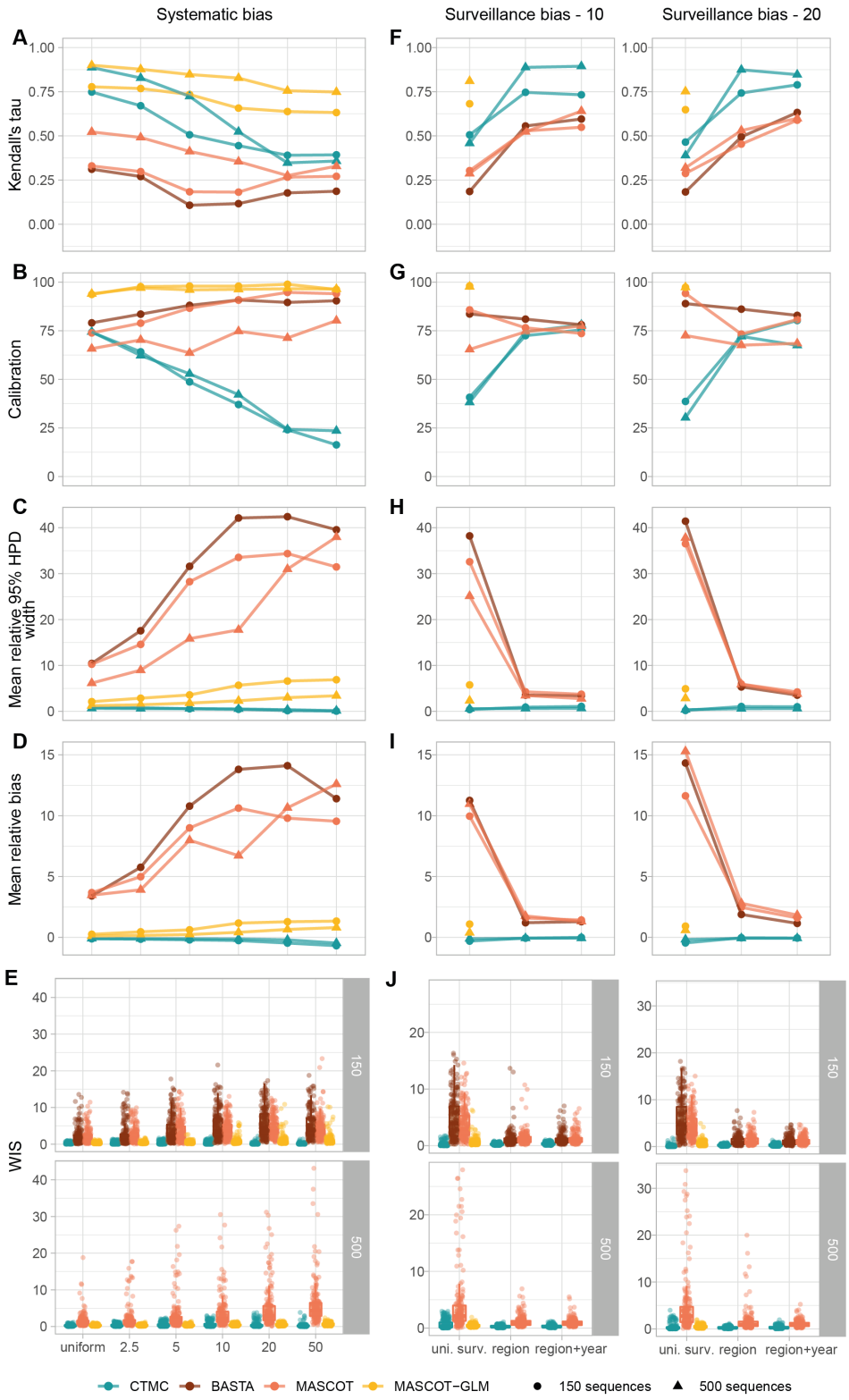


Figure D17: Impact and mitigation of spatial bias on the estimation of the total migration counts. (**A-E**) Impact of the increasing levels of spatial bias on the correlation, the calibration, the mean relative 95% HPD width, the mean relative bias, and the WIS between the simulated and the estimated total migration counts. (**F-J**) Mitigation of the impact of spatial bias on the correlation, the calibration, the mean relative 95% HPD width, the mean relative bias, and the WIS between the simulated and estimated total migration counts by using alternative sampling strategies. The mean relative bias and the mean relative 95% HPD width are not defined when the true value is null. We removed 612 out of 3,600 and 380 out of 3,600 simulated migration events in the small and large samples, respectively, due to true null values.

2.4. Introduction dates

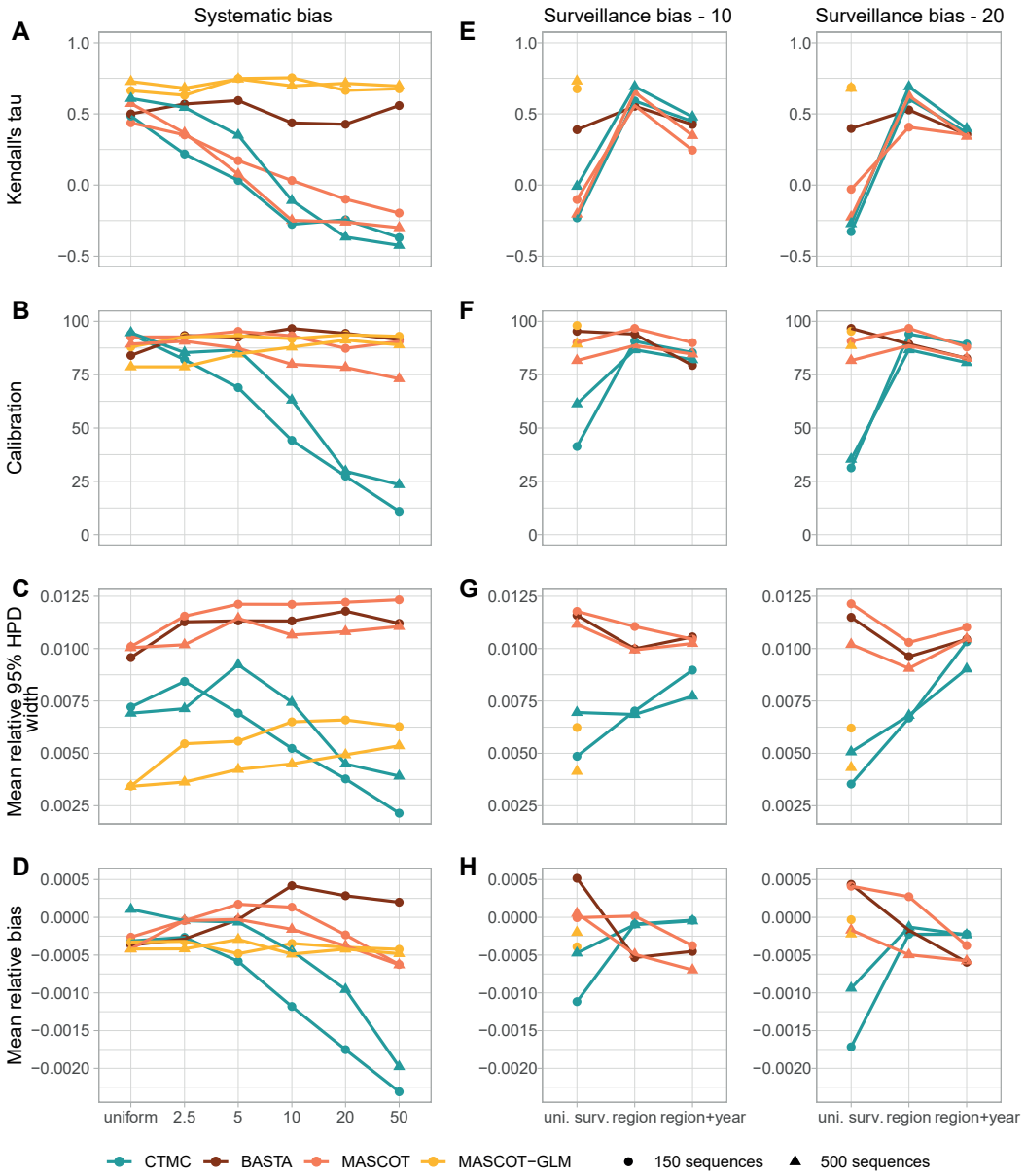


Figure D18: Impact and mitigation of spatial bias on the estimation of the introduction dates. (**A-D**) Impact of the increasing levels of spatial bias on correlation, calibration, mean relative 95% HPD width, and average relative error between the simulated and estimated introduction dates. (**E-H**) Mitigation of the impact of spatial bias on correlation, calibration, mean relative 95% HPD width, and average relative error between the simulated and estimated introduction dates by using alternative sampling strategies.

2.5. Root location

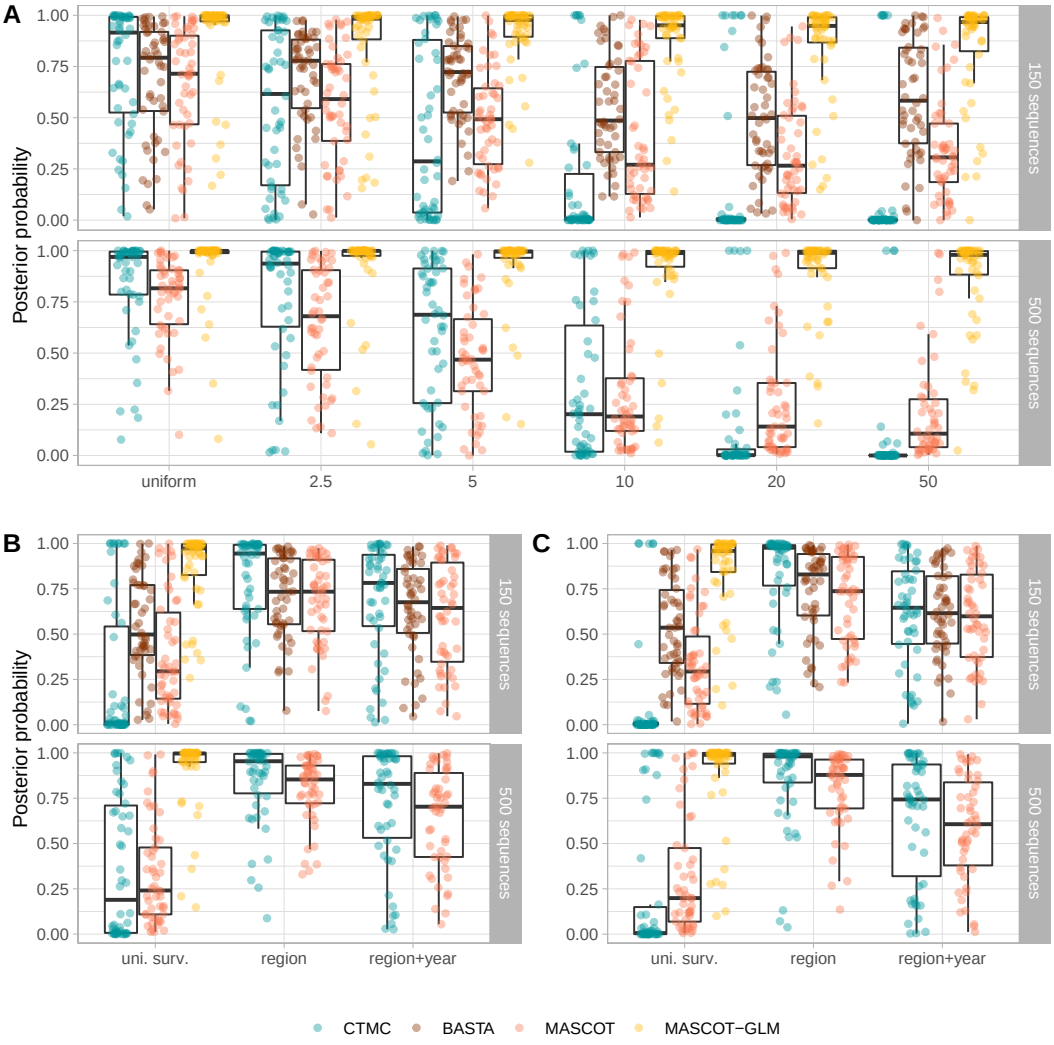


Figure D19: Impact and mitigation of spatial bias on the estimation of the root location. (**A**) Decreasing root state posterior probability with an increasing bias. (**B-C**) Mitigation of the effects of spatial bias using alternative sampling strategies with an underlying bias of 10 and 20, respectively. Each dot corresponds to the median root state posterior probability in one simulation (n = 50 per sampling protocol and sample size).

3. RABV spread in the Philippines

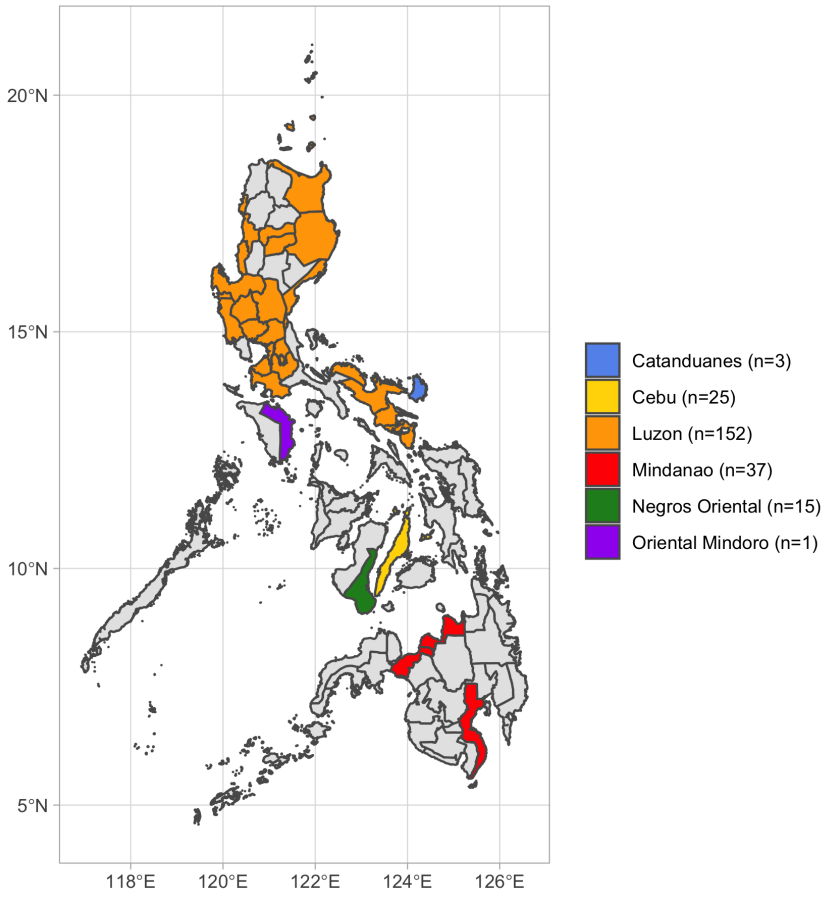


Figure D20: Sampling location and intensity of the RABV sequences in the Philippines, 2004-2010.

4. SARS-CoV-2 early spread

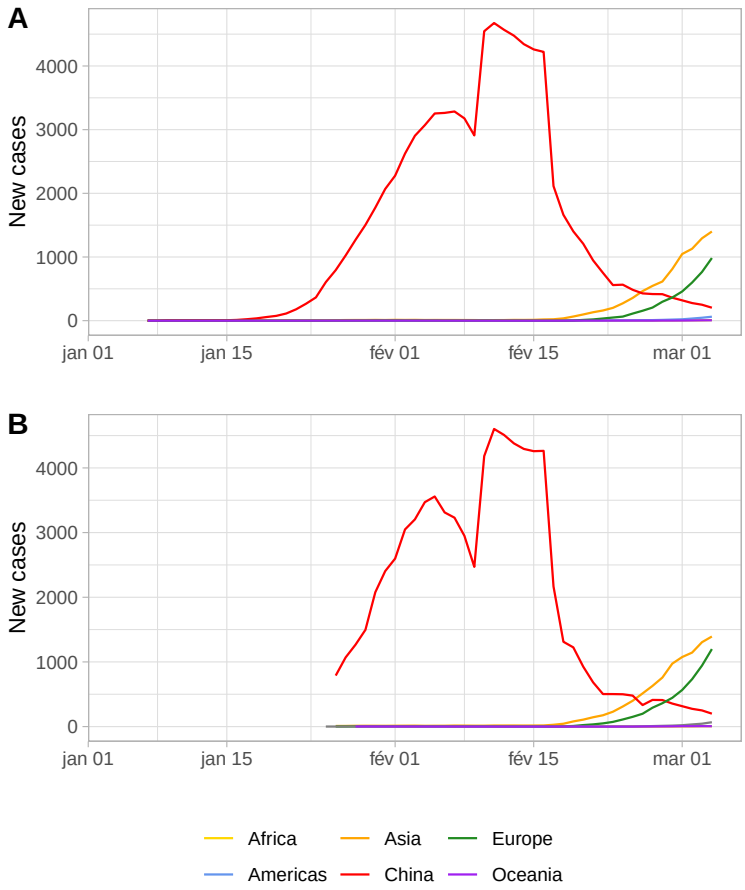
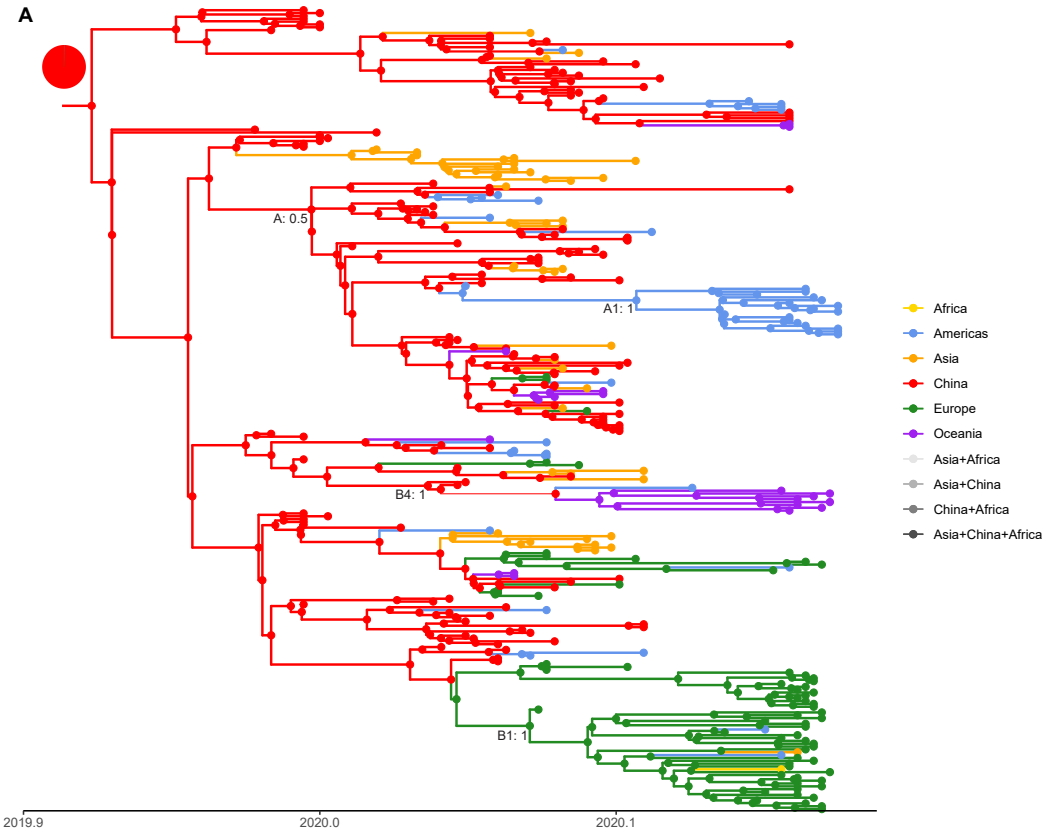
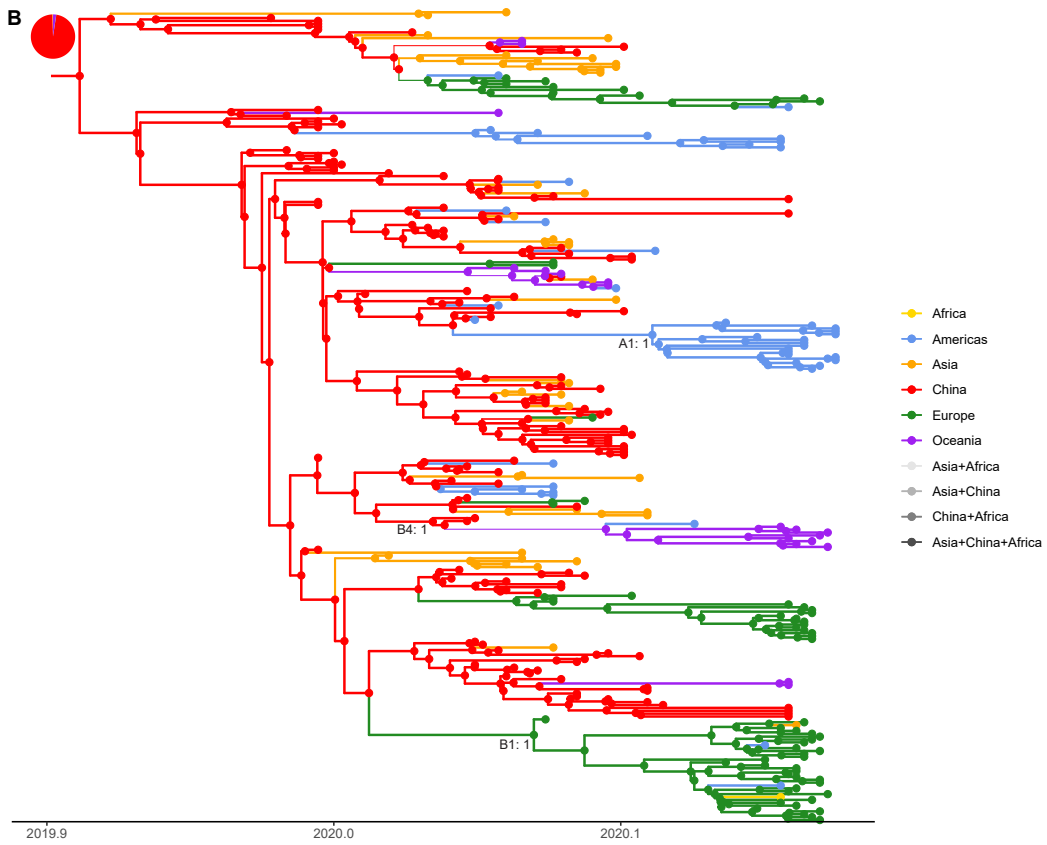
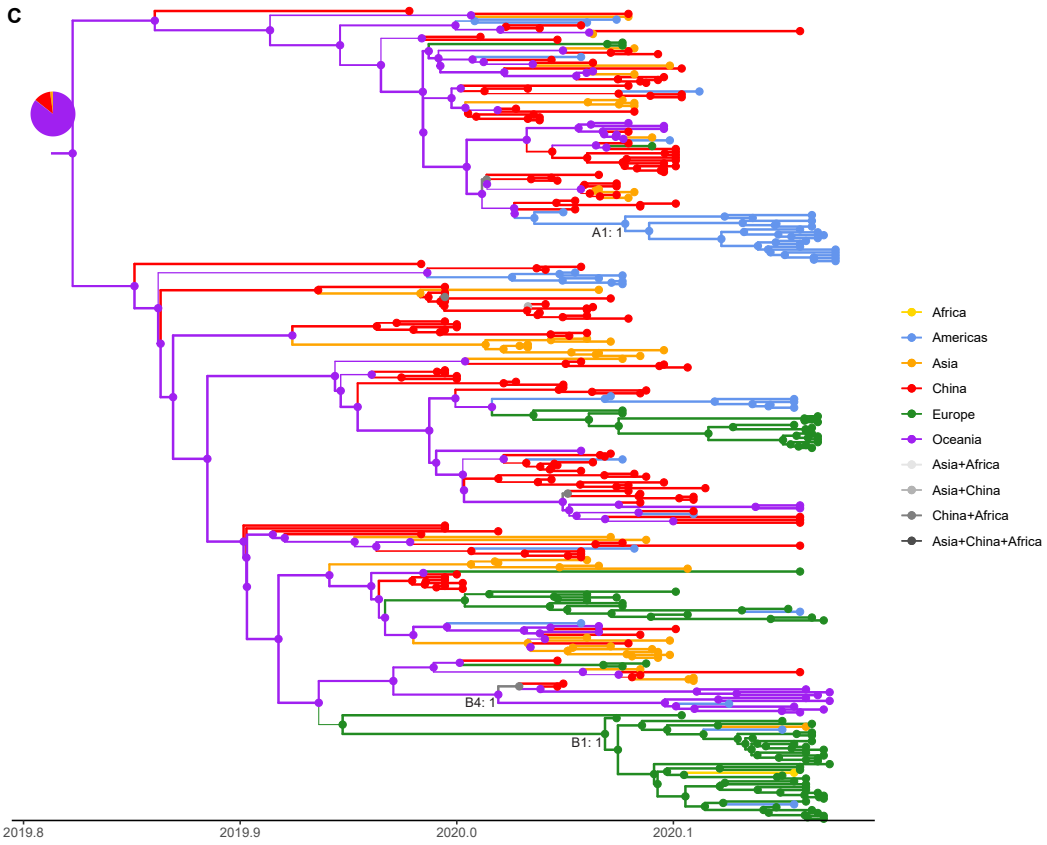
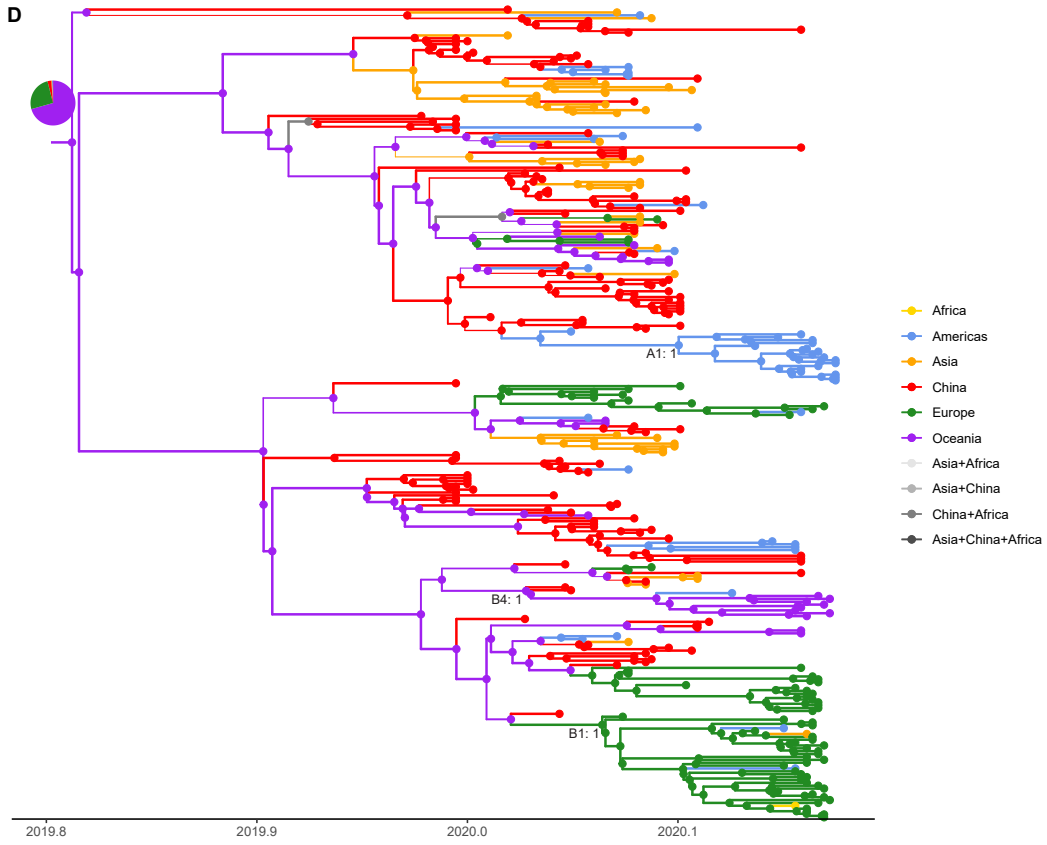


Figure D21: Case count data by continent used to inform MASCOT. (**A-B**) Case count data from the World Health Organization and Our World In Data, respectively. We report the moving average over a seven-day window.


A
A: 0.5
A1: 1
B4: 1
B1: 1
2019.9
2020.0
2020.1
Africa
Americas
Asia
China
Europe
Oceania
Asia+Africa
Asia+China
China+Africa
Asia+China+Africa


B
A1: 1
B4: 1
B1: 1
2019.9
2020.0
2020.1
Africa
Americas
Asia
China
Europe
Oceania
Asia+Africa
Asia+China
China+Africa
Asia+China+Africa


C
A1: 1
B4: 1
B1: 1
2019.8
2019.9
2020.0
2020.1
Africa
Americas
Asia
China
Europe
Oceania
Asia+Africa
Asia+China
China+Africa
Asia+China+Africa


D
A1: 1
B4: 1
B1: 1
2019.8
2019.9
2020.0
2020.1
Africa
Americas
Asia
China
Europe
Oceania
Asia+Africa
Asia+China
China+Africa
Asia+China+Africa


E
B1: 1
B4: 1
A1: 1
2019.9
2020.0
2020.1
Africa
Americas
Asia
China
Europe
Oceania
Asia+Africa
Asia+China
China+Africa
Asia+China+Africa

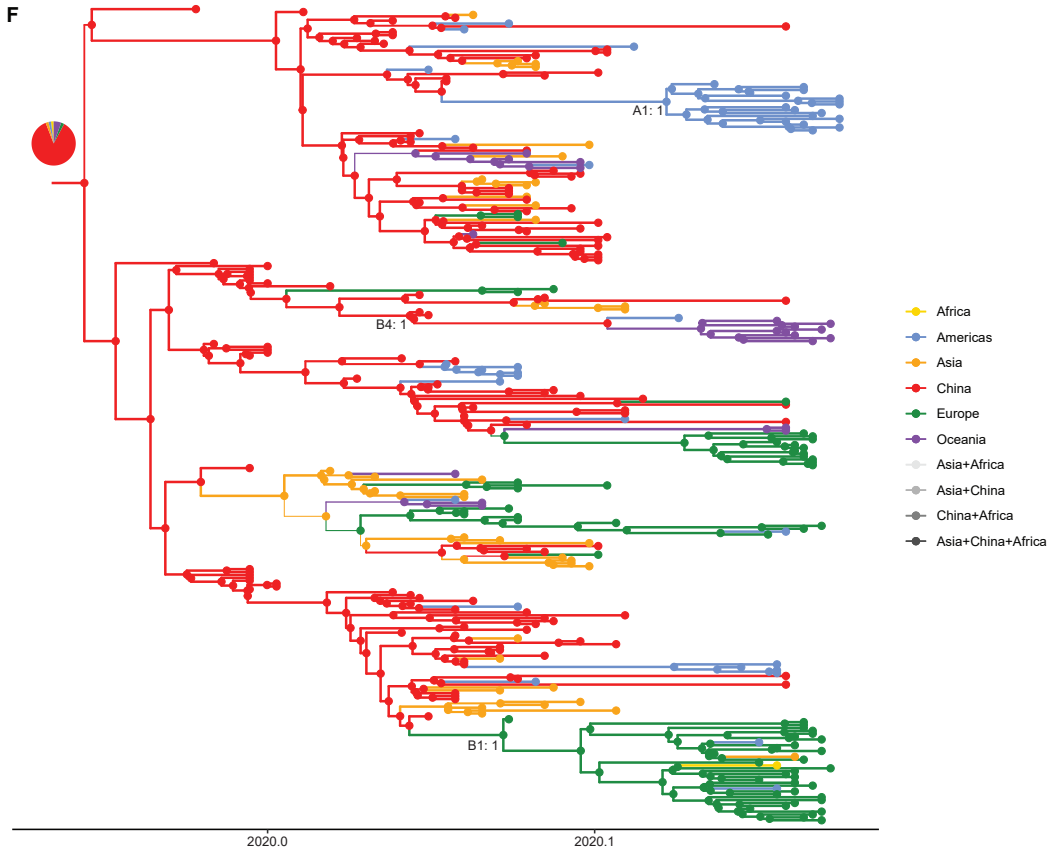


Figure D22: Maximum clade credibility (MCC) tree of SARS-CoV-2 genomes from the early stages of the pandemic for all the algorithms. (**A-F**) MCC trees for CTMC, BASTA (1st mode of evolutionary rate), BASTA (2nd mode of evolutionary rate), MASCOT, MASCOT-WID, and MASCOT-WHO respectively. The posterior support of lineages A, A1, B1, and B4 that were identified by (Lemey et al., 2020) are reported at the corresponding nodes when they gather the same sequences as in the original analysis. Lineages A1, B1, and B4 were estimated to be monophyletic with a high posterior support by all algorithms. However, BASTA, MASCOT, MASCOT-WID, and MASCOT-WHO did not infer monophyly for lineage A that is why we did not report its posterior supports on the corresponding MCC trees. Branch width is proportional to the maximal location probability of the parent node. Nodes and branches are colored by location with maximal probability. The root location probability distribution is reported in the pie chart.

Table D1: Predicted lineage location on the SARS-CoV-2 data. We report the predicted locations with maximal probability of the four SARS-CoV-2 lineages estimated by all algorithms along with their posterior probability. We did not report the location of lineage A predicted by BASTA, MASCOT, MASCOT-WID, and MASCOT-WHO because these algorithms did not infer a monophyletic lineage. Location of lineage A1 and B1 is estimated in a similar fashion by all algorithms. However, lineage B4 was estimated to be located in Oceania by MASCOT and the 2nd mode of BASTA, whereas it is estimated to be located in China by CTMC-TRAVEL and CTMC. When we add case count data from Our World In Data and the WHO, the estimated lineage location is China but the posterior probability for MASCOT-WID is lower than for CTMC-TRAVEL and CTMC.

	Lineage A1	**Lineage A**	**Lineage B1**	**Lineage B4**
CTMC-TRAVEL	Americas (1)	China (1)	Europe (0.761)	China (1)
CTMC	Americas (0.995)	China (1)	Europe (0.998)	China (1)
BASTA - 1st mode	Americas (0.958)	-	Europe (0.985)	China (0.983)
BASTA - 2nd mode	Americas (0.97)	-	Europe (0.977)	Oceania (0.851)
MASCOT	Americas (0.994)	-	Europe (0.993)	Oceania (0.781)
MASCOT-WID	Americas (0.937)	-	Europe (0.993)	China (0.446)
MASCOT-WHO	Americas (0.959)	-	Europe (0.977)	China (0.975)

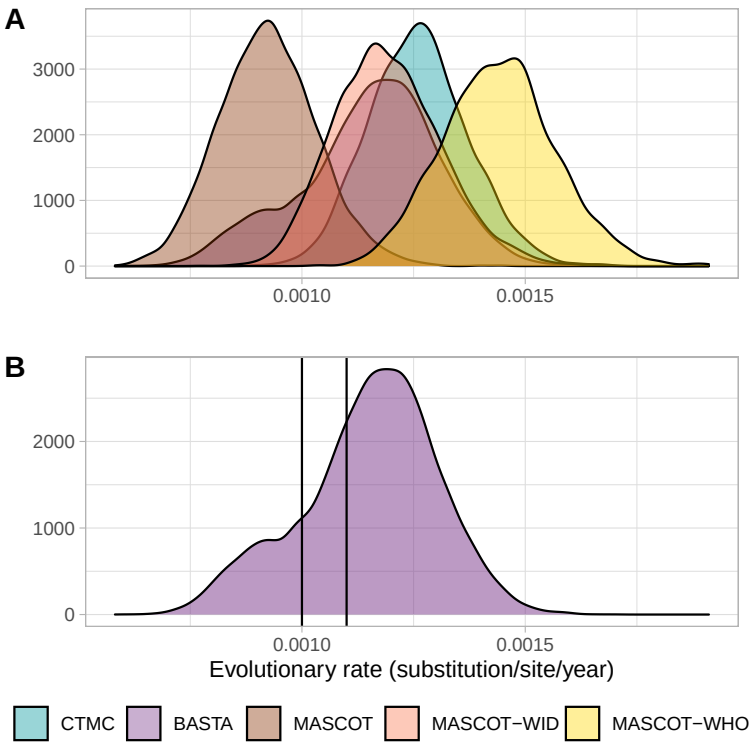


Figure D23: Posterior kernel density distribution of the evolutionary rate estimated on the SARS-CoV-2 data set. (**A**) Posterior kernel density distribution of the evolutionary rate estimated by all algorithms. BASTA posterior distribution is bimodal with its major mode close to the estimates of CTMC and MASCOT-WID, and the second mode close to the estimate of MASCOT. (**B**) Focus on the posterior kernel density distribution of BASTA. Due to the bimodality of the posterior density distribution, we split in two the tree posterior distribution according to the value of the evolutionary rate. The major mode corresponds to posterior samples with an evolutionary rate higher than 11^{-4} substitution.site^{-1}.year^{-1} and the minor mode to posterior samples with an evolutionary rate lower than 10^{-4} substitution.site^{-1}.year^{-1}.

5. Simulation framework of RABV epidemics

Table D2: Values of simulation parameters.

Notation	Parameter description	Value	Source
b	**Dog birth rate per day**	1/365	Assumption
d	**Dog death rate per day**	1/365	Assumption
β	**Rabies transmission rate**	3.2	Assumption
H_i	**Human population size per region**	**7 demes** Region 1: 4,917,672 Region 2: 2,208,003 Region 3: 11,913,790 Region 4: 5,773,588 Region 5: 3,431,383 Region 6: 5,023,878 Region 7: 672,319 **3 demes** Region 1: 10,557,059 Region 2: 17,687,379 Region 3: 5,696,197	WorldPop n.d.
r_d	**Dog:human ratio**	0.1	Assumption
C_s	**Scaling factor**	1.00e-08	Assumption
$v_{i \to j}$	**Contact matrix**	-	Radiation model (Simini et al., 2012; Golding et al., 2015)
γ	**Infectious period**	Discretized gamma distribution from 1 to 15 days $\Gamma(3, 1.1)$	Hampson et al. 2009
ε	**Incubation period**	$\Gamma(2, 11.055)$	Hampson et al. 2009
μ	**Mutation rate**	2.44e-4 subs.site^{-1}.yr^{-1}	Troupin et al. 2016
κ	**Transition/transversion ratio**	2	Assumption
π_A	**Base A frequency in the reference genome**	0.2852	Marston et al. 2013
π_C	**Base C frequency**	0.2198	Marston et al. 2013
π_G	**Base G frequency**	0.2313	Marston et al. 2013
π_T	**Base T frequency**	0.2638	Marston et al. 2013

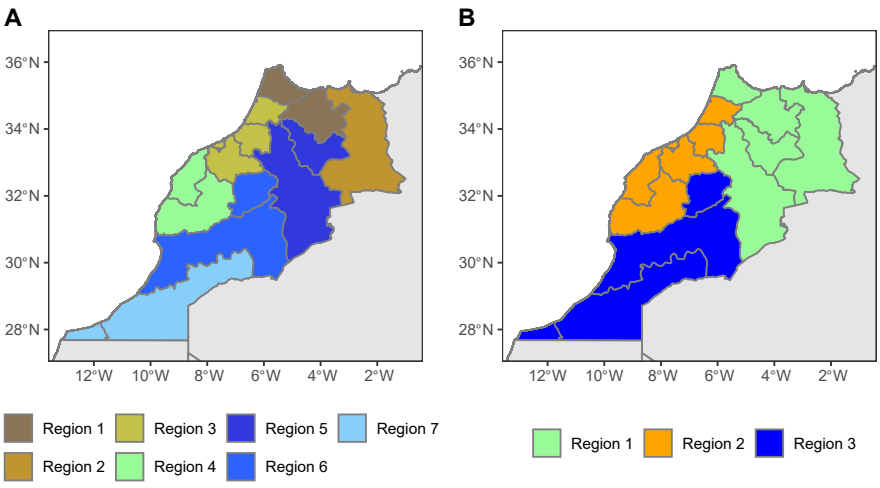


Figure D24: Partition of Morocco into three and seven arbitrary locations. (**A**) Partition of Morocco into seven arbitrary locations. The major results corresponding to this framework are presented in the main text. All epidemics started by the introduction of a single index case in Region 2. Regions 3 and 4 correspond to over-sampled locations. (**B**) Partition of Morocco into three arbitrary locations. The main results corresponding to this framework are presented exclusively in the Supplementary Materials. All epidemics started by the introduction of a single index case in Region 1. Region 2 corresponds to the over-sampled location. The gray outlines delineate the official regions downloaded from GADM and the colors indicate the arbitrary locations.

6. Bayesian inference

6.1. Simulation study

Table D3: Prior distributions used in the simulation study for the CTMC, BASTA, MASCOT, and MASCOT-GLM.

	Parameter	CTMC[a]	BASTA	MASCOT	MASCOT-GLM
HKY substitution model	**Transition-transversion ratio**	Lognormal(1, 1.25)	Lognormal(1, 1.25)	Lognormal(1, 1.25)	Lognormal(1, 1.25)
	Base frequencies	Dirichlet(α=1, sum=1)	Uniform([0:1])	Uniform([0:1])	Uniform([0:1])
Molecular clock	**Clock rate**	CTCM Rate reference (Ferreira and Suchard, 2008)	Lognormal(0, 4)	Lognormal(0, 4)	Lognormal(0, 4)
Spatial model	**Migration rate**	Exponential(1)	Exponential(1)	Exponential(1)	Exponential(1)
	Migration clock	CTMC Rate reference[7]	-	Set to 1	Exponential(1)
	Coefficient of migration predictor (GLM)	-	-	-	Normal(0,1)
	Root location frequency (CTMC)	Uniform([0:1])	-	-	-
	Deme population size (constant over time)	1/X[a]	Exponential(1)	Exponential(1)	-
	Deme size clock	-	-	-	Exponential(1)
	Coefficient of deme size predictor (CTMC)	-	-	-	Normal(0,1)
	Equal deme population sizes	-	Yes	Yes	-
BSSVS	**Sum of non-zero migration rates**	Poisson(n_{demes}-1)	Poisson(n_{demes}-1)	Poisson(n_{demes}-1)	-
	Sum of included predictors on migration rates	-	-	-	Poisson(1)
	Sum of included predictors on deme sizes	-	-	-	Poisson(1)

Abbreviations: BSSVS, Bayesian stochastic search variable selection; GLM, Generalized linear mode; HKY model, Hasegawa, Kishino, and Yano model.

[a] We used the default priors from BEAUTi v1.10.4.

Table D4: Indicative number of iterations per hour for the four discrete phylogeographic approaches according to the number of demes and genomes. The number of iterations is expressed in millions.

No. of demes	No. of genomes	CTMC	BASTA	MASCOT	MASCOT-GLM
3	**150**	14.85	0.67	5.23	2.4
	500	2.22	0.21	0.93	0.69
7	**150**	14.08	0.70	2.4	1.3
	500	1.85	0.12	0.32	0.23

Table D5: Number of chains excluded per algorithm for the main analysis on seven demes.

No. of genomes	CTMC	BASTA	MASCOT	MASCOT-GLM
150	0	17	6	0
500	0	NA	161	2

Abbreviations: NA, Not applicable.

Table D6: Number of chains excluded per algorithm for the supplementary analysis on three demes.

No. of genomes	CTMC	BASTA	MASCOT	MASCOT-GLM
150	0	0	0	0
500	0	NA	2	0

Abbreviations: NA, Not applicable.

6.2. Analysis of the SARS-CoV-2 data set

Table D7: List of priors used for each discrete phylogeographic approach on the RABV data set.

	Parameter	CTMC	BASTA	MASCOT
HKY substitution model	**Transition/transversion ratio**	Lognormal(1, 1.25)	Lognormal(1, 1.25)	Lognormal(1, 1.25)
	Base frequencies	Empirical	Empirical	Empirical
	Shape of the gamma rate of heterogeneity	Exponential(0.5)	Exponential(0.5)	Exponential(0.5)
Lognormal relaxed molecular clock	**Mean**	CTCM Rate reference	Lognormal(0.001, 1000)	Lognormal(0.001, 1000)
	Standard deviation	Exponential(1/3)	Exponential(1/3)	Exponential(1/3)
Spatial model	**Migration rate**	Exponential(1)	Exponential(1)	Exponential(1)
	Migration clock	CTMC Rate reference	-	-
	Region frequency (CTMC)	Uniform([0:1])	-	-
	Deme population size (constant over time)	Gamma(shape=0.001, scale=1000)	Exponential(1)	Exponential(1)
	Regions root frequencies	Uniform([0:1])	-	-
	Equal deme population sizes	-	Yes	Yes
BSSVS	**Sum of non-zero migration rates**	Poisson(5)	Poisson(5)	Poisson(5)

Abbreviations: BSSVS, Bayesian stochastic search variable selection; HKY model, Hasegawa, Kishino, and Yano model.

6.3. Analysis of the SARS-CoV-2 data set

Table D8: List of priors used for each discrete phylogeography algorithm on the SARS-CoV-2 data set.

	Parameter	CTMC	BASTA	MASCOT	MASCOT-WID MASCOT-WHO
HKY substitution model	**Transition/ transversion ratio**	Lognormal(1, 1.25)	Lognormal(1, 1.25)	Lognormal(1, 1.25)	Lognormal(1, 1.25)
	Base frequencies	Empirical	Empirical	Empirical	Empirical
	Proportion of invariant	Uniform([0,1])	Uniform([0,1])	Uniform([0,1])	Uniform([0,1])
	Shape of the gamma rate of heterogeneity	Exponential(0.5)	Exponential(0.5)	Exponential(0.5)	Exponential(0.5)
Strict molecular clock	**Evolutionary rate**	CTCM Rate reference	Lognormal(0, 4)	Lognormal(0,4)	Lognormal(0,4)
Exponential growth coalescent model	**Deme size**	Gamma(shape= 0.001, scale=1000)	-	-	-
	Exponential growth rate	Laplace(mean=0, scale=1)	-	-	-
Spatial model	**Migration rate**	Exponential(1)	Exponential(1)	Exponential(1)	-
	Migration clock	CTMC Rate reference	-	-	-
	Region frequency (CTMC)	Uniform([0:1])	-	-	-
	Regions root frequencies	Uniform([0:1])	-	-	-
	Deme size	-	Exponential(1)	Exponential(1)	-
	Equal deme population sizes	-	Yes	Yes	-
GLM spatial model	**Migration clock**	-	-	-	Exponential(1)
	Migration predictors scaler	-	-	-	Normal(0,1)
	Deme size clock	-	-	-	Exponential(1)
	Deme size predictors scaler	-	-	-	Normal(0,1)
BSSVS	**Sum of non-zero migration rates**	Poisson(5)	Poisson(5)	Poisson(5)	-
	Sum of non-zero deme size predictors	-	-	-	Poisson(1)
	Sum of non-zero migration rate predictors	-	-	-	Poisson(1)

Abbreviations: BSSVS, Bayesian stochastic search variable selection; GLM, Generalized linear model; HKY model, Hasegawa, Kishino, and Yano model.

Appendix E

Supplementary information on contact heterogeneity in households

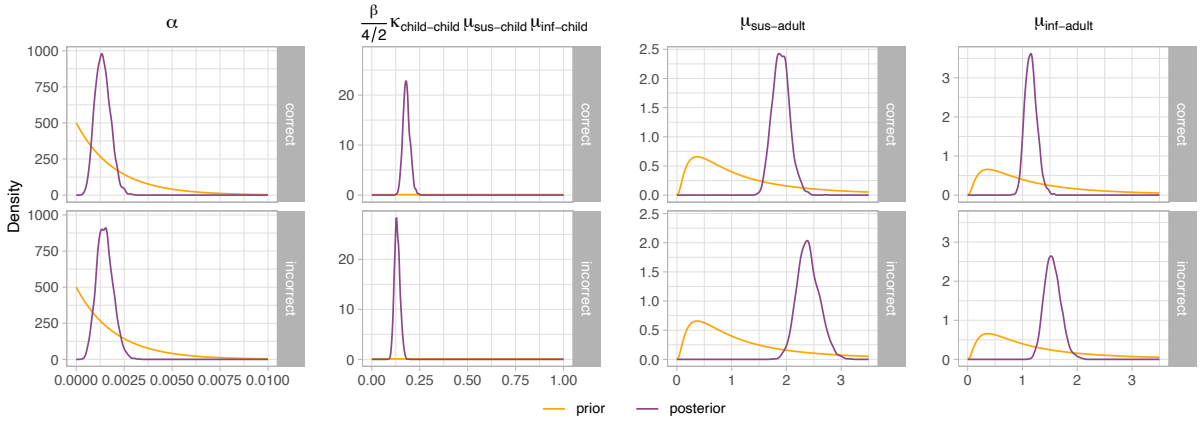


Figure E1: Example of prior and posterior distributions of model parameters for the SARS-CoV-2 infection scenario. Prior and posterior distributions of model parameters for the SARS-CoV-2 infection scenario. The examples depicted here correspond to one simulation inferred by the model that accounts for heterogeneous contact patterns ("correct"), and the model that assumes homogeneous mixing within households ("incorrect"). From left to right, the panels depict the hazard of infection in the community α, the infection rate in an heterosexual couple $\frac{\beta}{4/2}\ \kappa_{child,child}\ \mu_{sus,child}\ \mu_{inf,child}$, the relative susceptibility of adults $\mu_{sus,adult}$, and the relative infectivity of adults $\mu_{inf,adult}$.

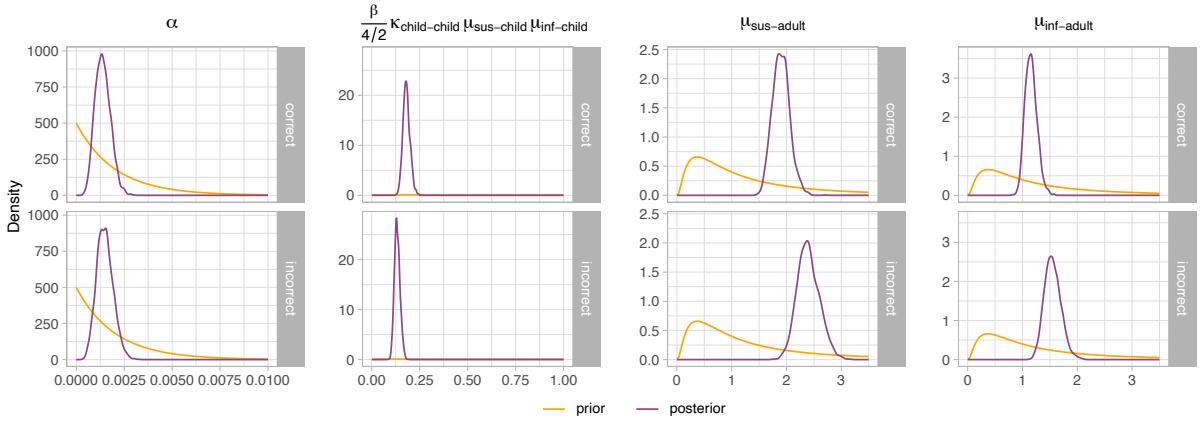


Figure E2: Example of prior and posterior distributions of model parameters for the influenza virus infection scenario. Prior and posterior distributions of model parameters for the influenza virus infection scenario. The examples depicted here correspond to one simulation inferred by the model that accounts for heterogeneous contact patterns ("correct"), and the model that assumes homogeneous mixing within households ("incorrect"). From left to right, the panels depict the hazard of infection in the community α, the infection rate in an heterosexual couple $\frac{\beta}{4/2}\ \kappa_{child,child}\ \mu_{sus,child}\ \mu_{inf,child}$, the relative susceptibility of adults $\mu_{sus,adult}$, and the relative infectivity of adults $\mu_{inf,adult}$.

Appendix F

Résumé étendu en français

1. Introduction

Les épidémies de maladies infectieuses correspondent à l'apparition et à la propagation rapide de maladies contagieuses dans une population. Elles sont causées par des pathogènes, des organismes qui parasitent un individu hôte pour se multiplier et se transmettre à d'autres individus. Il existe une très grande diversité de pathogènes qui appartiennent à des groupes taxonomiques aussi divers que les virus, les bactéries, les protozoaires, les arthropodes ou encore les helminthes.

Le processus de transmission dont les étapes principales d'infection, multiplication et propagation sont communes à tous les pathogènes dépend de plusieurs facteurs. Tout d'abord, la composition moléculaire des pathogènes détermine leur spectre d'hôtes. Ensuite, les dynamiques de transmission dans les populations hôtes dépendent des mécanismes d'évasion immunitaire déployés par les pathogènes comme l'évolution rapide de leurs antigènes ou l'expression alternative de gènes. D'autres facteurs liés à l'hôte interviennent aussi dans le processus de transmission, à savoir l'ensemble des facteurs génétiques qui contribuent au bon fonctionnement du système immunitaire au niveau individuel, l'immunité de groupe au niveau populationnel acquise suite à des épidémies passées ou des campagnes de vaccination, ainsi que la structure sociale et les dynamiques de mobilité de la population hôte. Enfin, des facteurs environnementaux peuvent jouer directement sur la stabilité des pathogènes dans l'environnement ou déterminer leur occurence spatiale et temporelle de manière indirecte en contrôlant les zones habitables par la population hôte. Le processus épidémique est donc multifactoriel mais aussi stochastique, c'est-à-dire que son succès est soumis à une part d'aléatoire. À cette complexité s'ajoute le caractère potentiellement explosif du processus épidémique en fonction du potentiel de transmission du pathogène et du niveau d'immunité pré-existante dans la population hôte. Cette complexité ne peut pas être appréhendée facilement et nécessite l'utilisation de modèles mathématiques qui ne sont autres que des représentations simplifiées de la réalité formalisées sous la forme d'équations.

Une grande variété de modèles a été développée pour décrire le processus épidémique. Nous distinguerons deux approches qui modélisent le processus de transmission sous deux angles différents et qui viennent de deux disciplines scientifiques différentes : la modélisation épidémiologique et la phylodynamique. La modélisation épidémiologique s'appuie sur des modèles mécanistiques de la dynamique de transmission inter-hôte. Les plus connus sont les modèles compartimentaux qui se déclinent sous une forme déterministe et une forme stochastique. Les modèles individus-centrés sont plus complexes mais permettent d'intégrer de multiples couches d'hétérogénéité entre les hôtes. De manière générale, la modélisation épidémiologique permet de prendre en compte des facteurs liés à l'hôte et à

l'environnement alors que l'approche phylodynamique modélise le processus de transmission du point de vue du pathogène. Certains pathogènes évoluent si rapidement que les processus évolutifs de sélection, dérive génétique, effet fondateur et migration se déroulent sur la même échelle de temps que le processus épidémique. La phylodynamique tire profit de cette concomitance et modélise l'émergence de nouveaux variants, l'évolution de la taille de la population de pathogènes, voire même la migration spatiale des lignées de pathogènes à partir de la phylogénie des pathogènes.

Bien que le développement de modèles théoriques permette de mieux comprendre le processus épidémique, l'estimation de paramètres clés de la transmission dans un contexte spécifique ne peut se faire que par leur ajustement à des données empiriques. Ces données ne reflètent que très rarement l'intégralité du processus de transmission car, d'une part, les chaînes de transmission exactes sont difficilement observables et, d'autre part, seule une partie des cas sont identifiés. Par ailleurs, plusieurs types de données peuvent être recueillies, des données sur le cas qui sont détaillées au niveau individuel ou agrégées au niveau populationnel, et les séquences génétiques des pathogènes, qui ne permettent pas d'accéder aux mêmes facettes du processus épidémique. En effet, les données individuelles permettent d'évaluer l'hétérogénéité entre les cas et les facteurs d'exposition, les données populationnelles permettent d'appréhender les dynamiques de transmission complexes tout comme les séquences génétiques.

L'ajustement des modèles à des données empiriques fait intervenir des méthodes statistiques qui permettent d'estimer les paramètres clés du processus de transmission. Les statistiques Bayésiennes sont particulièrement adaptées aux données épidémiques car, en plus de leur cadre conceptuel qui permet d'intégrer explicitement des connaissances à priori dans l'inférence, elles permettent de prendre en compte les incertitudes liées aux données manquantes et aux variables non observées et non observables comme la date d'infection des cas. Dans le cas spécifique de la phylodynamique, l'inférence Bayésienne permet d'estimer conjointement la phylogénie du pathogène, l'émergence de nouveaux variants, l'évolution de la taille de sa population et sa dynamique spatiale.

La modélisation épidémiologique et la phylodynamique sont des outils quantitatifs qui permettent de mieux comprendre les épidémies et d'informer les acteurs de la santé publique. Ces deux approches permettent d'identifier les déterminants de l'hôte impliqués dans le processus de transmission et de quantifier l'hétérogénéité entre les cas. Par exemple, la modélisation de la transmission dans les ménages a révélé que les enfants sont deux fois plus susceptibles que les adultes et la phylodynamique a permis de montrer qu'un événement de super-transmission au Massachusetts a été à l'origine de l'introduction d'une nouvelle lignée de coronavirus 2 du syndrome respiratoire aigu sévère (SARS-CoV-2) dans l'état américain. Les deux approches permettent de quantifier les facteurs de risque également à l'échelle de la population. La phylodynamique permet d'associer la dynamique de transmission à des covariables environnementales et à la structure spatiale et d'espèce des populations hôtes. La modélisation épidémiologique peut intégrer des données sur la structure spatiale et sociale de la population pour améliorer le réalisme des modèles. Elle permet également de traquer les dynamiques des variants de pathogènes. La modélisation en épidémiologie est également un outil considérable pour orienter les décisions de santé publique. L'analyse des données de surveillance permet de détecter rapidement et en temps réel une reprise épidémique ou son contrôle par modélisation épidémiologique. La modélisation épidémiologique permet aussi d'évaluer à postériori ou de manière prédictive l'impact des mesures de contrôle. La phylodynamique est encore peu utilisée en évaluation des interventions et nécessiterait le développe-

ment de nouvelles méthodes.

Ces exemples de l'utilisation de la modélisation épidémiologique et de la phylodynamique soulignent leur complémentarité. La modélisation épidémiologique est tout particulièrement adaptée à l'identification des facteurs individuels et environnementaux qui influent sur le processus épidémique, mais aussi à la prédiction à court et long terme et enfin à la conception et l'évaluation de mesures de contrôle. La phylodynamique quant à elle permet d'intégrer les déterminant liés aux pathogènes et, en confirmant des chaînes de transmission, apporte un éclairage sur la contribution de la transmission locale comparée aux importations. Il est intéressant de noter que la pandémie de la maladie à coronavirus 2019 (COVID-19) a été marquée par une certaine démocratisation des études interdisciplinaires qui combinent des données épidémiologiques individuelles et populationnelles à des séquences génétiques, le tout analysé par des techniques diverses de statistiques classiques, modélisation épidémiologique et/ou phylodynamique.

De nombreux défis subsistent concernant l'utilisation de la modélisation épidemiologique et de la phylodynamique. Ces défis sont liés à la prise en compte des biais d'échantillonnage dans le processus d'inférence et à la spécification des modèles. Un échantillon est biaisé lorsque sa composition ne reflète pas la dynamique d'une épidémie. Les biais d'échantillonnage sont particulièrement courant en phylodynamique car l'échantillonnage des séquences de pathogène se fait surtout de manière opportuniste et, quand il est planifié, des disparités spatiales subsistent. Les modèles de phylogéographie sont connus pour être sensibles à ces biais mais l'impact de ces biais et les moyens d'en atténuer les effets restent peu compris. Dans un tout autre registre, l'épidémiologie de la transmission des maladies respiratoires a largement bénéficié des études de leur transmission dans les ménages. Toutefois, les modèles de transmission dans les ménages font l'hypothèse que l'ensemble des membres du ménage entrent en contact à la même fréquence. Plusieurs études ont montré que cette hypothèse n'est pas vérifiée ce qui pourrait biaiser les estimations de susceptibilité et d'infectivité relative des membres du ménage en fonction de leur âge, sexe etc...

2. Objectifs de la thèse

Cette thèse a pour objectif principal d'explorer les contributions des données épidémiologiques individuelles dans les ménages et des séquences génétiques des pathogènes qui se trouvent aux deux extrémités du spectre de granularité des données dans la compréhension de la transmission des maladies infectieuses dans les populations. Ces données sont analysées par modélisation épidémiologique et phylodynamique respectivement. Le deuxième objectif de cette thèse est d'étudier les limites de ces deux approches, plus précisément, l'impact des biais d'échantillonnage sur les inférences de phylogéographie discrète et l'impact de contact hétérogènes dans les ménages sur l'estimation des paramètres de susceptibilité et d'infectivité relative des enfants par rapport aux adultes. Ces objectifs sont traités par le prisme de deux cas d'étude, le virus de la rage (RABV) et le SARS-CoV-2. La rage est une zoonose (maladie qui se transmet naturellement de l'animal à l'homme) tropicale négligée dont le réservoir principal sont les chiens domestiques, responsables de 99% des cas de rage humaine. Un vaccin contre la rage est disponible chez l'homme et l'animal mais elle reste endémique en Afrique et en Asie ce qui est dû en partie au manque de moyens financiers et d'implication politique. Par ailleurs, de nombreuses zones d'ombre subsistent concernant les dynamiques de transmission de la rage chez le chien. À l'opposé, les moyens qui ont été déployés pendant la pandémie de COVID-19 ont été sans précédent et la production

scientifique a été particulièrement prolifique. Cette mobilisation s'explique par l'impact considérable de la pandémie sur les systèmes de santé dans le monde et l'émergence récurrente de nouveaux variants mettant au défi l'efficacité des vaccins et des interventions non pharmaceutiques.

Dans l'Étude n°1, j'explore les contributions relatives de la modélisation épidémiologique et de la phylodynamique dans la compréhension de la transmission de la rage chez le chien et dans l'évaluation des stratégies de contrôle. Puis, je mets en pratique des approches de phylodynamique, plus précisément de phylogéographie, dans l'Étude n°2 afin de mieux caractériser la signature endémique de la rage chez le chien au Cambodge. Les limites de ces approches face au biais d'échantillonnage sont approfondies dans l'Étude n°3 par une étude de simulation. Dans l'Étude n°4, je mets également en pratique la modélisation épidémiologique mais cette fois-ci dans le contexte du SARS-CoV-2 afin d'estimer l'impact de la vaccination et des mesures de distanciation physique sur la transmission du virus dans les ménages. Enfin, j'étudie dans l'Étude n°5 comment la mauvaise spécification des contacts dans les ménages modifie l'estimation de la susceptibilité et de l'infectivité relative des enfants par rapport aux adultes.

3. Étude n°1: Revue exploratoire des études quantitatives de la transmission de la rage chez le chien

3.1. Contexte

La rage est une maladie fatale mais qui peut être prévenue par un vaccin efficace à 100% chez l'homme et chez l'animal. La rage est endémique en Afrique et en Asie dont la circulation est maintenue essentiellement par les chiens domestiques. Malgré de nombreux efforts, la situation épidémiologique des pays d'Afrique et d'Asie ne s'est pas améliorée ce qui est en partie dû à une connaissance partielle des dynamiques de la rage chez le chien. Dans cette étude, nous explorons les contributions des études quantitatives, plus précisément de la modélisation épidémiologique et de la phylodynamique dans la compréhension des dynamiques de la rage et dans l'évaluation des mesures de contrôle tout en soulignant les limites de ces méthodes et en identifiant les questions en suspens.

3.2. Méthodes

Nous avons réalisé une revue de la littérature de l'ensemble des études qui ont appliqué un modèle épidémiologique ou un modèle de phylodynamique à la transmission de la rage chez le chien. Nous avons utilisé une méthodologie rigoureuse de revue de la littérature scientifique, empruntée aux revues systématiques mais adaptée à l'étude d'articles hétérogènes dans un but d'exploration et de synthèse d'un domaine scientifique. Nous avons identifié sur les bases de données bibliographiques PubMed, Web of Science et Scopus un total de n=59 études dont n=30 études de modélisation épidémiologique, n=22 études de phylodynamique et n=7 études interdisciplinaires. Nous avons extrait systématiquement les informations suivantes : (*i*) l'approche de modélisation principale et ses hypothèses sous-jacentes, (*ii*) la source des données, (*iii*) les remarques sur les biais potentiels dans les données en lien avec les processus épidémiologiques et évolutifs, (*iv*) les résultats qualitatifs et quantitatives sur les dynamiques de la rage et (*v*) si réalisée, les résultats de l'étude de sensibilité.

3.3. Résultats

Les études de modélisation épidémiologique se concentrent en priorité sur l'évaluation des mesures de contrôle alors que l'identification des déterminants de la transmission est l'objectif premier des études de phylodynamique. La modélisation épidémiologique a permis de montrer que la vaccination est la stratégie de contrôle la plus efficace et doit atteindre une couverture vaccinale élevée dans les régions indemnes mais à haut risque d'introduction, alors qu'une couverture intermédiaire est suffisante dans les régions endémiques. La phylodynamique et les études interdisciplinaires ont mis en évidence le rôle central de l'hétérogénéité spatiale dans le maintien de la circulation de la rage. L'ensemble des méthodes utilisées sont limitées par la quantité des données et les biais d'échantillonnage sont généralement inconnus. De manière intéressante, les études interdisciplinaires donnent un cadre qui permet de formuler et tester des hypothèses sur les processus de transmission.

3.4. Limites de l'étude

L'identification des articles scientifiques, bien que se basant sur une procédure codifiée, est sensible à la combinaison de mots-clés utilisée. Puisque les résultats obtenus dans cette revue sont en accord avec les résultats de revues précédentes, nous pensons que nous n'avons pas manqué d'étude quantitative majeure sur la transmission de la rage.

3.5. Conclusion

La rage se transmet à bas bruit chez le chien mais est maintenue grâce à une forte hétérogénéité spatiale et sa dynamique est souvent impactée par l'homme. La modélisation épidémiologique a montré que la vaccination est la stratégie de contrôle la plus efficace mais il n'y a pas de consensus clair sur la fréquence des campagnes de vaccination. De nombreux défis restent à relever. En effet, le rôle des populations animales sauvages, de la structure sociale de la population de chiens domestiques et de leurs interactions avec l'homme restent peu caractérisés. Des études de terrain sont nécessaires pour recueillir ce type de données et informer les modèles épidémiologiques et phylodynamiques. Par ailleurs, la situation épidémiologique est inconnue dans la majorité des pays endémiques d'Asie et d'Afrique.

4. Étude n°2: Caractérisation de la circulation endémique du virus de la rage par phylodynamique

4.1. Contexte

La phylodynamique est un outil puissant en épidémiologie notamment pour étudier les maladies négligées ou les zoonoses pour lesquelles le recueil de séquences génétiques du pathogène est moins coûteux que le recueil de données sur les cas. De plus, la phylogéographie est un outil de choix pour étudier les dynamiques de transmission entre zones géographiques, et donc l'endémicité d'une maladie dans le sens d'une circulation intense dans une région non maintenue par des introductions. Le Cambodge est l'un des pays les plus touchés par la rage canine dans le monde mais sa circulation au sein du Cambodge et en Asie du Sud-Est reste méconnue. Dans cette étude, nous appliquons des approches de phylogéographie afin de caractériser l'endémicité de la rage au Cambodge.

4.2. Méthodes

Afin de mieux comprendre la dynamique de transmission de la rage en Asie du Sud-Est, 199 génomes de RABV isolés chez des chiens ont été séquencés et analysés conjointement à tous les gènes N de RABV d'Asie du Sud-Est disponibles sur GenBank. Les dynamiques spatiales entre deux localisations, le Cambodge d'une part et les autres pays d'Asie du Sud-Est d'autre part, sont reconstruites par phylogéographie discrète. Les dynamiques spatiales au sein du Cambodge sont quant à elles décrites par une analyse en phylogéographie continue des génomes et des gènes N isolés au Cambodge. Ces dynamiques sont comparées aux estimations par phylogéographie continue des dynamiques d'autres pays. Enfin, les facteurs du paysage associés à la vélocité des lignées virales ou à la localisation des lignées virales sont identifiés par phylodynamique du paysage.

4.3. Résultats

Les lignées circulant au Cambodge appartiennent essentiellement au même clade qui circule majoritairement au Cambodge et plus minoritairement au Vietnam. La transmission au sein du Cambodge n'est donc probablement pas maintenue par des introductions depuis les pays voisins ce qui corrobore l'hypothèse d'endémicité stricte de la rage au Cambodge. Au sein du pays, la région de Phnom Penh est la zone de transmission la plus active. Même les ancêtres directs des échantillons prélevés aux frontières du Cambodge sont prédits dans la région de Phnom Penh ce qui signifie qu'il y a des événements de transmission à longue distance. Ces événements à longue distance peuvent être plus rapides que les événements de transmission à courte distance car médiés par l'homme ; c'est pourquoi les dynamiques spatiales entre pays doivent être comparées à la même échelle spatiale. La comparaison avec d'autres pays très affectés par la rage montre que la transmission à petite échelle est similaire donc probablement médiée par le mouvement des chiens. Enfin, aucun facteur du paysage n'est associé avec la vélocité des lignées virales ce qui est en faveur de l'hypothèse d'endémicité stricte.

4.4. Limites de l'étude

La principale limite de cette étude concerne les biais d'échantillonnage. En effet, l'effort de séquençage déployé au Cambodge n'a pas d'équivalent dans les autres pays d'Asie du Sud-Est et la couverture spatiotemporelle n'est très certainement pas représentative des dynamiques de la rage dans toute l'Asie du Sud-Est y compris le Cambodge. Ces biais conduiraient à la sous-estimation des mouvements entre le Cambodge et les autres pays d'Asie du Sud-Est. Par ailleurs, le Nord-Est du Cambodge n'est quasiment pas échantillonné, une région où les ethnies ont une relation différente au chien et circulent beaucoup au Laos et au Vietnam. Comme ces ethnies ont peu de relation avec le reste du Cambodge, il est probable que les lignées virales qui y circulent aient une dynamique différente ce qui ne change pas nos conclusions sur le reste du Cambodge.

4.5. Conclusion

L'ensemble des techniques de phylogéographie nous a permis de mieux caractériser la transmission de la rage en Asie du Sud-Est et plus spécifiquement au Cambodge. Elle est marquée par peu de migrations avec ses voisins, évoluant quasiment en vase clos. Toutefois, les réimportations ne sont pas improbables

donc la stratégie d'élimination à long-terme ne doit pas se concentrer uniquement sur l'élimination locale mais investir dans un système de surveillance pour la détection rapide de nouvelles chaînes de transmission. Par ailleurs, la situation épidémiologique générale de la rage au Cambodge n'est que partiellement connue (incidence chez l'homme et l'animal). Il en va de même des facteurs contrôlant la transmission de la rage (rôle de l'homme dans le mouvement des chiens).

5. Étude n°3: Les biais d'échantillonnage en phylogéographie discrète

5.1. Contexte

L'échantillonnage des séquences génétiques pour une analyse phylodynamique se fait généralement de manière opportuniste et donc il n'est probablement pas représentatif du processus de transmission sous-jacent. L'impact de ces biais sur les estimations des modèles de phylogéographie a été discuté très tôt, notamment en ce qui concerne le modèle de chaîne de Markov en temps continu (CTMC) en phylogéographie discrète. Le modèle de coalescent structuré et ses approximations (MASCOT et BASTA) ont été proposés comme alternatives au CTMC car plus robustes face aux biais d'échantillonnage. Toutefois, les performances des trois algorithmes (CTMC, MASCOT, BASTA) n'ont jamais été comparées sur des données de séquences génétiques plus ou moins biaisées.

5.2. Méthodes

Pour mieux comprendre l'impact des biais sur les reconstructions phylogéographiques discrètes, nous avons simulé des épidémies de rage chez le chien au Maroc avec un modèle de métapopulation et nous avons échantillonné de manière plus ou moins biaisé spatialement les séquences génétiques associées aux cas de chiens. Nous avons ensuite comparé les dynamiques estimées par les trois algorithmes de phylogéographie discrète (CTMC, BASTA, MASCOT) aux dynamiques simulées. Plusieurs stratégies d'atténuation de l'impact des biais ont été testées : augmenter le nombre de séquences utilisées, maximiser la couverture spatiale ou spatiotemporelle de l'échantillon, intégrer des données sur le nombre de cas dans MASCOT (MASCOT-GLM). Enfin, tous les algorithmes (CTMC, BASTA, MASCOT, MASCOT-GLM) ont été testés sur un jeu de données empiriques de la rage aux Philippines et un autre sur le SARS-CoV-2 dans les premiers mois de la pandémie.

5.3. Résultats

Les performances du modèle de CTMC surpassent celles de BASTA et MASCOT en l'absence de biais, mais également avec un biais croissant, même si les performances du modèle de CTMC diminuent quand le biais augmente. Par ailleurs, les performances de BASTA et MASCOT changent peu avec le niveau de biais. La maximisation de la couverture spatiale ou spatiotemporelle améliore les prédictions de CTMC et, dans une moindre mesure, de BASTA et MASCOT. Il en va de même quand le nombre de séquences analysées augmente. L'ajout du nombre de cas dans le modèle de coalescent structuré (MASCOT-GLM) permet de totalement contrecarrer les biais d'échantillonnage et surpasse le modèle de CTMC. L'analyse

des jeux de données empiriques montre que lorsqu'une seule séquence génétique est disponible pour une des zones géographiques étudiées, BASTA et MASCOT reconstruisent une histoire de transmission spatiale très peu parsimonieuse. Ce problème est évité pour les données empiriques du SARS-CoV-2 en ajoutant le nombre de cas d'infection par zone géographique.

5.4. Limites de l'étude

Cette étude se focalise sur des biais d'échantillonnage importants qui peuvent conduire dans les échantillons analysés à l'absence de séquences des régions les moins affectées par la rage. Notre étude cumule donc l'impact des biais dus à la non-représentativité spatiale des échantillons et dus à l'absence de représentants de certaines régions. L'ensemble des analyses a été répliqué dans un contexte où la transmission a lieu entre trois régions et donc où la majorité des échantillons a des représentants des trois régions. Les résultats obtenus sont similaires. Enfin, l'analyse des données empiriques ne nous permet pas de conclure sur l'algorithme qui reconstruit les dynamiques de manière la plus fidèle.

5.5. Conclusion

Les biais d'échantillonnage en phylogéographie sont omniprésents et concernent autant les maladies humaines qu'animales. Toutefois, leur impact peut être atténué ou pris en compte à différents niveaux : au moment du recueil des données, du choix du modèle d'inférence, et de l'ajout de données épidémiologiques au modèle d'inférence. Nous conseillons aux chercheurs qui souhaitent réaliser une analyse de phylogéographie discrète d'utiliser une stratégie d'échantillonnage qui maximise la couverture spatiale et de comparer les reconstructions de plusieurs modèles sur des sous-échantillons différents lorsque cela est possible.

6. Étude n°4: Évaluation de l'effectivité vaccinale sur la transmission du SARS-CoV-2 dans les ménages

6.1. Contexte

Le recueil de données épidémiologiques individuelles dans les ménages permet d'appréhender la transmission des maladies respiratoires telles que la COVID-19 et de mesurer l'impact de la vaccination. En décembre 2020, alors qu'un vaccin était disponible et que les campagnes de vaccination en cours de planification partout dans le monde, un nouveau variant du SARS-CoV-2 a émergé, le variant Alpha. Il a fallu évaluer rapidement l'efficacité et l'effectivité vaccinale contre les formes sévères, l'infection et la transmission. Au moment de l'étude, le niveau de preuve concernant l'impact du vaccin contre la transmission si infecté était faible. Nous avons donc évalué l'effectivité du vaccin contre l'infection et la transmission, mais aussi l'impact des mesures de distanciation physique dans un contexte de transmission intra-ménagère.

6.2. Méthodes

Un total de 210 ménages israéliens a été identifié et suivi par le Sheba Medical Center à Tel Aviv entre décembre 2020 et avril 2021. Tous les membres du ménage étaient incités à réaliser des tests PCR pendant les dix jours suivants l'inclusion et des données sur les symptômes, les résultats des tests, le statut vaccinal et le comportement des individus ont été recueillies. Nous avons analysé ces données en utilisant un modèle de transmission du SARS-CoV-2 avec augmentation de données. L'effectivité vaccinale contre l'infection et la transmission, l'effectivité de la distanciation physique contre l'infection et la susceptibilité relative des enfants par rapport aux adolescents/adultes ont été estimées par inférence Bayésienne.

6.3. Résultats

Le taux d'attaque secondaire atteint environ 76% dans les ménages israéliens durant la vague Alpha. Une valeur particulièrement élevée qui serait due à la plus grande transmissibilité du variant Alpha par rapport au variant historique. L'effectivité vaccinale contre l'infection après deux doses de vaccin chez les adolescents/adultes est d'environ 79% et 75% contre la transmission. Toutefois, l'intervalle de crédibilité pour l'effectivité vaccinale contre la transmission s'étend de 23 à 94%. La distanciation physique réduit le risque d'infection de 88% chez les adolescents/adultes non vaccinés et réduit encore le risque d'infection chez les vaccinés. Enfin, les enfants seraient 50% moins susceptibles que les adolescents/adultes.

6.4. Limites de l'étude

La principale limite de l'étude est liée au nombre restreint d'individus ce qui a pour conséquence l'incertitude élevée autour de l'estimation de l'effectivité vaccinale contre la transmission. Les instructions concernant la réalisation des tests PCR étaient différentes entre les membres vaccinés et les non-vaccinés. Toutefois, les estimations sont peu impactées lorsque l'analyse est réduite aux ménages où tous les individus ont réalisé au moins un voire deux tests PCR durant le suivi. Enfin, l'effet de la distanciation physique est potentiellement sujet à caution et l'évolution des pratiques n'a pas été intégrée.

6.5. Conclusion

Après deux doses de vaccin, le risque d'infection et de transmission si infecté est considérablement diminué. La distanciation physique vis-à-vis du premier cas détecté dans le ménage confère une protection importante autant chez les adolescents/adultes que les enfants. Même si les contacts ménagers sont vaccinés, ils bénéficieraient d'une meilleure protection contre l'infection en évitant de manger et dormir avec le cas index. Ce résultat nécessite confirmation.

7. Étude n°5: Contacts hétérogènes et transmission intra-ménagère

7.1. Contexte

Dans le cadre des maladies respiratoires telles que la grippe et la COVID-19, les études de ménages sont idéales pour identifier les facteurs individuels qui influencent le processus de transmission et pour quantifier leur impact. Les modèles mathématiques de transmission intra-ménagère qui sont utilisés pour évaluer le rôle des enfants font l'hypothèse que l'ensemble des membres des ménages, parents et enfants, ont le même niveau de contact. Or, cette hypothèse n'est pas vérifiée par les données empiriques. Ici, nous explorons dans une étude de simulation comment les contacts hétérogènes dans les ménages peuvent conduire à une estimation biaisée de la susceptibilité et l'infectivité des enfants par rapport aux adultes.

7.2. Méthodes

Nous avons simulé des épidémies dans une base de données synthétique de 1,000 ménages dont la composition est dérivée d'une étude multicentrique européenne sur la transmission du variant historique du SARS-CoV-2 dans les ménages. Les taux de contact entre les membres des ménages sont tirés d'une étude belge qui montre que le couple mère-père a le plus de contacts, suivi du couple mère-enfant, puis enfant-enfant, et enfant le couple père-enfant. Nous avons testé deux scenarii : un scenario type COVID-19 où les enfants sont deux fois moins susceptibles et 20% moins infectieux que les adultes, et un scenario type grippe où les enfants sont deux fois plus susceptibles que les adultes. La susceptibilité et l'infectivité relatives des enfants par rapport aux adultes sont estimées par inférence Bayésienne et augmentation de données, soit en faisant l'hypothèse d'homogénéité des contacts, soit en précisant les niveaux réels d'hétérogénéité de contacts.

7.3. Résultats

Le modèle d'inférence qui prend en compte l'hétérogénéité de contacts entre les membres d'un même ménage estime correctement la susceptibilité relative et l'infectivité relative des enfants dans le scenario COVID-19 avec un biais relatif inférieur à 5%. Lorsque les contacts hétérogènes ne sont pas pris en compte, les deux paramètres sont estimés de 20% environ. Nous obtenu les mêmes résultats concernant la susceptibilité relative des enfants dans le scenario grippe. Cette sous-estimation compense le niveau de contact plus faible que les enfants ont avec les autres membres du ménage par rapport au couple mère-père.

7.4. Limites de l'étude

Le nombre de scenarii testés dans cette étude est relativement limité car nous n'avons exploré que deux scenarios de susceptibilité et d'infectivité relative des enfants. Par ailleurs, nous avons choisi la valeur des paramètres dans les simulations de manière à obtenir un taux d'attaque secondaire de 37% ce qui correspond à la fourchette haute des taux d'attaque secondaires estimés dans les ménages pour les infections au SARS-CoV-2 et aux virus influenza. Pour ces deux types d'infection, les taux d'attaque secondaires observés varient de quelques pourcents à 45%.

7.5. Conclusion

La nature hétérogène des contacts dans les ménages peut biaiser l'estimation des paramètres clés de la transmission comme la susceptibilité relative des enfants ou leur infectivité relative. La complexité des contacts dans les ménages doit donc être prise en compte dans les études futures. Toutefois, de nombreuses zones d'ombres subsistent concernant le changement des réseaux de contacts à la suite de l'introduction d'un nouveau cas dans les ménages. Les futures études de ménages devraient inclure non seulement le recueil de données sur la maladie mais aussi le recueil de données sur les réseaux de contacts dans les ménages. De telles études qui soulèvent de nouveaux défis quant à leur design et aux méthodes d'analyse des données.

8. Principaux résultats

Dans cette thèse, je montre la complexité de l'épidémiologie de la rage chez le chien qui dépend de la région du monde concernée. Alors que dans l'Étude n°1 je montre que de nombreux pays sont marqués par la co-circulation de lignées de virus de la rage et par de nombreux événements de transmissions entre pays voisins, la situation du Cambodge présentée dans l'Étude n°2 est marquée par la circulation d'une seule lignée virale, sans échange notable avec les autres pays. L'écologie des chiens, notamment leur relation à l'homme, est un autre facteur majeur des dynamiques de transmission qui varie très certainement en fonction des cultures. Peu de données sur les populations de chiens sont actuellement disponibles or, la meilleure compréhension des dynamiques de la rage à travers le monde nécessite la collecte de données de terrain sur les populations de chiens. Par exemple, de telles données pourraient améliorer le réalisme des modèles épidémiologiques de transmission de la rage. Enfin, même si je rappelle dans l'Étude n°1 que la vaccination des chiens est la mesure de contrôle la plus efficace pour contrôler la circulation de la rage, les campagnes de vaccination dans leur forme actuelle ne sont pas suffisantes et pourraient bénéficier du développement de la vaccination orale pour atteindre une couverture vaccinale plus élevée.

L'épidémiologie de la COVID-19 est très différente de celle de la rage. Son potentiel de transmission étant plus élevé, sa dynamique de transmission est souvent explosive ce qui nécessite une réponse rapide et adaptée. Ces exigences contractent le temps de la recherche ce qui n'est pas le cas pour la rage. La contraction du temps de recherche pour la COVID-19 est accentuée par l'émergence récurrente de nouveaux variants potentiellement plus transmissibles comme le variant Alpha et de l'impact des introductions dont un petit nombre suffit à l'établissement de nouvelles chaînes de transmission. Un point commun entre la COVID-19 et la rage concerne la conception des campagnes de vaccination dont la fréquence, les populations ciblées et le niveau de couverture doivent être optimisés. La conception de telles campagnes se base sur des estimations de l'effectivité vaccinale contre l'infection et la transmission ce que j'ai entrepris dans l'Étude n°4. Malheureusement, ces estimations sont devenues rapidement obsolètes suite à l'émergence du variant Delta qui a remplacé le variant Alpha en Israël en juillet 2021. Par ailleurs, des études ultérieures ont montré que l'effectivité vaccinale contre la transmission disparaît avec le temps. Dans l'Étude n°4, j'ai aussi estimé une susceptibilité plus faible des enfants par rapport aux adultes ce qui est en accord avec l'ensemble de la littérature. Toutefois, plusieurs facteurs de confusion liés à l'identification des cas dans les études de la littérature pourraient concourir à ce résultat et, comme démontré dans l'Étude n°5, l'hétérogénéité de contact dans les ménages pourrait aussi jouer un

rôle. Les futures études de ménages devraient intégrer des données sur les contacts entre les membres des ménages. Enfin, j'estime dans l'Étude n°4 une forte protection conférée par les mesures de distanciation physique dans un contexte ménager, même après vaccination. Bien que ces résultats demandent confirmation, il est fort probable que la distanciation physique soit moins efficace avec les derniers variants du SARS-CoV-2, les variants Omicron qui sont bien plus transmissibles que le variant Alpha.

Les agendas de recherche et les moyens investis dans la recherche de la COVID-19 et de la rage sont très différents, les données individuelles étant largement disponibles pour la première (Étude n°4) alors que les séquences génétiques de pathogènes restent une source majeure d'information pour la deuxième (Études n°1 et 2). Je montre dans cette thèse que ces deux types de données, analysées par des méthodologies très différentes, sont complémentaires. L'analyse phylodynamique des séquences génétiques de pathogènes permet d'étudier des processus à des échelles spatiales larges (Études n°2 et 3), et d'intégrer des covariables environnementales (Étude n°2). D'autre part, l'analyse des données individuelles grâce à des modèles épidémiologiques est particulièrement adaptée à l'évaluation des mesures de contrôle et à l'étude de l'hétérogénéité de transmission au niveau individuel (Étude n°4). Par ailleurs, je montre dans l'Étude n°3 que l'utilisation de modèles épidémiologiques peut servir à étudier les limites des approches de phylogéographie en prenant en compte des processus complexes comme le biais d'échantillonnage. Cette approche relativement nouvelle ouvre la voie à l'analyse comparative des modèles de phylogéographie. En effet, une dernière contribution de cette thèse concerne l'investigation des limites des méthodes quantitatives en épidémiologie des maladies infectieuses. L'Étude n°3 met en exergue l'importance de tester et de discuter l'impact des biais d'échantillonnage en phylogéographie discrète afin d'augmenter ou de moduler le niveau de preuve concernant les dynamiques spatiales. Les résultats de l'Étude n°5 préconisent quant à eux le recueil des contacts en même temps que les données liées à la transmission dans une étude de ménages. De telles études intégrées demandent une réflexion concernant le mode et la fréquence de recueil des données de contact.

9. Perspectives

De nombreuses questions subsistent concernant l'épidémiologique de la rage et du SARS-CoV-2, ainsi que sur les limites des méthodes de phylodynamique et de modélisation épidémiologique. Comment l'écologie des chiens modifie-t-elle la dynamique de la rage ? Comment cette écologie varie-t-elle dans le monde ? Comment l'hétérogénéité de l'immunité contre le SARS-CoV-2 due aux infections par des variants successifs et des vaccinations multiples peut-elle être estimée et quel est son impact sur les dynamiques de la COVID-19 ? Comment les modèles de phylodynamiques peuvent-ils être adaptés pour tester l'impact de mesures de contrôle ? Comment combiner données épidémiologiques et données génétiques dans un unique cadre de modélisation pour bénéficier de leur complémentarité ?

References

Abdelrahman, Z., M. Li, and X. Wang (2020). "Comparative Review of SARS-CoV-2, SARS-CoV, MERS-CoV, and Influenza A Respiratory Viruses". In: *Frontiers in Immunology* 11.

Abdul Taib, N. A., J. Labadin, and P. Piau (2019). "Model Simulation for the Spread of Rabies in Sarawak, Malaysia". In: *International Journal on Advanced Science, Engineering and Information Technology* 9.5, p. 1739.

Acevedo, M. A., F. P. Dillemuth, A. J. Flick, M. J. Faldyn, and B. D. Elderd (2019). "Virulence-driven trade-offs in disease transmission: A meta-analysis". In: *Evolution* 73.4, pp. 636–647.

Ahmed, K., P. Phommachanh, P. Vorachith, T. Matsumoto, P. Lamaningao, D. Mori, M. Takaki, B. Douangngeun, B. Khambounheuang, and A. Nishizono (2015). "Molecular Epidemiology of Rabies Viruses Circulating in Two Rabies Endemic Provinces of Laos, 2011–2012: Regional Diversity in Southeast Asia". In: *PLoS Neglected Tropical Diseases* 9.3, pp. 2011–2012.

"Aiming for elimination of dog-mediated human rabies cases by 2030" (2016). en. In: *Veterinary Record* 178.4, pp. 86–87.

Allen, L. J. S., V. L. Brown, C. B. Jonsson, S. L. Klein, S. M. Laverty, K. Magwedere, J. C. Owen, and P. Van Den Driessche (2012). "Mathematical Modeling of Viral Zoonoses in Wildlife". en. In: *Natural Resource Modeling* 25.1, pp. 5–51.

Allen, L. J. S. (2017). "A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis". en. In: *Infectious Disease Modelling* 2.2, pp. 128–142.

Alteri, C. et al. (2021). "Genomic epidemiology of SARS-CoV-2 reveals multiple lineages and early spread of SARS-CoV-2 infections in Lombardy, Italy". In: *Nature Communications* 12.1, p. 434.

Althaus, C. L. (2014). "Estimating the Reproduction Number of Ebola Virus (EBOV) During the 2014 Outbreak in West Africa". In: *PLoS Currents* 6, ecurrents.outbreaks.91afb5e0f279e7f29e7056095255b288.

Anderson, R. M. and R. M. May (1985). "Vaccination and herd immunity to infectious diseases". en. In: *Nature* 318.6044, pp. 323–329.

Andrews, N. et al. (2022). "Covid-19 Vaccine Effectiveness against the Omicron (B.1.1.529) Variant". In: *New England Journal of Medicine* 386.16, pp. 1532–1546.

Andronico, A., C. Tran Kiem, J. Paireau, T. Succo, P. Bosetti, N. Lefrancq, M. Nacher, F. Djossou, A. Sanna, C. Flamand, H. Salje, C. Rousseau, and S. Cauchemez (2021). "Evaluating the impact of curfews and other measures on SARS-CoV-2 transmission in French Guiana". en. In: *Nature Communications* 12.1, p. 1634.

Andronico, A. et al. (2017). "Real-Time Assessment of Health-Care Requirements During the Zika Virus Epidemic in Martinique". In: *American Journal of Epidemiology* 186.10, pp. 1194–1203.

Anothaisintawee, T., A. Julienne Genuino, M. Thavorncharoensap, S. Youngkong, W. Rattanavipapong, A. Meeyai, and U. Chaikledkaew (2019). "Cost-effectiveness modelling studies of all preventive measures against rabies: A systematic review". In: *Vaccine* 37, A146–A153.

Arechiga Ceballos, N., D. Karunaratna, and A. Aguilar Setien (2014). "Control of canine rabies in developing countries: key features and animal welfare implications". In: *Revue Scientifique et Technique de l'OIE* 33.1, pp. 311–321.

Ariën, K. K., G. Vanham, and E. J. Arts (2007). "Is HIV-1 evolving to a less virulent form in humans?" en. In: *Nature Reviews Microbiology* 5.2, pp. 141–151.

Ashcroft, P., J. S. Huisman, S. Lehtinen, J. A. Bouman, C. L. Althaus, R. R. Regoes, and S. Bonhoeffer (2020). "COVID-19 infectivity profile correction". In: *Swiss Medical Weekly* 150.32, pp. 3–5.

Ashcroft, P., S. Lehtinen, D. C. Angst, N. Low, and S. Bonhoeffer (2021). "Quantifying the impact of quarantine duration on COVID-19 transmission". In: *eLife* 10, pp. 1–33.

Attwood, S. W., S. C. Hill, D. M. Aanensen, T. R. Connor, and O. G. Pybus (2022). "Phylogenetic and phylodynamic approaches to understanding and combating the early SARS-CoV-2 pandemic". en. In: *Nature Reviews Genetics* 23.9, pp. 547–562.

Aw, D., A. B. Silva, and D. B. Palmer (2007). "Immunosenescence: emerging challenges for an ageing population". en. In: *Immunology* 120.4, pp. 435–446.

Ayres, D. L., M. P. Cummings, G. Baele, A. E. Darling, P. O. Lewis, D. L. Swofford, J. P. Huelsenbeck, P. Lemey, A. Rambaut, and M. A. Suchard (2019). "BEAGLE 3: Improved Performance, Scaling, and Usability for a High-Performance Computing Library for Statistical Phylogenetics". In: *Systematic Biology* 68.6, pp. 1052–1061.

Ayres, D. L., A. Darling, D. J. Zwickl, P. Beerli, M. T. Holder, P. O. Lewis, J. P. Huelsenbeck, F. Ronquist, D. L. Swofford, M. P. Cummings, A. Rambaut, and M. A. Suchard (2012). "BEAGLE: An Application Programming Interface and High-Performance Computing Library for Statistical Phylogenetics". In: *Systematic Biology* 61.1, pp. 170–173.

Azman, A. S., J. H. Stark, B. M. Althouse, C. J. Vukotich, S. Stebbins, D. S. Burke, and D. A. T. Cummings (2013). "Household transmission of influenza A and B in a school-based study of non-pharmaceutical interventions". en. In: *Epidemics* 5.4, pp. 181–186.

Baden, L. R. et al. (2021). "Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine". In: *New England Journal of Medicine* 384.5, pp. 403–416.

Baele, G., M. A. Suchard, A. Rambaut, and P. Lemey (2016). "Emerging Concepts of Data Integration in Pathogen Phylodynamics". In: *Systematic Biology* 66.1, syw054.

Baker, R. E., A. S. Mahmud, I. F. Miller, M. Rajeev, F. Rasambainarivo, B. L. Rice, S. Takahashi, A. J. Tatem, C. E. Wagner, L. F. Wang, A. Wesolowski, and C. J. E. Metcalf (2021). "Infectious disease in an era of global change". In: *Nature Reviews Microbiology 2021*, pp. 1–13.

Bansal, S., G. Chowell, L. Simonsen, A. Vespignani, and C. Viboud (2016). "Big Data for Infectious Disease Surveillance and Modeling". In: *The Journal of Infectious Diseases* 214.suppl_4, S375–S379.

Bardosh, K. (2014). "Global aspirations, local realities: the role of social science research in controlling neglected tropical diseases". In: *Infectious Diseases of Poverty* 3.1, p. 35.

Bärnighausen, T., D. E. Bloom, E. T. Cafiero, and J. C. O'Brien (2013). "Valuing the broader benefits of dengue vaccination, with a preliminary application to Brazil". en. In: *Seminars in Immunology*. Technologies and opportunities for vaccines 25.2, pp. 104–113.

Beerli, P. (2004). "Effect of unsampled populations on the estimation of population sizes and migration rates between sampled populations". In: *Molecular Ecology* 13.4, pp. 827–836.

Benenson, S., Y. Oster, M. J. Cohen, and R. Nir-Paz (2021). "BNT162b2 mRNA Covid-19 Vaccine Effectiveness among Health Care Workers". In: *New England Journal of Medicine*, pp. 1–2.

Benjathummarak, S., C. Fa-ngoen, C. Pipattanaboon, K. Boonha, P. Ramasoota, and P. Pitaksajjakul (2016). "Molecular genetic characterization of rabies virus glycoprotein gene sequences from rabid dogs in Bangkok and neighboring provinces in Thailand, 2013–2014". In: *Archives of Virology* 161.5, pp. 1261–1271.

Bernal, J. L., N. Andrews, C. Gower, J. Stowe, C. Robertson, E. Tessier, R. Simmons, S. Cottrell, R. Roberts, M. O'Doherty, K. Brown, C. Cameron, D. Stockton, J. McMenamin, and M. Ramsay (2021). "Early effectiveness of COVID-19 vaccination with BNT162b2 mRNA vaccine and ChAdOx1 adenovirus vector vaccine on symptomatic disease, hospitalisations and mortality in older adults in England". In: *medRxiv*, p. 2021.03.01.21252652.

Beyene, T. J., M. C. Fitzpatrick, A. P. Galvani, M. C. Mourits, C. W. Revie, N. Cernicchiaro, M. W. Sanderson, and H. Hogeveen (2019). "Impact of One-Health framework on vaccination cost-effectiveness: A case study of rabies in Ethiopia". In: *One Health* 8.August, p. 100103.

Beyer, H. L., K. Hampson, T. Lembo, S. Cleaveland, M. Kaare, and D. T. Haydon (2011). "Metapopulation dynamics of rabies and the efficacy of vaccination". In: *Proceedings of the Royal Society B: Biological Sciences* 278.1715, pp. 2182–2190.

— (2012). "The implications of metapopulation dynamics on the design of vaccination campaigns". In: *Vaccine* 30.6, pp. 1014–1022.

Bi, Q. et al. (2020). "Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study". In: *The Lancet Infectious Diseases* 20.8, pp. 911–919.

Bjørnstad, O. N. (2018). *Epidemics: Models and Data using R*. en. Springer International Publishing.

Bloom, D. E. and D. Cadarette (2019). "Infectious Disease Threats in the Twenty-First Century: Strengthening the Global Response". In: *Frontiers in Immunology* 10.

Bloomquist, E. W., P. Lemey, and M. A. Suchard (2010). "Three roads diverged? Routes to phylogeographic inference". In: *Trends in Ecology and Evolution* 25.11, pp. 626–632.

Bombara, C. B., S. Dürr, G. E. Machovsky-Capuska, P. W. Jones, and M. P. Ward (2017). "A preliminary study to estimate contact rates between free-roaming domestic dogs using novel miniature cameras". In: *PLoS ONE* 12.7, pp. 1–16.

Boskova, V., S. Bonhoeffer, and T. Stadler (2014). "Inference of Epidemiological Dynamics Based on Simulated Phylogenies Using Birth-Death and Coalescent Models". In: *PLoS Computational Biology* 10.11.

Bouckaert, R. et al. (2019). "BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis". In: *PLOS Computational Biology* 15.4. Ed. by M. Pertea, e1006650.

Bourhy, H., E. Nakouné, M. Hall, P. Nouvellet, A. Lepelletier, C. Talbi, L. Watier, E. C. Holmes, S. Cauchemez, P. Lemey, C. A. Donnelly, and A. Rambaut (2016). "Revealing the Micro-scale Signature of Endemic Zoonotic Disease Transmission in an African Urban Setting". In: *PLOS Pathogens* 12.4. Ed. by C. Parrish, e1005525.

Bourhy, H., J.-M. Reynes, E. J. Dunham, L. Dacheux, F. Larrous, V. T. Q. Huong, G. Xu, J. Yan, M. E. G. Miranda, and E. C. Holmes (2008). "The origin and phylogeography of dog rabies virus". In: *Journal of General Virology* 89.11, pp. 2673–2681.

Bracher, J., E. L. Ray, T. Gneiting, and N. G. Reich (2021). “Evaluating epidemic forecasts in an interval format”. In: *PLOS Computational Biology* 17.2. Ed. by V. E. Pitzer, e1008618.

Broban, A., M. C. Tejiokem, I. Tiembré, S. Druelles, and M. L’Azou (2018). “Bolstering human rabies surveillance in Africa is crucial to eliminating canine-mediated rabies”. In: *PLOS Neglected Tropical Diseases* 12.9. Ed. by D. Knobel, e0006367.

Brodin, P. and M. M. Davis (2017). “Human immune system variation”. en. In: *Nature Reviews Immunology* 17.1, pp. 21–29.

Bromham, L., S. Duchêne, X. Hua, A. M. Ritchie, D. A. Duchêne, and S. Y. Ho (2018). “Bayesian molecular dating: opening up the black box”. In: *Biological Reviews* 93.2, pp. 1165–1191.

Bromham, L. and D. Penny (2003). “The modern molecular clock”. en. In: *Nature Reviews Genetics* 4.3, pp. 216–224.

Brookes, V. J., S. Dürr, and M. P. Ward (2019). “Rabies-induced behavioural changes are key to rabies persistence in dog populations: Investigation using a network-based model”. In: *PLOS Neglected Tropical Diseases* 13.9. Ed. by T. S. Churcher, e0007739.

Brunker, K., S. Nadin-Davis, and R. Biek (2018a). “Genomic sequencing, evolution and molecular epidemiology of rabies virus”. In: *Revue Scientifique et Technique de l’OIE* 37.2, pp. 401–408.

Brunker, K., P. Lemey, D. A. Marston, A. R. Fooks, A. Lugelo, C. Ngeleja, K. Hampson, and R. Biek (2018b). “Landscape attributes governing local transmission of an endemic zoonosis: Rabies virus in domestic dogs”. In: *Molecular Ecology* 27.3, pp. 773–788.

Brunker, K., D. A. Marston, D. L. Horton, S. Cleaveland, A. R. Fooks, R. Kazwala, C. Ngeleja, T. Lembo, M. Sambo, Z. J. Mtema, L. Sikana, G. Wilkie, R. Biek, and K. Hampson (2015). “Elucidating the phylodynamics of endemic rabies virus in eastern Africa using whole-genome sequencing”. In: *Virus Evolution* 1.1, vev011.

Brunker, K. et al. (2020). “Rapid in-country sequencing of whole virus genomes to inform rabies elimination programmes”. In: *Wellcome Open Research* 5, p. 3.

Brynildsrud, O. B. et al. (2018). “Global expansion of Mycobacterium tuberculosis lineage 4 shaped by colonial migration and local adaptation”. In: *Science Advances* 4.10, pp. 5869–5886.

Buckee, C., A. Noor, and L. Sattenspiel (2021). “Thinking clearly about social aspects of infectious disease transmission”. In: *Nature* 595.7866, pp. 205–213.

Butera, Y. et al. (2021). “Genomic sequencing of SARS-CoV-2 in Rwanda reveals the importance of incoming travelers on lineage diversity”. In: *Nature Communications* 12.1, p. 5705.

Byambasuren, O., M. Cardona, K. Bell, J. Clark, M.-L. McLaws, and P. Glasziou (2020). “Estimating the extent of asymptomatic COVID-19 and its potential for community transmission: Systematic review and meta-analysis”. In: *Official Journal of the Association of Medical Microbiology and Infectious Disease Canada* 5.4, pp. 223–234.

Campbell, F., B. Archer, H. Laurenson-Schafer, Y. Jinnai, F. Konings, N. Batra, B. Pavlin, K. Vandemaele, M. D. Van Kerkhove, T. Jombart, O. Morgan, and O. le Polain de Waroux (2021). “Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021”. en. In: *Eurosurveillance* 26.24.

Candido, D. S. et al. (2020). “Evolution and epidemic spread of SARS-CoV-2 in Brazil”. In: *Science* 369.6508, pp. 1255–1260.

Cantaert, T. et al. (2019). “A 1-week intradermal dose-sparing regimen for rabies post-exposure prophylaxis (RESIST-2): an observational cohort study”. English. In: *The Lancet Infectious Diseases* 19.12, pp. 1355–1362.

Carnieli, P., R. de Novaes Oliveira, C. I. Macedo, and J. G. Castilho (2011). “Phylogeography of rabies virus isolated from dogs in Brazil between 1985 and 2006”. In: *Archives of Virology* 156.6, pp. 1007–1012.

Carnieli, P., H. B. C. Ruthner Batista, R. de Novaes Oliveira, J. G. Castilho, and L. F. P. Vieira (2013). “Phylogeographic dispersion and diversification of rabies virus lineages associated with dogs and crab-eating foxes (Cerdocyon thous) in Brazil”. In: *Archives of Virology* 158.11, pp. 2307–2313.

Carroll, M. J., A. Singer, G. C. Smith, D. P. Cowan, and G. Massei (2010). “The use of immunocontraception to improve rabies eradication in urban dog populations”. In: *Wildlife Research* 37.8, p. 676.

Cauchemez, S., A. Bhattarai, T. L. Marchbanks, R. P. Fagan, S. Ostroff, N. M. Ferguson, D. Swerdlow, S. V. Sodha, M. E. Moll, F. J. Angulo, R. Palekar, W. R. Archer, and L. Finelli (2011). “Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza”. In: *Proceedings of the National Academy of Sciences of the United States of America* 108.7, pp. 2825–2830.

Cauchemez, S., F. Carrat, C. Viboud, A. J. Valleron, and P. Y. Boëlle (2004). “A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data”. In: *Statistics in Medicine* 23.22, pp. 3469–3487.

Cauchemez, S., C. A. Donnelly, C. Reed, A. C. Ghani, C. Fraser, C. K. Kent, L. Finelli, and N. M. Ferguson (2009). “Household transmission of 2009 pandemic influenza A (H1N1) virus in the United States”. In: *New England Journal of Medicine* 361.27, pp. 2619–2627.

Cauchemez, S., N. M. Ferguson, A. Fox, L. Q. Mai, L. T. Thanh, P. Q. Thai, D. D. Thoang, T. N. Duong, L. N. Minh Hoa, N. Tran Hien, and P. Horby (2014). “Determinants of Influenza Transmission in South East Asia: Insights from a Household Cohort Study in Vietnam”. In: *PLoS Pathogens* 10.8, pp. 2–9.

Cauchemez, S., N. Hoze, A. Cousien, B. Nikolay, and Q. ten bosch (2019). “How Modelling Can Enhance the Analysis of Imperfect Epidemic Data”. en. In: *Trends in Parasitology* 35.5, pp. 369–379.

Cevik, M., M. Tate, O. Lloyd, A. E. Maraolo, J. Schafers, and A. Ho (2021). “SARS-CoV-2, SARS-CoV, and MERS-CoV viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and meta-analysis”. In: *The Lancet Microbe* 2.1, e13–e22.

Chaillon, A., S. Gianella, S. Dellicour, S. A. Rawlings, T. E. Schlub, M. F. De Oliveira, C. Ignacio, M. Porrachia, B. Vrancken, and D. M. Smith (2020). “HIV persists throughout deep tissues with repopulation from multiple anatomical sources”. In: *Journal of Clinical Investigation* 130.4, pp. 1699–1712.

Chaillon, A., S. Gianella, J. O. Wertheim, D. D. Richman, S. R. Mehta, and D. M. Smith (2014). “HIV Migration Between Blood and Cerebrospinal Fluid or Semen Over Time”. In: *The Journal of Infectious Diseases* 209.10, pp. 1642–1652.

Chen, J., L. Zou, Z. Jin, and S. Ruan (2015). “Modeling the Geographic Spread of Rabies in China”. In: *PLOS Neglected Tropical Diseases* 9.5. Ed. by C. E. Rupprecht, e0003772.

Chevalier, V., H. Davun, S. Sorn, P. Ly, V. Pov, and S. Ly (2021). “Large scale dog population demography, dog management and bite risk factors analysis: A crucial step towards rabies control in Cambodia”. In: *PLoS ONE* 16.7 July, pp. 1–18.

Choisy, M. (2010). “1. Changements environnementaux et santé : 3. Modélisation mathématique en épidémiologie”. In: *Ecologie de la santé et biodiversité*. Ed. by F. Thomas and M. Gauthier-Clerc. Licence Maîtrise Doctorat. Bruxelles: De Boeck, pp. 159–182.

Chung, E. et al. (2021). “Comparison of Symptoms and RNA Levels in Children and Adults With SARS-CoV-2 Infection in the Community Setting”. In: *JAMA Pediatrics* 175.10, e212025–e212025.

Cleaveland, S., S. Thumbi, M. Sambo, A. Lugelo, K. Lushasi, K. Hampson, and F. Lankester (2018). “Proof of concept of mass dog vaccination for thecontrol and elimination of canine rabies”. In: *Revue Scientifique et Technique de l’OIE* 37.2, pp. 559–568.

Cleaveland, S., H. Beyer, K. Hampson, D. Haydon, F. Lankester, T. Lembo, F.-X. Meslin, M. Morters, Z. Mtema, M. Sambo, and S. Townsend (2014). “The changing landscape of rabies epidemiology and control”. In: *Onderstepoort J Vet Res* 81.2, pp. 1–8.

Cleaveland, S. and K. Hampson (2017). “Rabies elimination research: juxtaposing optimism, pragmatism and realism”. In: *Proceedings of the Royal Society B: Biological Sciences* 284.1869, p. 20171880.

Coetzer, A., T. P. Scott, A. C. Amparo, S. Jayme, and L. H. Nel (2018). “Formation of the Asian Rabies Control Network (ARACON): A common approach towards a global good”. en. In: *Antiviral Research* 157, pp. 134–139.

Coleman, P. G. and C. Dye (1996). “Immunization coverage required to prevent outbreaks of dog rabies”. In: *Vaccine* 14.3, pp. 185–186.

Colombi, D., C. Poletto, E. Nakouné, H. Bourhy, and V. Colizza (2020). “Long-range movements coupled with heterogeneous incubation period sustain dog rabies at the national scale in Africa”. In: *PLOS Neglected Tropical Diseases* 14.5. Ed. by S. Recuenco, e0008317.

Constenla, D., C. Garcia, and N. Lefcourt (2015). “Assessing the Economics of Dengue: Results from a Systematic Review of the Literature and Expert Survey”. en. In: *PharmacoEconomics* 33.11, pp. 1107–1135.

Cori, A., C. A. Donnelly, I. Dorigatti, N. M. Ferguson, C. Fraser, T. Garske, T. Jombart, G. Nedjati-Gilani, P. Nouvellet, S. Riley, M. D. Van Kerkhove, H. L. Mills, and I. M. Blake (2017). “Key data for outbreak evaluation: building on the Ebola experience”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 372.1721, p. 20160371.

Cori, A., P. Nouvellet, T. Garske, H. Bourhy, E. Nakouné, and T. Jombart (2018). “A graph-based evidence synthesis approach to detecting outbreak clusters: An application to dog rabies”. In: *PLOS Computational Biology* 14.12. Ed. by M. Pascual, e1006554.

Dacheux, L., L. Dommergues, Y. Chouanibou, L. Doméon, C. Schuler, S. Bonas, D. Luo, C. Maufrais, C. Cetre-Sossah, E. Cardinale, H. Bourhy, and R. Métras (2019). “Co-circulation and characterization of novel African arboviruses (genus Ephemerovirus) in cattle, Mayotte island, Indian Ocean, 2017”. en. In: *Transboundary and Emerging Diseases* 66.6, pp. 2601–2604.

Dacheux, L., F. Larrous, R. Lavenir, A. Lepelletier, A. Faouzi, C. Troupin, J. Nourlil, P. Buchy, and H. Bourhy (2016). “Dual Combined Real-Time Reverse Transcription Polymerase Chain Reaction Assay for the Diagnosis of Lyssavirus Infection”. en. In: *PLOS Neglected Tropical Diseases* 10.7. Publisher: Public Library of Science, e0004812.

Dagan, N., N. Barda, E. Kepten, O. Miron, S. Perchik, M. A. Katz, M. A. Hernán, M. Lipsitch, B. Reis, and R. D. Balicer (2021). "BNT162b2 mRNA Covid-19 Vaccine in a Nationwide Mass Vaccination Setting". In: *New England Journal of Medicine* 384.15, pp. 1412–1423.

Darkaoui, S., F. Cliquet, M. Wasniewski, E. Robardet, N. Aboulfidaa, M. Bouslikhane, and O. Fassi-Fihri (2017). "A Century Spent Combating Rabies in Morocco (1911–2015): How Much Longer?" In: *Frontiers in Veterinary Science* 4.78, pp. 1–16.

Dattner, I., Y. Goldberg, G. Katriel, R. Yaari, N. Gal, Y. Miron, A. Ziv, R. Sheffer, Y. Hamo, and A. Huppert (2021). "The role of children in the spread of COVID-19: Using household data from Bnei Brak, Israel, to estimate the relative susceptibility and infectivity of children". In: *PLOS Computational Biology* 17.2. Ed. by J. Lloyd-Smith, e1008559.

Davies, N. G. et al. (2020a). "Age-dependent effects in the transmission and control of COVID-19 epidemics". In: *Nature Medicine* 26.8, pp. 1205–1211.

Davies, N. G. et al. (2020b). "Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study". In: *The Lancet Public Health* 5.7, e375–e385.

Davies, N. G. et al. (2021). "Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England". In: *Science* 372.6538, eabg3055.

De la Puente-Arévalo, M. et al. (2022). "Ranging patterns and factors associated with movement in free-roaming domestic dogs in urban Malawi". en. In: *Ecology and Evolution* 12.1, e8498.

De la Puente-León, M., M. Z. Levy, A. M. Toledo, S. Recuenco, J. Shinnick, and R. Castillo-Neyra (2020). "Spatial Inequality Hides the Burden of Dog Bites and the Risk of Dog-Mediated Human Rabies". In: *The American Journal of Tropical Medicine and Hygiene* 103.3, pp. 1247–1257.

De Maio, N., C.-H. Wu, K. M. O'Reilly, and D. Wilson (2015). "New Routes to Phylogeography: A Bayesian Structured Coalescent Approximation". In: *PLOS Genetics* 11.8. Ed. by J. K. Pritchard, e1005421.

Dearlove, B. L., F. Xiang, and S. D. W. Frost (2017). "Biased phylodynamic inferences from analysing clusters of viral sequences". In: *Virus Evolution* 3.2, vex020.

Debin, M., C. Turbelin, T. Blanchon, I. Bonmarin, A. Falchi, T. Hanslik, D. Levy-Bruhl, C. Poletto, and V. Colizza (2013). "Evaluating the Feasibility and Participants' Representativeness of an Online Nationwide Surveillance System for Influenza in France". en. In: *PLOS ONE* 8.9, e73675.

Degeling, C., V. Brookes, T. Lea, and M. Ward (2018). "Rabies response, One Health and more-than-human considerations in Indigenous communities in northern Australia". In: *Social Science & Medicine* 212.March, pp. 60–67.

Dellicour, S., G. Baele, G. Dudas, N. R. Faria, O. G. Pybus, M. A. Suchard, A. Rambaut, and P. Lemey (2018a). "Phylodynamic assessment of intervention strategies for the West African Ebola virus outbreak". In: *Nature Communications* 9.1, p. 2222.

Dellicour, S., M. S. Gill, N. R. Faria, A. Rambaut, O. G. Pybus, M. A. Suchard, and P. Lemey (2021a). "Relax, Keep Walking - A Practical Guide to Continuous Phylogeographic Inference with BEAST". In: *Molecular Biology and Evolution* 38.8, pp. 3486–3493.

Dellicour, S., S. L. Hong, B. Vrancken, A. Chaillon, M. S. Gill, M. T. Maurano, S. Ramaswami, P. Zappile, C. Marier, G. W. Harkins, G. Baele, R. Duerr, and A. Heguy (2021b). "Dispersal dynamics

of SARS-CoV-2 lineages during the first epidemic wave in New York City". In: *PLOS Pathogens* 17.5. Ed. by A. S. Lauring, e1009571.

Dellicour, S., R. Rose, N. R. Faria, P. Lemey, and O. G. Pybus (2016a). "SERAPHIM: studying environmental rasters and phylogenetically informed movements". In: *Bioinformatics* 32.20, pp. 3204–3206.

Dellicour, S., R. Rose, N. R. Faria, L. F. P. Vieira, H. Bourhy, M. Gilbert, P. Lemey, and O. G. Pybus (2017). "Using Viral Gene Sequences to Compare and Explain the Heterogeneous Spatial Dynamics of Virus Epidemics". In: *Molecular Biology and Evolution* 34.10, pp. 2563–2571.

Dellicour, S., R. Rose, and O. G. Pybus (2016b). "Explaining the geographic spread of emerging epidemics: a framework for comparing viral phylogenies and environmental landscape data". In: *BMC Bioinformatics* 17.1, p. 82.

Dellicour, S., C. Troupin, F. Jahanbakhsh, A. Salama, S. Massoudi, M. K. Moghaddam, G. Baele, P. Lemey, A. Gholami, and H. Bourhy (2019). "Using phylogeographic approaches to analyse the dispersal history, velocity and direction of viral lineages — Application to rabies virus spread in Iran". In: *Molecular Ecology* 28.18, pp. 4335–4350.

Dellicour, S., B. Vrancken, N. S. Trovão, D. Fargette, and P. Lemey (2018b). "On the importance of negative controls in viral landscape phylogeography". In: *Virus Evolution* 4.2, vey023.

Dellicour, S. et al. (2020). "Epidemiological hypothesis testing using a phylogeographic and phylodynamic framework". In: *Nature Communications* 11.1, p. 5620.

Dellicour, S. et al. (2021c). "A Phylodynamic Workflow to Rapidly Gain Insights into the Dispersal History and Dynamics of SARS-CoV-2 Lineages". In: *Molecular Biology and Evolution* 38.4. Ed. by N. Stuart, pp. 1608–1613.

Dibia, I. N., B. Sumiarto, H. Susetya, A. A. G. Putra, H. Scott-Orr, and G. N. Mahardika (2015). "Phylogeography of the current rabies viruses in Indonesia". In: *Journal of Veterinary Science* 16.4, p. 459.

Dietz, K. (1967). "Epidemics and Rumours: A Survey". In: *Journal of the Royal Statistical Society. Series A (General)* 130.4, pp. 505–528.

Disease Control and Prevention, C. for (2006). "Principles of Epidemiology in Public Health Practice, Third Edition: An Introduction". en. In: p. 511.

Drummond, A. J., S. Y. Ho, M. J. Phillips, and A. Rambaut (2006). "Relaxed phylogenetics and dating with confidence". In: *PLoS Biology* 4.5, pp. 699–710.

Drummond, A. J. and A. Rambaut (2007). "BEAST : Bayesian evolutionary analysis by sampling trees". In: *BMC Evolutionary Biology* 7.214, pp. 1–8.

Drummond, A. J., O. G. Pybus, A. Rambaut, R. Forsberg, and A. G. Rodrigo (2003). "Measurably evolving populations". en. In: *Trends in Ecology & Evolution* 18.9, pp. 481–488.

Dudas, G., L. M. Carvalho, A. Rambaut, and T. Bedford (2018). "MERS-CoV spillover at the camel-human interface". In: *eLife* 7, pp. 1–22.

Dudas, G. et al. (2017). "Virus genomes reveal factors that spread and sustained the Ebola epidemic". In: *Nature* 544.7650, pp. 309–315.

Duong, V., A. Tarantola, S. Ong, C. Mey, R. Choeung, S. Ly, H. Bourhy, P. Dussart, and P. Buchy (2016). "Laboratory diagnostics in dog-mediated rabies: an overview of performance and a proposed strategy for various settings". en. In: *International Journal of Infectious Diseases* 46, pp. 107–114.

Dupanloup, I., S. Schneider, and L. Excoffier (2002). “A simulated annealing approach to define the genetic structure of populations”. en. In: *Molecular Ecology* 11.12, pp. 2571–2581.

Dürr, S. and M. P. Ward (2015). “Development of a Novel Rabies Simulation Model for Application in a Non-endemic Environment”. In: *PLOS Neglected Tropical Diseases* 9.6. Ed. by J. Zinsstag, e0003876.

Eddelbuettel, D. and J. J. Balamuta (2018). “Extending R with C++: A Brief Introduction to Rcpp”. In: *The American Statistician* 72.1, pp. 28–36.

Emch, M., C. Feldacker, M. S. Islam, and M. Ali (2008). “Seasonality of cholera from 1974 to 2005: a review of global patterns”. In: *International Journal of Health Geographics* 7.1, p. 31.

Endo, A., M. Uchida, A. J. Kucharski, and S. Funk (2019). “Fine-scale family structure shapes influenza transmission risk in households: Insights from primary schools in Matsumoto city, 2014/15”. In: *PLOS Computational Biology* 15.12, e1007589.

Ewing, G. and A. Rodrigo (2006). “Estimating Population Parameters using the Structured Serial Coalescent with Bayesian MCMC Inference when some Demes are Hidden”. In: *Evolutionary Bioinformatics* 2, p. 117693430600200.

Eyre, D. W., D. Taylor, M. Purver, D. Chapman, T. Fowler, K. B. Pouwels, A. S. Walker, and T. E. Peto (2021). *The impact of SARS-CoV-2 vaccination on Alpha & Delta variant transmission*. en.

Faria, N. R. et al. (2017). “Establishment and cryptic transmission of Zika virus in Brazil and the Americas”. In: *Nature* 546.7658, pp. 406–410.

Faria, N. R. et al. (2019). “Distinct rates and patterns of spread of the major HIV-1 subtypes in Central and East Africa”. In: *PLOS Pathogens* 15.12. Ed. by R. Swanstrom, e1007976.

Faria, N. R. et al. (2016). “Zika virus in the Americas: Early epidemiological and genetic findings”. In: *Science* 352.6283, pp. 345–349.

Fazio, R. H., B. C. Ruisch, C. A. Moore, J. A. Granados Samayoa, S. T. Boggs, and J. T. Ladanyi (2021). “Social distancing decreases an individual’s likelihood of contracting COVID-19”. In: *Proceedings of the National Academy of Sciences* 118.8, e2023131118.

Featherstone, L. A., J. M. Zhang, T. G. Vaughan, and S. Duchene (2022). “Epidemiological inference from pathogen genomes: A review of phylodynamic models and applications”. In: *Virus Evolution* 8.1.

Feikin, D. R., M. M. Higdon, L. J. Abu-Raddad, N. Andrews, R. Araos, Y. Goldberg, M. J. Groome, A. Huppert, K. L. O’Brien, P. G. Smith, A. Wilder-Smith, S. Zeger, M. D. Knoll, and M. K. Patel (2022). “Duration of effectiveness of vaccines against SARS-CoV-2 infection and COVID-19 disease: results of a systematic review and meta-regression”. English. In: *The Lancet* 399.10328, pp. 924–944.

Ferguson, E. A., K. Hampson, S. Cleaveland, R. Consunji, R. Deray, J. Friar, D. T. Haydon, J. Jimenez, M. Pancipane, and S. E. Townsend (2015). “Heterogeneity in the spread and control of infectious disease: consequences for the elimination of canine rabies”. In: *Scientific Reports* 5.1, p. 18232.

Ferguson, N. M., C. Fraser, C. A. Donnelly, A. C. Ghani, and R. M. Anderson (2004). “Public Health Risk from the Avian H5N1 Influenza Epidemic”. In: *Science* 304.5673, pp. 968–969.

Ferreira, M. A. and M. A. Suchard (2008). “Bayesian analysis of elapsed times in continuous-time Markov chains”. In: *Canadian Journal of Statistics* 36.3, pp. 355–368.

Ferretti, L., A. Ledda, C. Wymant, L. Zhao, V. Ledda, L. Abeler-Dörner, M. Kendall, A. Nurtay, H.-Y. Cheng, T.-C. Ng, H.-H. Lin, R. Hinch, J. Masel, A. M. Kilpatrick, and C. Fraser (2020). *The timing of COVID-19 transmission*. en.

Fisher, C. R., D. G. Streicker, and M. J. Schnell (2018). "The spread and evolution of rabies virus: conquering new frontiers". In: *Nature Reviews Microbiology* 16.4, pp. 241–255.

FitzJohn, R. and T. Fischer (2022). *odin: ODE Generation and Integration*.

Fitzpatrick, M. C., K. Hampson, S. Cleaveland, L. A. Meyers, J. P. Townsend, and A. P. Galvani (2012). "Potential for Rabies Control through Dog Vaccination in Wildlife-Abundant Communities of Tanzania". In: *PLoS Neglected Tropical Diseases* 6.8. Ed. by R. Reithinger, e1796.

Flaxman, S. et al. (2020). "Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe". In: *Nature* 584.7820, pp. 257–261.

Franco, N., P. Coletti, L. Willem, L. Angeli, A. Lajot, S. Abrams, P. Beutels, C. Faes, and N. Hens (2022). "Inferring age-specific differences in susceptibility to and infectiousness upon SARS-CoV-2 infection based on Belgian social contact data". en. In: *PLOS Computational Biology* 18.3, e1009965.

Frost, S. D., O. G. Pybus, J. R. Gog, C. Viboud, S. Bonhoeffer, and T. Bedford (2015). "Eight challenges in phylodynamic inference". In: *Epidemics* 10, pp. 88–92.

Gallup, J. L. and J. D. Sachs (2001). "The economic burden of malaria". en. In: *The American Journal of Tropical Medicine and Hygiene* 64.1_suppl, pp. 85–96.

Ganyani, T., C. Kremer, D. Chen, A. Torneri, C. Faes, J. Wallinga, and N. Hens (2020). "Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020". en. In: *Eurosurveillance* 25.17.

Gardy, J. L. and N. J. Loman (2018). "Towards a genomics-informed, real-time, global pathogen surveillance system". en. In: *Nature Reviews Genetics* 19.1, pp. 9–20.

Gascuel, O. and M. Steel (2020). "A Darwinian Uncertainty Principle". In: *Systematic Biology* 69.3, pp. 521–529.

Gier, B. de, S. Andeweg, R. Joosten, R. ter Schegget, N. Smorenburg, J. van de Kassteele, S. J. Hahné, S. van den Hof, H. E. de Melker, and M. J. Knol (2021). "Vaccine effectiveness against SARS-CoV-2 transmission and infections among household and other close contacts of confirmed cases, the Netherlands, February to May 2021". In: *Eurosurveillance* 26.31, p. 2100640.

Gigante, C. M. et al. (2020). "Portable Rabies Virus Sequencing in Canine Rabies Endemic Countries Using the Oxford Nanopore MinION". In: *Viruses* 12.11, p. 1255.

Gilks, W. R., S. Richardson, and D. Spiegelhalter (1995). *Markov Chain Monte Carlo in Practice*. en. CRC Press.

Gill, M. S., P. Lemey, S. N. Bennett, R. Biek, and M. A. Suchard (2016). "Understanding Past Population Dynamics: Bayesian Coalescent-Based Modeling with Covariates". In: *Systematic Biology* 65.6, pp. 1041–1056.

Gill, M. S., P. Lemey, N. R. Faria, A. Rambaut, B. Shapiro, and M. A. Suchard (2013). "Improving Bayesian Population Dynamics Inference: A Coalescent-Based Model for Multiple Loci". In: *Molecular Biology and Evolution* 30.3, pp. 713–724.

Gire, S. K. et al. (2014). "Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak". In: *Science* 345.6202, pp. 1369–1372.

Glatman-Freedman, A., I. Portelli, S. K. Jacobs, J. I. Mathew, J. E. Slutzman, L. R. Goldfrank, and S. W. Smith (2012). "Attack Rates Assessment of the 2009 Pandemic H1N1 Influenza A in Children and Their Contacts: A Systematic Review and Meta-Analysis". en. In: *PLOS ONE* 7.11. Publisher: Public Library of Science, e50228.

Goeyvaerts, N., E. Santermans, G. Potter, A. Torneri, K. Van Kerckhove, L. Willem, M. Aerts, P. Beutels, and N. Hens (2018). "Household members do not contact each other at random: Implications for infectious disease modelling". In: *Proceedings of the Royal Society B: Biological Sciences* 285.1893.

Goldberg, Y., M. Mandel, Y. Woodbridge, R. Fluss, I. Novikov, R. Yaari, A. Ziv, L. Freedman, and A. Huppert (2021). "Protection of previous SARS-CoV-2 infection is similar to that of BNT162b2 vaccine protection: A three-month nationwide experience from Israel". In: *medRxiv*.

Golding, N., A. Schofield, and M. U. G. Kraemer (2015). "Movement: Functions for the analysis of movement data in disease modelling and mapping." In: *R Packag. version 0.2*.

Grenfell, B. T., O. G. Pybus, J. R. Gog, J. L. Wood, J. M. Daly, J. A. Mumford, and E. C. Holmes (2004). "Unifying the Epidemiological and Evolutionary Dynamics of Pathogens". In: *Science* 303.5656, pp. 327–332.

Griffiths, R. C., S. Tavare, W. F. Bodmer, and P. J. Donnelly (1994). "Sampling theory for neutral alleles in a varying environment". In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 344.1310, pp. 403–410.

Grubaugh, N. D., J. T. Ladner, P. Lemey, O. G. Pybus, A. Rambaut, E. C. Holmes, and K. G. Andersen (2019a). "Tracking virus outbreaks in the twenty-first century". en. In: *Nature Microbiology* 4.1, pp. 10–19.

Grubaugh, N. D. et al. (2019b). "Travel Surveillance and Genomics Uncover a Hidden Zika Outbreak during the Waning Epidemic". In: *Cell* 178.5, 1057–1071.e11.

Guindon, S. and N. De Maio (2021). "Accounting for spatial sampling patterns in Bayesian phylogeography". In: *Proceedings of the National Academy of Sciences* 118.52, e2105273118.

Guo, Z., X. Tao, C. Yin, N. Han, J. Yu, H. Li, H. Liu, W. Fang, J. Adams, J. Wang, G. Liang, Q. Tang, and S. Rayner (2013). "National Borders Effectively Halt the Spread of Rabies: The Current Rabies Epidemic in China Is Dislocated from Cases in Neighboring Countries". In: *PLoS Neglected Tropical Diseases* 7.1. Ed. by C. E. Rupprecht, e2039.

Gupta, S., N. Ferguson, and R. Anderson (1998). "Chaos, Persistence, and Evolution of Strain Structure in Antigenically Diverse Infectious Agents". In: *Science* 280.5365, pp. 912–915.

Haas, E. J., F. J. Angulo, J. M. McLaughlin, E. Anis, S. R. Singer, F. Khan, N. Brooks, M. Smaja, G. Mircus, K. Pan, J. Southern, D. L. Swerdlow, L. Jodar, Y. Levy, and S. Alroy-Preis (2021). "Impact and effectiveness of mRNA BNT162b2 vaccine against SARS-CoV-2 infections and COVID-19 cases, hospitalisations, and deaths following a nationwide vaccination campaign in Israel: an observational study using national surveillance data". In: *The Lancet* 397.10287, pp. 1819–1829.

Hadfield, J., C. Megill, S. M. Bell, J. Huddleston, B. Potter, C. Callender, P. Sagulenko, T. Bedford, and R. A. Neher (2018). "NextStrain: Real-time tracking of pathogen evolution". In: *Bioinformatics* 34.23, pp. 4121–4123.

Hampson, K., K. De Balogh, and J. McGrane (2019a). "Lessons for rabies control and elimination programmes: a decade of One Health experience from Bali, Indonesia". In: *Revue Scientifique et Technique de l'OIE* 38.1, pp. 213–224.

Hampson, K., B. Abela-Ridder, O. Bharti, L. Knopf, M. Léchenne, R. Mindekem, A. Tarantola, J. Zinsstag, and C. Trotter (2019b). "Modelling to inform prophylaxis regimens to prevent human rabies". In: *Vaccine* 37, A166–A173.

Hampson, K., J. Dushoff, J. Bingham, G. Brückner, Y. H. Ali, and A. Dobson (2007). "Synchronous cycles of domestic dog rabies in sub-Saharan Africa and the impact of control efforts". In: *Proceedings of the National Academy of Sciences* 104.18, pp. 7717–7722.

Hampson, K., J. Dushoff, S. Cleaveland, D. T. Haydon, M. Kaare, C. Packer, and A. Dobson (2009). "Transmission Dynamics and Prospects for the Elimination of Canine Rabies". In: *PLoS Biology* 7.3. Ed. by C. E. Rupprecht, e1000053.

Hampson, K. et al. (2015). "Estimating the Global Burden of Endemic Canine Rabies". In: *PLOS Neglected Tropical Diseases* 9.4. Ed. by M. S. Carvalho, e0003709.

Harris, R. J., J. A. Hall, A. Zaidi, N. J. Andrews, J. K. Dunbar, and G. Dabrera (2021). "Effect of Vaccination on Household Transmission of SARS-CoV-2 in England". In: *New England Journal of Medicine* 385.8, pp. 759–760.

Hart, W. S., S. Abbott, A. Endo, J. Hellewell, E. Miller, N. Andrews, P. K. Maini, S. Funk, and R. N. Thompson (2022). "Inference of the SARS-CoV-2 generation time using UK household data". In: *eLife* 11.

Hasanov, E., S. Zeynalova, M. Geleishvili, E. Maes, E. Tongren, E. Marshall, A. Banyard, L. M. Mcelhinney, A. M. Whatmore, A. R. Fooks, and D. L. Horton (2018). "Assessing the impact of public education on a preventable zoonotic disease: rabies". In: *Epidemiology and Infection* 146.2, pp. 227–235.

Hasegawa, M., H. Kishino, and T.-a. Yano (1985). "Dating of the human-ape splitting by a molecular clock of mitochondrial DNA". In: *Journal of Molecular Evolution* 22.2, pp. 160–174.

Hastings, W. K. (1970). "Monte Carlo Sampling Methods Using Markov Chains and Their Applications". In: *Biometrika* 57.1, pp. 97–109.

Hawkes, M. T., B. E. Lee, J. N. Kanji, N. Zelyas, K. Wong, M. Barton, S. Mukhi, and J. L. Robinson (2021). "Seasonality of Respiratory Viruses at Northern Latitudes". In: *JAMA Network Open* 4.9, e2124650.

Hayman, D. T. S., N. Johnson, D. L. Horton, J. Hedge, P. R. Wakeley, A. C. Banyard, S. Zhang, A. Alhassan, and A. R. Fooks (2011). "Evolutionary History of Rabies in Ghana". In: *PLoS Neglected Tropical Diseases* 5.4. Ed. by J. Zinsstag, e1001.

Hazelbag, C. M., J. Dushoff, E. M. Dominic, Z. E. Mthombothi, and W. Delva (2020). "Calibration of individual-based models to epidemiological data: A systematic review". en. In: *PLOS Computational Biology* 16.5, e1007893.

He, W.-T. et al. (2022). "Phylogeography Reveals Association between Swine Trade and the Spread of Porcine Epidemic Diarrhea Virus in China and across the World". In: *Molecular Biology and Evolution* 39.2. Ed. by M. Barlow.

Heaney, A. K., K. A. Alexander, and J. Shaman (2020). "Ensemble forecast and parameter inference of childhood diarrhea in Chobe District, Botswana". en. In: *Epidemics* 30, p. 100372.

Heesterbeek, H. et al. (2015). "Modeling infectious disease dynamics in the complex landscape of global health". In: *Science* 347.6227, aaa4339.

Hens, N., M. Aerts, C. Faes, Z. Shkedy, O. Lejeune, P. Van Damme, and P. Beutels (2010). "Seventy-five years of estimating the force of infection from current status data". English. In: *Epidemiology and Infection* 138.6, pp. 802–812.

Herlihy, D. and S. H. Cohn (1997). *The Black Death and the Transformation of the West*. en. Harvard University Press.

Hewins, B., M. Rahman, J. F. Bermejo-Martin, A. A. Kelvin, C. D. Richardson, S. Rubino, A. Kumar, P. Ndishimye, A. Toloue Ostadgavahi, A. Mahmud-Al-Rafat, and D. J. Kelvin (2022). "Alpha, Beta, Delta, Omicron, and SARS-CoV-2 Breakthrough Cases: Defining Immunological Mechanisms for Vaccine Waning and Vaccine-Variant Mismatch". en. In: *Frontiers in Virology* 2, p. 849936.

Hill, V., C. Ruis, S. Bajaj, O. G. Pybus, and M. U. G. Kraemer (2021). "Progress and challenges in virus genomic epidemiology". English. In: *Trends in Parasitology* 37.12, pp. 1038–1049.

Ho, S. Y. W. and S. Duchêne (2014). "Molecular-clock methods for estimating evolutionary rates and timescales". en. In: *Molecular Ecology* 23.24, pp. 5947–5965.

Hodcroft, E. B., N. De Maio, R. Lanfear, D. R. MacCannell, B. Q. Minh, H. A. Schmidt, A. Stamatakis, N. Goldman, and C. Dessimoz (2021a). "Want to track pandemic variants faster? Fix the bioinformatics bottleneck". In: *Nature* 591.7848, pp. 30–33.

Hodcroft, E. B. et al. (2021b). "Spread of a SARS-CoV-2 variant through Europe in the summer of 2020". en. In: *Nature* 595.7869, pp. 707–712.

Hoekstra, R., R. D. Morey, J. N. Rouder, and E.-J. Wagenmakers (2014). "Robust misinterpretation of confidence intervals". en. In: *Psychonomic Bulletin & Review* 21.5, pp. 1157–1164.

Hong, S. L., P. Lemey, M. A. Suchard, and G. Baele (2021). "Bayesian Phylogeographic Analysis Incorporating Predictors and Individual Travel Histories in BEAST". In: *Current Protocols* 1.4, pp. 1–16.

Horton, D. L., L. M. McElhinney, C. M. Freuling, D. A. Marston, A. C. Banyard, H. Goharrriz, E. Wise, A. C. Breed, G. Saturday, J. Kolodziejek, E. Zilahi, M. F. Al-Kobaisi, N. Nowotny, T. Mueller, and A. R. Fooks (2015). "Complex Epidemiology of a Zoonotic Disease in a Culturally Diverse Region: Phylogeography of Rabies Virus in the Middle East". In: *PLOS Neglected Tropical Diseases* 9.3. Ed. by C. E. Rupprecht, e0003569.

Hou, Q., Z. Jin, and S. Ruan (2012). "Dynamics of rabies epidemics and the impact of control efforts in Guangdong Province, China". In: *Journal of Theoretical Biology* 300, pp. 39–47.

Hozé, N., I. Diarra, A. K. Sangaré, B. Pastorino, L. Pezzi, B. Kouriba, I. Sagara, A. Dabo, A. Djimdé, M. A. Thera, O. K. Doumbo, X. de Lamballerie, and S. Cauchemez (2021a). "Model-based assessment of Chikungunya and O'nyong-nyong virus circulation in Mali in a serological cross-reactivity context". In: *Nature Communications* 12.1.

Hozé, N., J. Paireau, N. Lapidus, C. T. Kiem, H. Salje, G. Severi, M. Touvier, M. Zins, X. d. Lamballerie, D. Lévy-Bruhl, F. Carrat, and S. Cauchemez (2021b). "Monitoring the proportion of the population infected by SARS-CoV-2 using age-stratified hospitalisation and serological data: a modelling study". English. In: *The Lancet Public Health* 6.6, e408–e415.

Huang, J., S. Ruan, Y. Shu, and X. Wu (2019). "Modeling the Transmission Dynamics of Rabies for Dog, Chinese Ferret Badger and Human Interactions in Zhejiang Province, China". In: *Bulletin of Mathematical Biology* 81.4, pp. 939–962.

Hudson, E. G., V. J. Brookes, S. Dürr, and M. P. Ward (2019a). "Modelling targeted rabies vaccination strategies for a domestic dog population with heterogeneous roaming patterns". In: *PLOS Neglected Tropical Diseases* 13.7. Ed. by S. Recuenco, e0007582.

Hudson, E. G., V. J. Brookes, M. P. Ward, and S. Dürr (2019b). "Using roaming behaviours of dogs to estimate contact rates: the predicted effect on rabies spread". In: *Epidemiology and Infection* 147, e135.

Huisman, J. S. et al. (2022a). "Wastewater-Based Estimation of the Effective Reproductive Number of SARS-CoV-2". In: *Environmental Health Perspectives* 130.5, p. 057011.

Huisman, J. S., J. Scire, D. C. Angst, J. Li, R. A. Neher, M. H. Maathuis, S. Bonhoeffer, and T. Stadler (2022b). "Estimation and worldwide monitoring of the effective reproductive number of SARS-CoV-2". In: *eLife* 11. Ed. by M. P. Davenport, e71345.

Hussaini, N., K. Okuneye, and A. B. Gumel (2017). "Mathematical analysis of a model for zoonotic visceral leishmaniasis". en. In: *Infectious Disease Modelling* 2.4, pp. 455–474.

Ishikawa, S. A., A. Zhukova, W. Iwasaki, and O. Gascuel (2019). "A Fast Likelihood Method to Reconstruct and Visualize Ancestral Scenarios". In: *Molecular Biology and Evolution* 36.9, pp. 2069–2085.

Izda, V., M. A. Jeffries, and A. H. Sawalha (2021). "COVID-19: A review of therapeutic strategies and vaccine candidates". In: *Clinical Immunology* 222.108634, p. 108634.

Jacquot, M., K. Nomikou, M. Palmarini, P. Mertens, and R. Biek (2017). "Bluetongue virus spread in Europe is a consequence of climatic, landscape and vertebrate host factors as revealed by phylogeographic inference". In: *Proceedings of the Royal Society B: Biological Sciences* 284.1864, p. 20170919.

Jayaweera, M., H. Perera, B. Gunawardana, and J. Manatunge (2020). "Transmission of COVID-19 virus by droplets and aerosols: A critical review on the unresolved dichotomy". en. In: *Environmental Research* 188, p. 109819.

Jing, Q.-L. et al. (2020). "Household secondary attack rate of COVID-19 and associated determinants in Guangzhou, China: a retrospective cohort study". In: *The Lancet Infectious Diseases* 20.10, pp. 1141–1150.

Jones, J. E., V. Le Sage, and S. S. Lakdawala (2021). "Viral and host heterogeneity and their effects on the viral life cycle". en. In: *Nature Reviews Microbiology* 19.4, pp. 272–282.

Juan Zhang, Zhen Jin, Guiquan Sun, Xiangdong Sun, and Shigui Ruan (2012). "Spatial spread of rabies in China". In: *Journal of Applied Analysis & Computation* 2.1, pp. 111–126.

Kadowaki, H., K. Hampson, K. Tojinbara, A. Yamada, and K. Makita (2018). "The risk of rabies spread in Japan: a mathematical modelling assessment". In: *Epidemiology and Infection* 146.10, pp. 1245–1252.

Kaleta, T. et al. (2022). "Antibody escape and global spread of SARS-CoV-2 lineage A.27". In: *Nature Communications* 13.1, p. 1152.

Kalkauskas, A., U. Perron, Y. Sun, N. Goldman, G. Baele, S. Guindon, and N. De Maio (2021). "Sampling bias and model choice in continuous phylogeography: Getting lost on a random walk". In: *PLOS Computational Biology* 17.1. Ed. by S. L. Kosakovsky Pond, e1008561.

Kass, R. E. and A. E. Raftery (1995). "Bayes Factors". In: *Journal of the American Statistical Association* 90.430, pp. 773–795.

Katoh, K. and D. M. Standley (2013). “MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability”. In: *Molecular Biology and Evolution* 30.4, pp. 772–780.

Keeling, M. J. and P. Rohani (2008). *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press.

Kermack, W. O., A. G. McKendrick, and G. T. Walker (1927). “A contribution to the mathematical theory of epidemics”. In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 115.772, pp. 700–721.

Kim, H., R. G. Webster, and R. J. Webby (2018). “Influenza Virus: Dealing with a Drifting and Shifting Pathogen”. In: *Viral Immunology* 31.2, pp. 174–183.

Kingman, J. F. C. (1982). “On the Genealogy of Large Populations”. In: *Journal of Applied Probability* 19, pp. 27–43.

Kissler, S. M. et al. (2021). *Viral dynamics of SARS-CoV-2 variants in vaccinated and unvaccinated individuals*. en.

Kitala, P. M., J. J. McDermott, P. G. Coleman, and C. Dye (2002). “Comparison of vaccination strategies for the control of dog rabies in Machakos District, Kenya”. In: *Epidemiology and Infection* 129.1, pp. 215–222.

Kittisiam, T., W. Phimpraphai, S. Kasemsuwan, and K. K. Thakur (2021). “Analyses of Contact Networks of Community Dogs on a University Campus in Nakhon Pathom, Thailand”. en. In: *Veterinary Sciences* 8.12, p. 299.

Knobel, D. L., T. Lembo, M. Morters, S. E. Townsend, S. Cleaveland, and K. Hampson (2013). “Dog Rabies and Its Control”. In: *Rabies*. Vol. 1921. Elsevier, pp. 591–615.

Koelle, K., X. Rodó, M. Pascual, M. Yunus, and G. Mostafa (2005). “Refractory periods and climate forcing in cholera dynamics”. en. In: *Nature* 436.7051, pp. 696–700.

Kruschke, J. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. en. Academic Press.

Kühnert, D., T. Stadler, T. G. Vaughan, and A. J. Drummond (2016). “Phylodynamics with Migration: A Computational Framework to Quantify Population Structure from Genomic Data”. In: *Molecular Biology and Evolution* 33.8, pp. 2102–2116.

Laager, M., M. Léchenne, K. Naissengar, R. Mindekem, A. Oussiguere, J. Zinsstag, and N. Chitnis (2019). “A metapopulation model of dog rabies transmission in N’Djamena, Chad”. In: *Journal of Theoretical Biology* 462, pp. 408–417.

Laager, M., C. Mbilo, E. A. Madaye, A. Naminou, M. Léchenne, A. Tschopp, S. K. Naïssengar, T. Smieszek, J. Zinsstag, and N. Chitnis (2018). “The importance of dog population contact network structures in rabies transmission”. In: *PLOS Neglected Tropical Diseases* 12.8. Ed. by C. E. Rupprecht, e0006680.

Lau, L. L. H., H. Nishiura, H. Kelly, D. K. M. Ip, G. M. Leung, and B. J. Cowling (2012). “Household Transmission of 2009 Pandemic Influenza A (H1N1): A Systematic Review and Meta-analysis”. en-US. In: *Epidemiology* 23.4, pp. 531–542.

Layan, M., S. Dellicour, G. Baele, S. Cauchemez, and H. Bourhy (2021). “Mathematical modelling and phylodynamics for the study of dog rabies dynamics and control: A scoping review”. en. In: *PLOS Neglected Tropical Diseases* 15.5, e0009449.

Layan, M., N. F. Müller, S. Dellicour, N. D. Maio, H. Bourhy, S. Cauchemez, and G. Baele (2022). *Impact and mitigation of sampling bias to determine viral spread: evaluating discrete phylogeography through CTMC modeling and structured coalescent model approximations*. en.

Lazzarini, L., L. Barzon, F. Foglia, V. Manfrin, M. Pacenti, G. Pavan, M. Rassu, G. Capelli, F. Montarsi, S. Martini, F. Zanella, M. T. Padovan, F. Russo, and F. Gobbi (2020). "First autochthonous dengue outbreak in Italy, August 2020". en. In: *Eurosurveillance* 25.36, p. 2001606.

Lei, H., Y. Li, S. Xiao, C.-H. Lin, S. L. Norris, D. Wei, Z. Hu, and S. Ji (2018). "Routes of transmission of influenza A H1N1, SARS CoV, and norovirus in air cabin: Comparative analyses". en. In: *Indoor Air* 28.3, pp. 394–403.

Lei, H., X. Xu, S. Xiao, X. Wu, and Y. Shu (2020). "Household transmission of COVID-19-a systematic review and meta-analysis". In: *Journal of Infection* 81.6, pp. 979–997.

Lembo, T., K. Hampson, D. T. Haydon, M. Craft, A. Dobson, J. Dushoff, E. Ernest, R. Hoare, M. Kaare, T. Mlengeya, C. Mentzel, and S. Cleaveland (2008). "Exploring reservoir dynamics: a case study of rabies in the Serengeti ecosystem". en. In: *Journal of Applied Ecology* 45.4, pp. 1246–1257.

Lembo, T., K. Hampson, M. T. Kaare, E. Ernest, D. Knobel, R. R. Kazwala, D. T. Haydon, and S. Cleaveland (2010). "The Feasibility of Canine Rabies Elimination in Africa: Dispelling Doubts with Data". In: *PLoS Neglected Tropical Diseases* 4.2. Ed. by C. E. Rupprecht, e626.

Lemey, P., S. L. Hong, V. Hill, G. Baele, C. Poletto, V. Colizza, Á. O'Toole, J. T. McCrone, K. G. Andersen, M. Worobey, M. I. Nelson, A. Rambaut, and M. A. Suchard (2020). "Accommodating individual travel history and unsampled diversity in Bayesian phylogeographic inference of SARS-CoV-2". In: *Nature Communications* 11.1, p. 5110.

Lemey, P., A. Rambaut, T. Bedford, N. Faria, F. Bielejec, G. Baele, C. A. Russell, D. J. Smith, O. G. Pybus, D. Brockmann, and M. A. Suchard (2014). "Unifying Viral Genetics and Human Transportation Data to Predict the Global Transmission Dynamics of Human Influenza H3N2". In: *PLoS Pathogens* 10.2.

Lemey, P., A. Rambaut, A. J. Drummond, and M. A. Suchard (2009a). "Bayesian Phylogeography Finds Its Roots". In: *PLoS Computational Biology* 5.9. Ed. by C. Fraser, e1000520.

Lemey, P., A. Rambaut, J. J. Welch, and M. A. Suchard (2010). "Phylogeography takes a relaxed random walk in continuous space and time". In: *Molecular Biology and Evolution* 27.8, pp. 1877–1885.

Lemey, P., M. Salemi, and A.-M. Vandamme (2009b). *The phylogenetic handbook*. Cambridge University Press.

Lemey, P. et al. (2021). "Untangling introductions and persistence in COVID-19 resurgence in Europe". en. In: *Nature* 595.7869, pp. 713–717.

Lemieux, J. E. et al. (2021). "Phylogenetic analysis of SARS-CoV-2 in Boston highlights the impact of superspreading events". In: *Science* 371.6529, eabe3261.

Lessler, J., M. K. Grabowski, K. H. Grantz, E. Badillo-Goicoechea, C. J. E. Metcalf, C. Lupton-Smith, A. S. Azman, and E. A. Stuart (2021). "Household COVID-19 risk and in-person schooling". In: *Science* 372.6546, pp. 1092–1097.

Leung, N. H. L. (2021). "Transmissibility and transmission of respiratory viruses". en. In: *Nature Reviews Microbiology* 19.8, pp. 528–545.

Leung, T. and S. A. Davis (2017). "Rabies Vaccination Targets for Stray Dog Populations". In: *Frontiers in Veterinary Science* 4.April, pp. 1–10.

Li, A. J., N. Sreenivasan, U. R. Siddiqi, S. Tahmina, K. Penjor, L. Sovann, A. Gunesekera, J. D. Blanton, L. Knopf, and T. B. Hyde (2019). "Descriptive assessment of rabies post-exposure prophylaxis procurement, distribution, monitoring, and reporting in four Asian countries: Bangladesh, Bhutan, Cambodia, and Sri Lanka, 2017–2018". en. In: *Vaccine*. Scientific and Operational Updates on Rabies 37, A14–A19.

Li, F. et al. (2021). "Household transmission of SARS-CoV-2 and risk factors for susceptibility and infectivity in Wuhan: a retrospective observational study". In: *The Lancet Infectious Diseases* 21.5, pp. 617–628.

Lipsitch, M., F. Krammer, G. Regev-Yochay, Y. Lustig, and R. D. Balicer (2022). "SARS-CoV-2 breakthrough infections in vaccinated individuals: measurement, causes and impact". en. In: *Nature Reviews Immunology* 22.1, pp. 57–65.

Liu, P., Y. Song, C. Colijn, and A. MacPherson (2022). *The impact of sampling bias on viral phylogeographic reconstruction*. en.

Liu, Y., R. M. Eggo, and A. J. Kucharski (2020). "Secondary attack rate and superspreading events for SARS-CoV-2". In: *The Lancet* 395.10227, e47.

Lloyd-Smith, J. O., S. J. Schreiber, P. E. Kopp, and W. M. Getz (2005). "Superspreading and the effect of individual variation on disease emergence". en. In: *Nature* 438.7066, pp. 355–359.

Lloyd-Smith, J. O., D. George, K. M. Pepin, V. E. Pitzer, J. R. C. Pulliam, A. P. Dobson, P. J. Hudson, and B. T. Grenfell (2009). "Epidemic Dynamics at the Human-Animal Interface". In: *Science* 326.5958, pp. 1362–1367.

Long, J. S., B. Mistry, S. M. Haslam, and W. S. Barclay (2019). "Host and viral determinants of influenza A virus species specificity". en. In: *Nature Reviews Microbiology* 17.2, pp. 67–81.

Longini, E. M., J. S. Koopman, M. Haber, G. A. Cotsonis, and I. M. (Longini (1988). "Stastical inference for infectious diseases and risk-specific household and community transmission parameters". In: *American Journal of Epidemiology* 128.4.

Longini JR., I. M., J. S. Koopman, A. S. Monto, and J. P. Fox (1982). "Estimating household and community transmission parameters for influenza". In: *American Journal of Epidemiology* 115.5, pp. 736–751.

Lordan, R. et al. (2021). "Considerations for the Safe Operation of Schools During the Coronavirus Pandemic". In: *Frontiers in Public Health* 9.

Lu, L. et al. (2021). "Adaptation, spread and transmission of SARS-CoV-2 in farmed minks and associated humans in the Netherlands". In: *Nature Communications* 12.1, p. 6802.

Lu, R. et al. (2020). "Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding". In: *The Lancet* 395.10224, pp. 565–574.

Luo, D.-S., B. Li, X.-R. Shen, R.-D. Jiang, Y. Zhu, J. Wu, Y. Fan, H. Bourhy, B. Hu, X.-Y. Ge, Z.-L. Shi, and L. Dacheux (2021). "Characterization of Novel Rhabdoviruses in Chinese Bats". en. In: *Viruses* 13.1, p. 64.

Ly, S., P. Buchy, N. Y. Heng, S. Ong, N. Chhor, H. Bourhy, and S. Vong (2009). "Rabies Situation in Cambodia". en. In: *PLOS Neglected Tropical Diseases* 3.9, e511.

Ma, C. et al. (2017). "Re-emerging of rabies in Shaanxi province, China, from 2009 to 2015". In: *Journal of Medical Virology* 89.9, pp. 1511–1519.

Ma, X. et al. (2019). “Analysis of error profiles in deep next-generation sequencing data”. In: *Genome Biology* 20.1, p. 50.

Madewell, Z. J., Y. Yang, I. M. Longini, M. E. Halloran, and N. E. Dean (2020). “Household Transmission of SARS-CoV-2”. In: *JAMA Network Open* 3.12, e2031756.

Magee, D. and M. Scotch (2018). “The effects of random taxa sampling schemes in Bayesian virus phylogeography”. In: *Infection, Genetics and Evolution* 64.June, pp. 225–230.

Mahardika, G. N. K., N. Dibia, N. S. Budayanti, N. M. Susilawathi, K. Subrata, A. E. Darwinata, F. S. Wignall, J. A. Richt, W. A. Valdivia-Granda, and A. A. R. Sudewi (2014). “Phylogenetic analysis and victim contact tracing of rabies virus from humans and dogs in Bali, Indonesia”. In: *Epidemiology and Infection* 142.6, pp. 1146–1154.

Mäkinen, T. M., R. Juvonen, J. Jokelainen, T. H. Harju, A. Peitso, A. Bloigu, S. Silvennoinen-Kassinen, M. Leinonen, and J. Hassi (2009). “Cold temperature and low humidity are associated with increased occurrence of respiratory tract infections”. en. In: *Respiratory Medicine* 103.3, pp. 456–462.

Mancy, R., M. Rajeev, A. Lugelo, K. Brunker, S. Cleaveland, E. A. Ferguson, K. Hotopp, R. Kazwala, M. Magoto, K. Rysava, D. T. Haydon, and K. Hampson (2022). “Rabies shows how scale of transmission can enable acute infections to persist at low prevalence”. In: *Science* 376.6592, pp. 512–516.

Marston, D. A., L. M. McElhinney, R. J. Ellis, D. L. Horton, E. L. Wise, S. L. Leech, D. David, X. de Lamballerie, and A. R. Fooks (2013). “Next generation sequencing of viral RNA genomes”. In: *BMC Genomics* 14.1, p. 444.

Martínez-Baz, I., A. Miqueleiz, I. Casado, A. Navascués, C. Trobajo-Sanmartín, C. Burgui, M. Guevara, C. Ezpeleta, and J. Castilla (2021). “Effectiveness of COVID-19 vaccines in preventing SARS-CoV-2 infection and hospitalisation, Navarre, Spain, January to April 2021”. In: *Eurosurveillance* 26.21, p. 2100438.

Mastin, A. J., F. v. d. Bosch, T. R. Gottwald, V. A. Chavez, and S. R. Parnell (2017). “A method of determining where to target surveillance efforts in heterogeneous epidemiological systems”. en. In: *PLOS Computational Biology* 13.8, e1005712.

Mathieu, E., H. Ritchie, E. Ortiz-Ospina, M. Roser, J. Hasell, C. Appel, C. Giattino, and L. Rodés-Guirao (2021). “A global database of COVID-19 vaccinations”. In: *Nature Human Behaviour* 5.7, pp. 947–953.

Mavian, C., T. K. Paisie, M. T. Alam, C. Browne, V. M. Beau De Rochars, S. Nembrini, M. N. Cash, E. J. Nelson, T. Azarian, A. Ali, J. G. Morris, and M. Salemi (2020). “Toxigenic Vibrio cholerae evolution and establishment of reservoirs in aquatic ecosystems”. In: *Proceedings of the National Academy of Sciences* 117.14, pp. 7897–7904.

Mbilo, C. et al. (2021). “Dog rabies control in West and Central Africa: A review”. In: *Acta Tropica* 224, p. 105459.

McAloon, C., Á. Collins, K. Hunt, A. Barber, A. W. Byrne, F. Butler, M. Casey, J. Griffin, E. Lane, D. McEvoy, P. Wall, M. Green, L. O’Grady, and S. J. More (2020). “Incubation period of COVID-19: a rapid systematic review and meta-analysis of observational research”. In: *BMJ Open* 10.8, e039652.

McMahan, C. S., S. Self, L. Rennert, C. Kalbaugh, D. Kriebel, D. Graves, C. Colby, J. A. Deaver, S. C. Popat, T. Karanfil, and D. L. Freedman (2021). “COVID-19 wastewater epidemiology: a model to estimate infected populations”. English. In: *The Lancet Planetary Health* 5.12, e874–e881.

McRae, B. H. (2006). “Isolation by Resistance”. en. In: *Evolution* 60.8, pp. 1551–1561.

Meng, S., Y. Sun, X. Wu, J. Tang, G. Xu, Y. Lei, J. Wu, J. Yan, X. Yang, and C. E. Rupprecht (2011). "Evolutionary dynamics of rabies viruses highlights the importance of China rabies transmission in Asia". In: *Virology* 410.2, pp. 403–409.

Meredith, L. W. et al. (2020). "Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study". en. In: *The Lancet Infectious Diseases* 20.11, pp. 1263–1271.

Mey, C., A. Metlin, V. Duong, S. Ong, S. In, P. F. Horwood, J.-M. Reynes, H. Bourhy, A. Tarantola, and P. Buchy (2016). "Evidence of two distinct phylogenetic lineages of dog rabies virus circulating in Cambodia". en. In: *Infection, Genetics and Evolution* 38, pp. 55–61.

Milne, I., G. Stephen, M. Bayer, P. J. Cock, L. Pritchard, L. Cardle, P. D. Shaw, and D. Marshall (2013). "Using Tablet for visual exploration of second-generation sequencing data". In: *Briefings in Bioinformatics* 14.2, pp. 193–202.

Minh, B. Q., H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. von Haeseler, and R. Lanfear (2020). "IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era". In: *Molecular Biology and Evolution* 37.5, pp. 1530–1534.

Minin, V. N., E. W. Bloomquist, and M. A. Suchard (2008). "Smooth Skyride through a Rough Skyline: Bayesian Coalescent-Based Inference of Population Dynamics". In: *Molecular Biology and Evolution* 25.7, pp. 1459–1471.

Mistry, D., M. Litvinova, A. Pastore y Piontti, M. Chinazzi, L. Fumanelli, M. F. C. Gomes, S. A. Haque, Q.-H. Liu, K. Mu, X. Xiong, M. E. Halloran, I. M. Longini, S. Merler, M. Ajelli, and A. Vespignani (2021). "Inferring high-resolution human mixing patterns for disease modeling". en. In: *Nature Communications* 12.1, p. 323.

Miura, F., C. E. v. Ewijk, J. A. Backer, M. Xiridou, E. Franz, E. O. d. Coul, D. Brandwagt, B. v. Cleef, G. v. Rijckevorsel, C. Swaan, S. v. d. Hof, and J. Wallinga (2022). "Estimated incubation period for monkeypox cases confirmed in the Netherlands, May 2022". In: *Eurosurveillance* 27.24, p. 2200448.

Mollentze, N., J. Weyer, W. Markotter, K. le Roux, and L. H. Nel (2013). "Dog rabies in southern Africa: regional surveillance and phylogeographical analyses are an important component of control and elimination strategies". In: *Virus Genes* 47.3, pp. 569–573.

Mollentze, N., L. H. Nel, S. Townsend, K. le Roux, K. Hampson, D. T. Haydon, and S. Soubeyrand (2014). "A Bayesian approach for inferring the dynamics of partially observed endemic infectious diseases from space-time-genetic data". In: *Proceedings of the Royal Society B: Biological Sciences* 281.1782, p. 20133251.

Moon, S. et al. (2015). "Will Ebola change the game? Ten essential reforms before the next pandemic. The report of the Harvard-LSHTM Independent Panel on the Global Response to Ebola". English. In: *The Lancet* 386.10009, pp. 2204–2221.

Morel, B., P. Barbera, L. Czech, B. Bettisworth, L. Hübner, S. Lutteropp, D. Serdari, E.-G. Kostaki, I. Mamais, A. M. Kozlov, P. Pavlidis, D. Paraskevis, and A. Stamatakis (2021). "Phylogenetic Analysis of SARS-CoV-2 Data Is Difficult". In: *Molecular Biology and Evolution* 38.5. Ed. by H. Malik, pp. 1777–1791.

Morters, M. K., O. Restif, K. Hampson, S. Cleaveland, J. L. N. Wood, and A. J. K. Conlan (2013). "Evidence-based control of canine rabies: a critical review of population density reduction". In: *Journal of Animal Ecology* 82.1. Ed. by M. Boots, pp. 6–14.

Muinde, P., J. M. Bettridge, F. M. Sousa, S. Dürr, I. R. Dohoo, J. Berezowski, T. Mutwiri, C. O. Odinga, E. M. Fèvre, and L. C. Falzon (2021). "Who let the dogs out? Exploring the spatial ecology of free-roaming domestic dogs in western Kenya". en. In: *Ecology and Evolution* 11.9, pp. 4218–4231.

Müller, N. F. et al. (2021). "Viral genomes reveal patterns of the SARS-CoV-2 outbreak in Washington State". In: *Science Translational Medicine* 13.595, p. 202.

Müller, N. F., G. Dudas, and T. Stadler (2019). "Inferring time-dependent migration and coalescence patterns from genetic sequence and predictor data in structured populations". In: *Virus Evolution* 5.2, pp. 1–10.

Müller, N. F., D. Rasmussen, and T. Stadler (2018). "MASCOT: parameter and state inference under the marginal structured coalescent approximation". In: *Bioinformatics* 34.22. Ed. by J. Kelso, pp. 3843–3848.

Müller, N. F., D. A. Rasmussen, and T. Stadler (2017). "The Structured Coalescent and Its Approximations". In: *Molecular Biology and Evolution* 34.11, pp. 2970–2981.

Munitz, A., M. Yechezkel, Y. Dickstein, D. Yamin, and M. Gerlic (2021). "BNT162b2 vaccination effectively prevents the rapid rise of SARS-CoV-2 variant B.1.1.7 in high-risk populations in Israel". In: *Cell Reports Medicine* 2.5, p. 100264.

Murray, C. J. et al. (2022). "Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis". en. In: *The Lancet* 399.10325, pp. 629–655.

Neal, P. and T. Kypraios (2015). "Exact Bayesian inference via data augmentation". en. In: *Statistics and Computing* 25.2, pp. 333–347.

Nguyen, A. K., D. V. Nguyen, G. C. Ngo, T. T. Nguyen, S. Inoue, A. Yamada, X. K. Dinh, D. V. Nguyen, T. X. Phan, B. Q. Pham, H. T. Nguyen, and H. T. Nguyen (2011). "Molecular epidemiology of rabies virus in Vietnam (2006-2009)". In: *Japanese Journal of Infectious Diseases* 64.5, pp. 391–396.

Nguyen, H. K. L., S. V. Nguyen, A. P. Nguyen, P. M. V. Hoang, T. T. Le, T. C. Nguyen, H. T. Hoang, C. D. Vuong, L. t. T. Tran, and M. Q. Le (2017). "Surveillance of Severe Acute Respiratory Infection (SARI) for Hospitalized Patients in Northern Vietnam, 2011–2014". In: *Japanese Journal of Infectious Diseases* 70.5, pp. 522–527.

Nourbakhsh, S. et al. (2022). "A wastewater-based epidemic model for SARS-CoV-2 with application to three Canadian cities". en. In: *Epidemics* 39, p. 100560.

Omran, A. R. (2005). "The Epidemiologic Transition: A Theory of the Epidemiology of Population Change". en. In: *The Milbank Quarterly* 83.4, pp. 731–757.

Opgen-Rhein, R., L. Fahrmeir, and K. Strimmer (2005). "Inference of demographic history from genealogical trees using reversible jump Markov chain Monte Carlo". In: *BMC Evolutionary Biology* 5.1, p. 6.

Ortega, N. R. S., P. C. Sallum, and E. Massad (2000). "Fuzzy Dynamical Systems in Epidemic Modeling". In: *Kybernetes* 29.2, pp. 201–218.

Pavoine, S., S. Ollier, D. Pontier, and D. Chessel (2008). "Testing for phylogenetic signal in phenotypic traits: New matrices of phylogenetic proximities". en. In: *Theoretical Population Biology* 73.1, pp. 79–91.

Pawlowski, C., P. Lenehan, A. Puranik, V. Agarwal, A. Venkatakrishnan, M. J. Niesen, J. C. O'Horo, A. Virk, M. D. Swift, A. D. Badley, J. Halamka, and V. Soundararajan (2021). "FDA-authorized

mRNA COVID-19 vaccines are effective per real-world evidence synthesized across a multi-state health system". In: *Med* 2.8, 979–992.e8.

Perez, L. J. et al. (2022). "The early SARS-CoV-2 epidemic in Senegal was driven by the local emergence of B.1.416 and the introduction of B.1.1.420 from Europe". In: *Virus Evolution* 8.1, pp. 1–12.

Perez-Saez, J., J. Lessler, E. C. Lee, F. J. Luquero, E. B. Malembaka, F. Finger, J. P. Langa, S. Yennan, B. Zaitchik, and A. S. Azman (2022). "The seasonality of cholera in sub-Saharan Africa: a statistical modelling study". English. In: *The Lancet Global Health* 10.6, e831–e839.

Petersen, E. et al. (2018). "Emerging infections—an increasingly important topic: review by the Emerging Infections Task Force". en. In: *Clinical Microbiology and Infection* 24.4, pp. 369–375.

Pipes, L., H. Wang, J. P. Huelsenbeck, and R. Nielsen (2021). "Assessing Uncertainty in the Rooting of the SARS-CoV-2 Phylogeny". In: *Molecular Biology and Evolution* 38.4. Ed. by H. Malik, pp. 1537–1543.

Pitzer, V. E. and T. Cohen (2020). "Household studies provide key insights on the transmission of, and susceptibility to, SARS-CoV-2". English. In: *The Lancet Infectious Diseases* 20.10, pp. 1103–1104.

Plessis, L. du and T. Stadler (2015). "Getting to the root of epidemic spread with phylodynamic analysis of genomic data". en. In: *Trends in Microbiology* 23.7, pp. 383–386.

Polack, F. P. et al. (2020). "Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine". In: *New England Journal of Medicine* 383.27, pp. 2603–2615.

Polonsky, J. A. et al. (2019). "Outbreak analytics: A developing data science for informing the response to emerging pathogens". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 374.1776.

Ponsich, A., F. Goutard, S. Sorn, and A. Tarantola (2016). "A prospective study on the incidence of dog bites and management in a rural Cambodian, rabies-endemic setting". en. In: *Acta Tropica* 160, pp. 62–67.

Pritchard, E. et al. (2021). "Impact of vaccination on new SARS-CoV-2 infections in the United Kingdom". In: *Nature Medicine* 27.8, pp. 1370–1378.

Prunas, O., J. L. Warren, F. W. Crawford, S. Gazit, T. Patalon, D. M. Weinberger, and V. E. Pitzer (2022). "Vaccination with BNT162b2 reduces transmission of SARS-CoV-2 to household contacts in Israel". In: *Science* 375.6585, pp. 1151–1154.

Pullano, G. et al. (2021). "Underdetection of cases of COVID-19 in France threatens epidemic control". en. In: *Nature* 590.7844, pp. 134–139.

Pybus, O. G., A. Rambaut, and P. H. Harvey (2000). "An integrated framework for the inference of viral population history from reconstructed genealogies". In: *Genetics* 155.3, pp. 1429–1437.

Pybus, O. G. and A. Rambaut (2009). "Evolutionary analysis of the dynamics of viral infectious disease". In: *Nature Reviews Genetics* 10.8, pp. 540–550.

Pybus, O. G., M. A. Suchard, P. Lemey, F. J. Bernardin, A. Rambaut, F. W. Crawford, R. R. Gray, N. Arinaminpathy, S. L. Stramer, M. P. Busch, and E. L. Delwart (2012). "Unifying the spatial epidemiology and molecular evolution of emerging epidemics". In: *Proceedings of the National Academy of Sciences* 109.37, pp. 15066–15071.

Quilty, B. J. et al. (2021). "Quarantine and testing strategies in contact tracing for SARS-CoV-2: a modelling study". In: *The Lancet Public Health* 6.3, e175–e183.

Rambaut, A., A. J. Drummond, D. Xie, G. Baele, and M. A. Suchard (2018). “Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7”. In: *Systematic Biology* 67.5. Ed. by E. Susko, pp. 901–904.

Rambaut, A., E. C. Holmes, Á. O’Toole, V. Hill, J. T. McCrone, C. Ruis, L. du Plessis, and O. G. Pybus (2020). “A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology”. en. In: *Nature Microbiology* 5.11, pp. 1403–1407.

Rattanavipapong, W., M. Thavorncharoensap, S. Youngkong, A. J. Genuino, T. Anothaisintawee, U. Chaikledkaew, and A. Meeyai (2019). “The impact of transmission dynamics of rabies control: Systematic review”. In: *Vaccine* 37, A154–A165.

Regev-Yochay, G. et al. (2021). “Decreased infectivity following BNT162b2 vaccination: A prospective cohort study in Israel”. In: *The Lancet Regional Health - Europe* 7.5, p. 100150.

Richardson, E. J. et al. (2018). “Gene exchange drives the ecological success of a multi-host bacterial pathogen”. In: *Nature Ecology & Evolution* 2.9, pp. 1468–1478.

Riley, J. C. (2005). “Estimates of Regional and Global Life Expectancy, 1800-2001”. In: *Population and Development Review* 31.3, pp. 537–543.

Ritchie, H., E. Mathieu, L. Rodés-Guirao, C. Appel, C. Giattino, E. Ortiz-Ospina, J. Hasell, B. Macdonald, D. Beltekian, and M. Roser (2020). “Coronavirus Pandemic (COVID-19)”. In: *Our World in Data*.

Roche, B., A. Garchitorena, and D. Roiz (2020). “The impact of lockdown strategies targeting age groups on the burden of COVID-19 in France”. In: *Epidemics* 33, p. 100424.

Rocklöv, J. and R. Dubrow (2020). “Climate change: an enduring challenge for vector-borne disease prevention and control”. en. In: *Nature Immunology* 21.5, pp. 479–483.

Rosenbaum, L. (2020). “The Untold Toll — The Pandemic’s Effects on Patients without Covid-19”. In: *New England Journal of Medicine* 382.24, pp. 2368–2371.

Rupprecht, C. E., I. V. Kuzmin, G. Yale, T. Nagarajan, and F.-X. Meslin (2019). “Priorities in applied research to ensure programmatic success in the global elimination of canine rabies”. In: *Vaccine* 37, A77–A84.

Sagulenko, P., V. Puller, and R. A. Neher (2018). “TreeTime: Maximum-likelihood phylodynamic analysis”. In: *Virus Evolution* 4.1, vex042.

Saito, M., H. Oshitani, J. R. C. Orbina, K. Tohma, A. S. de Guzman, T. Kamigaki, C. S. Demetria, D. L. Manalo, A. Noguchi, S. Inoue, and B. P. Quiambao (2013). “Genetic Diversity and Geographic Distribution of Genetically Distinct Rabies Viruses in the Philippines”. In: *PLoS Neglected Tropical Diseases* 7.4. Ed. by C. E. Rupprecht, e2144.

Salathé, M. (2018). “Digital epidemiology: what is it, and where is it going?” In: *Life Sciences, Society and Policy* 14.1, p. 1.

Salje, H. et al. (2021). “Reconstructing unseen transmission events to infer dengue dynamics from viral sequences”. In: *Nature Communications* 12.1, p. 1810.

Salo, J., M. Hägg, M. Kortelainen, T. Leino, T. Saxell, M. Siikanen, and L. Sääksvuori (2021). *The indirect effect of mRNA-based Covid-19 vaccination on unvaccinated household members*. en.

Sanyaolu, A., C. Okorie, A. Marinkovic, R. Patidar, K. Younis, P. Desai, Z. Hosein, I. Padda, J. Mangat, and M. Altaf (2020). “Comorbidity and its Impact on Patients with COVID-19”. en. In: *SN Comprehensive Clinical Medicine* 2.8, pp. 1069–1076.

SE, B. (2021). *Real-World Evidence Confirms High Effectiveness of Pfizer-BioNTech COVID-19 Vaccine and Profound Public Health Impact of Vaccination One Year After Pandemic Declared*.

Shah, A. S., C. Gribben, J. Bishop, P. Hanlon, D. Caldwell, R. Wood, M. Reid, J. McMenamin, D. Goldberg, D. Stockton, S. Hutchinson, C. Robertson, P. M. McKeigue, H. M. Colhoun, and D. A. McAllister (2021). "Effect of Vaccination on Transmission of SARS-CoV-2". In: *New England Journal of Medicine* 385.18, pp. 1718–1720.

Simini, F., M. C. González, A. Maritan, and A.-L. Barabási (2012). "A universal model for mobility and migration patterns". In: *Nature* 484.7392, pp. 96–100.

Smith, D. L., K. E. Battle, S. I. Hay, C. M. Barker, T. W. Scott, and F. E. McKenzie (2012). "Ross, Macdonald, and a Theory for the Dynamics and Control of Mosquito-Transmitted Pathogens". en. In: *PLOS Pathogens* 8.4, e1002588.

Smith, D. R. M., A. Duval, K. B. Pouwels, D. Guillemot, J. Fernandes, B.-T. Huynh, L. Temime, L. Opatowski, and on behalf of the AP-HP/Universities/Inserm COVID-19 research collaboration (2020). "Optimizing COVID-19 surveillance in long-term care facilities: a modelling study". In: *BMC Medicine* 18.1, p. 386.

Sor, S., M. Higuchi, M. A. B. Sarker, and N. Hamajima (2018). "Knowledge of rabies and dog-related behaviors among people in Siem Reap Province, Cambodia". In: *Tropical Medicine and Health* 46.1, p. 20.

Sparkes, J., G. Körtner, G. Ballard, and P. J. S. Fleming (2022). "Spatial and temporal activity patterns of owned, free-roaming dogs in coastal eastern Australia". en. In: *Preventive Veterinary Medicine* 204, p. 105641.

Sparkes, J., S. McLeod, G. Ballard, P. J. Fleming, G. Körtner, and W. Y. Brown (2016). "Rabies disease dynamics in naïve dog populations in Australia". In: *Preventive Veterinary Medicine* 131, pp. 127–136.

Stadler, T. and S. Bonhoeffer (2013). "Uncovering epidemiological dynamics in heterogeneous host populations using phylogenetic methods". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 368.1614.

Stadler, T., D. Kühnert, D. A. Rasmussen, and L. du Plessis (2014). "Insights into the Early Epidemic Spread of Ebola in Sierra Leone Provided by Viral Sequence Data". In: *PLoS Currents*.

Stan Development Team (2020). *RStan: the R interface to Stan*.

Starnini, M., B. Lepri, A. Baronchelli, A. Barrat, C. Cattuto, and R. Pastor-Satorras (2017). "Robust Modeling of Human Contact Networks Across Different Scales and Proximity-Sensing Techniques". en. In: *Social Informatics*. Ed. by G. L. Ciampaglia, A. Mashhadi, and T. Yasseri. Lecture Notes in Computer Science. Cham: Springer International Publishing, pp. 536–551.

Stern, A., P. Nickel, T. F. Meyer, and M. So (1984). "Opacity determinants of Neisseria gonorrhoeae: Gene expression and chromosomal linkage to the gonococcal pilus gene". en. In: *Cell* 37.2, pp. 447–456.

Stoler, N. and A. Nekrutenko (2021). "Sequencing error profiles of Illumina sequencing instruments". In: *NAR Genomics and Bioinformatics* 3.1, lqab019.

Strimmer, K. and O. G. Pybus (2001). "Exploring the Demographic History of DNA Sequences Using the Generalized Skyline Plot". In: *Molecular Biology and Evolution* 18.12, pp. 2298–2305.

Suchard, M. A., P. Lemey, G. Baele, D. L. Ayres, A. J. Drummond, and A. Rambaut (2018). "Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10". In: *Virus Evolution* 4.1, pp. 1–5.

Suchard, M. A., R. E. Weiss, and J. S. Sinsheimer (2005). "Models for Estimating Bayes Factors with Applications to Phylogeny and Tests of Monophyly". In: *Biometrics* 61.3, pp. 665–673.

Sukumaran, J. and M. T. Holder (2010). "DendroPy: a Python library for phylogenetic computing". In: *Bioinformatics* 26.12, pp. 1569–1571.

Susilawathi, N. M., A. E. Darwinata, I. B. Dwija, N. S. Budayanti, G. A. Wirasandhi, K. Subrata, N. K. Susilarini, R. A. Sudewi, F. S. Wignall, and G. N. Mahardika (2012). "Epidemiological and clinical features of human rabies cases in Bali 2008-2010". In: *BMC Infectious Diseases* 12.1, p. 81.

Talbi, C., E. C. Holmes, P. de Benedictis, O. Faye, E. Nakouné, D. Gamatié, A. Diarra, B. O. Elmamy, A. Sow, E. V. Adjogoua, O. Sangare, W. G. Dundon, I. Capua, A. A. Sall, and H. Bourhy (2009). "Evolutionary history and dynamics of dog rabies virus in western and central Africa". In: *Journal of General Virology* 90.4, pp. 783–791.

Talbi, C., P. Lemey, M. A. Suchard, E. Abdelatif, M. Elharrak, N. Jalal, A. Faouzi, J. E. Echevarría, S. Vazquez Morón, A. Rambaut, N. Campiz, A. J. Tatem, E. C. Holmes, and H. Bourhy (2010). "Phylodynamics and Human-Mediated Dispersal of a Zoonotic Virus". In: *PLoS Pathogens* 6.10. Ed. by M. Emerman, e1001166.

Tande, A. J., B. D. Pollock, N. D. Shah, G. Farrugia, A. Virk, M. Swift, L. Breeher, M. Binnicker, and E. F. Berbari (2022). "Impact of the Coronavirus Disease 2019 (COVID-19) Vaccine on Asymptomatic Infection Among Patients Undergoing Preprocedural COVID-19 Molecular Screening". In: *Clinical Infectious Diseases* 74.1, pp. 59–65.

Tarantola, A., S. Ly, M. Chan, S. In, Y. Peng, C. Hing, C. N. Taing, C. Phoen, S. Ly, S. Cauchemez, P. Buchy, P. Dussart, H. Bourhy, and J.-Y. Mary (2019). "Intradermal rabies post-exposure prophylaxis can be abridged with no measurable impact on clinical outcome in Cambodia, 2003–2014". en. In: *Vaccine*. Scientific and Operational Updates on Rabies 37, A118–A127.

Tarantola, A., S. Ly, S. In, S. Ong, Y. Peng, N. Heng, and P. Buchy (2015). "Rabies Vaccine and Rabies Immunoglobulin in Cambodia: Use and Obstacles to Use". In: *Journal of Travel Medicine* 22.5, pp. 348–352.

Tarkoma, S., S. Alghnam, and M. D. Howell (2020). "Fighting pandemics with digital epidemiology". English. In: *eClinicalMedicine* 26.

Tavare, S. (1986). "Some probabilistic and statistical problems in the analysis of DNA sequences". English. In: *Some mathematical questions in biology / DNA sequence analysis edited by Robert M. Miura*.

Taylor, L. H., K. Hampson, A. Fahrion, B. Abela-Ridder, and L. H. Nel (2017a). "Difficulties in estimating the human burden of canine rabies". In: *Acta Tropica* 165, pp. 133–140.

Taylor, L. H., R. M. Wallace, D. Balaram, J. M. Lindenmayer, D. C. Eckery, B. Mutonono-Watkiss, E. Parravani, and L. H. Nel (2017b). "The Role of Dog Population Management in Rabies Elimination—A Review of Current Approaches and Future Opportunities". In: *Frontiers in Veterinary Science* 4.JUL.

Tazawa, K., A. N. Lewis, S. Ly, Y. Peng, V. Duong, F. Lohr, A. D. Gibson, W. Phimpraphai, and L. Gamble (2022). "Evaluation of the Risk of Rabies in Human Victims through Implementation of

Integrated Bite Case Management System in Phnom Penh, Cambodia". en. In: *International Journal of Infectious Diseases*. Abstracts from the Eighth International Meeting on Emerging Diseases and Surveillance, IMED 2021 116, S105.

Team, W. E. R. (2014). "Ebola Virus Disease in West Africa — The First 9 Months of the Epidemic and Forward Projections". In: *New England Journal of Medicine* 371.16, pp. 1481–1495.

Tettelin, H. et al. (2000). "Complete Genome Sequence of Neisseria meningitidis Serogroup B Strain MC58". In: *Science* 287.5459, pp. 1809–1815.

Thanapongtharm, W., S. Kasemsuwan, V. Wongphruksasoong, K. Boonyo, T. Pinyopummintr, A. Wiratsudakul, M. Gilbert, and K. Leelahapongsathon (2021). "Spatial Distribution and Population Estimation of Dogs in Thailand: Implications for Rabies Prevention and Control". In: *Frontiers in Veterinary Science* 0, p. 1571.

Thompson, H. A. et al. (2021). "Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Setting-specific Transmission Rates: A Systematic Review and Meta-analysis". In: *Clinical Infectious Diseases* 73.3, e754–e764.

Thompson, R. N., J. E. Stockwin, R. D. van Gaalen, J. A. Polonsky, Z. N. Kamvar, P. A. Demarsh, E. Dahlqwist, S. Li, E. Miguel, T. Jombart, J. Lessler, S. Cauchemez, and A. Cori (2019). "Improved inference of time-varying reproduction numbers during infectious disease outbreaks". en. In: *Epidemics* 29, p. 100356.

Tian, H. et al. (2018). "Transmission dynamics of re-emerging rabies in domestic dogs of rural China". In: *PLOS Pathogens* 14.12. Ed. by M. J. Schnell, e1007392.

Tiembré, I., A. Broban, J. Bénié, M. Tetchi, S. Druelles, and M. L'Azou (2018). "Human rabies in Côte d'Ivoire 2014-2016: Results following reinforcements to rabies surveillance". In: *PLOS Neglected Tropical Diseases* 12.9. Ed. by C. E. Rupprecht, e0006649.

Tohma, K., M. Saito, C. S. Demetria, D. L. Manalo, B. P. Quiambao, T. Kamigaki, and H. Oshitani (2016). "Molecular and mathematical modeling analyses of inter-island transmission of rabies into a previously rabies-free island in the Philippines". In: *Infection, Genetics and Evolution* 38, pp. 22–28.

Tohma, K., M. Saito, T. Kamigaki, L. T. Tuason, C. S. Demetria, J. R. C. Orbina, D. L. Manalo, M. E. Miranda, A. Noguchi, S. Inoue, A. Suzuki, B. P. Quiambao, and H. Oshitani (2014). "Phylogeographic analysis of rabies viruses in the Philippines". In: *Infection, Genetics and Evolution* 23, pp. 86–94.

Townsend, S. E., T. Lembo, S. Cleaveland, F. X. Meslin, M. E. Miranda, A. A. G. Putra, D. T. Haydon, and K. Hampson (2013a). "Surveillance guidelines for disease elimination: A case study of canine rabies". In: *Comparative Immunology, Microbiology and Infectious Diseases* 36.3, pp. 249–261.

Townsend, S. E. et al. (2013b). "Designing Programs for Eliminating Canine Rabies from Islands: Bali, Indonesia as a Case Study". In: *PLoS Neglected Tropical Diseases* 7.8. Ed. by C. E. Rupprecht, e2372.

Tricco, A. C. et al. (2018). "PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation". In: *Annals of Internal Medicine* 169.7, pp. 467–473.

Troupin, C., L. Dacheux, M. Tanguy, C. Sabeta, H. Blanc, C. Bouchier, M. Vignuzzi, S. Duchene, E. C. Holmes, and H. Bourhy (2016). "Large-Scale Phylogenomic Analysis Reveals the Complex Evolutionary History of Rabies Virus in Multiple Carnivore Hosts". In: *PLOS Pathogens* 12.12. Ed. by C. Parrish, e1006041.

Tsang, T. K., V. J. Fang, D. K. Ip, R. A. Perera, H. C. So, G. M. Leung, J. S. Peiris, B. J. Cowling, and S. Cauchemez (2019). "Indirect protection from vaccinating children against influenza in households". In: *Nature Communications* 10.1, pp. 25–28.

Tsang, T. K., L. L. Lau, S. Cauchemez, and B. J. Cowling (2016). "Household Transmission of Influenza Virus". In: *Trends in Microbiology* 24.2, pp. 123–133.

Vaughan, T. G., D. Kühnert, A. Popinga, D. Welch, and A. J. Drummond (2014). "Efficient Bayesian inference under the structured coalescent". In: *Bioinformatics* 30.16, pp. 2272–2279.

Velander, L., J. Fogelberg, V. Putthana, A. Keosengthong, and J. F. Lindahl (2022). "Rabies Vaccination in Dogs in Laos: Owner Knowledge and Serological Status of Dogs". In: *Pathogens 2022, Vol. 11, Page 69* 11.1, p. 69.

Verberk, J. D. M., M. L. A. de Hoog, I. Westerhof, S. van Goethem, C. Lammens, G. Ieven, E. de Bruin, D. Eggink, J. A. Bielicki, S. Coenen, J. van Beek, M. J. M. Bonten, H. Goossens, and P. C. J. L. Bruijning-Verhagen (2022). "Transmission of SARS-CoV-2 within households: a remote prospective cohort study in European countries". en. In: *European Journal of Epidemiology* 37.5, pp. 549–561.

Viana, R. et al. (2022). "Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa". en. In: *Nature* 603.7902, pp. 679–686.

Viboud, C., O. N. Bjørnstad, D. L. Smith, L. Simonsen, M. A. Miller, and B. T. Grenfell (2006). "Synchrony, waves, and spatial hierarchies in the spread of influenza". In: *Science* 312.5772, pp. 447–451.

Vignuzzi, M., J. K. Stone, J. J. Arnold, C. E. Cameron, and R. Andino (2006). "Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population". en. In: *Nature* 439.7074, pp. 344–348.

Viner, R. M. et al. (2021). "Susceptibility to SARS-CoV-2 Infection Among Children and Adolescents Compared With Adults". In: *JAMA Pediatrics* 175.2, p. 143.

Volk, A. A. and J. A. Atkinson (2013). "Infant and child death in the human environment of evolutionary adaptation". en. In: *Evolution and Human Behavior* 34.3, pp. 182–192.

Volz, E. et al. (2021). "Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England". In: *Nature* 593.7858, pp. 266–269.

Volz, E. M., K. Koelle, and T. Bedford (2013). "Viral Phylodynamics". In: *PLoS Computational Biology* 9.3. Ed. by S. Wodak, e1002947.

Volz, E. M., S. L. Kosakovsky Pond, M. J. Ward, A. J. Leigh Brown, and S. D. Frost (2009). "Phylodynamics of infectious disease epidemics". In: *Genetics* 183.4, pp. 1421–1430.

VoPham, T., M. D. Weaver, G. Adamkiewicz, and J. E. Hart (2021). "Social Distancing Associations with COVID-19 Infection and Mortality Are Modified by Crowding and Socioeconomic Status". en. In: *International Journal of Environmental Research and Public Health* 18.9, p. 4680.

Vos, T. et al. (2020). "Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019". English. In: *The Lancet* 396.10258, pp. 1204–1222.

Vrancken, B., B. Zhao, X. Li, X. Han, H. Liu, J. Zhao, P. Zhong, Y. Lin, J. Zai, M. Liu, D. M. Smith, S. Dellicour, and A. Chaillon (2020). "Comparative Circulation Dynamics of the Five Main HIV Types in China". In: *Journal of Virology* 94.23. Ed. by G. Silvestri, pp. 683–703.

Vu, A. H., T. T. Nguyen, D. V. Nguyen, G. C. Ngo, T. Q. Pham, S. Inoue, A. Nishizono, T. D. Nguyen, and A. K. T. Nguyen (2021). "Rabies-infected dogs at slaughterhouses: A potential risk of rabies transmission via dog trading and butchering activities in Vietnam". en. In: *Zoonoses and Public Health* 68.6, pp. 630–637.

Wang, C. C., K. A. Prather, J. Sznitman, J. L. Jimenez, S. S. Lakdawala, Z. Tufekci, and L. C. Marr (2021). "Airborne transmission of respiratory viruses". In: *Science* 373.6558, eabd9149.

Wang, L., X. Wu, J. Bao, C. Song, and J. Du (2019). "Phylodynamic and transmission pattern of rabies virus in China and its neighboring countries". In: *Archives of Virology* 164.8, pp. 2119–2129.

Warembourg, C. et al. (2021). "Comparative Study of Free-Roaming Domestic Dog Management and Roaming Behavior Across Four Countries: Chad, Guatemala, Indonesia, and Uganda". In: *Frontiers in Veterinary Science* 8.

Welburn, S. C., P. G. Coleman, and J. Zinsstag (2017). "Rabies Control: Could Innovative Financing Break the Deadlock?" In: *Frontiers in Veterinary Science* 4.MAR, pp. 1–8.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wilkinson, E. et al. (2021). "A year of genomic surveillance reveals how the SARS-CoV-2 pandemic unfolded in Africa". In: *Science* 374.6566, pp. 423–431.

Willem, L., F. Verelst, J. Bilcke, N. Hens, and P. Beutels (2017). "Lessons from a decade of individual-based models for infectious disease transmission: a systematic review (2006-2015)". In: *BMC Infectious Diseases* 17.1, p. 612.

Wilson-Aggarwal, J. K., L. Ozella, M. Tizzoni, C. Cattuto, G. J. F. Swan, T. Moundai, M. J. Silk, J. A. Zingeser, and R. A. McDonald (2019). "High-resolution contact networks of free-ranging domestic dogs Canis familiaris and implications for transmission of infection". In: *PLOS Neglected Tropical Diseases* 13.7. Ed. by S. Recuenco, e0007565.

Wilson-Aggarwal, J. K., C. E. Goodwin, T. Moundai, M. K. Sidouin, G. J. Swan, M. Léchenne, and R. A. McDonald (2021). "Spatial and temporal dynamics of space use by free-ranging domestic dogs Canis familiaris in rural Africa". In: *Ecological Applications*.

World Health Organization (WHO) (2018). *WHO Expert Consultation on Rabies. Third report.* Tech. rep. Geneva: World Health Organization, p. 195.

— (2021a). *Weekly epidemiological update on COVID-19 - 13 October 2021*. Tech. rep. Geneva: World Health Organization.

— (n.d.). *Covid-19 cases and deaths by continent*.

— (2021b). *Cross-Country Analysis*.

World Health Organization (WHO), Food and Agriculture Organization of the United Nations (FAO), World Organisation for Animal Health (OIE), and Global Alliance for Rabies Control (GARC) (2018). *Zero by 30: the global strategic plan to end human deaths from dog-mediated rabies by 2030*. Tech. rep. Geneva: World Health Organization, p. 59.

— (2019). *United Against Rabies Collaboration First annual progress report: global strategic plan to end human deaths from dog-mediated rabies by 2030*. Tech. rep. Geneva: World Health Organization, pp. 1–38.

WorldPop (n.d.). *WorldPop project*.

Wymant, C. et al. (2022). "A highly virulent variant of HIV-1 circulating in the Netherlands". In: *Science* 375.6580, pp. 540–545.

Yale, G., M. Lopes, S. Isloor, J. R. Head, S. Mazeri, L. Gamble, K. Dukpa, G. Gongal, and A. D. Gibson (2022). "Review of Oral Rabies Vaccination of Dogs and Its Application in India". In: *Viruses 2022, Vol. 14, Page 155* 14.1, p. 155.

Yang, J., N. F. Müller, R. Bouckaert, B. Xu, and A. J. Drummond (2019). "Bayesian phylodynamics of avian influenza A virus H9N2 in Asia with time-dependent predictors of migration". In: *PLOS Computational Biology* 15.8. Ed. by S. Cobey, e1007189.

Yang, Y. and E. Kenah (2022). "Understanding how fast SARS-CoV-2 variants transmit from household studies". English. In: *The Lancet Infectious Diseases* 22.5, pp. 564–565.

Yang, Z. (1994). "Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods". In: *Journal of Molecular Evolution* 39.3, pp. 306–314.

Yao, H.-W., Y. Yang, K. Liu, X.-L. Li, S.-Q. Zuo, R.-X. Sun, L.-Q. Fang, and W.-C. Cao (2015). "The Spatiotemporal Expansion of Human Rabies and Its Probable Explanation in Mainland China, 2004-2013". In: *PLOS Neglected Tropical Diseases* 9.2. Ed. by C. E. Rupprecht, e0003502.

Yu, J., H. Li, Q. Tang, S. Rayner, N. Han, Z. Guo, H. Liu, J. Adams, W. Fang, X. Tao, S. Wang, and G. Liang (2012). "The Spatial and Temporal Dynamics of Rabies in China". In: *PLoS Neglected Tropical Diseases* 6.5. Ed. by M. S. Carvalho, e1640.

Zhang, J., Z. Jin, G.-Q. Sun, X.-D. Sun, and S. Ruan (2012). "Modeling Seasonal Rabies Epidemics in China". In: *Bulletin of Mathematical Biology* 74.5, pp. 1226–1251.

Zhang, J., Z. Jin, G.-Q. Sun, T. Zhou, and S. Ruan (2011). "Analysis of Rabies in China: Transmission Dynamics and Control". In: *PLoS ONE* 6.7. Ed. by P.-J. Cardona, e20891.

Zhang, J., M. Litvinova, Y. Liang, Y. Wang, W. Wang, S. Zhao, Q. Wu, S. Merler, C. Viboud, A. Vespignani, M. Ajelli, and H. Yu (2020). "Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China". In: *Science* 368.6498, pp. 1481–1486.

Zhang, Y., B. Vrancken, Y. Feng, S. Dellicour, Q. Yang, W. Yang, Y. Zhang, L. Dong, O. G. Pybus, H. Zhang, and H. Tian (2017). "Cross-border spread, lineage displacement and evolutionary rate estimation of rabies virus in Yunnan Province, China". In: *Virology Journal* 14.1, p. 102.

Zhu, Y., C. J. Bloxham, K. D. Hulme, J. E. Sinclair, Z. W. M. Tong, L. E. Steele, E. C. Noye, J. Lu, Y. Xia, K. Y. Chew, J. Pickering, C. Gilks, A. C. Bowen, and K. R. Short (2021). "A Meta-analysis on the Role of Children in Severe Acute Respiratory Syndrome Coronavirus 2 in Household Transmission Clusters". In: *Clinical Infectious Diseases* 72.12, e1146–e1153.

Zinsstag, J., S. Dürr, M. A. Penny, R. Mindekem, F. Roth, S. M. Gonzalez, S. Naissengar, and J. Hattendorf (2009). "Transmission dynamics and economics of rabies control in dogs and humans in an African city". In: *Proceedings of the National Academy of Sciences* 106.35, pp. 14996–15001.

Zinsstag, J. et al. (2017). "Vaccination of dogs in an African city interrupts rabies transmission and reduces human exposure". In: *Science Translational Medicine* 9.421.

Zuckerkandl, E. and L. Pauling (1965). "Evolutionary Divergence and Convergence in Proteins". en. In: *Evolving Genes and Proteins*. Ed. by V. Bryson and H. J. Vogel. Academic Press, pp. 97–166.

Zuckerkandl, E. and L. Pauling (1962). *Molecular disease, evolution and genetic heterogeneity*. New York.

PLOS NEGLECTED TROPICAL DISEASES

RESEARCH ARTICLE

Mathematical modelling and phylodynamics for the study of dog rabies dynamics and control: A scoping review

Maylis Layan[1,2]*, **Simon Dellicour**[3,4], **Guy Baele**[4], **Simon Cauchemez**[1‡], **Hervé Bourhy**[5,6‡]

1 Mathematical Modelling of Infectious Diseases Unit, Institut Pasteur, UMR2000, CNRS, Paris, France, **2** Sorbonne Université, Paris, France, **3** Spatial Epidemiology Lab (SpELL), Université Libre de Bruxelles, Bruxelles, Belgium, **4** Department of Microbiology, Immunology and Transplantation, Rega Institute, KU Leuven, Leuven, Belgium, **5** Lyssavirus Epidemiology and Neuropathology Unit, Institut Pasteur, Paris, France, **6** WHO Collaborating Centre for Reference and Research on Rabies, Institut Pasteur, Paris, France

‡ These authors are joint senior contribution on this work.
* maylis.layan@pasteur.fr

Citation: Layan M, Dellicour S, Baele G, Cauchemez S, Bourhy H (2021) Mathematical modelling and phylodynamics for the study of dog rabies dynamics and control: A scoping review. PLoS Negl Trop Dis 15(5): e0009449. https://doi.org/10.1371/journal.pntd.0009449

Editor: Monique Léchenne, Environment and Sustainability Institute, UNITED KINGDOM

Received: December 17, 2020

Accepted: May 5, 2021

Published: May 27, 2021

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files. Supporting Information files are also available online at https://mlayan.github.io/RabiesScopingReview/ and archived on the open-access repository Zenodo (DOI: 10.5281/zenodo.4743553ss).

Funding: SD is supported by the *Fonds National de la Recherche Scientifique* (FNRS, Belgium). The funder had no role in study design, data collection

Abstract

Background

Rabies is a fatal yet vaccine-preventable disease. In the last two decades, domestic dog populations have been shown to constitute the predominant reservoir of rabies in developing countries, causing 99% of human rabies cases. Despite substantial control efforts, dog rabies is still widely endemic and is spreading across previously rabies-free areas. Developing a detailed understanding of dog rabies dynamics and the impact of vaccination is essential to optimize existing control strategies and developing new ones. In this scoping review, we aimed at disentangling the respective contributions of mathematical models and phylodynamic approaches to advancing the understanding of rabies dynamics and control in domestic dog populations. We also addressed the methodological limitations of both approaches and the remaining issues related to studying rabies spread and how this could be applied to rabies control.

Methodology/principal findings

We reviewed how mathematical modelling of disease dynamics and phylodynamics have been developed and used to characterize dog rabies dynamics and control. Through a detailed search of the PubMed, Web of Science, and Scopus databases, we identified a total of $n = 59$ relevant studies using mathematical models ($n = 30$), phylodynamic inference ($n = 22$) and interdisciplinary approaches ($n = 7$). We found that despite often relying on scarce rabies epidemiological data, mathematical models investigated multiple aspects of rabies dynamics and control. These models confirmed the overwhelming efficacy of massive dog vaccination campaigns in all settings and unraveled the role of dog population structure and frequent introductions in dog rabies maintenance. Phylodynamic approaches successfully disentangled the evolutionary and environmental determinants of rabies dispersal and consistently reported support for the role of reintroduction events and human-

and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

mediated transportation over long distances in the maintenance of rabies in endemic areas. Potential biases in data collection still need to be properly accounted for in most of these analyses. Finally, interdisciplinary studies were determined to provide the most comprehensive assessments through hypothesis generation and testing. They also represent new avenues, especially concerning the reconstruction of local transmission chains or clusters through data integration.

Conclusions/significance

Despite advances in rabies knowledge, substantial uncertainty remains regarding the mechanisms of local spread, the role of wildlife in dog rabies maintenance, and the impact of community behavior on the efficacy of control strategies including vaccination of dogs. Future integrative approaches that use phylodynamic analyses and mechanistic models within a single framework could take full advantage of not only viral sequences but also additional epidemiological information as well as dog ecology data to refine our understanding of rabies spread and control. This would represent a significant improvement on past studies and a promising opportunity for canine rabies research in the frame of the One Health concept that aims to achieve better public health outcomes through cross-sector collaboration.

Author summary

Rabies is a fatal yet vaccine-preventable zoonotic disease. Domestic dog populations are known to constitute the predominant reservoir of rabies in developing countries, causing 99% of human rabies cases. Despite valuable efforts to control rabies spread, the last two decades have seen only a limited reduction in the global rabies disease burden. Dog rabies is still endemic in Africa, Asia, and the Middle East, in part due to remaining knowledge gaps on dog rabies dynamics. We conducted an in-depth review of phylodynamic approaches and mathematical models used to study the spread and control of rabies in domestic dogs. We identified 59 relevant studies which used mathematical models (30), phylodynamic approaches (22), or interdisciplinary approaches (7). Our study revealed that these approaches disentangled different aspects of rabies spread and control. Mathematical models support the role of dog population heterogeneity as a key driver of rabies spread, and the overwhelming efficacy of dog vaccination campaigns to control rabies. Phylodynamic studies confirm the role of frequent reintroduction events and human-mediated transportation over long distances in rabies maintenance. Interdisciplinary studies represent a powerful tool to generate and test hypotheses on rabies spread. Finally, we identified new avenues which represent a promising opportunity for canine rabies research to achieve more impactful public health outcomes.

Introduction

Background

Rabies is a viral zoonosis affecting the central nervous system of mammals that is almost always fatal to humans. Domestic dogs represent the main reservoir of rabies virus (RABV) worldwide. They are responsible for 99% of human rabies cases [1]. In-depth understanding of

dog ecology and host-pathogen interactions is necessary to characterize rabies dynamics and design appropriate control measures. Rabies is a vaccine-preventable disease in both human and canine populations, and dog vaccination is the most cost-effective control measure [2]. Strong evidence is available for the efficacy of dog rabies elimination programs in endemic areas [3–7], notably in South America where massive dog vaccination campaigns in the 1980s alleviated the burden of canine rabies. Regardless, there has been only little improvement of the global burden since the successes in South America. Dog rabies is still endemic in Africa, Asia, and the Middle East [8,9].

In 2015, the World Health Organization (WHO), the Global Alliance for Rabies Control (GARC), the World Organization for Animal Health (OIE) and the Food and Agriculture Organization of the United Nations (FAO) launched a comprehensive framework targeting the global elimination of dog-mediated human rabies by 2030 [10]. Effective One Health interventions such as the improvement of the current prophylaxis in both humans [11,12] and dogs should enable reaching this goal.

Despite valuable efforts in several endemic countries [9,13,14], control strategies have not stopped rabies from circulating due to inadequate political, economic, and social responses. Weak interest from veterinary services, lack of sustainable resources and political neglect [15] prevent most endemic countries to reach the 70% vaccination coverage recommended by the WHO[9]. Moreover, rabies infections continue to spread, notably in previously rabies-free areas in countries such as Indonesia [16–18] and the Philippines [19,20]. In this resource-limited context, in-depth knowledge of the mechanisms underlying rabies dynamics (environmental drivers of spread, impact of dog density, impact of dog behavior, etc.) would be a key asset to limiting the spread of this vaccine-preventable disease, notably by aiding to design more effective vaccination campaigns that are robust to resurgence in the long-term. The development of novel methodologies to better understand rabies epidemiology and transmission dynamics therefore constitutes a promising avenue of research.

Objectives

In this scoping review, we focused on the insights of two quantitative approaches applied to the study of rabies: mathematical modelling of infectious diseases and phylodynamics. The former is a field of research that exploits epidemiological data to unravel the spread of diseases in populations, assess the impact of interventions, support policy making, and optimize control strategies. The latter studies the interactions between epidemiological, immunological, and evolutionary processes from the analysis of viral genetic sequence data [21]. Within phylodynamics, phylogeographic inference specifically aims at reconstructing the dispersal history and dynamics of viral lineages in space and time. Here, we assessed the uses and respective contributions of both approaches, as well as their limitations and the remaining knowledge gaps concerning rabies dispersal and control in domestic dog populations.

Methods

Search strategy

This review follows the guidelines of the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) statement for scoping reviews [22]. In this review, we screened PubMed, Web of Science and Scopus databases on the 2nd of June, 2020 using the following combination of terms ["rabies" AND ("dog" OR "canine") AND ("modelling" OR "modeling" OR "phylogeography" OR "phylodynamics") AND "dynamics"] along with the "all fields" option and without restriction on publication year. The "all fields" option enabled to apply the search terms for their appearance in the title, abstract

and keywords. Only English-written papers published in scientific journals were considered. All data were searched and screened by the same researcher (ML). The search strategy identified 65, 94 and 768 publications in PubMed, Web of Science and Scopus databases respectively, which corresponded to 797 unique records. In addition, references of selected publications were screened manually, leading to the identification and inclusion of two additional studies [23,24]. Finally, the paper of Colombi et al. [25], which was not identified in the databases nor in the references, was also included (Fig 1).

Selection of studies

In total, 797 records were included and processed manually in a multi-stage procedure. At each selection step, a conservative approach was taken to ensure the best sensitivity level. Firstly, studies were selected based on their title using the following inclusion criteria: mathematical models of dog and human rabies assessing the impact of control strategies, the risk of rabies importation, the drivers of rabies spread or models estimating epidemiological parameters, cost-effectiveness studies, phylodynamic studies including RABV isolated from dogs, and broad studies on new phylodynamic or mathematical models. Indeed, rabies has often been used as a model disease in phylodynamics and mathematical modelling, and a reference to rabies might not appear directly in the title or the abstract. The following exclusion criteria were used: reviews, studies strictly on wildlife rabies, dog ecology and population dynamics, conservation biology, and evolutionary analyses for diagnostic purposes. Secondly, studies were selected based on their abstract with a refined set of exclusion criteria to exclude statistical

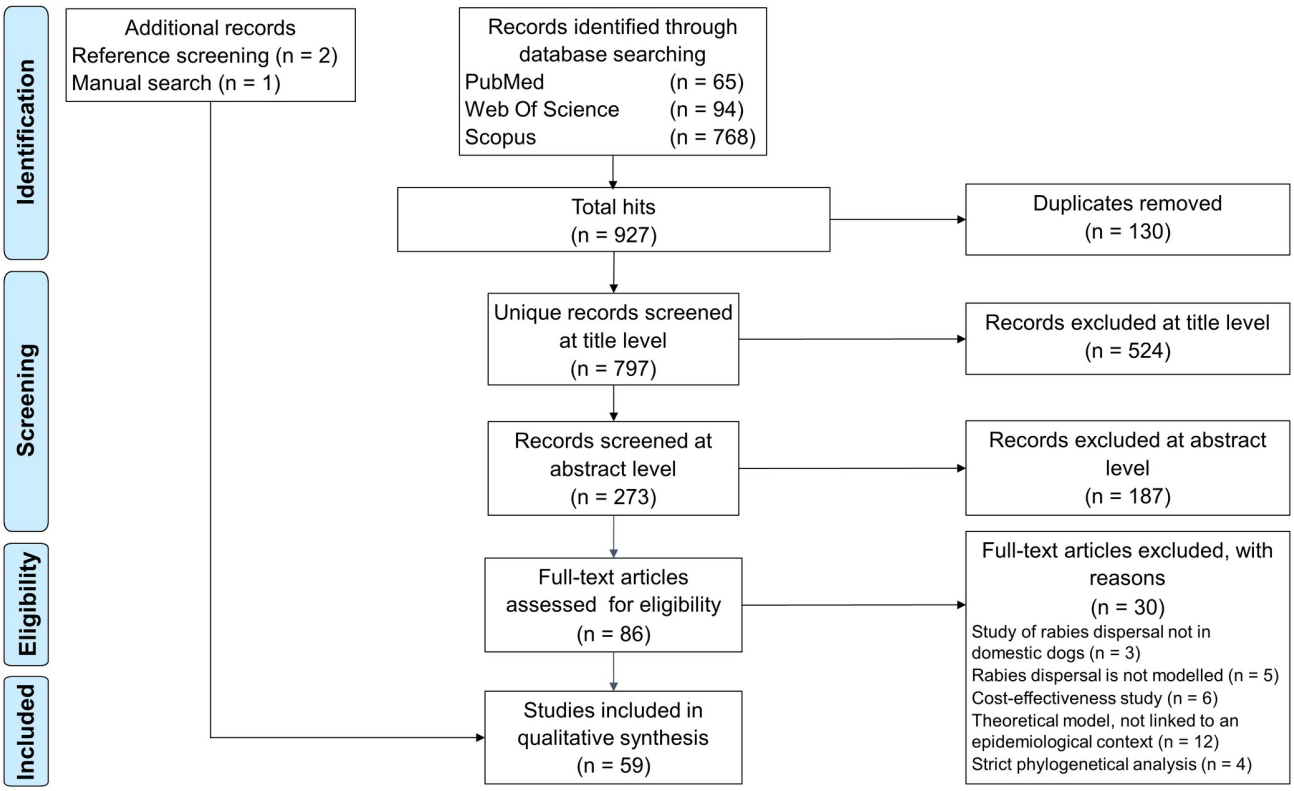


Fig 1. PRISMA-ScR Flow Diagram showing the number of identified and selected records along the multi-stage selection process. Scopus accounted for most of the records as it retrieved 71% ($n = 46$) of PubMed records and 79% ($n = 74$) of Web of Science records.

https://doi.org/10.1371/journal.pntd.0009449.g001

analyses of epidemiological data, cost-effectiveness studies with no focus on rabies dynamics, experimental rabies cross-species transmission which did not incorporate a modelling aspect and studies on the evolutionary processes of RABV. Finally, studies went through a full-text reading step to verify that their content matched our selection criteria. At this step, theoretical models which were not grounded in a specific epidemiological context were excluded (Fig 1).

Data extraction and analysis

Selected studies were classified into three categories based on their methodology: mathematical models, phylodynamic and interdisciplinary studies. Most phylodynamic studies identified in this review correspond to phylogeographic analyses, where the main focus is on inferring the spread of a pathogen over time using location data associated with the available sequence data. The interdisciplinary category covers studies either integrating epidemiological and genetic data in a unified modelling framework or mixing modelling approaches with phylodynamics. Data were systematically charted in an Excel spreadsheet designed to retrieve: *i*) the main modelling strategy with its assumptions; *ii*) the data source; *iii)* remarks about potential bias of the data in relation to the underlying evolutionary and epidemiological processes; *iv*) the qualitative and quantitative results concerning the dynamics of dog rabies; and *v*) if performed, the sensitivity analysis determining the robustness of the methodology to parameter values or potential biases.

Results

General characteristics of selected studies

Our selection procedure identified 59 studies that meet our selection criteria with 30 mathematical models [16,23–51], 22 phylodynamic studies [17,19,52–71], and 7 interdisciplinary studies [20,72–77], all published between 1996 and 2020 (Figs 1 and 2A and 2B). Mathematical models were first published followed by phylodynamic and interdisciplinary studies (Fig 2B). This timeline can be explained by the recent developments of Bayesian phylodynamic, and in particular phylogeographic, models in BEAST [78–80]. Africa and Asia are the most studied continents in the three methodological categories, while China accounts for most of the Asian studies (Fig 2C). Oceania is not represented in the interdisciplinary and phylodynamic categories since it is a rabies-free area (Fig 2A).

Topics addressed by the studies

Phylodynamic studies are homogeneous in terms of methodologies (essentially phylogeographic studies) and research goals. They predominantly focus on unraveling the dispersal dynamics of rabies at the regional and country levels ($n = 16$) [17,19,52–58,61,63,64,67,68,70,71]. In four of them, the authors deciphered the role of lineage introduction in rabies maintenance or emergence [59,60,62,66]. In recent years, researchers have been trying to identify external factors impacting the spatial dynamics of RABV spread ($n = 5$) [63,64,68,69,71] (Fig 2D and S1 Table). Contrary to phylodynamic studies, the modelling category gathers a diverse panel of models with aims that cover the implementation of new mathematical methodologies ($n = 2$) [42,46], the characterization of rabies dynamics ($n = 11$) [26,27,31,32,40,41,44,47–49,51], the identification of factors driving the resurgence or maintenance of rabies ($n = 9$) [16,23,25,33–35,37,38,43], the assessment of control strategies efficacy ($n = 18$) [16,23,24,27–29,31,33–36,42–45,49–51], the risk assessment of rabies introduction and the evaluation of outbreak preparedness in rabies-free areas ($n = 3$) [30,36,42], and cost-effectiveness studies ($n = 2$) [39,48] (Fig 2D and S1 Table). Finally, interdisciplinary studies mainly focused on

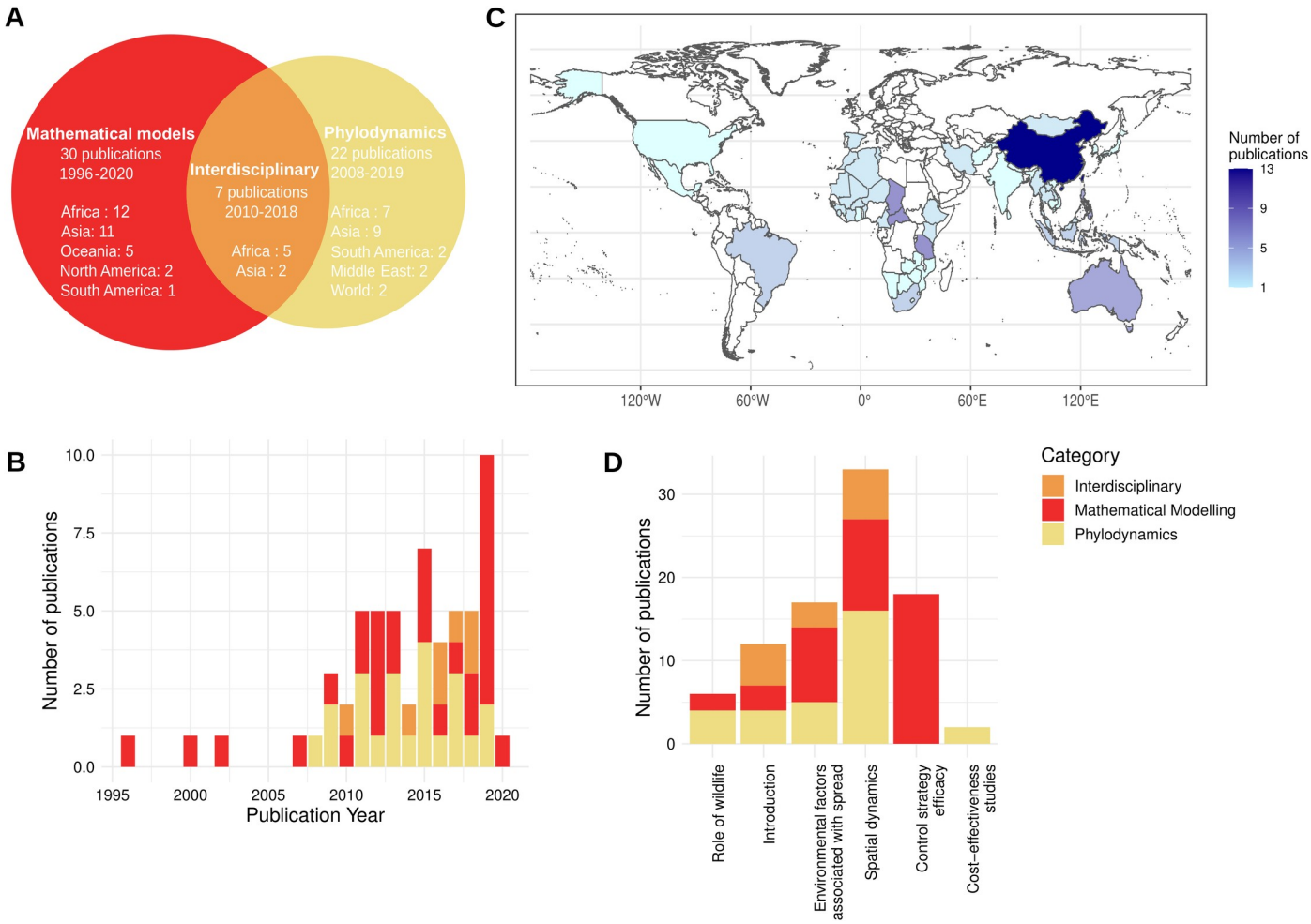


Fig 2. General characteristics of the selected dog rabies studies. (A) Classification of the included publications with the total number of studies, the publication time span, and the number of publications per continent of study. Asia and Africa account for up to 78% of the included studies. (B) Number of publications per year and per methodological category. Mathematical models were the first studies to be published followed by phylodynamic and interdisciplinary studies. (C) Number of publications per country of study. Each publication was attributed to one or multiple countries based on the origin of the RABV genetic sequences, rabid case data or dog ecology data. For phylodynamic studies, countries were not considered if their genetic data were included only in regular phylogenetic tree reconstructions. Similarly, two studies which described rabies dynamics at the global scale [52,65] were not considered in this figure. In our collected records, China accounts for most Asian studies. Spain appears on the map because Ceuta and Melilla, which are Spanish enclaves in North Africa, are represented in two datasets of RABV genetic sequences [68,72]. (D) Number of studies per topic and methodological category. The World Bank, https://datacatalog.worldbank.org/dataset/world-bank-official-boundaries, CC-BY 4.0.

https://doi.org/10.1371/journal.pntd.0009449.g002

rabies dynamics in endemic areas ($n = 6$) [20,72–74,76,77] and the identification of environmental factors influencing rabies spread and maintenance such as recurrent reintroductions ($n = 3$) [72,75,77]. Two of these used genetic and epidemiological data of dog rabies in a unified modelling approach [73,76], whereas the others analyzed sequences through regular phylogenetic approaches and completed their analysis with a mathematical model [20,72,74,75,77] (Fig 2D and S1 Table).

Potential sources of bias in the data

Data source (active/passive surveillance), resolution (number and length of RABV sequences, incidence per country/region, etc.) and representativity influence the level of evidence of the

studies on the underlying epidemiological and evolutionary processes. In particular, recorded cases collected through passive surveillance systems are expected to underestimate the disease burden and to be potentially spatiotemporally biased [8,81]. Similarly, genetic sequences collected from publicly available databases such as GenBank often lack precise metadata (e.g., sampling time and location) and/or are of short length.

In our text corpus of phylodynamic and interdisciplinary studies, passive surveillance systems and GenBank represent the main sources of RABV genetic sequence data (S2–S4 Tables). By combining these two data sources, researchers have generally managed to increase the spatiotemporal coverage of their dataset. This however does not guarantee a good representativity of the epidemic process. Active surveillance was mostly used to collect dog specimens from animal markets in China ($n = 2$) [58,60] and thorough contact tracing after biting events in China and Tanzania ($n = 2$) [63,66]. On average, the datasets analyzed in these studies contained 183 sequences spanning from approximatively 3% to 100% of the RABV genome length. Short sequences containing the N gene constitute the most common type of data. They are less informative than whole genomes which were only generated and analyzed in recent years across four studies [63,65,69,71] (S2 Table).

In studies from the modelling and interdisciplinary categories, authors generally simulated rabies epidemics ($n = 24$) [20,23,25,28–36,38,40,42–47,49–51,72], and thus predominantly relied on publicly available estimates of the natural history of rabies, dog demographics and dog ecology (S3 and S4 Tables). When models were fitted to incidence data ($n = 13$) [16,24,26,27,37,39,41,48,73–77], human and/or dog case data from passive surveillance systems were used, or bite incidence data from thorough active surveillance. In general, there was a lack of data on dog rabies cases (available in 10 studies; [16,24,26,27,37,48,73–77]) and estimates on dog demographics and ecology integrating the local specificities of host ecology were available in only seven studies [27,37,39,41,48,75,77]. Access to local data is crucial since differences in rabies spread [27] and dog carrying capacities [39] were estimated between areas of the same country. We would expect these differences to be more pronounced across different countries. To overcome the lack of epidemiological data on dog rabies, one study used serological data (from vaccination campaigns) to model the dynamics of rabies [46], and another study [36] based its analyses on historical records in Japan from the 1950s. Similarly, most Australian studies [30,42–44] took the perspective of dog ecology data since Australia is free of rabies. This way, the authors explored the impact of dog population structure and dog roaming behavior on rabies dynamics.

Description of the models

In studies using phylodynamic approaches, the geographical dispersal of rabies was studied using either parsimony ($n = 4$) [52,54,55,58], Bayesian discrete phylogeography ($n = 18$) [17,19,20,53,56,57,59,60,62–64,66,67,70–72,77,82], or Bayesian continuous phylogeography ($n = 6$) [61,68,69,71,74,77] (S2–S4 Tables). All Bayesian phylogeographic studies were carried out in BEAST 1 [79] with discrete trait analysis (DTA) to perform a phylogeographic reconstruction based on discrete/discretized sampling locations (e.g. provinces or countries) or with continuous trait analysis to perform a phylogeographic reconstruction based on spatially-explicit sampling location data (latitude and longitude coordinates). Several methodologies take advantage of such phylogeographic inferences to investigate the impact of external factors on the dispersal of viruses: a generalized linear model (GLM) extension of DTA developed by Lemey et al. [83] to test predictors of dispersal transition frequencies among discrete locations which was implemented by Brunker et al. [69]; and *post hoc* statistical approaches developed by Dellicour et al. [71,84,85] to investigate the impact of environmental factors on the dispersal

velocity, direction, or frequency of viral lineages in continuous phylogeographic frameworks which were applied in four rabies studies [68,69,71,77]. Finally, Zinsstag et al. [75] were the only authors to implement a birth-death model in BEAST 2 [80] to reconstruct the effective reproduction ratio (R) along vaccination campaigns and compare it to estimates obtained with a modelling approach (S4 Table).

Compared to phylodynamics, mathematical models display a large diversity of specifications and parametrizations. Compartmental models (n = 18) [20,23,24,26,27,33,34,39–41,45–49,51,75,77] are the most represented models, followed by agent-based (n = 8) [16,30,31,35,36,42–44] and metapopulation (n = 5) [25,28,32,37,50] models. Other model types such as network models or branching processes are also represented [29,38,73,74,76] (S3 and S4 Tables). The development of new dog rabies models builds upon the literature since 15 models out of the 37 identified were adapted from previously published dog rabies or wildlife rabies models (S3 and S4 Tables). This is the case notably for compartmental models which correspond to the simplest models of rabies dynamics. Metapopulation, agent-based, and other model types are more complex, in that these approaches often integrate spatial dynamics of dog rabies [25,30,32,35–38,42,43,72,73,76].

Population structure can be integrated in any modelling framework under the form of contact heterogeneity, age-structured populations, roaming behavior, or individual heterogeneity. In compartmental models, population structure is integrated either as a set of strata (stray dogs, owned free-roaming dogs, owned confined dogs) interacting together [33], or by specifying a structured next-generation matrix from which R is generally derived [34]. Such models are also referred to as multi-host models and may integrate other hosts: humans [32,39,40,48,49,51,86], cattle [39], wildlife [27,41]. In agent-based and network models, population structure is defined at the individual level using spatial kernels [16,25,30,31,36,42], individual contact rates [30,35,44], vaccination status [30,36], life span, infectious period [16,31,44], etc.

Sensitivity analyses

Sensitivity analyses are commonly used to assess the robustness of inference to both data representativity and model specifications, and to identify the most influential parameters on specific model outputs. In our text corpus, no sensitivity analyses were found to be carried out in phylodynamic studies which can be attributed to the relatively small number of sequences analyzed in those studies. In contrast, sensitivity analyses were commonly performed in mathematical models, either to unravel the key parameters influencing rabies dynamics or to verify the robustness of the results to model assumptions. Dog ecology parameters such as birth rate and carrying capacities are often reported as key parameters on rabies dynamics predictions although they are not estimated using local data. Transmission rates are also determinant in model predictions (S3 Table). In spatially explicit studies, mobility parameters also have a strong impact on model inferences. Finally, the impact of under-reporting was tested only in interdisciplinary studies, two of which reported a strong impact of the reporting rate on model inference [20,76] whereas the other two were robust to a change in this parameter [74,75] (S4 Table).

Insights into dog rabies dynamics and its drivers from phylodynamic and modelling studies

Phylogeographic analyses have aimed to unravel the spatial dynamics of dog rabies at the global and regional scales and showed that dog RABV lineages cluster spatially at the global scale, except for one lineage, referred to as the cosmopolitan lineage, which is largely

distributed across the world [52]. At the regional and country scales, there is co-circulation of dog-related lineages, notably in China [55,58,64,66,70], in the Middle East [62,71], as well as in Western and Central Africa [54]. However, each lineage exhibits a strong geographical structure. In the case of country-specific lineages, various studies have suggested that transboundary movements are not a major force of rabies dispersal [19,53,54,59,60,68]. All study categories unraveled the role of human-mediated movements in rabies spread. Overall, phylogeographic analyses provided evidence for the effect of anthropogenic factors: major roads are associated with rabies dispersal in North Africa [72], and RABV lineages tended to preferentially circulate within populated areas in North Africa [68] and the Middle East [71]. Other factors are associated with rabies spread in Yunnan (China, Tables 1 and S5). These results may reflect the intimate link between rabies dynamics, host ecology and dog-human interactions. Mathematical models highlighted the short length of canine rabies transmission chains [31,73,76] and unraveled the importance of long-range human movements in disease spread [25,32]. Finally, interdisciplinary approaches highlighted the crucial role of long-distance transmission events likely due to humans in rabies dynamics in North Africa [72] and also showed that main roads act as barriers to dog rabies dispersal in an urban setting in Africa [35].

Phylodynamic studies showed that introduction through infected dog movement is the major force of rabies spread towards disease-free areas, as Indonesia [16–18] and the Philippines [19,20] have recently experienced, and also represents a driver of rabies spread in endemic areas where frequent reintroductions counteract local rabies elimination after vaccination campaigns [74,75]. In these settings, phylodynamic analysis constitutes a powerful tool to confirm introduction events [19,56,59,72,74,75]. Multiple mathematical models have also shown that frequent reintroductions drive rabies persistence in endemic areas [31,37,73,76].

Population structure constitutes another driving force of rabies maintenance as explored in simulation studies integrating dog ecology data in Australian [30,42–44], Japanese [36], Tanzania [28,50] and Chadian [35] settings. Rabies-induced behavioral changes were shown to contribute to rabies persistence in small dog populations [44] as well as differential roaming behavior, contact rates between dog strata and the structure of contact networks [30,34–36,44].

The contribution of wildlife to canine rabies spread and maintenance is rarely addressed in phylodynamic studies because viruses isolated from wildlife specimens often correspond to dog-related lineages [19,56,64,66,70] or because of insufficient sampling efforts when it comes to wildlife [58] (S1 Table). Nevertheless, specific RABV lineages were shown to circulate both in wildlife and domestic dogs in the Middle East and Tanzania with complex interspecies transmissions [62,65,69,71]. A phylodynamic study at the global scale showed that host shifts from dogs to wildlife with adaptation to the new host were common in RABV history [65], which may explain why different lineages circulate in dogs and wild foxes in Brazil [61], in dogs and ferret badgers in Asia [65] and in dogs and mongooses in South Africa [65] with rare interspecies transmission events. By incorporating direct interspecies transmission, mathematical modeling studies showed that dog population contributes to sustained rabies circulation in wildlife instead of the other way around [27,41]. Similarly, the proximity to wildlife was shown to not impact rabies spread in dogs in the model of Beyer et al. [28].

Finally, mathematical models and phylodynamics provide convenient estimates of a range of parameters on rabies dispersal dynamics (lineage dispersal velocities, diffusion coefficients; Table 1), rabies evolutionary processes and dog ecology. For example, the evolutionary rate was homogeneously estimated to be between 1×10^{-4} and 5×10^{-4} substitutions per site per year across RABV genes and lineages, except for the Asian lineage which is estimated to evolve faster (Fig 3A). The time to the most recent common ancestor (TMRCA) is also frequently

Table 1. Estimated parameters in phylodynamic models.

Location	Sampling window	Viral lineages	RABV sequence	Migration rate (migrations. year^{-1}, 95% HPD)	Velocity[a] (km.year^{-1}, 95% HPD)	Diffusion coefficient (km^2.year^{-1}, 95% HPD)	Factors facilitating viral spread[b]	Factors impeding viral spread[b]	Reference
Bangui, Central African Republic	1986–2012	Africa 1 Africa 2	N P M G intergenic G-L	-	v = 0.9 (0.65–1.2)	-	-	-	Bourhy et al., 2016 [74]
Serengeti district, Tanzania	2004–2013	Africa 1b	Whole-genome	-	v = 4.46 (3.22–5.88) Coefficient of variation M = 3.10	-	Dog presence	Elevation Rivers	Brunker et al., 2018 [69]
Yunnan province, China	2008–2015	SEA-1 SEA-2 SEA-3	N G	-	v = 57.5 (39.2–85.1) $v_{weighted}$ = 23.4 (2.4–32.6)	D = 1733 (1082–2928) $D_{weighted}$ = 1064 (116–1638)	Forest coverage (but with a tendency to spread towards areas associated with relatively low forest coverage)	-	Tian et al., 2018 [77]
North and Northeast regions, Brazil	2002–2005	-	N	-	$v_{overall}$ = 12.88[c] v_{dogs} = 30.5[c] $v_{\textit{cerdocyon thous}}$ = 9.0[c]	-	-	-	Carnieli et al., 2013 [61]
Algeria	2001–2008	Africa 1	N P intergenic G-L	-	$v_{great\ circle\ distances}$ = 26 (18–34) $v_{road\ distances}$ = 33 (23–43)	-	Major roads	-	Talbi et al., 2010 [72]
Algeria	2001–2008	Africa 1	N P intergenic G-L	-	$v_{wavefront}$ ~ 15[c]	D = 2874 (1900–5420) $D_{weighted}$ = 1305 (1086–1574)	Grasslands Urban areas	Elevation	Dellicour et al., 2017 [68]
Morocco	2004–2008	Africa 1	N P intergenic G-L	-	$v_{great\ circle\ distances}$ = 42 (26–58) $v_{road\ distances}$ = 51 (34–72)	-	Major roads	-	Talbi et al., 2010 [72]
Morocco	2004–2008	Africa 1	N P intergenic G-L	-	$v_{wavefront}$ ~ 22[c]	D = 2874 (1900–5420) $D_{weighted}$ = 1305 (1086–1574)	Grasslands Urban areas	Elevation	Dellicour et al., 2017 [68]
Iran	2008–2015	-	Whole-genome	-	v = 55.5 (38.9–142.4) $v_{weighted}$ = 18.1 (16.3–20.8)	D = 2676 (1935–5066) $D_{weighted}$ = 1643 (1356–2325)	(Tendency to spread towards and preferentially circulate within accessible areas associated with relatively higher human population density)	(Tendency to avoid circulating in barren vegetation areas and to avoid spreading towards grasslands)	Dellicour et al., 2019 [71]

(*Continued*)

Table 1. (Continued)

Location	Sampling window	Viral lineages	RABV sequence	Migration rate (migrations. year^{-1}, 95% HPD)	Velocity[a] (km.year^{-1}, 95% HPD)	Diffusion coefficient (km^2.year^{-1}, 95% HPD)	Factors facilitating viral spread[b]	Factors impeding viral spread[b]	Reference
China	1983–2016	Arctic-like 2 Central Asian 1 SEA-1 SEA-2 SEA-3 SEA-5	N	5.81e-3 (3.92e-3–7.77e-3)	-	-	-	-	Wang et al., 2019 [70]

Abbreviations: HPD, Highest Posterior Density; SEA-1, South-East Asia 1; SEA-2, South-East Asia 2; SEA-3, South-East Asia 3.

The sampling window and the spatial scale of the studies are highly variable. Thus, it is not possible to directly compare the velocity and diffusion coefficients amongst the different study settings.

a Depending on the study, estimates of RABV lineage velocity or diffusivity were obtained by estimating different dispersal statistics. Talbi et al. [72] reconstructed for each branch of the phylogenetic tree the expected number of migrations between two locations using a discrete phylogeographic model. The authors multiplied these estimates by the great-circle distance between the two locations, and thus, obtained the expected distance travelled within the time elapsed on each branch. Carnieli et al. [61], Bourhy et al. [74], Brunker et al. [68], Tian et al. [77], and Dellicour et al. [70] estimated the mean branch velocity using continuous phylogeographic reconstructions. Finally, Dellicour et al. [67] estimated the temporal evolution of the wavefront velocity that corresponds to the distance between the reconstructed epidemic origin and the maximal epidemic wavefront. While the mean branch velocity (v) and diffusion coefficient (D) are estimates of the dispersal velocity and of the diffusion coefficient averaged over all tree branches, respectively, their weighted average counterparts involve a weighting by branch time resulting in lower-variance estimates [70].

b Depending on the study, the impact of environmental factors on dispersal of viral lineages were investigated using different approaches. Talbi et al. [72] simulated random or conditional dispersal of RABV in northern Africa along phylogenetic trees reconstructed by phylogeographic inference and compared simulated dispersal patterns with the observed spread. Brunker et al. [68] parametrized a generalized linear model (GLM) in a discrete phylogeographic framework with resistance distances derived from landscape data between clusters of rabies cases. Dellicour et al. [67] and Tian et al. [77] assessed which environmental factors are associated with RABV velocity using continuous phylogeographic inference and post hoc statistical analyses. Dellicour et al. [70] and Tian et al. [77] also identified factors associated with the direction of spread using phylogeographic reconstructions and subsequent post hoc analyses.

c 95% Highest Posterior Density (HPD) intervals are not specified in the original publications.

https://doi.org/10.1371/journal.pntd.0009449.t001

estimated in phylodynamic studies (S2 Table) which is generally more recent than suggested by historical records. R, the expected number of secondary infections, is often estimated by fitting case data to mathematical models (Fig 3B) or by computing its value based on the choice of parameters value (S6 Table). Its estimate ranges from 0.80 to 3.36 according to the setting but it is generally estimated to be between 1 and 2, corresponding to a low-grade transmission with frequent stochastic extinctions. Other parameters such as the dog-to-dog transmission rate, the introduction rate or the dog carrying capacity are also frequently estimated (S6 Table).

Effective control strategies

Interdisciplinary and modelling studies generally assessed the impact of past or potential control strategies to eliminate dog rabies. The specifications of the explored control strategies depended on the economic situation of the country in which the study was supposed to be performed, as well as the model type. Dog vaccination was the most studied control measure ($n = 28$) [16,23,24,26–28,30–37,39–41,43–51,75,77], whereas culling ($n = 7$) [30,33,42,45,48,49,51], dog confinement or movement ban ($n = 4$) [30,31,36,42], control of dog birth rate ($n = 4$) [40,45,49,86] and community behavior ($n = 1$) [31] were rarely modelled. Culling was shown to be effective in two compartmental model studies [45,51] while

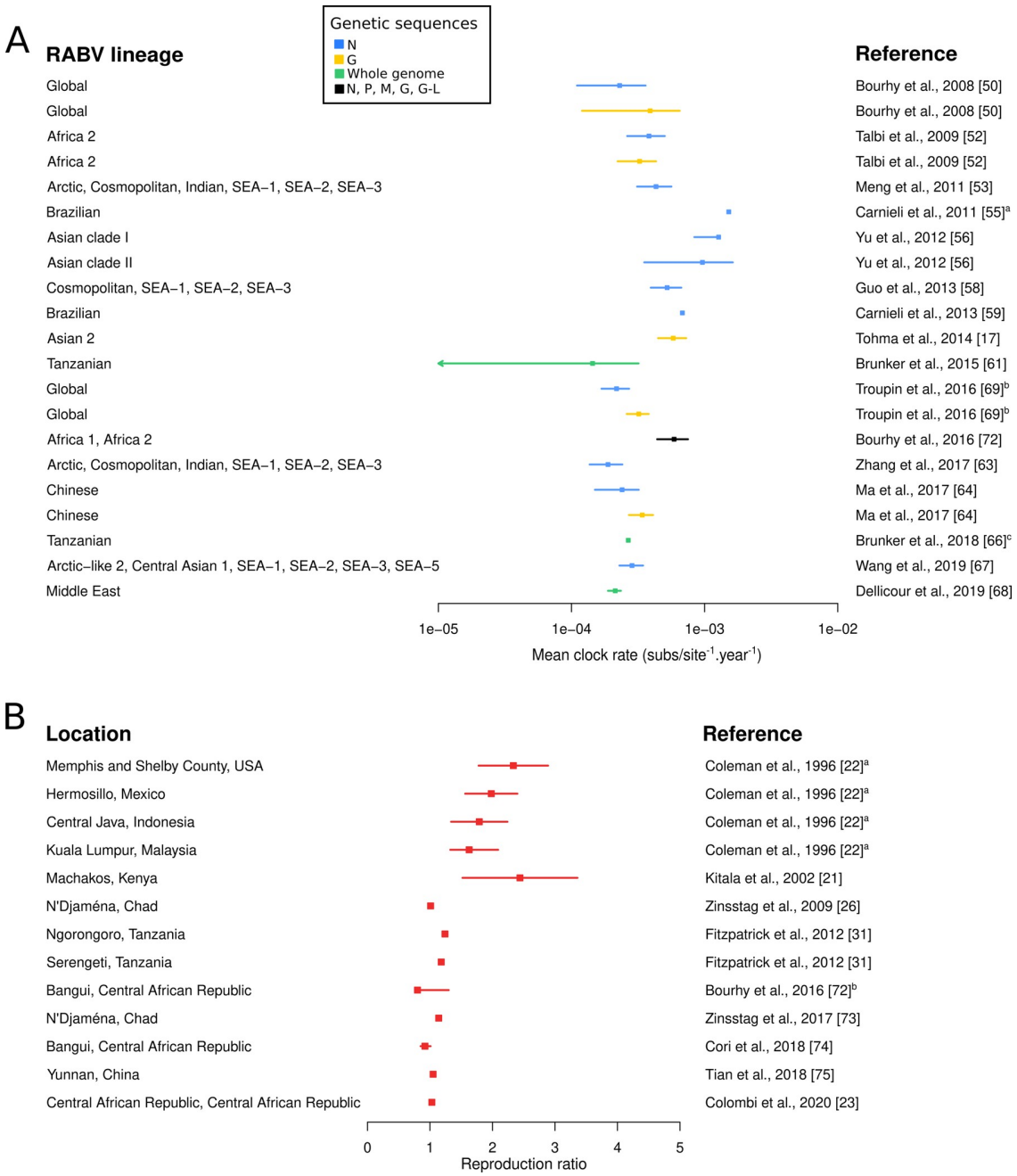


Fig 3. Estimates of the mean evolutionary rate of RABV and the reproduction ratio of canine rabies in the included studies. (A) Bayesian credibility intervals (mean and 95% Highest Posterior Density, HPD) of the mean evolutionary rate of canine RABV per genetic sequence and RABV lineage. [a]The estimate corresponds to the upper bound of the 95% HPD. [b]The dot corresponds to the median and the interval to the 95% HPD interval. cThe 95% HPD was not specified in the original publication. (B) Estimates of the reproduction ratio of dog rabies per control strategy or geographical location. The dot corresponds to the mean and the interval to the 95% confidence interval unless stated otherwise. [a] The interval corresponds to the standard error. [b] The authors estimated the effective reproduction ratio along time. Here, the value range of the median monthly point estimate is depicted.

https://doi.org/10.1371/journal.pntd.0009449.g003

vaccination was generally found to be the most effective strategy. Vaccination coverage strongly depends on the setting: 90% or complete dog vaccination coverages are recommended

in rabies-free areas with high surveillance and control capacities whereas lower coverages associated with complementary strategies are recommended in endemic areas (Table 2). Nevertheless, the efficacy of vaccination strategies is mitigated by new introductions due to neighboring transmission or long-distance movements mediated by humans [25,29,31,37,74,75,87], notably in low vaccinated populations [32]. In this case, reactive vaccination strategies [16] or dog movement bans [25] constitute alternative effective measures. However, Ferguson et al. [31] evaluated the impact of new introductions in vaccinated areas, and concluded that vaccination coverages were robust to rabies introduction in their specific setting. Similarly, Beyer et al. [50] suggested that the spatial structure of dog population had more impact than rabies introduction on the efficacy of vaccination campaigns. In terms of vaccination coverage, successful vaccination campaigns should target homogenous coverage since hidden pockets of rabies transmission might jeopardize control efforts [16,23,29,31]. In terms of campaign frequency, the efficacy of pluriannual compared to annual vaccination campaigns is difficult to evaluate as it results from many factors including the number of vaccination pulses, the time interval between each pulse, dog birth rate and the introduction rate of infectious animals [23,28,45].

Recent studies [28,34–36,42,43] proposed targeting at-risk dog populations, such as explorers and roaming dogs, to improve the efficacy of vaccination campaigns (Table 2). However, the sensitivity analysis of Laager et al. [37] showed that population structure did not impact the efficacy of vaccination strategies. There are conflicting results concerning stray dog vaccination which was either less efficient than owned dog vaccination [51] or dependent on population composition [34].

Several studies also suggested an impact of dog birth rate reduction on the incidence of rabies [23,26,40,41,45,49]. However, the cost and feasibility of dog population management strategies such as sterilization render this unfeasible in many settings [88]. Dog confinement, which is generally spontaneously put in place by local communities during rabies outbreaks, may improve elimination prospects but, when implemented, the level of confinement is not sufficient to reach elimination [25,30,31]. Concerning the rabies burden in humans, some studies recalled the importance of public awareness (Table 2) and proper PEP coverage to reduce the number of human cases, even though it does not impact rabies circulation in dogs [26,35,36,41]. All these findings confirmed and justified the strategic plan that provides a phased, all-inclusive, intersectoral approach to eliminate human deaths from rabies recently launched by United Against Rabies, in a collaboration between four partners: WHO, FAO, OIE and GARC [13].

Discussion

Insights on rabies epidemiology and control

In this review, we assessed the respective contributions of mathematical modelling and phylodynamics to the understanding of rabies spread and control in dog populations. Contrary to phylodynamic studies, mathematical modelling approaches were multi-faceted and mainly addressed the efficacy of control strategies and, less frequently, the drivers of rabies spread. They revealed the crucial role of frequent introductions and the potential role of dog population structure in disease dispersal and maintenance, as well as the overwhelming efficacy of dog vaccination campaigns over other control strategies. Certain studies also estimated key parameters of rabies dynamics and dog ecology, such as dog birth rate or dog carrying capacity. On the other hand, phylodynamic studies mostly focused on the description of viral dynamics at the global, regional, and local scales, and recently tested which environmental factors are impacting RABV spread. These approaches consistently unraveled the occurrence of long-distance movements suspected to be human-mediated and confirmed the role of humans

Table 2. Recommended control strategies in mathematical modelling studies.

Epidemiological context	Recommended control strategy	Specificities of the recommended control strategie	Location	Reference
Introduction in previously rabies-free areas	Reactive dog vaccination	Followed by a 2-year monitoring period		Townsend et al., 2013 [29]
		Until all targeted dogs are vaccinated	Northern Peninsula Area and Elcho Island, Australia	Dürr et al., 2015 [30]
	90% dog vaccination coverage		Northern Australia and New South Wales, Australia	Sparkes et al., 2016 [33]
			Kubin, Saibai and Warraber divisions, Australia	Brookes et al., 2019 [44]
	Targeted dog vaccination campaigns	Vaccination of free-roaming dogs	Northern Peninsula Area, Australia	Hudson et al., 2019 [43]
	Integrated approach	Mandatory dog vaccination Dog owner awareness Dog registration, Capture of free-roaming dogs Quarantine of imported animals	Ibaraki and Hokkaido prefectures, Japan	Kadowaki et al., 2018 [36]
Endemic areas	90% dog vaccination coverage		Lemuna-bilbilo and bishoftu districts, Ethiopia	Beyene et al., 2019 [39]
	75% dog vaccination coverage	Stray dog management	Guangdong, China	Hou et al., 2012 [51]
	70% dog vaccination coverage	Annual vaccination (or biannual vaccination with a 60% coverage)	Machakos district, Kenya	Kitala et al., 2002 [23]
			N'Djaména, Chad	Zinsstag et al., 2009 [48]
		Even coverage	Bali, Indonesia	Townsend et al., 2013 [16]
			Serengeti and Ngorongoro districts, Tanzania	Fitzpatrick et al., 2012 [27]
	≥50% dog vaccination coverage	≥ 50% fertility control coverage		Carroll et al., 2010 [45]
			Sarawak state, Malaysia	Taib et al., 2019 [40]
	Even dog vaccination coverage		Region IV, Philippines	Ferguson et al., 2015 [31]
	Targeted dog vaccination campaigns	Frequent dog vaccination campaigns targeting the reduction in metapopulation risk	Serengeti district, Tanzania	Beyer et al., 2012 [28]
		Stray dog vaccination coverage based on dog population composition		Leung et al., 2017 [34]
		Vaccination based on social and roaming behaviors Public awareness Locally reactive interventions Reporting of 60% of cases by the surveillance system	N'Djaména, Chad	Laager et al., 2018 [35]
	Dog population management	Dog vaccination Public awareness	China	Zhang et al., 2012 [26]
		Massive dog vaccination campaigns in urban areas Dog movement bans	Central African Republic	Colombi et al., 2020 [25]
		Dog vaccination	China	Zhang et al., 2011 [49]

The efficacy of control strategies on dog rabies dynamics has been addressed in only a subset of the currently available mathematical modelling studies. Studies presented in this table compared several control strategies or different dog vaccination coverages on rabies elimination prospects. The optimal control strategy inherently depends on the epidemiological context (endemic or introduction in previously rabies-free areas), the setting (local surveillance and vaccination capacities), the assumptions of the dog rabies model and the control strategies tested by the researchers. Here, we report the strategies recommended by the authors which include quantitative and qualitative criteria such as the estimated impact of public awareness on rabid dog detection and management. Three studies [35,40,51] are not grounded in a specific geographical area. Using simulated scenarios, they test the impact of control strategies according to the time to detection [35], dog population structure [40] and the use of immunocontraceptives [51].

https://doi.org/10.1371/journal.pntd.0009449.t002

in rabies dispersal dynamics in Africa and the Middle East. A third kind of studies either combined phylodynamics and mathematical modelling or presented new models integrating epidemiological and genetic data. In the former approach, hypotheses on rabies spread were generated and tested in the same epidemiological context, and thus, confirmed the impact of introductions and human movements in a low-grade transmission process characterized by small clusters and frequent stochastic extinctions. The latter approaches aimed at reconstructing local transmission chains or clusters, opening new horizons on data integration and the study of rabies (Fig 4). Unfortunately, a large number of endemic countries is still not, or poorly studied. Data collection and/or model formulation are still needed in Russia, and most of Africa, and South-East Asia.

The limitations of our review should be acknowledged. In preliminary analyses, we noticed a high variability in record selection according to the combination of search terms, and certainly due to the ambiguous use of specific terms such as phylodynamics in the literature. Since the studies selected in this review are mainly in line with previous reviews [82,89,90], we argue that we retrieved a large part of the available studies on rabies dynamics and control.

Open questions in rabies epidemiology and control

In this review, we summarized the findings of mathematical modelling and phylodynamics on the factors that impact rabies spread. Nevertheless, the full picture of rabies epidemiology remains unclear. First, the role of dog roaming behavior, and dog contact networks in dog rabies spread should be further investigated. In this review, we identified seven studies [33–35,38,42–44], all situated in Australia and Africa, showing that highly connected dogs or free roaming dogs participate in a large part in the spread of the disease. By specifically targeting this type of dogs, vaccination campaigns could be more effective according to Leung et al., 2017 [34], Laager et al., 2018 [35], and Hudson et al., 2019 [43]. Yet only one study combined contact data with epidemiological data [35]. The ecological and behavioral drivers of rabies circulation in domestic dogs are still not fully understood. If stray dogs do constitute a major driver of rabies dispersal, this could have direct implications on the field concerning stray dog population management for example.

Additionally, the role of wildlife and other host species remains unclear [91]. Even though the circulation of dog rabies seems predominant in dog populations, there are too few studies addressing the dynamics of RABV in wildlife and dogs. Furthermore, the interactions between dogs and other carnivore species are expected to change from location to location. Indeed, the interactions between dog populations and wild carnivores depend on the abundance of wild populations and the frequency of contacts between the dog reservoir and wildlife [27,41]. Better understanding the role of wildlife could also have direct implications on local policies such as increasing public awareness, notably in rural areas or strengthening wildlife surveillance systems for rabies.

At a broader scale, the spatial dynamics of rabies are still poorly understood. Urban areas were first thought to be hubs of rabies transmission but recent studies have shown that rabies could be eliminated temporarily at the city-level through mass dog vaccination campaigns [37,74,75]. These case studies suggest that urban areas are not hubs of rabies transmission but part of the complex spatial heterogeneity of dog ecology and dog movement. By exploring the dynamics of dog rabies circulation in urban, peri urban and rural areas, rabies research could see an improved understanding of rabies ecology. This could have direct implications on the design of vaccination campaigns, by prioritizing vaccination campaigns in hubs of rabies transmission, followed by locations with intermediate and low transmission.

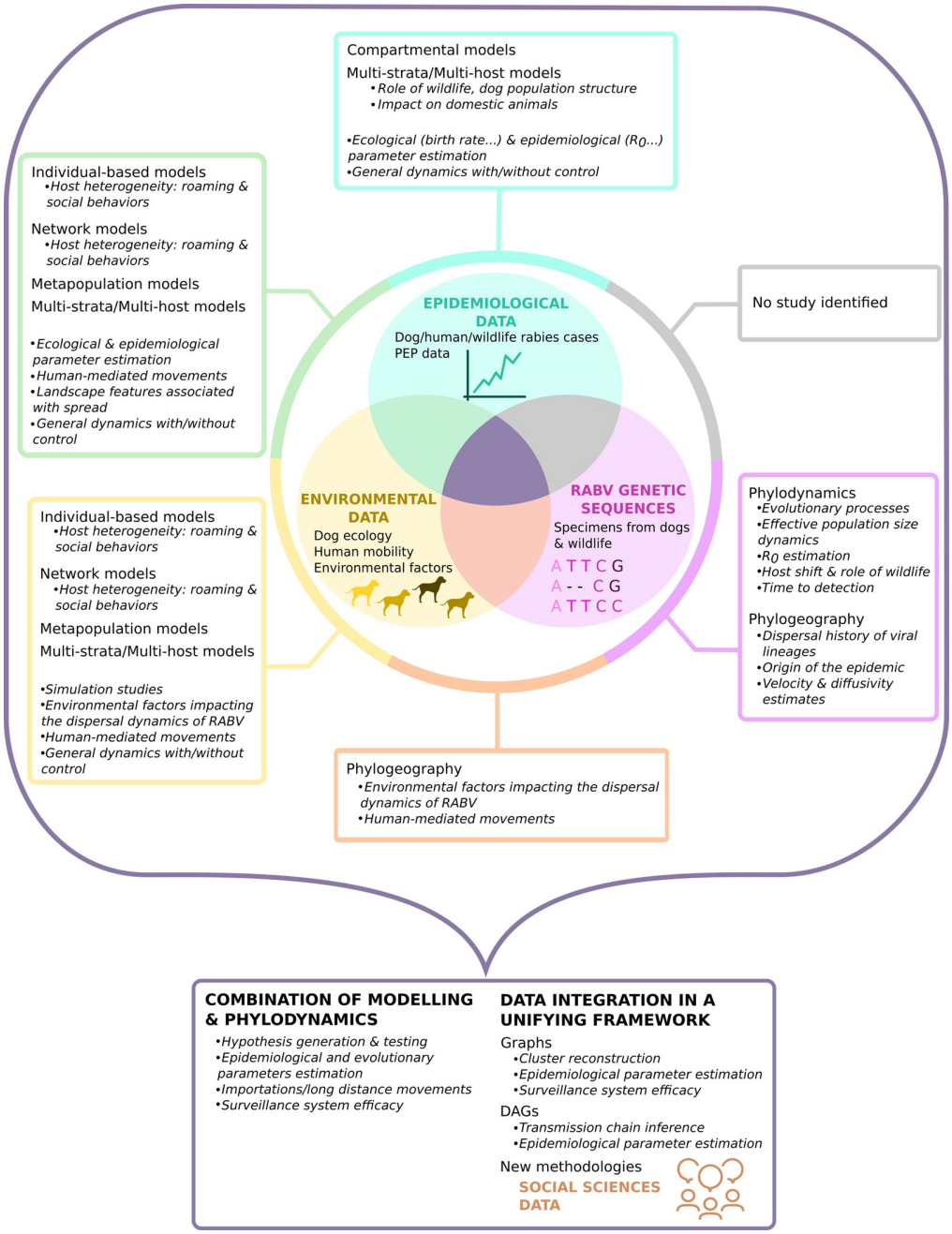


Fig 4. Visual summary of the uses of epidemiological data, environmental data and RABV genetic sequences for the study of rabies dynamics and control. Epidemiological data, environmental data, RABV genetic sequences and social sciences data are highlighted in cyan, yellow, pink, and brown, respectively. The section corresponding to models combining epidemiological data and RABV genetic sequences only is colored in grey since no study that meets this criterion has been identified using our search strategy. Models and their contributions to the understanding of rabies spread and control are detailed in the colored tags. Models using multiple types of data are colored with the intersection color of the corresponding data types. In our text corpus, few studies combined epidemiological, ecological, and genetic data in a unified framework.

https://doi.org/10.1371/journal.pntd.0009449.g004

Finally, there is extensive evidence of the efficacy of dog vaccination to control the spread of rabies in both human and dog populations. We showed in this review that higher coverages

are recommended in rabies-free areas than in endemic areas, however, the practicalities of vaccination campaigns are rarely addressed. As a neglected tropical disease, rabies control programs are designed and deployed in resource-limited contexts. Thus, high, and even intermediate vaccination coverages cannot be achieved at once over a large area. The periodicity, the spatial prioritization, the targeted populations, and the association with other control strategies (dog population management, dog movement ban. . .) are interesting modalities that could be tested in models and could substantially improve resource allocation.

Future directions of mathematical modelling and phylodynamics for rabies research

There is an evident lack of extensive and adequate databases possibly due to restricted data collection, data accessibility and/or data analysis capacity in resource-limited settings [92,93]. This constitutes the main weakness of mathematical modelling and phylodynamic studies that we identified in this review (Table 3). Epidemiological and ecological (census data, population structure, contact networks) data are needed to account for local specificities in terms of modeling interactions between rabies virus (RABV), dog reservoir, domestic animals, wildlife reservoir and human populations. Similarly, there is a need for longer RABV genetic sequences and more thorough sampling to discriminate fine-scale migration events and better characterize the interactions between RABV lineages [63,82,94]. Improving operational data collection is nevertheless challenging. Genomic surveillance relies on laboratory infrastructures, supply chains and expertise, all of which are costly and generally lacking in low- and middle-income countries. New portable sequencing technologies combined with bioinformatics workflows could accelerate capacity building through portability and affordability [94,95]. In parallel, potential sampling bias effects should not be overlooked [53,96] since they may hide a part of disease dynamics such as silent spread in deprived rural areas. Additionally, many endemic countries with high human incidence (Russia, Malaysia, Cambodia, Burma, Niger, Mozambique, etc.) [8] remain largely unstudied using quantitative approaches. This represents an opportunity for data collection, rabies dynamics characterization and control strategy optimization. Besides filling knowledge gaps, improving the availability of epidemiological, ecological, and genetic data offers an opportunity to strengthen countries' veterinary surveillance capacities [15] and enhance the impact assessment of control strategies, two pillars of the 2030 strategic elimination plan.

Other data types such as social sciences data could help identify knowledge gaps and refine control measures to be tested further using mathematical models. For example, there is little quantitative evidence of the impact of community response on the efficacy of control measures [91], although it is key to human rabies prevention [97,98] and it is expected to change over rabies outbreaks and affect rabies dynamics. By bridging the two disciplines, alternative control strategies that are both effective and adapted to community preferences could be designed [99] (Fig 4).

Table 3. Strengths and weaknesses of phylodynamics and mathematical modelling studies identified in this review for the study of rabies.

	Strengths	Weaknesses
Phylodynamics	• Homogeneous methodology which facilitates the comparison of rabies dynamics in different areas • Recent advances in phylogeographical models	• Small datasets and short genetic sequences • Studies generally remain desriptive in terms of environmental factors contributing to rabies spread • Large room to apply other models (such as models implemented in BEAST 2) • The potential impact of reporting biases is barely addressed
Mathematical modelling	• Diversity of models that explore multiple aspects of rabies spread • Assessment of rabies control strategies efficacy • Integration of the waning of vaccine-induced immunity	• Mostly simulation studies, models are rarely fitted to dog rabies data • Mostly deterministic models with strong assumptions (homogeneous mixing of dogs, absence of dog population structure, absence of individual heterogeneity) • No direct comparison of rabies dynamics due to the diversity of models

https://doi.org/10.1371/journal.pntd.0009449.t003

Finally, novel methodologies combining genetic, epidemiological, and environmental data in a comprehensive analysis framework are promising tools for the rabies field. Indeed, the interdisciplinary studies identified in this review exploited the complementarity of genetic and epidemiological information to efficiently generate and test hypotheses on the mechanisms of rabies dynamics [20,72,73,76,77], and the limitations of control strategies [74,75]. These new avenues represent a significant improvement on past studies and a promising opportunity for canine rabies research in the frame of a One Health concept that aims to achieve better public health outcomes through cross-sector collaboration.

Conclusions

In this review, we highlighted the need for more epidemiological, ecological, and genetic data to better characterize rabies dynamics and to get practical information on the drivers of disease transmission. We think that the development of new methodologies integrating genetic and epidemiological data, or the combined use of mathematical models and phylodynamics, constitutes a promising approach that could ultimately contribute to the improvement of the efficacy of control measures including vaccination campaigns and help optimizing the allocation of resources in a context of limited funding.

Supporting information

S1 Table. General characteristics of the included studies.
(XLSX)

S2 Table. Description of the phylogeographic models with an emphasis on data source and potential sources of bias.
(XLSX)

S3 Table. Description of the mathematical models with their key quantitative results.
(XLSX)

S4 Table. Description of the interdisciplinary studies combining phylodynamics and mathematical modelling or integrating epidemiological and genetic data.
(XLSX)

S5 Table. Detailed list of the estimated parameters in phylodynamic models.
(XLSX)

S6 Table. Detailed list of the estimated parameters in mathematical models.
(XLSX)

S1 Text. Rabies epidemiological situation and methodologies implemented to study rabies dispersal and control at the continent level.
(PDF)

S1 PRISMA Checklist. Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) Checklist.
(PDF)

Acknowledgments

We would like to thank Dr Nathanaël Hozé for his informed advice related to scoping reviews and data visualization.

Author Contributions

Conceptualization: Maylis Layan, Simon Dellicour, Guy Baele, Simon Cauchemez, Hervé Bourhy.

Data curation: Maylis Layan.

Formal analysis: Maylis Layan, Simon Dellicour, Guy Baele.

Investigation: Maylis Layan.

Methodology: Maylis Layan, Simon Cauchemez, Hervé Bourhy.

Supervision: Simon Cauchemez, Hervé Bourhy.

Validation: Simon Dellicour, Guy Baele, Simon Cauchemez, Hervé Bourhy.

Visualization: Maylis Layan, Simon Dellicour, Hervé Bourhy.

Writing – original draft: Maylis Layan.

Writing – review & editing: Maylis Layan, Simon Dellicour, Guy Baele, Simon Cauchemez, Hervé Bourhy.

References

1. World Health Organization (WHO). WHO Expert Consultation on Rabies. Third report. World Heal Organ Tech Rep Ser. Geneva; 2018. https://doi.org/10.1007/s13398-014-0173-7.2
2. Anothaisintawee T, Julienne Genuino A, Thavorncharoensap M, Youngkong S, Rattanavipapong W, Meeyai A, et al. Cost-effectiveness modelling studies of all preventive measures against rabies: A systematic review. Vaccine. 2019; 37: A146–A153. https://doi.org/10.1016/j.vaccine.2018.11.071 PMID: 30554795
3. Cleaveland S, Hampson K. Rabies elimination research: Juxtaposing optimism, pragmatism and realism. Proc R Soc B Biol Sci. 2017;284. https://doi.org/10.1098/rspb.2017.1880 PMID: 29263285
4. Lembo T, Hampson K, Kaare MT, Ernest E, Knobel D, Kazwala RR, et al. The feasibility of canine rabies elimination in Africa: Dispelling doubts with data. PLoS Negl Trop Dis. 2010;4. https://doi.org/10.1371/journal.pntd.0000626 PMID: 20186330
5. Aréchiga Ceballos N, Karunaratna D, Aguilar Setién A. Control of canine rabies in developing countries: Key features and animal welfare implications. OIE Rev Sci Tech. 2014; 33: 311–321. https://doi.org/10.20506/rst.33.1.2278 PMID: 25000804
6. Cleaveland S, Beyer H, Hampson K, Haydon D, Lankester F, Lembo T, et al. The changing landscape of rabies epidemiology and control. Onderstepoort J Vet Res. 2014; 81: 1–8. https://doi.org/10.4102/ojvr.v81i2.731 PMID: 25005807
7. Cleaveland S, Thumbi SM, Sambo M, Lugelo A, Lushasi K, Hampson K, et al. Proof of concept of mass dog vaccination for the control and elimination of canine rabies. Rev Sci Tech l'OIE. 2018; 37: 559–568. https://doi.org/10.20506/rst.37.2.2824 PMID: 30747125
8. Hampson K, Coudeville L, Lembo T, Sambo M, Kieffer A, Attlan M, et al. Estimating the Global Burden of Endemic Canine Rabies. PLoS Negl Trop Dis. 2015; 9: e0003709. https://doi.org/10.1371/journal.pntd.0003709 PMID: 25881058
9. Mbilo C, Coetzer A, Bonfoh B, Angot A, Bebay C, Cassama B, et al. Dog rabies control in West and Central Africa: A review. Acta Trop. 2020; 105459. https://doi.org/10.1016/j.actatropica.2020.105459 PMID: 32404295
10. World Health Organization (WHO), Food and Agriculture Organization of the United Nations (FAO), World Organisation for Animal Health (OIE), Global Alliance for Rabies Control (GARC). Zero by 30: the global strategic plan to end human deaths from dog-mediated rabies by 2030. WHO. Geneva; 2018. Available: www.who.int
11. Hampson K, Ventura F, Steenson R, Mancy R, Trotter C, Cooper L, et al. The potential effect of improved provision of rabies post-exposure prophylaxis in Gavi-eligible countries: a modelling study. Lancet Infect Dis. 2019; 19: 102–111. https://doi.org/10.1016/S1473-3099(18)30512-7 PMID: 30472178
12. Hampson K, Abela-Ridder B, Bharti O, Knopf L, Léchenne M, Mindekem R, et al. Modelling to inform prophylaxis regimens to prevent human rabies. Vaccine. 2019; 37: A166–A173. https://doi.org/10.1016/j.vaccine.2018.11.010 PMID: 30528846

13. World Health Organization (WHO), Food and Agriculture Organization of the United Nations (FAO), World Organisation for Animal Health (OIE), Global Alliance for Rabies Control (GARC). United Against Rabies Collaboration First annual progress report: global strategic plan to end human deaths from dog-mediated rabies by 2030. Geneva; 2019.
14. Tiembré I, Broban A, Bénié J, Tetchi M, Druelles S, L'Azou M. Human rabies in Côte d'Ivoire 2014–2016: Results following reinforcements to rabies surveillance. PLoS Negl Trop Dis. 2018;12. https://doi.org/10.1371/journal.pntd.0006649 PMID: 30188890
15. Welburn SC, Coleman PG, Zinsstag J. Rabies control: Could innovative financing break the deadlock? Front Vet Sci. 2017; 4: 1–8. https://doi.org/10.3389/fvets.2017.00001 PMID: 28154816
16. Townsend SE, Sumantra IP, Pudjiatmoko, Bagus GN, Brum E, Cleaveland S, et al. Designing Programs for Eliminating Canine Rabies from Islands: Bali, Indonesia as a Case Study. PLoS Negl Trop Dis. 2013;7. https://doi.org/10.1371/journal.pntd.0002372 PMID: 23991233
17. Dibia IN, Sumiarto B, Susetya H, Putra AAG, Scott-Orr H, Mahardika GN. Phylogeography of the current rabies viruses in Indonesia. J Vet Sci. 2015; 16: 459–466. https://doi.org/10.4142/jvs.2015.16.4.459 PMID: 25643792
18. Mahardika GNK, Dibia N, Budayanti NS, Susilawathi NM, Subrata K, Darwinata AE, et al. Phylogenetic analysis and victim contact tracing of rabies virus from humans and dogs in Bali, Indonesia. Epidemiol Infect. 2014; 142: 1146–1154. https://doi.org/10.1017/S0950268813002021 PMID: 23958065
19. Tohma K, Saito M, Kamigaki T, Tuason LT, Demetria CS, Orbina JRC, et al. Phylogeographic analysis of rabies viruses in the Philippines. Infect Genet Evol. 2014; 23: 86–94. https://doi.org/10.1016/j.meegid.2014.01.026 PMID: 24512808
20. Tohma K, Saito M, Demetria CS, Manalo DL, Quiambao BP, Kamigaki T, et al. Molecular and mathematical modeling analyses of inter-island transmission of rabies into a previously rabies-free island in the Philippines. Infect Genet Evol. 2016; 38: 22–28. https://doi.org/10.1016/j.meegid.2015.12.001 PMID: 26656835
21. Volz EM, Koelle K, Bedford T. Viral Phylodynamics. PLoS Comput Biol. 2013;9. https://doi.org/10.1371/journal.pcbi.1002947 PMID: 23555203
22. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. Ann Intern Med. 2018; 169: 467–473. https://doi.org/10.7326/M18-0850 PMID: 30178033
23. Kitala PM, McDermott JJ, Coleman PG, Dye C. Comparison of vaccination strategies for the control of dog rabies in Machakos District, Kenya. Epidemiol Infect. 2002; 129: 215–222. https://doi.org/10.1017/s0950268802006957 PMID: 12211590
24. Coleman PG, Dye C. Immunization coverage required to prevent outbreaks of dog rabies. Vaccine. 1996; 14: 185–186. https://doi.org/10.1016/0264-410x(95)00197-9 PMID: 8920697
25. Colombi D, Poletto C, Nakouné E, Bourhy H, Colizza V. Long-range movements coupled with heterogeneous incubation period sustain dog rabies at the national scale in Africa. PLoS Negl Trop Dis. 2020; 14: e0008317. https://doi.org/10.1371/journal.pntd.0008317 PMID: 32453756
26. Zhang J, Jin Z, Sun GQ, Sun XD, Ruan S. Modeling Seasonal Rabies Epidemics in China. Bull Math Biol. 2012; 74: 1226–1251. https://doi.org/10.1007/s11538-012-9720-6 PMID: 22383117
27. Fitzpatrick MC, Hampson K, Cleaveland S, Meyers LA, Townsend JP, Galvani AP. Potential for Rabies Control through Dog Vaccination in Wildlife-Abundant Communities of Tanzania. PLoS Negl Trop Dis. 2012;6. https://doi.org/10.1371/journal.pntd.0001796 PMID: 22928056
28. Beyer HL, Hampson K, Lembo T, Cleaveland S, Kaare M, Haydon DT. The implications of metapopulation dynamics on the design of vaccination campaigns. Vaccine. 2012; 30: 1014–1022. https://doi.org/10.1016/j.vaccine.2011.12.052 PMID: 22198516
29. Townsend SE, Lembo T, Cleaveland S, Meslin FX, Miranda ME, Putra AAG, et al. Surveillance guidelines for disease elimination: A case study of canine rabies. Comp Immunol Microbiol Infect Dis. 2013; 36: 249–261. https://doi.org/10.1016/j.cimid.2012.10.008 PMID: 23260376
30. Dürr S, Ward MP. Development of a novel rabies simulation model for application in a non-endemic environment. PLoS Negl Trop Dis. 2015; 9: 1–22. https://doi.org/10.1371/journal.pntd.0003876 PMID: 26114762
31. Ferguson EA, Hampson K, Cleaveland S, Consunji R, Deray R, Friar J, et al. Heterogeneity in the spread and control of infectious disease: Consequences for the elimination of canine rabies. Sci Rep. 2015; 5: 1–13. https://doi.org/10.1038/srep18232 PMID: 26667267
32. Chen J, Zou L, Jin Z, Ruan S. Modeling the Geographic Spread of Rabies in China. PLoS Negl Trop Dis. 2015; 9: 1–18. https://doi.org/10.1371/journal.pntd.0003772 PMID: 26020234
33. Sparkes J, McLeod S, Ballard G, Fleming PJS, Körtner G, Brown WY. Rabies disease dynamics in naïve dog populations in Australia. Prev Vet Med. 2016; 131: 127–136. https://doi.org/10.1016/j.prevetmed.2016.07.015 PMID: 27544262

34. Leung T, Davis SA. Rabies Vaccination Targets for Stray Dog Populations. Front Vet Sci. 2017; 4: 1–10. https://doi.org/10.3389/fvets.2017.00001 PMID: 28154816

35. Laager M, Mbilo C, Madaye EA, Naminou A, Léchenne M, Tschopp A, et al. The importance of dog population contact network structures in rabies transmission. PLoS Negl Trop Dis. 2018; 12: 1–18. https://doi.org/10.1371/journal.pntd.0006680 PMID: 30067733

36. Kadowaki H, Hampson K, Tojinbara K, Yamada A, Makita K. The risk of rabies spread in Japan: a mathematical modelling assessment. Epidemiol Infect. 2018; 146: 1245–1252. https://doi.org/10.1017/S0950268818001267 PMID: 29781416

37. Laager M, Léchenne M, Naissengar K, Mindekem R, Oussiguere A, Zinsstag J, et al. A metapopulation model of dog rabies transmission in N'Djamena, Chad. J Theor Biol. 2019; 462: 408–417. https://doi.org/10.1016/j.jtbi.2018.11.027 PMID: 30500602

38. Wilson-Aggarwal JK, Ozella L, Tizzoni M, Cattuto C, Swan GJF, Moundai T, et al. High-resolution contact networks of free-ranging domestic dogs Canis familiaris and implications for transmission of infection. PLoS Negl Trop Dis. 2019; 13: 1–19. https://doi.org/10.1371/journal.pntd.0007565 PMID: 31306425

39. Beyene TJ, Fitzpatrick MC, Galvani AP, Mourits MCM, Revie CW, Cernicchiaro N, et al. Impact of One-Health framework on vaccination cost-effectiveness: A case study of rabies in Ethiopia. One Heal. 2019; 8: 100103. https://doi.org/10.1016/j.onehlt.2019.100103 PMID: 31528684

40. Taib NAA, Labadin J, Piau P. Model simulation for the spread of rabies in Sarawak, Malaysia. Int J Adv Sci Eng Inf Technol. 2019; 9: 1739–1745. https://doi.org/10.18517/ijaseit.9.5.10230

41. Huang J, Ruan S, Shu Y, Wu X. Modeling the Transmission Dynamics of Rabies for Dog, Chinese Ferret Badger and Human Interactions in Zhejiang Province, China. Bull Math Biol. 2019; 81: 939–962. https://doi.org/10.1007/s11538-018-00537-1 PMID: 30536160

42. Hudson EG, Brookes VJ, Ward MP, Dürr S. Using roaming behaviours of dogs to estimate contact rates: The predicted effect on rabies spread. Epidemiol Infect. 2019;147. https://doi.org/10.1017/S0950268819000189 PMID: 30869048

43. Hudson EG, Brookes VJ, Dürr S, Ward MP. Modelling targeted rabies vaccination strategies for a domestic dog population with heterogeneous roaming patterns. PLoS Negl Trop Dis. 2019; 13: 1–15. https://doi.org/10.1371/journal.pntd.0007582 PMID: 31283780

44. Brookes VJ, Dürr S, Ward MP. Rabies-induced behavioural changes are key to rabies persistence in dog populations: Investigation using a network-based model. PLoS Negl Trop Dis. 2019; 13: 1–19. https://doi.org/10.1371/journal.pntd.0007739 PMID: 31545810

45. Carroll MJ, Singer A, Smith GC, Cowan DP, Massei G. The use of immunocontraception to improve rabies eradication in urban dog populations. Wildl Res. 2010; 37: 676–687. https://doi.org/10.1071/WR10027

46. Ortega NRS, Sallum PC, Massad E. Fuzzy dynamical systems in epidemic modeling. Stud Fuzziness Soft Comput. 2000; 232: 181–206. https://doi.org/10.1007/978-3-540-69094-8_9

47. Hampson K, Dushoff J, Bingham J, Brückner G, Ali YH, Dobson A. Synchronous cycles of domestic dog rabies in sub-Saharan Africa and the impact of control efforts. Proc Natl Acad Sci U S A. 2007; 104: 7717–7722. https://doi.org/10.1073/pnas.0609122104 PMID: 17452645

48. Zinsstag J, Dürr S, Penny MA, Mindekem R, Roth F, Menendez Gonzalez S, et al. Transmission dynamics and economics of rabies control in dogs and humans in an African city. Proc Natl Acad Sci U S A. 2009; 106: 14996–15001. https://doi.org/10.1073/pnas.0904740106 PMID: 19706492

49. Zhang J, Jin Z, Sun GQ, Zhou T, Ruan S. Analysis of rabies in China: transmission dynamics and control. PLoS One. 2011;6. https://doi.org/10.1371/journal.pone.0020891 PMID: 21789166

50. Beyer HL, Hampson K, Lembo T, Cleaveland S, Kaare M, Haydon DT. Metapopulation dynamics of rabies and the efficacy of vaccination. Proc R Soc B Biol Sci. 2011; 278: 2182–2190. https://doi.org/10.1098/rspb.2010.2312 PMID: 21159675

51. Hou Q, Jin Z, Ruan S. Dynamics of rabies epidemics and the impact of control efforts in Guangdong Province, China. J Theor Biol. 2012; 300: 39–47. https://doi.org/10.1016/j.jtbi.2012.01.006 PMID: 22273729

52. Bourhy H, Reynes JM, Dunham EJ, Dacheux L, Larrous F, Huong VTQ, et al. The origin and phylogeography of dog rabies virus. J Gen Virol. 2008; 89: 2673–2681. https://doi.org/10.1099/vir.0.2008/003913-0 PMID: 18931062

53. Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots. PLoS Comput Biol. 2009;5. https://doi.org/10.1371/journal.pcbi.1000520 PMID: 19779555

54. Talbi C, Holmes EC, de Benedictis P, Faye O, Nakouné E, Gamatié D, et al. Evolutionary history and dynamics of dog rabies virus in western and central Africa. J Gen Virol. 2009; 90: 783–791. https://doi.org/10.1099/vir.0.007765-0 PMID: 19264663

55. Meng S, Sun Y, Wu X, Tang J, Xu G, Lei Y, et al. Evolutionary dynamics of rabies viruses highlights the importance of China rabies transmission in Asia. Virology. 2011; 410: 403–409. https://doi.org/10.1016/j.virol.2010.12.011 PMID: 21195445

56. Hayman DTS, Johnson N, Horton DL, Hedge J, Wakeley PR, Banyard AC, et al. Evolutionary history of rabies in Ghana. PLoS Negl Trop Dis. 2011;5. https://doi.org/10.1371/journal.pntd.0001001 PMID: 21483707

57. Carnieli P, de Novaes Oliveira R, Macedo CI, Castilho JG. Phylogeography of rabies virus isolated from dogs in Brazil between 1985 and 2006. Arch Virol. 2011; 156: 1007–1012. https://doi.org/10.1007/s00705-011-0942-y PMID: 21327782

58. Yu J, Li H, Tang Q, Rayner S, Han N, Guo Z, et al. The spatial and temporal dynamics of rabies in China. PLoS Negl Trop Dis. 2012;6. https://doi.org/10.1371/journal.pntd.0001640 PMID: 22563518

59. Mollentze N, Weyer J, Markotter W, Le Roux K, Nel LH. Dog rabies in southern Africa: Regional surveillance and phylogeographical analyses are an important component of control and elimination strategies. Virus Genes. 2013; 47: 569–573. https://doi.org/10.1007/s11262-013-0974-3 PMID: 23996607

60. Guo Z, Tao X, Yin C, Han N, Yu J, Li H, et al. National Borders Effectively Halt the Spread of Rabies: The Current Rabies Epidemic in China Is Dislocated from Cases in Neighboring Countries. PLoS Negl Trop Dis. 2013;7. https://doi.org/10.1371/journal.pntd.0002039 PMID: 23383359

61. Carnieli P, Ruthner Batista HBC, de Novaes Oliveira R, Castilho JG, Vieira LFP. Phylogeographic dispersion and diversification of rabies virus lineages associated with dogs and crab-eating foxes (Cerdocyon thous) in Brazil. Arch Virol. 2013; 158: 2307–2313. https://doi.org/10.1007/s00705-013-1755-y PMID: 23749047

62. Horton DL, McElhinney LM, Freuling CM, Marston DA, Banyard AC, Goharrriz H, et al. Complex Epidemiology of a Zoonotic Disease in a Culturally Diverse Region: Phylogeography of Rabies Virus in the Middle East. PLoS Negl Trop Dis. 2015; 9: 1–17. https://doi.org/10.1371/journal.pntd.0003569 PMID: 25811659

63. Brunker K, Marston DA, Horton DL, Cleaveland S, Fooks AR, Kazwala R, et al. Elucidating the phylodynamics of endemic rabies virus in eastern Africa using whole-genome sequencing. Virus Evol. 2015; 1: 1–11. https://doi.org/10.1093/ve/vev001 PMID: 27774275

64. Yao HW, Yang Y, Liu K, Lou Li X, Zuo SQ, Sun RX, et al. The Spatiotemporal Expansion of Human Rabies and Its Probable Explanation in Mainland China, 2004–2013. PLoS Negl Trop Dis. 2015; 9: 2004–2013. https://doi.org/10.1371/journal.pntd.0003502 PMID: 25692883

65. Troupin C, Dacheux L, Tanguy M, Sabeta C, Blanc H, Bouchier C, et al. Large-Scale Phylogenomic Analysis Reveals the Complex Evolutionary History of Rabies Virus in Multiple Carnivore Hosts. PLoS Pathog. 2016; 12: e1006041. https://doi.org/10.1371/journal.ppat.1006041 PMID: 27977811

66. Zhang Y, Vrancken B, Feng Y, Dellicour S, Yang Q, Yang W, et al. Cross-border spread, lineage displacement and evolutionary rate estimation of rabies virus in Yunnan Province, China. Virol J. 2017; 14: 1–8. https://doi.org/10.1186/s12985-016-0669-1 PMID: 28081705

67. Ma C, Hao X, Deng H, Wu R, Liu J, Yang Y, et al. Re-emerging of rabies in Shaanxi province, China, from 2009 to 2015. J Med Virol. 2017; 89: 1511–1519. https://doi.org/10.1002/jmv.24769 PMID: 28112421

68. Dellicour S, Rose R, Faria NR, Vieira LFP, Bourhy H, Gilbert M, et al. Using Viral Gene Sequences to Compare and Explain the Heterogeneous Spatial Dynamics of Virus Epidemics. Mol Biol Evol. 2017; 34: 2563–2571. https://doi.org/10.1093/molbev/msx176 PMID: 28651357

69. Brunker K, Lemey P, Marston DA, Fooks AR, Lugelo A, Ngeleja C, et al. Landscape attributes governing local transmission of an endemic zoonosis: Rabies virus in domestic dogs. Mol Ecol. 2018; 27: 773–788. https://doi.org/10.1111/mec.14470 PMID: 29274171

70. Wang L, Wu X, Bao J, Song C, Du J. Phylodynamic and transmission pattern of rabies virus in China and its neighboring countries. Arch Virol. 2019. https://doi.org/10.1007/s00705-019-04297-8 PMID: 31147766

71. Dellicour S, Troupin C, Jahanbakhsh F, Salama A, Massoudi S, Moghaddam MK, et al. Using phylogeographic approaches to analyse the dispersal history, velocity and direction of viral lineages—Application to rabies virus spread in Iran. Mol Ecol. 2019; 28: 4335–4350. https://doi.org/10.1111/mec.15222 PMID: 31535448

72. Talbi C, Lemey P, Suchard MA, Abdelatif E, Elharrak M, Jalal N, et al. Phylodynamics and Human-mediated dispersal of a zoonotic virus. PLoS Pathog. 2010;6. https://doi.org/10.1371/journal.ppat.1001166 PMID: 21060816

73. Mollentze N, Nel LH, Townsend S, le Roux K, Hampson K, Haydon DT, et al. A bayesian approach for inferring the dynamics of partially observed endemic infectious diseases from space-time-genetic data. Proc R Soc B Biol Sci. 2014;281. https://doi.org/10.1098/rspb.2013.3251 PMID: 24619442

74. Bourhy H, Nakouné E, Hall M, Nouvellet P, Lepelletier A, Talbi C, et al. Revealing the Micro-scale Signature of Endemic Zoonotic Disease Transmission in an African Urban Setting. PLoS Pathog. 2016; 12: e1005525. https://doi.org/10.1371/journal.ppat.1005525 PMID: 27058957

75. Zinsstag J, Lechenne M, Laager M, Mindekem R, Naïssengar S, Oussiguéré A, et al. Vaccination of dogs in an African city interrupts rabies transmission and reduces human exposure. Sci Transl Med. 2017;9. https://doi.org/10.1126/scitranslmed.aaf6984 PMID: 29263230

76. Cori A, Nouvellet P, Garske T, Bourhy H, Nakouné E, Jombart T. A graph-based evidence synthesis approach to detecting outbreak clusters: An application to dog rabies. PLoS Comput Biol. 2018;14. https://doi.org/10.1371/journal.pcbi.1006554 PMID: 30557340

77. Tian H, Feng Y, Vrancken B, Cazelles B, Tan H, Gill MS, et al. Transmission dynamics of re-emerging rabies in domestic dogs of rural China. PLoS Pathog. 2018; 14: 1–18. https://doi.org/10.1371/journal.ppat.1007392 PMID: 30521641

78. Baele G, Suchard MA, Rambaut A, Lemey P. Emerging concepts of data integration in pathogen phylodynamics. Syst Biol. 2017; 66: e47–e65. https://doi.org/10.1093/sysbio/syw054 PMID: 28173504

79. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. 2018; 4: 1–5. https://doi.org/10.1093/ve/vey016 PMID: 29942656

80. Bouckaert R, Vaughan TG, Barido-Sottan J, Duchêne S, Fourment M, Gavryushkina A, et al. BEAST 2.5: An Advanced Software Platform for Bayesian Evolutionary Analysis. PLoS Comput Biol. 2019; 15: e1006650. https://doi.org/10.1371/journal.pcbi.1006650 PMID: 30958812

81. De la Puente-León M, Levy M, Toledo A, Recuenco S, Shinnick J, Castillo-Neyra R. Spatial Inequality Hides the Burden of Dog Bites and the Risk of Dog-Mediated Human Rabies. Am J Trop Med Hyg. 2020; 00: 1–10. https://doi.org/10.4269/ajtmh.20-0180 PMID: 32662391

82. Brunker K, Nadin-Davis S, Biek R. Genomic sequencing, evolution and molecular epidemiology of rabies virus. Rev Sci Tech. 2018; 37: 401–408. https://doi.org/10.20506/rst.37.2.2810 PMID: 30747139

83. Lemey P, Rambaut A, Bedford T, Faria N, Bielejec F, Baele G, et al. Unifying Viral Genetics and Human Transportation Data to Predict the Global Transmission Dynamics of Human Influenza H3N2. PLoS Pathog. 2014; 10: e1003932. https://doi.org/10.1371/journal.ppat.1003932 PMID: 24586153

84. Dellicour S, Rose R, Pybus OG. Explaining the geographic spread of emerging epidemics: A framework for comparing viral phylogenies and environmental landscape data. BMC Bioinformatics. 2016;17. https://doi.org/10.1186/s12859-015-0864-x PMID: 26729273

85. Dellicour S, Baele G, Dudas G, Faria NR, Pybus OG, Suchard MA, et al. Phylodynamic assessment of intervention strategies for the West African Ebola virus outbreak. Nat Commun. 2018;9. https://doi.org/10.1038/s41467-017-01881-x PMID: 29339724

86. Zhang J, Jin Z, Sun G, Sun X, Ruan S. Spatial spread of rabies in China. J Appl Anal Comput. 2012; 2: 111–126. https://doi.org/10.11948/2012008

87. Knobel DL, Lembo T, Morters M, Townsend SE, Cleaveland S, Hampson K. Dog Rabies and Its Control. Third Edit. Rabies. Elsevier Inc.; 2013. https://doi.org/10.1016/B978-0-12-396547-9.00017–1

88. Taylor LH, Wallace RM, Balaram D, Lindenmayer JM, Eckery DC, Mutonono-Watkiss B, et al. The role of dog population management in rabies elimination-A review of current approaches and future opportunities. Front Vet Sci. 2017;4. https://doi.org/10.3389/fvets.2017.00004 PMID: 28197407

89. Rattanavipapong W, Thavorncharoensap M, Youngkong S, Genuino AJ, Anothaisintawee T, Chaikledkaew U, et al. The impact of transmission dynamics of rabies control: Systematic review. Vaccine. 2019; 37: A154–A165. https://doi.org/10.1016/j.vaccine.2018.11.035 PMID: 30528329

90. Fisher CR, Streicker DG, Schnell MJ. The spread and evolution of rabies virus: Conquering new frontiers. Nat Rev Microbiol. 2018; 16: 241–255. https://doi.org/10.1038/nrmicro.2018.11 PMID: 29479072

91. Rupprecht CE, Kuzmin I V., Yale G, Nagarajan T, Meslin FX. Priorities in applied research to ensure programmatic success in the global elimination of canine rabies. Vaccine. 2019; 37: A77–A84. https://doi.org/10.1016/j.vaccine.2019.01.015 PMID: 30685249

92. Aiming for elimination of dog-mediated human rabies cases by 2030. Vet Rec. 2016. https://doi.org/10.1136/vr.i51 PMID: 26795858

93. Hampson K, De Balogh K, Mcgrane J. Lessons for rabies control and elimination programmes: a decade of One Health experience from Bali, Indonesia. Rev Sci Tech. 2019; 38: 213–224. https://doi.org/10.20506/rst.38.1.2954 PMID: 31564729

94. Brunker K, Jaswant G, Thumbi SM, Lushasi K, Lugelo A, Czupryna AM, et al. Rapid in-country sequencing of whole virus genomes to inform rabies elimination programmes. Wellcome Open Res. 2020; 5: 1–30. https://doi.org/10.12688/wellcomeopenres.15518.2 PMID: 32090172

95. Gigante CM, Yale G, Condori RE, Costa NC, Hampson K, Thumbi SM, et al. Portable Rabies Virus Sequencing in Canine Rabies Endemic Countries Using the Oxford Nanopore MinION. 2020.

96. Ishikawa SA, Zhukova A, Iwasaki W, Gascuel O, Pupko T. A Fast Likelihood Method to Reconstruct and Visualize Ancestral Scenarios. Mol Biol Evol. 2019; 36: 2069–2085. https://doi.org/10.1093/molbev/msz131 PMID: 31127303

97. Hasanov E, Zeynalova S, Geleishvili M, Maes E, Tongren E, Marshall E, et al. Assessing the impact of public education on a preventable zoonotic disease: Rabies. Epidemiol Infect. 2018; 146: 227–235. https://doi.org/10.1017/S0950268817002850 PMID: 29271331

98. Bardosh K. Global aspirations, local realities: The role of social science research in controlling neglected tropical diseases. Infect Dis Poverty. 2014; 3: 1–15. https://doi.org/10.1186/2049-9957-3-1 PMID: 24401663

99. Degeling C, Brookes V, Lea T, Ward M. Rabies response, One Health and more-than-human considerations in indigenous communities in northern Australia. Soc Sci Med. 2018; 212: 60–67. https://doi.org/10.1016/j.socscimed.2018.07.006 PMID: 30005225

American Journal of Epidemiology

Vol. 191, No. 7
https://doi.org/10.1093/aje/kwac042
Advance Access publication:
March 3, 2022

Original Contribution

Impact of BNT162b2 Vaccination and Isolation on SARS-CoV-2 Transmission in Israeli Households: An Observational Study

Maylis Layan, Mayan Gilboa, Tal Gonen, Miki Goldenfeld, Lilac Meltzer, Alessio Andronico, Nathanaël Hozé, Simon Cauchemez*, and Gili Regev-Yochay*

* Correspondence to Dr. Simon Cauchemez, Mathematical Modelling of Infectious Diseases Unit, Institut Pasteur, 25-28 rue du Docteur Roux, 75014 Paris, France (e-mail: simon.cauchemez@pasteur.fr); or Dr. Gili Regev-Yochay, Infection Prevention and Control Unit, Sheba Medical Center, Derech Sheba 2, Ramat Gan, Israel (e-mail: gili.regev@sheba.health.gov.il).

Initially submitted September 2, 2021; accepted for publication March 1, 2022.

Several studies have characterized the effectiveness of vaccines against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections. However, estimates of their impact on transmissibility remain limited. Here, we evaluated the impact of isolation and vaccination (7 days after the second dose) on SARS-CoV-2 transmission within Israeli households. From December 2020 to April 2021, confirmed cases were identified among health-care workers of the Sheba Medical Centre and their family members. Recruited households were followed up with repeated PCR for at least 10 days after case confirmation. Data were analyzed using a data augmentation Bayesian framework. A total of 210 households with 215 index cases were enrolled; 269 out of 667 (40%) susceptible household contacts developed a SARS-CoV-2 infection. Of those, 170 (63%) developed symptoms. Compared with unvaccinated and unisolated adult/teenager (aged >12 years) contacts, vaccination reduced the risk of infection among unisolated adult/teenager contacts (relative risk (RR) = 0.21, 95% credible interval (CrI): 0.08, 0.44), and isolation reduced the risk of infection among unvaccinated adult/teenager (RR = 0.12, 95% CrI: 0.06, 0.21) and child contacts (RR = 0.17, 95% CrI: 0.08, 0.32). Infectivity was reduced in vaccinated cases (RR = 0.25, 95% CrI: 0.06, 0.77). Within households, vaccination reduces both the risk of infection and of transmission if infected. When contacts were unvaccinated, isolation also led to important reductions in the risk of transmission.

COVID-19; household; infectious disease transmission; physical distancing; SARS-CoV-2; vaccination; vaccine effectiveness

Abbreviations: CrI, credible interval; COVID-19, coronavirus disease 2019; HCW, health-care workers; PCR, polymerase chain reaction; SAR, secondary attack rate; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a highly transmissible virus that was first detected in Wuhan China in December 2019 (1, 2). It is the cause of coronavirus disease 2019 (COVID-19), which has spread through the world, leading to a pandemic that had infected at least 250 million people and caused more than 5 million deaths worldwide by November 10, 2021 (3). The advent of novel coronavirus disease 2019 (COVID-19) vaccines has been an important breakthrough in the management of the pandemic. To determine how vaccination may modify epidemic dynamics, it is essential to estimate its effectiveness with respect to infection, transmission, and disease severity. Multiple studies have shown that COVID-19 vaccines are effective at reducing both the risk of infection (4–8) and the risk of developing severe symptoms (4, 8–10) in the general population.

Documenting vaccine impact on transmission is more challenging, stemming from the difficulty of thoroughly documenting chains of transmission and accounting for the ways different types of contacts may lead to different risks of transmission (11). Households represent the perfect environment to evaluate factors affecting transmission such as vaccination because the probability of SARS-CoV-2 transmission among household members is high, ranging between 14%

and 32% (12–14). Beyond the evaluation of vaccine effectiveness, understanding how vaccines affect household transmission is also important to determine how recommendations should evolve with vaccines. For example, should isolation precautions be maintained in partially vaccinated households (15)? A number of studies have shown that vaccines provide indirect protection against household transmission (16–20). However, none of these studies evaluated how isolation affected the outcome, and for some of the studies (16–19), the passive nature of surveillance may have led to underestimating household transmission rates.

During the first months of 2021, Israel underwent its third pandemic wave due to the rise of the Alpha variant that quickly accounted for 90% of infections (21). Concomitantly, vaccination was extended to all adults older than age 16 years, making Israel one of the first countries to reach high vaccination coverage in their population, with 60% of the total population being vaccinated by March 22, 2021 (3). During this period, we followed SARS-CoV-2 transmission in the households of 12,518 health-care workers (HCWs) of the Sheba Medical Center, the largest medical center in Israel. Here, we describe dynamics of transmission in these households and evaluate the impact vaccination and isolation measures had on these dynamics.

METHODS

Study design and study population

All HCWs, regardless of their vaccination status, were required to use an electronic questionnaire to report daily any COVID-19 related symptom they, or a member of their household, had. SARS-CoV-2 polymerase chain reaction testing (PCR) was readily available, and HCWs were encouraged to be tested for any mild symptom or suspected exposures. All HCWs were instructed to notify the infection prevention and control unit if one of their household members was SARS-CoV-2 positive. All SARS-CoV-2–detected HCWs as well as those with a positive SARS-CoV-2 household member were immediately contacted as part of the epidemiologic investigation for contact tracing and were provided with instructions regarding isolation precautions. All unvaccinated household members (i.e., those that did not receive the 2 vaccine doses at least 7 days before the detection of the COVID-19 patient) were required to perform 2 PCR tests in the 10 days after the diagnosis of the positive COVID-19 patient. Vaccinated household members were encouraged to perform 2 PCR tests during the 10 days after detection. Household members were not required to test a second time if they had a positive test (Web Table 1 in Web Appendix 1, available at https://doi.org/10.1093/aje/kwac042). Unvaccinated HCW contacts were isolated at home, whereas vaccinated HCWs were instructed to perform a PCR test every day they reported to the hospital for work.

Between December 31, 2020, and April 26, 2021, the HCWs who were SARS-CoV-2–positive or reported a positive household member were contacted at least 10 days after detection and were offered enrollment in the study. Those who agreed, and gave their consent, answered a telephone interview.

Data and sample collection

Data collected during the phone interview included the age and gender of the HCW's household members, their vaccination status, information about prior COVID-19 infections, their COVID-19 PCR test dates and results, their symptoms (i.e., fever, cough, myalgia, headache, congestion, diarrhea, vomiting, anosmia, or ageusia), the number of rooms and bathrooms in the household, and the degree to which isolation precautions were adhered to (Web Appendix 2). At the time of the study, only individuals 16 years old or older were eligible for vaccination.

The household member who had the first positive PCR test was defined as the index case. When multiple household members had a positive PCR test on the same day, they were defined as co-index cases. We defined complete isolation as complete separation in sleeping and eating between household contacts and index case(s) (i.e., they did not spend any time in the same room) and whether a separate bathroom was provided for the index case(s). Partial isolation was defined if one of the above was violated, but masks were continuously used, and eating was consistently separate.

For HCWs, nasopharyngeal swabs were collected by trained personnel, and reverse-transcriptase quantitative PCR analysis was performed using the Allplex 2019-nCov RT-qPCR assay (Seegene Inc., Seoul, South Korea) and expressed by cycle threshold (Ct). Other household members reported the results of their COVID-19 test(s) performed by their health-care providers.

Clinical outcome

Confirmed SARS-CoV-2 infections were defined by a positive PCR test (i.e., with a Ct value lower than 40). Symptomatic cases were defined as confirmed cases with the presence of at least 1 symptom from among the following: fever, cough, myalgia, headache, congestion, diarrhea, vomiting, anosmia, or ageusia. Contacts who reported at least 1 of the above-mentioned symptoms but were not confirmed because they performed no PCR test ($n = 6$) or a single test at inclusion ($n = 2$) were also considered as symptomatic cases. Asymptomatic cases were defined as confirmed cases who did not report any symptom over the follow-up period of the household.

Statistical analysis

We evaluated transmission in households using 2 metrics: the secondary attack rate (SAR), defined as the proportion of susceptible household contacts that are infected after the index case is detected (22), and the person-to-person probability of transmission, defined as the per-capita probability that an infected individual transmits to a susceptible household contact. The first metric includes tertiary (i.e., household contacts infected by a household member that is not the index case) and community cases (i.e., household contacts infected in the community) contrary to the second metric. In both cases, we assumed that individuals who reported past infection of SARS-CoV-2 confirmed by PCR over the year preceding the detection of the household index

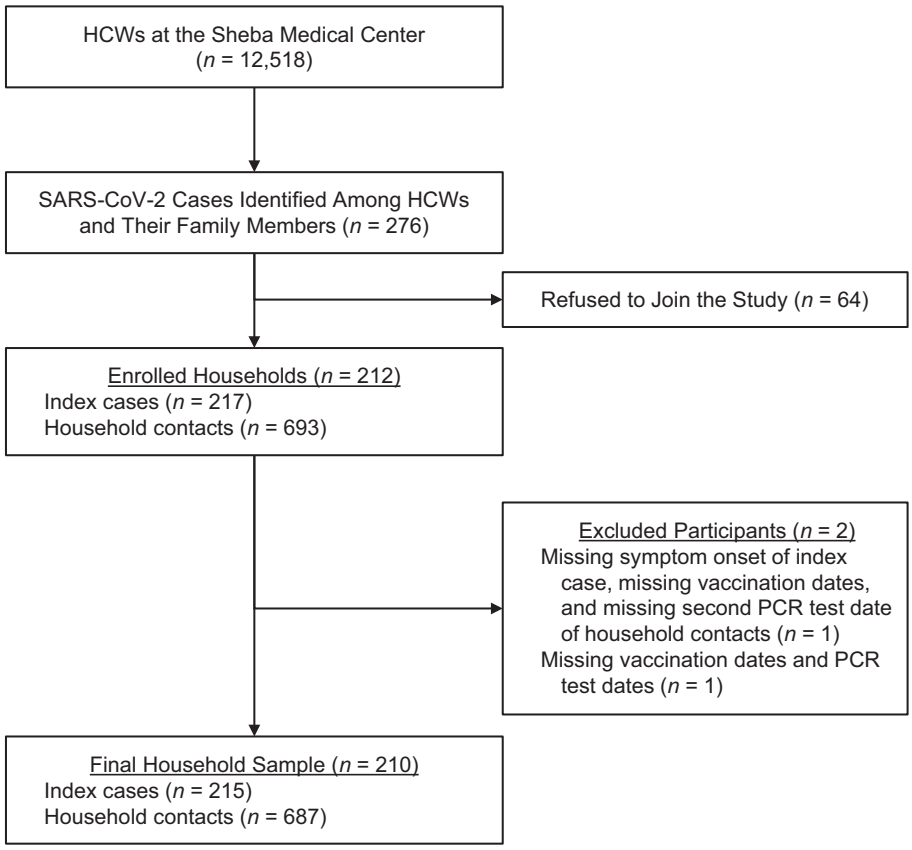


Figure 1. Flow chart of the households included in our analysis, Ramat Gan, Israel, 2020–2021. HCW, health-care worker; PCR, polymerase chain reaction; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

case (n = 20) were protected from infection and therefore, did not count as susceptible household contacts.

Baseline characteristics of the index cases and household contacts were described according to their vaccination status. All individuals older than 12 years were considered as adults/teenagers. We calculated the SAR for different categories of household contacts: unisolated and unvaccinated adults/teenagers, unisolated and vaccinated adults/teenagers, isolated and unvaccinated adults/teenagers, vaccinated and isolated adults/teenagers, unisolated children, and isolated children. Here, isolation corresponds to complete or partial isolation between household contacts and the index case. We also defined the SAR of vaccinated and unvaccinated index cases as the proportion of infected household contacts in households with vaccinated or unvaccinated index cases, respectively. In a sensitivity analysis, the SAR calculation was restricted to households in which a single index case was identified (Web Table 2 in Web Appendix 3). We also report the 95% confidence interval of the SAR.

We developed a statistical model to evaluate the effect of age, isolation precautions, BNT162b2 vaccination, and household size on SARS-CoV-2 transmission dynamics in households (Web Appendix 4). The model uses the sequence of symptom onset dates and positive molecular test dates to estimate the person-to-person risk of transmission within the household while accounting for the community hazard of infection (i.e., household contacts infected outside the household) and the possibility of tertiary transmissions (i.e., household contacts infected by a member of the household that is not the index case) (23). The person-to-person risk of transmission is decomposed into the baseline person-to-person risk of infection depending on household size, the relative infectivity of the infector depending on their vaccination status (reference group: unvaccinated cases), and the relative susceptibility of the infectee depending on their age, isolation behavior, and vaccination status. The relative susceptibility is estimated separately for unisolated children, isolated children, isolated and unvaccinated adults/teenagers, unisolated and vaccinated adults/teenagers, and adults/teenagers that are both isolated and vaccinated, considering the group of adults/teenagers that are unisolated and unvaccinated as the reference group. None of the children were vaccinated at the time of the study. This formulation accommodates the potential confounding effects between the 3 variables characterizing household contacts (i.e., being vaccinated, being isolated, or being a child). We assumed that individuals whose isolation behavior was missing (n = 6) did not comply with isolation precautions.

Model parameters were estimated using Bayesian Markov chain Monte Carlo sampling with data augmentation (23)

Table 1. Characteristics of the Index Cases According to Age, Ramat Gan, Israel, 2020–2021

Characteristic	Adult/Teenager Index Cases[a] (*n* = 191)			Child Index Cases (*n* = 24)			All Index Cases (*n* = 215)		
	No.	%	Median (IQR)	No.	%	Median (IQR)	No.	%	Median (IQR)
Male sex	76	40		14	58		90	42	
Age, years[b]	36 (14)			6 (4)			32 (16)		
Cluster size[c]			2 (1–3)			2 (1–3)			2 (1–3)
Symptom status									
Symptomatic	172	90		10	42		182	85	
Asymptomatic	19	10		14	58		33	15	
Vaccination									
Vaccinated	15	8		N/A	N/A		15	7	
Days from second dose to detection			44 (13–59)			N/A			44 (13–59)

Abbreviation: IQR, interquartile range; N/A, not applicable.
[a] Individuals aged >12 years were considered adults/teenagers.
[b] Values are expressed as mean (standard deviation).
[c] Number of secondary cases among the susceptible contacts of the index case(s).

(Web Appendix 5). Data were augmented with the probable date of infection of confirmed cases. For symptomatic cases, the date of infection was reconstructed from the date of symptom onset, using the probabilistic distribution of the incubation period (24). For asymptomatic cases, we assumed that the date of infection could occur up to 10 days prior to their molecular detection based on a meta-analysis (25).

Since the study was conducted during the vaccine rollout, participants were enrolled at varying stages of their vaccination process. We assumed that vaccines reach their full effect 7 days after receiving a second dose (4, 9, 10). Cases were therefore considered vaccinated if their symptom onset (or if unknown, the date of their first positive PCR test) occurred $\geq$7 days after the second dose. Similarly, household contacts were considered vaccinated if their exposure to the index case (starting with symptom onset or, in its absence, from the date of first positive PCR of the index case) occurred $\geq$7 days after the second dose. In a sensitivity analysis, we investigated how parameter estimates changed under the assumption that vaccination is effective $\geq$15 days after the first dose. We also assessed how estimates changed when the analysis was restricted to households in which all negative contacts had performed at least 1 or 2 PCR tests in the 10 days following the detection of the index case. In the baseline scenario, we assumed that asymptomatic cases are 40% less infectious than symptomatic cases based on a meta-analysis (26), and we investigated whether assuming the same level of infectivity in asymptomatic and symptomatic cases modified our estimates. Finally, in our baseline analysis, we chose a log-normal with log-mean = 0 and log-standard deviation = 1 prior distribution for the relative infectivity and relative susceptibility parameters and explored smaller and larger values (log-standard deviation = 0.7 or 2) in a sensitivity analysis.

We compared the observed and expected distributions of the number of cases per household size to assess the goodness-of-fit of the model (Web Table 3 in Web Appendix 6). We report the posterior median and the 95% credible interval (CrI) of estimated parameters. We also report the posterior probability that isolated and vaccinated adult/teenager contacts are less susceptible than vaccinated adult/teenager contacts that do not isolate. To measure the strength of evidence of a reduced susceptibility in isolated contacts among vaccinated ones, we report the associated Bayes factor. Here, it directly corresponds to the posterior odds of a reduced susceptibility in isolated contacts among vaccinated ones. Additional details are available in Web Appendix 1–6.

Ethics

The study was approved by the Sheba Medical Center institutional review board committee (approval #8130-21).

RESULTS

All 12,518 HCWs employed by the Sheba Medical Center were eligible to join the study. Between December 19 and April 28, 2021, 91% of the Sheba Medical Center personnel received both doses of the BNT162b2 vaccine, and a rapid and significant decrease in newly detected cases was observed among HCWs.

From December 31, 2020, to April 26, 2021, 276 SARS-CoV-2 cases were identified among HCWs of the Sheba Medical Center and their household members (Figure 1). Of these, 212 agreed to participate, gave their consent, and were enrolled in the study with their household members. Two households were excluded due to missing vaccination

Table 2. Characteristics of the Household Contacts According to Age, Ramat Gan, Israel, 2020–2021

Characteristic	Adult/Teenager Household Contacts[a] (*n* = 494)			Child Household Contacts (*n* = 193)			All Household Contacts (*n* = 687)		
	No.	%	Median (IQR)	No.	%	Median (IQR)	No.	%	Median (IQR)
Male sex	242	49		109	56		351	51	
Age, years[b]	36 (17)[c]			6 (4)			27 (20)		
Infection and symptom status									
Past infection	16	3		4	2		20	3	
Not infected	304	62		94	49		398	58	
Symptomatic	127	26		41	21		168	24	
Asymptomatic	46	9		53	27		99	14	
Symptomatic (missing onset)	1	0		1	1		2	0	
Vaccination									
Vaccinated	125	25		N/A	N/A		125	18	
Days from second dose to exposure			23 (14–36)			N/A			23 (14–36)
Isolation									
Partial	115	23		32	17		147	21	
Complete	227	46		58	30		285	41	
Missing	5	1		1	1		6	1	

Abbreviation: IQR, interquartile range; N/A, not applicable
[a] Individuals aged >12 years were considered adults/teenagers.
[b] Values are expressed as mean (standard deviation).
[c] Missing age for 5 adult/teenager contacts.

status, dates of PCR test, and/or symptom onset. In total, we analyzed data from 210 households with 215 index cases, including 4 co-index cases, and their 687 household contacts. The median household size was 4 (interquartile range, 3–5). Mean age was 32 years among index cases (Table 1) and 27 years among household contacts (Table 2). Age was missing for 5 adult/teenager contacts, and isolation behavior was missing for 6 contacts. There was a slight over-representation of females among index cases (58%), and 191 index cases (89%) were adults/teenagers, of whom 15 (8%) were vaccinated. None of the 24 child index cases were vaccinated. Among the 494 adult/teenager household contacts, 125 (25%) were vaccinated. Of these, 83 (17%) also complied with isolation precautions. Among the 369 unvaccinated adult/teenager contacts, 259 (70%) isolated during the study. None of the 193 child household contacts were vaccinated and 47% of them (n = 90) isolated during the study period (Table 2). In the following, we refer to susceptible contacts (i.e., contacts that did not report SARS-CoV-2 infection over the preceding year) as contacts.

A total of 269 out of 667 (40%) household contacts developed a SARS-CoV-2 infection. Of those, 170 (63%) developed symptoms (Table 2). The SAR varied with the characteristics of the contacts. Among the 105 adult/teenager contacts who were unisolated and unvaccinated, 80 (76%) were infected by SARS-CoV-2 (Table 3). This proportion dropped to 28% (11 out of 40) among those who were unisolated and vaccinated, 29% (71 out of 245) among those who were isolated but unvaccinated, and 11% (9 out of 83) among those who were isolated and vaccinated; 65% (66 out of 101) of child contacts who were unisolated got infected by SARS-CoV-2. This proportion declined to 33% (29 out of 87) for isolated child contacts. The proportion of asymptomatic cases varied from 26% (46 out of 174) among adult/teenager contact cases to 56% (53 out of 95) among child contact cases (Table 2).

The SAR also varied with the vaccination status of the index case regardless of the contacts' characteristics. Among the 622 household contacts whose index case was unvaccinated, 261 (42%) developed a SARS-CoV-2 infection (Table 3). This proportion dropped to 19% (8 out of 42) among household contacts whose index case was vaccinated. Finally, the SAR was relatively invariant with household size: 31%, 40%, 32%, and 32% for households of size 2, 3, 4, and 5, respectively (Web Figure 1 in Web Appendix 6).

Our statistical model makes it possible to perform a multivariate analysis of the drivers of SARS-CoV-2 transmission in households. We estimate that, relative to adult/teenager

Table 3. Observed Secondary Attack Rates According to the Type of Contact, Ramat Gan, Israel, 2020–2021

Type	No. of Infected Contacts	No. of Susceptible Contacts	SAR	
			%	95% CI
Contacts[a]				
Unisolated and unvaccinated adult/teenager	80	105	76	67, 84
Isolated and unvaccinated adult/teenager	71	245	29	23, 35
Unisolated but vaccinated adult/teenager	11	40	28	15, 44
Isolated and vaccinated adult/teenager	9	83	11	5, 20
Unisolated child	66	101	65	55,75
Isolated child	29	87	33	24, 44
Index[b]				
Vaccinated	8	42	19	9, 34
Unvaccinated	261	622	42	38, 46

Abbreviations: CI, confidence interval; SAR, secondary attack rate.
[a] Isolation is missing for 1 child contact and for 5 adult contacts.
[b] The last 2 rows correspond to the SAR among the household contacts of vaccinated (n = 14 households) and unvaccinated index cases (n = 195 households). One household was excluded from this analysis because its co-index cases did not have the same vaccination status.

contacts who were unisolated and unvaccinated, the relative risk of being infected was 0.21 (95% CrI: 0.08, 0.44) among adult/teenager household contacts who were vaccinated but unisolated (Figure 2A, Web Table 4 in Web Appendix 7). It was 0.12 (95% CrI: 0.06, 0.21) among household contacts who did isolate and were unvaccinated, and 0.07 (95% CrI: 0.03, 0.16) among household contacts who were both isolated and vaccinated. Isolation might reduce the risk of infection among vaccinated contacts (96% posterior probability, Bayes factor = 23) with a relative risk of 0.34 (95% CrI: 0.11, 1.14). Relative to adult/teenager contacts who were unisolated and unvaccinated, the relative risk of infection was 0.50 (95% CrI: 0.32, 0.77) for child contacts that did not isolate, and 0.17 (95% CrI: 0.08, 0.31) for those that did. We estimate that the risk of transmission from vaccinated cases was 0.25 (95% CrI: 0.06, 0.77) times that of unvaccinated cases (Figure 2B, Web Table 4 in Web Appendix 7).

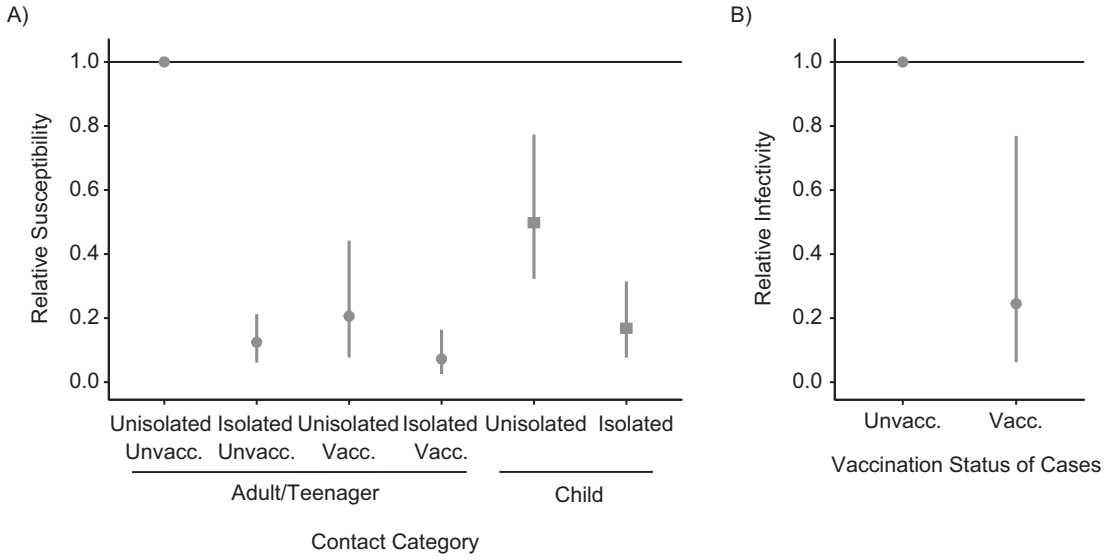


Figure 2. Estimates of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) transmission parameters within households, Ramat Gan, Israel, 2020–2021. A) Estimated relative susceptibility of isolated and unvaccinated adults/teenagers, unisolated but vaccinated adults/teenagers, isolated and vaccinated adults/teenagers, unisolated children, and isolated children. The reference group is the group of adults/teenagers that were unisolated and unvaccinated. B) Estimated relative infectivity of vaccinated cases compared with unvaccinated cases. The posterior median and its associated 95% Bayesian credible interval are reported.

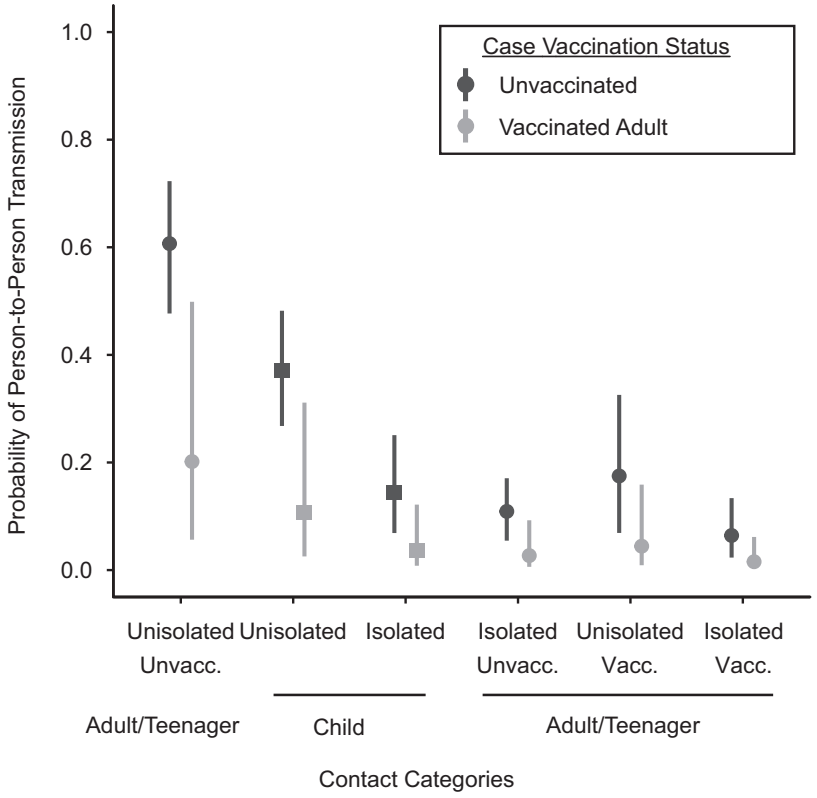


Figure 3. Person-to-person probability of transmission within households according to the characteristics of the case and of the contact, Ramat Gan, Israel, 2020–2021. Estimated person-to-person probability of transmission within households of size 4, decomposed by the age, isolation behavior, and vaccination status of the contact as well as the vaccination status of the case. The posterior median and its associated 95% Bayesian credible interval are reported.

Overall, we estimate that, in a household of size 4, the person-to-person probability of SARS-CoV-2 transmission is 61% (95% CrI: 48, 72) between an unvaccinated case and an unvaccinated and unisolated adult/teenager. This probability drops to 4% (95% CrI: 1, 16) between 2 vaccinated adults/teenagers who do not follow isolation rules (Figure 3, Web Table 5 in Web Appendix 7). The person-to-person probability of transmission from an unvaccinated case to a child who does not isolate is 37% (95% CrI: 27, 48). This probability drops to 11% (95% CrI: 3, 31) if the case is vaccinated and to 14% (95% CrI: 7, 25) if the child contact is isolated.

In general, our estimates of relative susceptibility and relative infectivity were robust to model assumptions (Figure 4). When the analysis was restricted to households in which all contacts performed at least 1 or 3 PCR tests in the 10 days following the recruitment of the index case, the relative susceptibility of vaccinated adult/teenager contacts who did not isolate was slightly higher compared with the baseline scenario. It increased from 0.21 (95% CrI: 0.08, 0.44) in the baseline scenario to 0.28 (95% CrI: 0.09, 0.66) in the analysis with at least 1 PCR and 0.32 (95% CrI: 0.09, 0.83) with at least 2 PCR tests (Web Table 4 in Web Appendix 7). In the alternative scenarios, the number of individuals included was substantially lower, increasing CrIs (Web Figures 2 and 3, Web Tables 6–9 in Web Appendix 8). Similarly, the relative susceptibility of vaccinated adult/teenager contacts who did isolate increased from 0.07 (95% CrI: 0.03, 0.16) in the baseline scenario to 0.12 (95% CrI: 0.04, 0.28) in the analysis with at least 1 PCR, and 0.13 (95% CrI: 0.04, 0.32) in the one with at least 2 PCR tests. Consequently, the posterior probability that isolated and vaccinated adult/teenager contacts were less susceptible than vaccinated adult/teenager contacts that did not isolate dropped from 96% to 88% with 1 PCR and 89% with 2 PCR tests. Still, the statistical support was high with a Bayes factor equal to 7 and 8, respectively. Relative infectivity and relative susceptibility were slightly sensitive to their prior distribution (Web Table 10 in Web Appendix 8). When the log-standard deviation increased, estimates were pulled towards lower values.

DISCUSSION

We evaluated the impact of BNT162b2 vaccination on case infectivity and the mitigating effect of age, isolation from the index case, and BNT162b2 vaccination on susceptibility to infection in household settings. Our approach accounts for infections in the community, potential tertiary infections within the households, the reduced infectivity of asymptomatic cases, potential misidentification of

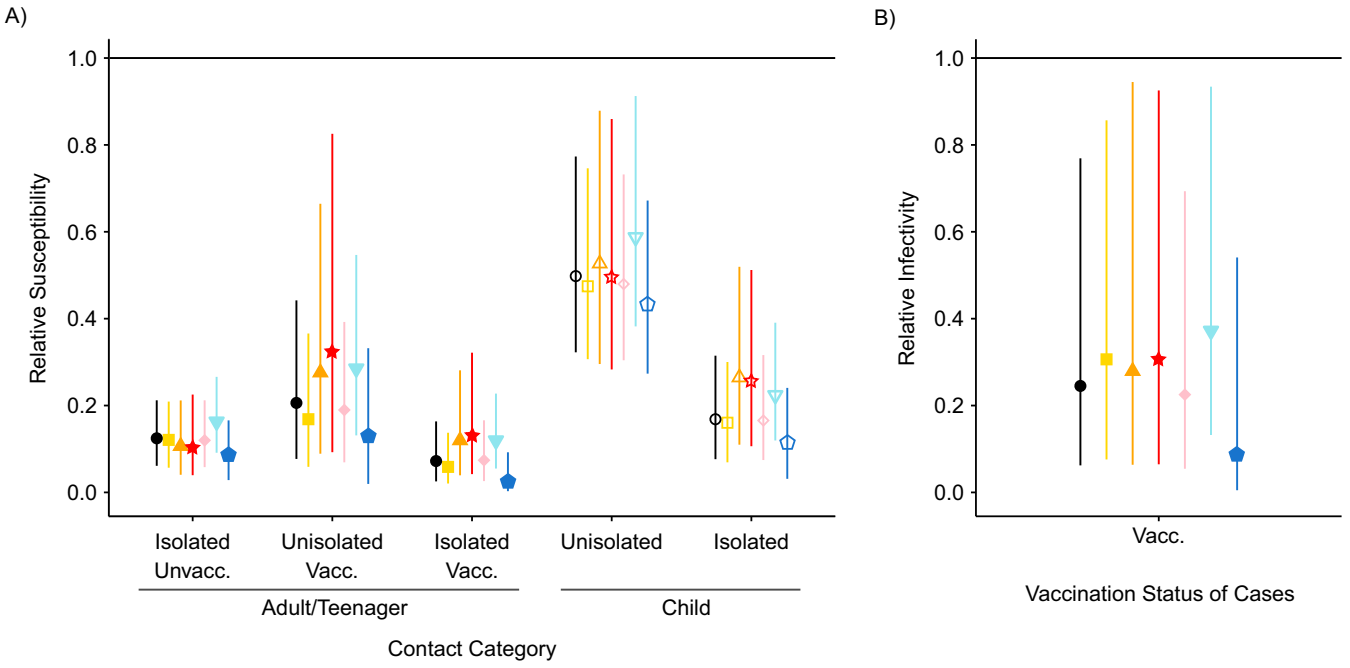


Figure 4. Impact of model assumptions on the estimation of the relative susceptibility and relative infectivity parameters, Ramat Gan, Israel, 2020–2021. A) Estimates of the relative susceptibility of household contacts for the baseline and sensitivity analysis scenarios. B) Estimates of the relative infectivity of vaccinated cases compared with unvaccinated ones for the baseline and sensitivity analysis scenarios. In the baseline scenario (black circle), we assumed that vaccination was effective from 7 days after the second dose, the relative infectivity of asymptomatic cases compared with symptomatic cases was equal to 60%, and the log-standard deviation of the relative infectivity and relative susceptibility prior distributions was equal to 1. Sensitivity analysis scenarios: yellow square, vaccination is effective ≥15 days after the dose; orange triangle, 1 polymerase chain reaction (PCR) test for all negative contacts; red star, 2 PCR tests for all negative contacts; pink diamond, 100% infectivity of asymptomatic cases; blue inverted triangle, relative parameter prior with log-standard deviation = 0.7; blue pentagon, relative parameter prior with log-standard deviation = 2. The posterior median and its associated 95% Bayesian credible interval are reported.

the index case(s), and varying follow-up periods between households.

In our analysis, the SAR in unvaccinated adult/teenager contacts who did not isolate was estimated at around 76%, which is substantially higher than previous estimates obtained in household settings (12–14, 18, 27, 28). In meta-analyses (12–14), the average SAR ranged between 14% and 32%; however, in some studies, it could be as high as 90% (13). Most of these studies date back to the time when historical lineages were still dominant. In contrast, our study took place when the Alpha variant represented up to 90% of infections in Israel (21). Our higher estimate could be at least partly explained by the fact that the Alpha variant is substantially more transmissible than historical lineages (21, 29–31).

In agreement with previous reports, we found that children are less susceptible to SARS-CoV-2 infections than adults/teenagers (12–14, 32). We further estimated that, 7 days after their second dose, vaccinated adults/teenagers benefit from a 79% reduction in the risk of infection compared with unvaccinated adults/teenagers. We show, consistent with previous studies (21, 33), that BNT162b2 vaccination is highly effective against infection by the Alpha variant. In general population studies, vaccine effectiveness for symptomatic infections ranged from 57% 14 days after the first dose (4) to 89% (4), and 97% 7 days after the second dose (9). For asymptomatic infections, vaccine effectiveness against infection was 79% 10 days after the first dose (5) and 94% 14 days after the second dose (7). Our estimate of vaccine effectiveness in household settings is lower than those obtained in the general population. This is consistent with estimates obtained in households (19, 20, 33) and might in part be explained by the elevated contact rates in households that may favor transmission. Additionally, studies in the general population are less suitable to detect all asymptomatic cases compared with the household setting. This might lead general population studies to overestimate vaccine effectiveness against asymptomatic infections if vaccinated contacts are less often tested than unvaccinated ones. On another note, we estimate a vaccine effectiveness against transmission of 75% (95% CrI: 23, 94), which is in line with other studies in household settings (18–20).

To our knowledge, this is the first study estimating the effect of isolation on SARS-CoV-2 transmission in households that are partially vaccinated. We showed that isolation precautions markedly reduce the overall infection risk in both adult/teenager and child contacts even when considering partial physical distancing measures. We estimated a similar reduction of infection in adult/teenager contacts that were vaccinated but did not isolate. There was a signal in

the data that isolation also benefited vaccinated individuals, although credible intervals were larger, and further investigations are required to confirm this finding.

Our study has several limitations. First, household studies such as ours may be affected by multiple sources of bias. On the one hand, we may overestimate the SAR if we are more likely to detect households with multiple cases. On the other hand, we might underestimate it if some asymptomatic, or paucisymptomatic, cases are missed during follow-up. Second, we estimated an important reduction of infectivity in vaccinated cases with 2 doses compared with unvaccinated cases as previously shown (18–20, 34). However, this is associated with important uncertainty due to the small number of cases (15 vaccinated index cases and 21 vaccinated secondary cases). Thus, more data are needed to reduce the size of credible intervals. Third, we assumed that vaccination was effective from 7 days after the second dose (or 15 days after the first dose in our sensitivity analysis; see Web Table 11 in Web Appendix 8). In practice, the effect of the vaccine is likely to be progressive, which might push down estimates of effectiveness since individuals with early partial protection would be considered to be unvaccinated. However, excluding households with the early-vaccinated index cases did not affect our estimates (Web Figure 4 and Web Table 12 in Web Appendix 8). The limited number of households does not make it possible to dissociate early vs. full protection conferred by the vaccine nor to investigate the infectivity of children relative to adults/teenagers. Fourth, testing instructions were different for vaccinated and unvaccinated household contacts, as well as HCWs and non-HCWs. Most vaccinated contacts were HCWs at the Sheba Medical Center who complied with testing instructions to go back to work, leading to high testing rates in vaccinated individuals, with 67% having at least 2 PCR tests and 70% having 1 positive PCR or at least 2 PCR tests in the 10 days following case detection (Web Table 1 in Web Appendix 1). Among unvaccinated contacts, 49% had at least 2 PCR tests and 79% had 1 positive PCR or at least 2 PCR tests in the 10 days following case detection. This higher testing rate is notably due to the high proportion of single positive tests (30%). These differential testing behaviors and positivity rates between vaccinated, unvaccinated, HCW, and non-HCW contacts make it difficult to anticipate the directionality of a potential bias. When restricting our evaluation to households where all negative contacts were tested at least once or twice, estimates remained relatively similar to the baseline values. In the analysis with at least 2 tests for all negative contacts, we observed a slight reduction in the point estimate for vaccine effectiveness against infection that remained difficult to interpret given the very broad credible intervals (17%–91%). Fifth, the measurement of isolation precautions may be subject to recall bias and/or overreporting, as they represent a socially desirable behavior. The timing and evolution of isolation precautions were not measured, and thus not integrated in our model. Nevertheless, our estimate of isolation effectiveness is consistent with a 10-day period of quarantine in modeling studies (35). Finally, we estimate vaccine effectiveness against infection and transmission in a context where the Alpha variant was dominant. These estimates are very likely to be different for the Delta variant (36) that was first reported in October 2020 and rapidly became dominant worldwide (37).

To conclude, vaccination with 2 doses substantially reduces the risk of transmission and the risk of infection in households. Isolation from the index case while sleeping and eating provides a high level of protection to unvaccinated household members, whether they are adults/teenagers or children. Household contacts of COVID-19 patients should ideally isolate, or at least refrain from significant contact, with household cases. This may also be the case for vaccinated household members, although larger studies are required to confirm this finding.

ACKNOWLEDGMENTS

Author affiliations: Mathematical Modelling of Infectious Diseases Unit, Institut Pasteur, Université de Paris Cité, UMR2000, Centre national de la recherche scientifique (CNRS), Paris, France (Maylis Layan, Alessio Andronico, Nathanaël Hozé, Simon Cauchemez); Collège Doctoral, Sorbonne Université, Paris, France (Maylis Layan); Infectious Disease Unit, Sheba Medical Centre, Ramat-Gan, Israel (Mayan Gilboa); Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel (Mayan Gilboa, Tal Gonen, Miki Goldenfeld, Lilac Meltzer, Gili Regev-Yochay); and Infection Prevention and Control Unit, Sheba Medical Center, Ramat-Gan, Israel (Tal Gonen, Miki Goldenfeld, Lilac Meltzer, Gili Regev-Yochay).

M.L. and M.G. contributed equally to the work as first authors. S.C. and G.R.-Y. contributed equally as senior authors.

This work was supported by Sheba Medical Center, Ramat-Gan, Israel. S.C. acknowledges financial support from the Investissement d'Avenir program, the Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases program (grant ANR-10-LABX-62-IBEID), the National Research Agency (ANR) through the ANR-Flash call for COVID-19 (grant ANR-20-COVI-0018), the EMERGEN project (ANRS0151), Haute Autorité de Santé (HAS), the INCEPTION project (grant PIA/ANR-16-CONV-0005), the European Union's Horizon 2020 research and innovation program (grants RECOVER 101003589 and VEO 874735), AXA, and Groupama.

Data and code are available online at https://github.com/mlayan/VaccineEffectivenessSheba.

We acknowledge the Infection Prevention and Control Unit team for their devoted work, the extensive epidemiologic investigations and contact tracing from which the data was derived. We thank the Sheba management for their support of this study.

Conflict of interest: none declared.

REFERENCES

1. Lu R, Zhao X, Li J, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for

virus origins and receptor binding. *Lancet.* 2020;395(10224):565–574.
2. Izda V, Jeffries MA, Sawalha AH. COVID-19: a review of therapeutic strategies and vaccine candidates. *Clin Immunol.* 2021;222:108634.
3. Mathieu E, Ritchie H, Ortiz-Ospina E, et al. A global database of COVID-19 vaccinations. *Nat Hum Behav.* 2021;5(7):947–953.
4. Dagan N, Barda N, Kepten E, et al. BNT162b2 mRNA Covid-19 vaccine in a nationwide mass vaccination setting. *N Engl J Med.* 2021;384(15):1412–1423.
5. Tande AJ, Pollock BD, Shah ND, et al. Impact of the coronavirus disease 2019 (COVID-19) vaccine on asymptomatic infection among patients undergoing preprocedural COVID-19 molecular screening. *Clin Infect Dis.* 2021;1–7.
6. Pawlowski C, Lenehan P, Puranik A, et al. FDA-authorized mRNA COVID-19 vaccines are effective per real-world evidence synthesized across a multi-state health system. *Med.* 2021;2(8):979–992.e8.
7. BioNTech SE. Real-world evidence confirms high effectiveness of Pfizer-BioNTech COVID-19 vaccine and profound public health impact of vaccination one year after pandemic declared. 2021; https://www.globenewswire.com/fr/news-release/2021/03/11/2191164/0/en/Real-World-Evidence-Confirms-High-Effectiveness-of-Pfizer-BioNTech-COVID-19-Vaccine-and-Profound-Public-Health-Impact-of-Vaccination-One-Year-After-Pandemic-Declared.html. Accessed June 7, 2021.
8. Martínez-Baz I, Miqueleiz A, Casado I, et al. Effectiveness of COVID-19 vaccines in preventing SARS-CoV-2 infection and hospitalisation, Navarre, Spain, January to April 2021. *Eurosurveillance.* 2021;26(21):2100438.
9. Haas EJ, Angulo FJ, McLaughlin JM, et al. Impact and effectiveness of mRNA BNT162b2 vaccine against SARS-CoV-2 infections and COVID-19 cases, hospitalisations, and deaths following a nationwide vaccination campaign in Israel: an observational study using national surveillance data. *Lancet.* 2021;397(10287):1819–1829.
10. Goldberg Y, Mandel M, Woodbridge Y, et al. Protection of previous SARS-CoV-2 infection is similar to that of BNT162b2 vaccine protection: a three-month nationwide experience from Israel [preprint]. *medRxiv*. 2021. https://doi.org/10.1101/2021.04.20.21255670. Accessed November 19, 2021.
11. Bi Q, Wu Y, Mei S, et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *Lancet Infect Dis.* 2020;20(8):911–919.
12. Madewell ZJ, Yang Y, Longini IM, et al. Household transmission of SARS-CoV-2: a systematic review and meta-analysis. *JAMA Netw Open.* 2020;3(12):e2031756.
13. Lei H, Xu X, Xiao S, et al. Household transmission of COVID-19—a systematic review and meta-analysis. *J Infect.* 2020;81(6):979–997.
14. Thompson HA, Mousa A, Dighe A, et al. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) setting-specific transmission rates: a systematic review and meta-analysis. *Clin Infect Dis.* 2021;73(3):e754–e764.
15. World Health Organization. Cross-country analysis. https://analysis.covid19healthsystem.org/index.php/2020/05/19/how-do-measures-for-isolation-quarantine-and-contact-tracing-differ-among-countries/. Accessed June 7, 2021.
16. Shah ASV, Gribben C, Bishop J, et al. Effect of vaccination on transmission of SARS-CoV-2. *N Engl J Med.* 2021;385(18):1718–1720.
17. Salo J, Hägg M, Kortelainen M, et al. The indirect effect of mRNA-based Covid-19 vaccination on unvaccinated household members [preprint]. *medRxiv.* 2021. https://doi.org/10.1101/2021.05.27.21257896. Accessed January 31, 2022.
18. Harris RJ, Hall JA, Zaidi A, et al. Effect of vaccination on household transmission of SARS-CoV-2 in England. *N Engl J Med.* 2021;385(8):759–760.
19. Prunas O, Warren JL, Crawford FW, et al. Vaccination with BNT162b2 reduces transmission of SARS-CoV-2 to household contacts in Israel. *Science.* 2022;375(6585):1151–1154.
20. de Gier B, Andeweg S, Joosten R, et al. Vaccine effectiveness against SARS-CoV-2 transmission and infections among household and other close contacts of confirmed cases, the Netherlands, February to May 2021. *Eurosurveillance*. 2021;26(31):2100640.
21. Munitz A, Yechezkel M, Dickstein Y, et al. BNT162b2 vaccination effectively prevents the rapid rise of SARS-CoV-2 variant B.1.1.7 in high-risk populations in Israel. *Cell Reports Med.* 2021;2(5):100264.
22. Liu Y, Eggo RM, Kucharski AJ. Secondary attack rate and superspreading events for SARS-CoV-2. *Lancet.* 2020;395(10227):e47.
23. Cauchemez S, Carrat F, Viboud C, et al. A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. *Stat Med.* 2004;23(22):3469–3487.
24. McAloon C, Collins Á, Hunt K, et al. Incubation period of COVID-19: a rapid systematic review and meta-analysis of observational research. *BMJ Open.* 2020;10(8):1–9.
25. Cevik M, Tate M, Lloyd O, et al. SARS-CoV-2, SARS-CoV, and MERS-CoV viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and meta-analysis. *The Lancet Microbe.* 2021;2(1):e13–e22.
26. Byambasuren O, Cardona M, Bell K, et al. Estimating the extent of asymptomatic COVID-19 and its potential for community transmission: systematic review and meta-analysis. *Off J Assoc Med Microbiol Infect Dis Canada.* 2020;5(4):223–234.
27. Jing Q-L, Liu M-J, Zhang Z-B, et al. Household secondary attack rate of COVID-19 and associated determinants in Guangzhou, China: a retrospective cohort study. *Lancet Infect Dis.* 2020;20(10):1141–1150.
28. Li F, Li YY, Liu MJ, et al. Household transmission of SARS-CoV-2 and risk factors for susceptibility and infectivity in Wuhan: a retrospective observational study. *Lancet Infect Dis.* 2021;21(5):617–628.
29. Kissler SM, Fauver JR, Mack C, et al. Viral dynamics of SARS-CoV-2 variants in vaccinated and unvaccinated individuals [preprint]. *medRxiv.* 2021. https://doi.org/10.1101/2021.02.16.21251535. Accessed November 19, 2021.
30. Davies NG, Kucharski AJ, Eggo RM, et al. Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. *Lancet Public Health.* 2020;5(7):e375–e385.
31. Volz E, Mishra S, Chand M, et al. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature.* 2021;593(7858):266–269.
32. Viner RM, Mytton OT, Bonell C, et al. Susceptibility to SARS-CoV-2 infection among children and adolescents compared with adults. *JAMA Pediatr.* 2021;175(2):143–156.

33. Pritchard E, Matthews PC, Stoesser N, et al. Impact of vaccination on new SARS-CoV-2 infections in the United Kingdom. *Nat Med.* 2021;27(8):1370–1378.
34. Regev-Yochay G, Amit S, Bergwerk M, et al. Decreased infectivity following BNT162b2 vaccination: a prospective cohort study in Israel. *Lancet Reg Heal - Eur.* 2021;7(5): 100150.
35. Ashcroft P, Lehtinen S, Angst DC, et al. Quantifying the impact of quarantine duration on covid-19 transmission. *Elife.* 2021;10:1–33.
36. Eyre DW, Taylor D, Purver M, et al. The impact of SARS-CoV-2 vaccination on alpha and Delta variant transmission [preprint]. *medRxiv*. 2021. https://doi.org/10.1101/2021.09.28.21264260. Accessed November 19, 2021.
37. World Health Organization. Weekly epidemiological update on COVID-19 - 13 October 2021. (COVID-19 Weekly Epidemiological Update, edition 61). https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---13-october-2021. Accessed November 19, 2021.

American Journal of Epidemiology

Vol. 00, No. 00

Correction

CORRECTION TO: IMPACT OF BNT162B2 VACCINATION AND ISOLATION ON SARS-COV-2 TRANSMISSION IN ISRAELI HOUSEHOLDS: AN OBSERVATIONAL STUDY

In the article "Impact of BNT162b2 Vaccination and Isolation on SARS-CoV-2 Transmission in Israeli Households: An Observational Study" by Layan et al. (1), there was an error in the code of the Markov chain Monte Carlo inference model that led to minor errors in the estimates of transmission parameters. In our analysis, inference is performed in continuous time. However, in our code, the probability of infection of detected cases was expressed in discrete time, $1-\exp\left(-\lambda\left(t_{infection}\right)\right)$, instead of continuous time, $\lambda\left(t_{infection}\right)$. We have now corrected this error.

Table 1 shows a comparison of the uncorrected and corrected median estimates and 95% credible intervals reported at the end of the abstract and in the last 3 paragraphs of the results.

Overall, the correction of the inference model code had only a very minor impact on estimates. It slightly modified the posterior probabilities that isolated and vaccinated adult/teenager contacts are less susceptible than are unisolated and vaccinated adult/teenager contacts in the sensitivity analyses. We reported a posterior probability of 88% with a Bayes factor (BF) of 7 with 1 polymerase chain reaction (PCR) test and a posterior probability of 89% with a BF of 8 with 2 PCR tests, whereas the correct estimates are 90% probability and a BF equal to 9 with 1 PCR test and a posterior probability of 91% and a BF of 10 with 2 PCR tests.

Finally, we reported in the discussion that the reduction in the risk of infection in vaccinated adults/teenagers compared to that in unvaccinated adults/teenagers was 79% instead of 80% and that the vaccine effectiveness against transmission was 75% (95% credible interval: 23, 94) instead of 76% (95% credible interval: 31, 94). We also referred to the broad credible interval of vaccine effectiveness against infection that was 17%–91% instead of 24%–91%.

The code published on GitHub has been corrected. The Web Material detailing all parameters values have been corrected as well.

The authors regret these errors.

Reference

1. Layan M, Gilboa M, Gonen T, et al. Impact of BNT162b2 vaccination and isolation on SARS-CoV-2 transmission in Israeli households: an observational study. *Am J Epidemiol.* 2022;191(7):1224–1234.

https://doi.org/10.1093/aje/kwac119; Advance Access publication:

Table 1. Comparison of the Uncorrected and Corrected Median Estimates and 95% Credible Intervals

Parameter	Scenario	Uncorrected		Corrected	
		Median	95% CrI	Median	95% CrI
Relative susceptibility (reference: unisolated and unvaccinated adult/teenager)					
Unisolated and vaccinated adult/teenager	Baseline	0.21	0.08–0.44	0.20	0.08–0.41
	1 PCR test	0.28	0.09–0.66	0.26	0.08–0.61
	2 PCR tests	0.32	0.09–0.83	0.30	0.09–0.76
Isolated and unvaccinated adult/teenager	Baseline	0.12	0.06–0.21	0.12	0.06–0.20
	1 PCR test	0.12	0.04–0.28	0.11	0.04–0.25
	2 PCR tests	0.13	0.04–0.32	0.12	0.04–0.27
Isolated and vaccinated adult/teenager	Baseline	0.07	0.03–0.16	0.07	0.02–0.15
Unisolated child	Baseline	0.50	0.32–0.77	0.48	0.31–0.73
Isolated child	Baseline	0.17	0.08–0.32	0.16	0.07–0.30
Relative susceptibility (reference: unisolated and vaccinated adult/teenager)					
Isolated and vaccinated adult/teenager	Baseline	0.34	0.11–0.14	0.34	0.10–0.13
Relative infectivity (reference: unisolated and unvaccinated adult/teenager)					
Vaccinated case	Baseline	0.25	0.06–0.77	0.25	0.06–0.69
Transmission probabilities, %					
Unvaccinated case to unvaccinated and unisolated adult/teenager	Baseline	61	48–72	66	53–76
Vaccinated case to unisolated and vaccinated adult/teenager	Baseline	4	1–16	5	1–15
Unvaccinated case to unisolated child	Baseline	37	27–48	40	29–50
Vaccinated case to unisolated child	Baseline	11	3–31	11	3–30
Unvaccinated case to isolated child	Baseline	14	7–25	16	8–26

Abbreviations: CrI, credible interval; PCR, polymerase chain reaction.