Quantifying SARS-CoV-2 and mpox transmission patterns through phylodynamic inference

Miguel Ignacio Paredes

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2024

Reading Committee:

Trevor Bedford, Chair

M. Elizabeth Halloran

John T. McCrone

Anna Wald

Program Authorized to Offer Degree:
Epidemiology

University of Washington

Abstract

Quantifying SARS-CoV-2 and mpox transmission patterns through phylodynamic inference

Miguel Ignacio Paredes

Chair of the Supervisory Committee:
Trevor Bedford
Department of Epidemiology

Emerging infectious diseases represent an urgent public health challenge. Unequal coverage of public health surveillance as well as asymptomatic spread, however, limit our ability to respond to outbreaks in a precise and timely manner. In this dissertation, I describe how genomic epidemiology can aid traditional public health investigations of emerging infectious disease dynamics. I begin by describing how matching epidemiological and genomic data from a genomic surveillance system allows for the quantification of variant-specific effects of SARS-CoV-2 infection on the risk of hospitalization, and how vaccination modifies that risk. The subsequent chapters represent phylodynamic studies of mpox and SARS-CoV-2, which show how incorporating epidemiological and mobility information into phylodynamic analyses allows for more precise examination of within- and between-region transmission dynamics, both on a global and local scale. I use these phylodynamic models to investigate the impact of infection control measures, such as stay-at-home orders or vaccination campaigns, on curbing disease spread. Collectively, this dissertation highlights the utility of robustly joining epidemiological and genomic data to augment outbreak response, especially in support of marginalized communities that are especially vulnerable to emerging pathogens.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Science is a collaborative effort; similarly, the work described herein would not have been possible without the help and support of so many people.

I want to first start off by thanking my PhD mentor, Trevor Bedford. Trevor, you have been a guiding example of what it means to be both brilliant and kind. Your focus on careful, impactful science that is open and accessible to all has and will continue to serve as a model for my career. Your support often extended beyond my science and helped me along this journey more than I can describe. Thank you tremendously for guiding me as a scientist and for believing in me, even when I sometimes wavered in doing so myself.

I am beyond grateful for the support and guidance of my supervisory committee: Betz Halloran, Erick Matsen, JT McCrone, and Anna Wald. Betz and Anna - thank you for inspiring me way before we even met, for believing in me as a scientist and a future clinician, and for letting me stand on the shoulders of giants. Anna, you have served as a role model of a physician-scientist and have pushed me to always think of the community and social aspects of infectious diseases. Erick, since my rotation in your lab in 2020 you have been a thoughtful mentor and an enthusiastic supporter of my work – thank you so much. And JT, I am so appreciative of all the time spent with me troubleshooting my models, chatting about big and exciting science ideas, and trying to find ways to sneak silly puns into our work. Thank you for making my science so much better and bit more lighthearted.

I also want to thank all the members of the Bedford lab, both past and present. I am so incredibly lucky to have been surrounded by such a brilliant group of individuals who constantly strive to improve public health through thoughtful, careful, caring work. You

all have taught me that science is very much a team sport – and it's so much more fun that way. To the Bedford lab alumni that become both mentors and friends: thank you for being inspiring role models to follow and to learn from. To Alli Black, you have been one of my few role models that center epidemiology in "genomic epidemiology" and our monthly meetings made me a better public health scientist, demystified the trial and tribulations of graduate school, and made this journey so much more joyous. To Nicola Müller, a large part of this dissertation would not have been possible without your guidance. Thank you for patiently answering every single one of my incessant questions. If I can profess to work on phylodynamics, it's thanks to you. To Louise Moncla, thank you for your giving advice, for believing in my work, and for teaching me to "fail upwards", which throughout this journey has ended up teaching me to believe in myself as well.

It has taken a village to support and guide me to this point. To previous science mentors – Brian Ramos, Roko Rasin, Nenad Sestan, Andrew Tebbenkamp, Nancy Gore Saravia, Alejandro Vargas, Yusuf Ransome, and Albert Ko – thank you for taking a chance on me. You all saw the potential for a budding scientist in me and somehow, helped me see it too. Thank you for teaching me to ask impactful questions, to design careful experiments, and to live life fully along the way.

A special thanks to my MD-PhD cohort, who started this really long journey with me and have been a beautiful source of inspiration, camaraderie, so many laughs, and even a shoulder to cry on. Thanks for making this really long journey a little less lonely, and a lot more fun along the way.

To my friends, both nearby and far away – The Gaggle, Alex Willcox, Liliana Cruz, Savannah Perez, Kar Jin Ong, Azan Virji, Becky Byler, Rumya Raghavan, Alero Egbe, Ted Gobillot, Alec Gibson, and so many others – you all have supported this science journey through every single stage. This work, journey, and so much more would have been impossible

without you all picking me up after every fall and celebrating every single one of my small victories. It takes a village and I'm so glad you all make up mine.

To Christian, thank you for being my partner-in-crime, a soft shoulder to support myself on, and a constant source of many, many smiles. Thank you for putting up with, and even often encouraging, all the winding paths this science journey has taken us. Every single part of my life so much brighter, sillier, and full of joy because I live it hand-in-hand with you. Thank you, for everything.

To my extended family, *mil gracias por ser una fuente tan grande de apoyo desde lejos. Gracias por celebrar cada uno de mis logros; ninguno de ellos serían posibile sin el amor de todos ustedes.*

And finally, to my family, to my parents and sister, *gracias.* It is thanks to every sacrifice, big or small – from migrating across continents in search of safety and opportunity to staying up late to help me with a science project – that I was even able to dream this dream and make it a reality. Everything I do is thanks to and in honor of you.

Chapter 1

INTRODUCTION

1.1 Emerging Infectious Diseases

In late 2019, an outbreak of severe pneumonia of unknown etiology arose in Wuhan, China. The causative virus, which was quickly discovered to be a novel coronavirus and named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), disseminated rapidly around the world, disrupting daily life and devastating communities globally [26]. In 2022, individuals in the United Kingdom without a history of travel to endemic regions were diagnosed with mpox (formally monkeypox), a disease characterized by fever and rash that was historically restricted to Western and Central Africa and caused limited human-to-human transmission. In just a few months, mpox had spread throughout Europe and the Americas, establishing itself within queer sexual networks and causing severe pain and stigmatizing lesions to thousands [146].

As highlighted by SARS-CoV-2 and mpox, emerging infectious diseases represent one of the most urgent public health challenges of the twenty-first century. Emerging infectious diseases are those that have appeared and affected a population for the first time, or have existed previously but are now spreading rapidly, either in terms of new cases within the same population or spread to new areas and communities[60]. There is no dearth of recent examples – SARS-CoV-2, MERS, Ebola, Zika – all of which have had a substantial impact on human health despite ongoing pandemic and epidemic preparedness [18, 42, 41, 48].

Most emerging infectious diseases are zoonotic in origin, meaning that they arose from

human interaction with animal reservoir hosts. Human population growth, rapid urbanization and deforestation, increased global connectivity and travel, wars, conflict, displacement, microbial evolution, and climate change, among others, all increase the risk of spillover of an infectious pathogen into humans from infected animals [71]. Given that emerging infectious diseases often represent a novel introduction into a human population, the lack of preexisting immunity to the pathogen results in significant morbidity and mortality. In addition to the lack of preexisting immunity, emerging infectious diseases exploit structural systems of social inequity, whereby the most marginalized communities are often those that bear a disproportionate burden of disease [49]. The multi-factorial system that gives rise to emerging pathogens makes predicting the emergence and impact of new infectious diseases a complex and difficult task, but the significant human cost of epidemics demands that our efforts to detect and control these diseases be rapid, robust, scaleable, and evidence-based [60, 71]. Detection of an emerging pathogen requires robust public health surveillance, or the ongoing and systematic collection of samples or data, analysis, and subsequent dissemination regarding a health event [116]. Detection, however, is often the beginning. Upon emergence and detection of a pathogen, it is imperative to characterize the identity of the pathogen, where it arose from, where it is spreading to, who it is affecting, and how it might affect others in the future, among other questions. Historically, these questions were addressed solely through the use of case data, but unequal coverage of public health surveillance around the world limits the ability of case counts to inform outbreak response in a precise and timely manner [60, 71]. The need for new methods to tackle emerging infectious has led public health to focus on molecular diagnostics and surveillance, notably pathogen genome sequencing. Pathogen genomic data, when paired with traditional epidemiological information, has the ability to augment public health surveillance, allowing for earlier detection of emerging pathogens and greater precision in outbreak investigations.

1.2 Genomic Epidemiology

The science of using pathogen genomics to understand infectious disease emergence and transmission is termed "genomic epidemiology" [71]. Pathogen genomes can be used on a small scale to identify person-to-person transmission events, or on a population scale to understand the overall transmission dynamics of an epidemic (Fig 1.1). While genomic epidemiology has been leveraged for a wide range of pathogens, it has historically been most informative for viral outbreaks [154], which will be the focus of this dissertation.

As viruses replicate within a host, they accumulate mutations on their genomes due to errors arising from their respective polymerases (for example, RNA viruses use RNA- dependent RNA polymerases which lack proofreading abilities, allowing for the rapid accumulation of mutations) [87]. When an infection event occurs, the within-host viral diversity of the infector is sampled and transmitted to the infectee, resulting in both infected individuals sharing mutations that arose in the infector [103, 39].

While the probability of mutation on a pathogen genome is generally dependent on specific pathogen factors, many mutations are detrimental to the survival of the pathogen and thus are never or very rarely observed, resulting in the majority of genetic diversity observed being a consequence of neutral or nearly neutral evolution. When looking at a large population of viruses over time, the rate of mutational accumulation on the pathogen genome is often relatively constant [154]. Estimation of the evolutionary rate at which a pathogen accumulates mutations on its genome is not just affected by pathogen factors but also by host immune selective pressure and even the rate at which pathogens are sampled in a population [154, 17, 50].

The signal of evolution of a pathogen population throughout time is termed the "molecular clock" [71]. The molecular clock translates the number of observed mutations into calendar time, allowing us to estimate the amount of time that has needed to elapse in order to explain the observed genetic diversity. Through the estimate of elapsed time, we

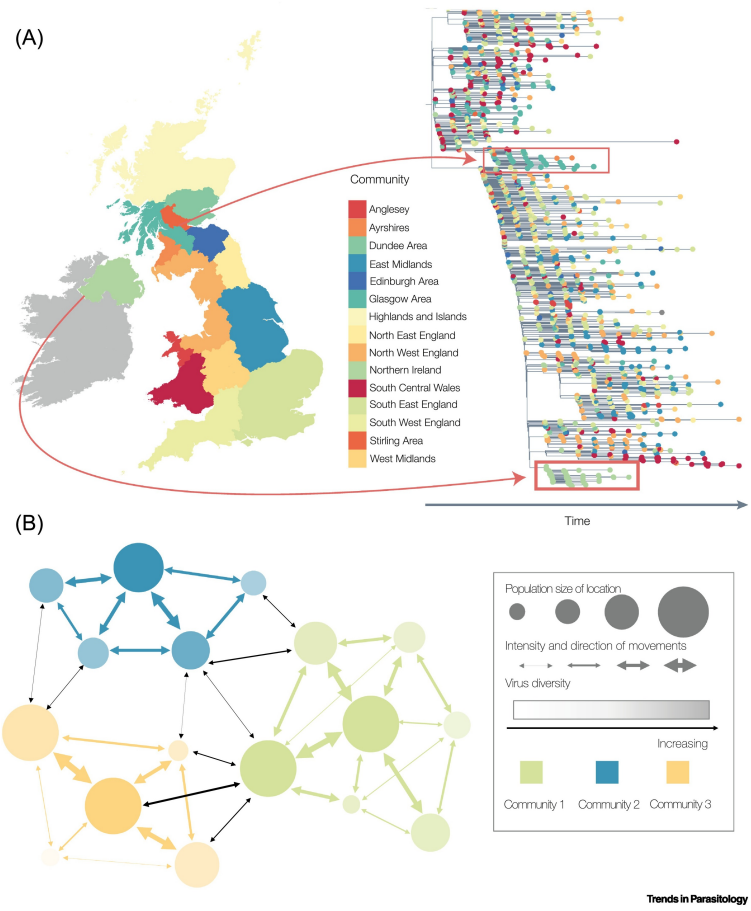

Figure 1.1: Genomic epidemiology at various scales. A) Maps of hypothetical groupings of network communities in the UK with corresponding (illustrative) phylogenetic tree. Areas with the same color represent communities that are grouped. The two highlighted areas show sections of the tree where the virus has entered a community and then spread within the community. (B) Hypothetical representation of network communities. Each circle represents a location, colored by movement community. Arrows represent direction and intensity of movements between locations, and locations are grouped into communities based on these movement patterns. Circle size represents the population size within each location. Circle shading represents the virus diversity in each location in a hypothetical outbreak – well-connected locations with large population sizes have the highest diversity and poorly-connected locations with small population sizes have the lowest., Adapted from Fig 1 of Hill et al. [76]

can estimate important information for outbreak investigation such as when our sample of sequenced viruses last shared a common genetic ancestor, which is informative of when the outbreak first occurred [71].

Given the relatively constant evolutionary rate, we can group sampled viruses based on genomic similarity and reconstruct their ancestral genetic relationships via a viral phylogeny. The fast mutation rate of viruses results in the rate of evolution and the transmission dynamics of a pathogen falling on comparable timescales, such that the viral phylogeny also contains information about transmission throughout the course of an outbreak [155].

1.3 Phylodynamics

The study of using phylogenies to understand the epidemiological, immunological and evolutionary processes influencing patterns of genetic variation in pathogen populations is termed "phylodynamics" [68]. Phylodynamic analyses have allowed for the estimation of pathogen population size, quantification of viral importations into and between regions, identification of consequential mutations, and investigations into within host evolution, among others [11]. Phylodynamic estimation of transmission dynamics within and between different populations is often modelled via the coalescent.

1.3.1 The Coalescent

We can derive pathogen population size as a function of genetic diversity in a population modeled by the coalescent. Consider an idealized Wright-Fisher model of haploid organisms that assumes a neutral population of constant size N and exhibits complete mixing and discrete generations. When we look forward in time, given that reproductive success is random in this idealized population, we can expect genetic drift – the random loss of genetic lineages. Looking backwards in time, however, we can see lineages *coalesce* whenever two or more individuals were produced by the same parent, creating the basis for the *coalescent*

[82, 156].

In a Wright-Fisher model with a constant population size of size N, there are N possible parents, or ancestors, from which to choose from, each with a probability of $1/N$ to be chosen. To have a coalescent event, however, it is required for two lineages to coalesce, or choose the same parent. If we have a sample of size n, then we have $\frac{n(n-1)}{2}$ possible pairs of random lineages that could pick the same ancestry with a probability of $1/N$. Thus, the probability of any coalescent event happening in a generation, also known as the coalescent rate, would be the product:

$$\frac{n(n-1)}{2} * \frac{1}{N} = \binom{n}{2} * \frac{1}{N}.$$

From this equation, we can see that the coalescent rate is inversely proportional to the size of the population N, meaning that a larger population size would result in a smaller coalescent rate. This principle can be extended to pathogen genomics. If the genetic diversity of a viral population is large, we then expect that it takes longer for two randomly selected lineages to coalesce, thus by knowing the coalescent rate, under the aforementioned assumptions, we can derive the pathogen population size. Changes of population size, via the coalescent rate, result in changes in the inferred genealogy. This forms the basis of subsequently developed demographic models in phylogenetics that derive the changes in population size over time via the changes in coalescent rates observed in a viral genealogy over time (See Figure 1.2).

In reality, the Wright-Fisher assumptions are rarely met, resulting in a smaller estimated population size deemed the *effective population size*, N_e. Formally, N_e is defined to be the size of a Wright Fisher population that would produce the same coalescent rate as the population of interest. By extending the Wright Fisher model to account for unequal offspring distributions, we can express N_e as N divided by the variance of the offspring distribution σ^2. If a population size changes over time, then the coalescent rate will also be a function of time. We can modify the above equation for the coalescent to account for both an unequal

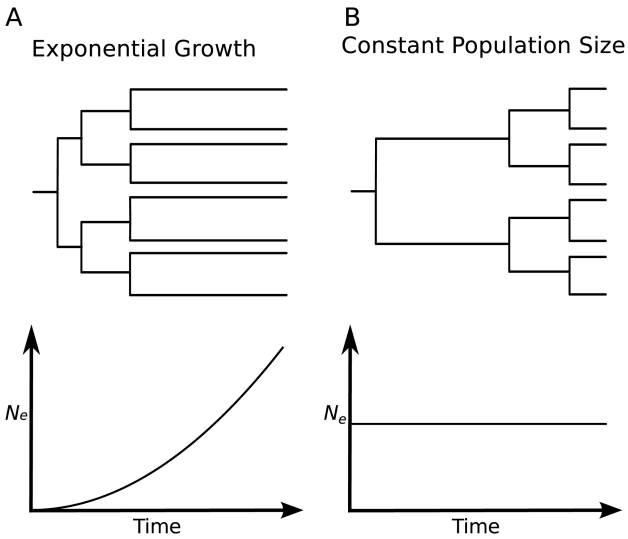

Figure 1.2: Impact of population size changes on viral phylogenies. The figure shows idealized caricatures of virus phylogenies that show the effects of changes in viral population size. Adapted from Figure 1 of Volz et al. [154]

offspring distribution and the time of each Wright Fisher generation [155], as

$$\lambda = \binom{n}{2} * \frac{1}{N_e \tau}$$

Based on this formation, we can see that both N_e and τ drive the coalescent rate and subsequently, only the product $N_e\tau$ can be derived directly from the inferred genealogy of pathogen genomes. To derive N_e, we then have to divide $N_e\tau$ by τ. τ represents the generation time, often derived in terms of years per generation, which in pathogen genomics is represented by the transmission interval, or the time between successive infections. This is different from the serial interval, which represents the time between symptom onset of an infector-infectee pair, which is easier to observe in practice but is often dependent on host factors [90].

The above formulation assumes a population with equal mixing, which is often unrealistic. The structured coalescent was subsequently developed in order to account for different population dynamics both within and between different subpopulations.

1.3.2 The Structured Coalescent

Briefly, the structured coalescent is a statistical model that describes the genealogy of individuals samples from a structured population that evolves according to a migration matrix [153]. A structured population refers to the existence of subpopulations, or demes, that are stable in size over time. Some of the key assumptions are that migration occurs at a constant rate over time, there is no substructure *within* demes, and that within each deme, individuals are sampled at random. A key strength of the structured coalescent is that it models both demographic processes within and between subpopulations, allowing for the estimation of both effective population sizes and migration rates between demes [101].

The study and reconstruction of evolutionary history and spatial diffusion between demes from sequence information is called phylogeography, whereby inferring the location of an ancestral node can provide insights into its dispersal over time. The most common method for phylogeography is called discrete trait analysis (DTA), which models the evolution of a discrete state on a phylogeny in the same way that sequence evolution is modeled [92]. DTA relies on key assumptions: each state evolves independent of each other and the evolution along different branches of the phylogeny are independent as well [92, 165]. For these assumptions to hold, the tips of the phylogeny must be sampled at random and in proportion to the relative size of each state. It also implies that the migration process is independent of the demographic tree building process, ignoring their interdependence. There's a strong consequence to this as it causes DTA to treat the relative sampling intensities of each different state as informative of the migration parameters [101].

In reality, it is extremely rare to have perfectly random sampling such that the number of samples for each state are proportional to the relative size of each state. Sampling intensity can be affected by health system priorities, surveillance networks, preferential sampling via contact tracing and public health investigations, and convenience, among others. Nonrandom sampling violates the above assumption and can impact inference of migration history among

states.

Unlike DTA, the structured coalescent does not assume that the migration process and the tree generating process are independent and thus are instead modeled jointly [101]. This interdependence results in structured coalescent approaches being more robust to sampling bias, which can commonly arise though differential sampling [101].

The structured coalescent, however, requires that the state of any ancestral lineage in the phylogeny at any time be inferred, which is computationally expensive due to MCMC (Markov chain Monte Carlo) sampling. As the number of states or demes increases, MCMC convergence becomes more difficult, greatly limiting the number of sequences or demes that can be analyzed [113].

1.3.3 Approximations of the Structured Coalescent

In an attempt to overcome these computational limitations, various approximations of the structured coalescent have been developed. Most approaches marginalize over all possible migration histories by calculating the probability of each lineage being in each state rather than explicitly assigning lineages to particular states. One of these approximations is called BASTA (BAyesian STructured coalescent Approximation) [101]. An assumption of many of these approximations is that the state of a lineage evolves independently of the coalescent process between time events, meaning that changes in the probabilities of lineages being in a certain state are only dependent on the migration rates and are not dependent on other lineages in the phylogeny. The independence assumption has been shown to lead to bias in both parameter estimation, as well as inference of ancestral states primarily when migration is slow compared with the coalescent rate [113]. When migration rates are low compared to the coalescent rates, the influence of the coalescent on lineage states increases. The independence assumption however, prevents coalescent information from being used in the estimation of lineage state probabilities, biasing the results.

An approximation of the structured coalescent that was developed to address the limitation of the independence assumption is called MASCOT (Marginal Approximation of the Structured COalsescenT) [112]. MASCOT explicitly includes information about the location of other lineages and their probability of coalescing when modeling the movement of a lineage. This is done using a backwards/forward approach that incorporates information from the entire tree to calculate the probability of internal nodes being in any state. MASCOT has been found to be more robust to sampling bias than BASTA and other approximations of the structured coalescent, allowing both more robust estimations of migration dynamics and ancestral states.

1.3.4 Integration of empirical predictors into phylodynamics

A recent extension of MASCOT now allows for the joint integration of genomic and epidemiological metadata. This is accomplished through data integration via a GLM (generalized linear model) approach. The addition of a GLM approach had previously been implemented for DTA in order to inform the inference of migration parameters [91]. This approach was then extended to MASCOT as DTA does not estimate the transmission dynamics in different subpopulations [111]. MASCOT-GLM allows for the estimation of time-varying Ne and migration rates from both predictor and sequence data via log-linear combinations of coefficients, indicators and time-varying predictors. Formally this can be expressed as:

$$N_e(t) = \beta_{N_e} e^{(\sum_{i=1}^{c} \beta^i_{N_e} \sigma^i_{N_e} p^i_{N_e}(t))}$$

where $\beta^i_{N_e}$ denotes the coefficients and the extent to which each predictor contributes to predictor effective population sizes, $\sigma^i_{N_e}$ is an indicator which can be 0/1 denoting if a predictor contributes at all, and c denotes the linear combination of p predictors of N_e. The same formulation is done for migration rates. The estimation of indicators is done via Bayesian Stochastic Search Variable Selection (BSSVS), which allows for both the identification of predictors and effect sizes that best explains the phylogeny while also reducing

overfitting. MASCOT-GLM has been found to be robust at estimating spatiotemporal parameters, even at high levels of sampling bias [88]. The integration of empirical predictors also allows for more robust estimation of within deme transmission dynamic.

1.4 About this dissertation

This dissertation focuses on the use of genomic sequencing to augment epidemiological investigation into tracking the spread of SARS-CoV-2 and monkeypox virus (MPXV) and estimating the impact of various public health interventions to control their dissemination at various scales. In Chapter 2, I present work done in collaboration with the Washington State Department of Health (WADOH) to estimate the risk of hospitalization following infection with one of seven variant SARS-CoV-2 viruses using a retrospective cohort study of cases with sequenced virus from Washington State. Specifically, we demonstrate how integrating pathogen genomes sequenced as part of a statewide sentinel surveillance system with traditional epidemiological data and methods allows us to analyze variant-specific effects on clinical disease severity and how vaccination modifies that risk.

In Chapters 3 and 4, I present two genomic epidemiological studies on two different pathogens – SARS-CoV-2 and mpox – and on two different scales – local and global. Chapter 3 discusses how to leverage genomic epidemiology and phylodynamics to understand heterogeneous transmission at a local, actionable level. Actionable public health, in which we attempt to *interrupt* transmission, happens at the *local* scale. National level data often do not effectively inform these local level interventions because their lack of granularity masks significant transmission heterogeneity. This mismatch has led to a dearth of information on SARS-CoV-2 transmission dynamics *within* counties and metropolitan areas, leading to a situation where the public health departments with the greatest capacity to stop transmission have the least detailed information about how to target their efforts. I employ phylodynamic methods to investigate local-scale spatiotemporal SARS-CoV-2 transmission dynamics in

King County, Washington, a diverse US county containing the metropolitan area of Seattle. Specifically, I leverage MASCOT-GLM to inform our estimates of transmission dynamics and the impact of population-wide non-pharmaceutical interventions via the joint integration of epidemiologic, demographic, mobility, and genomic data.

While Chapter 3 zoomed in to investigate local-scale epidemic dynamics, Chapter 4 zooms out to track mpox spread around Europe and the Americas during the 2022 mpox epidemic. In 2022, mpox evaded early detection and sparked a public health emergency of international concern, resulting in over more than 90,000 recorded cases worldwide in already marginalized and vulnerable communities [29]. Due to social stigmas around case reporting as well as sub-clinical disease manifestation, surveillance data alone can only provide limited insights into mpox transmission dynamics. I employ phylogeographic and phylodynamic methods to estimate changes in case detection rate, the impact of underdetection on transmission, and the role of introductions in promoting local community spread in various global regions. I also examine the impact of vaccination campaigns on epidemic growth and decay in North America as well as estimate the degree of transmission heterogeneity in the declining phase of the epidemic.

Ultimately, this dissertation demonstrates how the joint integration of genomic and epidemiological information can provide novel and more precise insights into viral epidemic spread and the evaluation of infection control measures than possible through the use of a single data source alone.

Chapter 2

ASSOCIATIONS BETWEEN SARS-COV-2 VARIANTS AND RISK OF COVID-19 HOSPITALIZATION AMONG CONFIRMED CASES IN WASHINGTON STATE: A RETROSPECTIVE COHORT STUDY

This work, which was co-led by Stephanie Lunn, was originally published in *Clinical Infectious Diseases* [123]

2.1 Abstract

Background: The COVID-19 pandemic is dominated by variant viruses; the resulting impact on disease severity remains unclear. Using a retrospective cohort study, we assessed the hospitalization risk following infection with seven SARS-CoV-2 variants.

Methods: Our study includes individuals with positive SARS-CoV-2 RT-PCR in the Washington Disease Reporting System with available viral genome data, from December 1, 2020 to January 14, 2022. The analysis was restricted to cases with specimens collected through sentinel surveillance. Using a Cox proportional hazards model with mixed effects, we estimated hazard ratios (HR) for hospitalization risk following infection with a variant, adjusting for age, sex, calendar week, and vaccination.

Findings: 58,848 cases were sequenced through sentinel surveillance, of which 1705 (2.9%) were hospitalized due to COVID-19. Higher hospitalization risk was found for infections with Gamma (HR 3.20, 95%CI 2.40-4.26), Beta (HR 2.85, 95%CI 1.56-5.23), Delta (HR 2.28 95%CI 1.56-3.34) or Alpha (HR 1.64, 95%CI 1.29-2.07) compared to infections with ancestral lineages; Omicron (HR 0.92, 95%CI 0.56-1.52) showed no significant difference

in risk. Following Alpha, Gamma, or Delta infection, unvaccinated patients show higher hospitalization risk, while vaccinated patients show no significant difference in risk, both compared to unvaccinated, ancestral lineage cases. Hospitalization risk following Omicron infection is lower with vaccination.

Conclusion: Infection with Alpha, Gamma, or Delta results in a higher hospitalization risk, with vaccination attenuating that risk. Our findings support hospital preparedness, vaccination, and genomic surveillance.

2.2 Introduction

Following initial detection, SARS-CoV-2 disseminated rapidly worldwide, with the first reported COVID-19 case in the United States detected in Washington State (WA) on January 19, 2020 [18]. During the third quarter of 2020, distinct phenotypic changes on the SARS-CoV-2 spike protein were identified, raising concerns about increased transmission or greater disease severity [27]. The first detections of these variant viruses in WA occurred on January 23, 2021, when the first two cases of Alpha were found in Snohomish County [159] .

Since the initial detection of the first cases of the Alpha variant, multiple SARS-CoV-2 variants have been reported in WA. In March 2021, the Washington State Department of Health (WADOH) partnered with multiple laboratories to establish a sentinel surveillance program to monitor the genomic epidemiology of SARS-CoV-2 [160]. Given the replacement of ancestral lineages due to increasingly greater effective reproductive numbers, variant viruses now represent the majority of sequenced cases in WA [160].

The rapid emergence of variant viruses has resulted in numerous studies reporting increased transmissibility [34, 3, 127, 67]. Previous studies have identified an increased risk of hospitalization for both Alpha and Delta in various regions around the world [117, 55, 14, 142]. However, these studies compared a single variant lineage to an ancestral lineage or to

a small aggregated subset of variant viruses, leaving a dearth of knowledge into how risk of severe disease differs among the various lineages.

To address this gap in knowledge regarding healthcare outcomes following infection with a variant lineage, we designed a retrospective cohort study analyzing epidemiologic and genomic data from WA in order to compare the risk of hospitalization among seven SARS-CoV-2 variants.

2.3 Results

The COVID-19 epidemic in WA shows a distinct trend in the lineage distribution over time (Fig 2.1). Early on, the epidemic was predominantly characterized by ancestral lineages, while by March 2021, SARS-CoV-2 variants gained predominance over ancestral lineages.

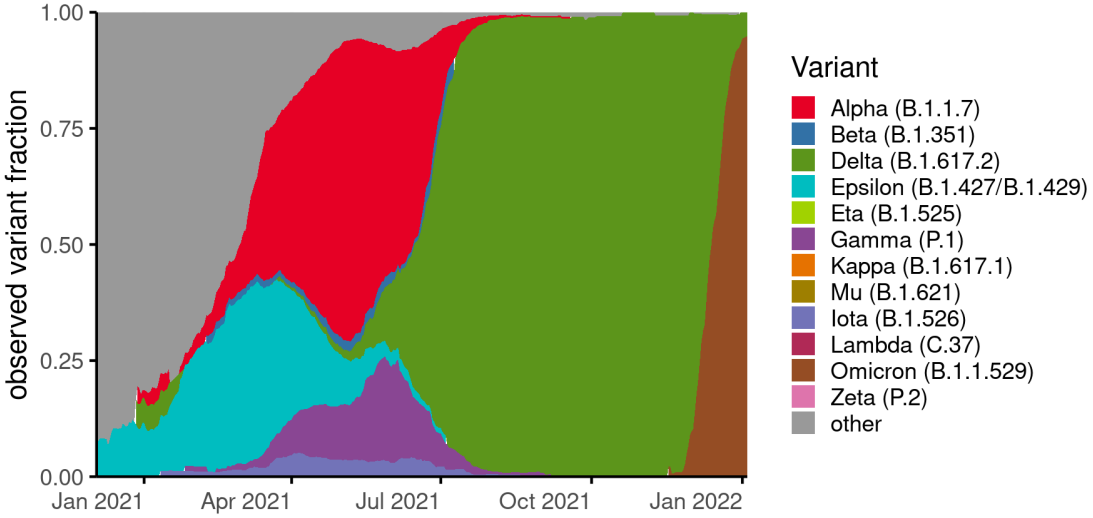

Figure 2.1: Changing proportion of infections due to variant lineages in Washington over study period. Variant fraction is calculated from a 21-day rolling average from our full sequenced dataset spanning from December 1, 2020 to January 14, 2022 and normalized to 100% to better observe changes in proportion of infections from variant lineages compared to total infections.

In this study, we included 63,639 cases with viral genome data available on WDRS, with

specimens collected from December 1, 2020 to January 14, 2022 (Fig. A.1). Of these, the final study population for the main analysis was restricted to 58,848 (92.3%) cases that were part of sentinel surveillance. The proportion of total cases in WA that were sequenced as part of sentinel surveillance over time is shown in Fig. S2.

Table 2.1 represents the general characteristics of the study population. The number of cases infected with a variant includes 8723 (14.8%) infected with Alpha, 231 (0.4%) with Beta, 2101 (3.6%) with Gamma, 33,107 (56.3%) with Delta, and 5362 (9.1%) with Omicron. 5178 (8.8%) individuals were infected with an ancestral lineage other than a variant of concern or interest as defined herein. Of the cases in the main analytic sample, 1705 (2.9%) cases were hospitalized.

Characteristics	**All**	**Ancestral**		**Alpha**		**Beta**		**Gamma**		**Delta**		**Epsilon**		**Iota**		**Omicron**	
Total N*	58,848	5178	(8.8)	8729	(14.8)	231	(0.4)	2103	(3.6)	33,115	(56.2)	3526	(6.0)	638	(1.1)	5362	(9.1)
Hospitalized (%)**	1705	117	(2.3)	233	(2.7)	11	(4.8)	112	(5.3)	1109	(3.3)	74	(2.1)	13	(2.0)	36	(0.7)
Median time (in days) to hospitalization (IQR)		6	(5)	6	(5)	7.5	(3.25)	6	(4.75)	6	(4.75)	5	(4)	3.5	(4)	7	(6)
Age (%)**																	
0-9	5661	418	(8.1)	888	(10.2)	22	(9.5)	198	(9.4)	3412	(10.3)	307	(8.7)	54	(8.5)	362	(6.8)
10 -19	8954	814	(15.7)	1561	(17.9)	52	(22.5)	272	(12.9)	4548	(13.7)	590	(16.7)	113	(17.7)	1004	(18.7)
20- 29	12340	1053	(20.3)	1948	(22.3)	51	(22.1)	527	(25.1)	6485	(19.6)	760	(21.6)	154	(24.1)	1362	(25.4)
30 - 39	11156	984	(19.0)	1640	(18.8)	39	(16.9)	432	(20.5)	6285	(19.0)	657	(18.6)	128	(20.1)	991	(18.5)
40 - 49	8345	741	(14.3)	1273	(14.6)	34	(14.7)	317	(15.1)	4583	(13.8)	524	(14.9)	93	(14.6)	780	(14.5)
50 - 59	6012	569	(11.0)	797	(9.1)	22	(9.5)	186	(8.8)	3529	(10.7)	383	(10.9)	56	(8.8)	470	(8.8)
60 - 69	3807	378	(7.3)	389	(4.5)	9	(3.9)	108	(5.1)	2429	(7.3)	209	(5.9)	25	(3.9)	260	(4.8)
70 - 79	1704	139	(2.7)	150	(1.7)	2	(0.9)	33	(1.6)	1223	(3.7)	62	(1.8)	11	(1.7)	84	(1.6)
80-89	681	60	(1.2)	68	(0.8)	0	(0.0)	22	(1.0)	462	(1.4)	31	(0.9)	3	(0.5)	35	(0.7)
90+	222	22	(0.4)	15	(0.2)	0	(0.0)	8	(0.4)	159	(0.5)	3	(0.1)	1	(0.2)	14	(0.3)
Sex (%)**																	
Female	28518	2410	(46.5)	4234	(48.5)	104	(45.0)	1026	(48.8)	16053	(48.5)	1666	(47.2)	292	(45.8)	2733	(51.0)
Male	29219	2595	(50.1)	4328	(49.6)	123	(53.2)	1040	(49.5)	16513	(49.9)	1739	(49.3)	330	(51.7)	2551	(47.6)
Other	58	5	(0.1)	9	(0.1)	0	(0.0)	1	(0.0)	36	(0.1)	2	(0.1)	0	(0.0)	5	(0.1)
Unknown	1087	168	(3.2)	158	(1.8)	4	(1.7)	36	(1.7)	513	(1.5)	119	(3.4)	16	(2.5)	73	(1.4)
Vaccination (%)**																	
No Vaccination to <21 post dose one	44845	5100	(98.5)	8412	(96.4)	221	(95.7)	1946	(92.5)	23108	(69.8)	3438	(97.5)	620	(97.2)	2000	(37.3)
≥21 days post dose one to <21 days post booster	12895	75	(1.4)	309	(3.5)	10	(4.3)	154	(7.3)	9434	(28.5)	86	(2.4)	18	(2.8)	2809	(52.4)
≥21 days post dose one	1142	0	(0.0)	2	(0.1)	0	(0.0)	1	(0.1)	565	(1.7)	2	(0.1)	0	(0.0)	553	(10.3)

Table 2.1: Study Cohort Characteristics by Variant of Concern/Variant of Interest. *denotes row percentages **denotes column percentage

In the adjusted model, we find a significant global effect of variant lineages on the hospitalization risk when compared to those cases infected with an ancestral virus (likelihood ratio test, p ¡0.001). The highest risks (Fig. 2.2) were found in cases infected with Gamma (HR 3.20, 95% CI 2.40-4.26) or Beta (HR 2.85, 95% CI 1.56-5.23). Cases with infection by Delta

(HR 2.28, 95% CI 1.56-3.34) or Alpha (HR 1.64, 95% CI 1.29-2.07) also showed a higher risk of hospitalization when compared to the reference. All other variants, including Omicron (HR 0.92, 95% CI 0.56-1.52) failed to show a significant difference in risk of hospitalization (Table 2.2).

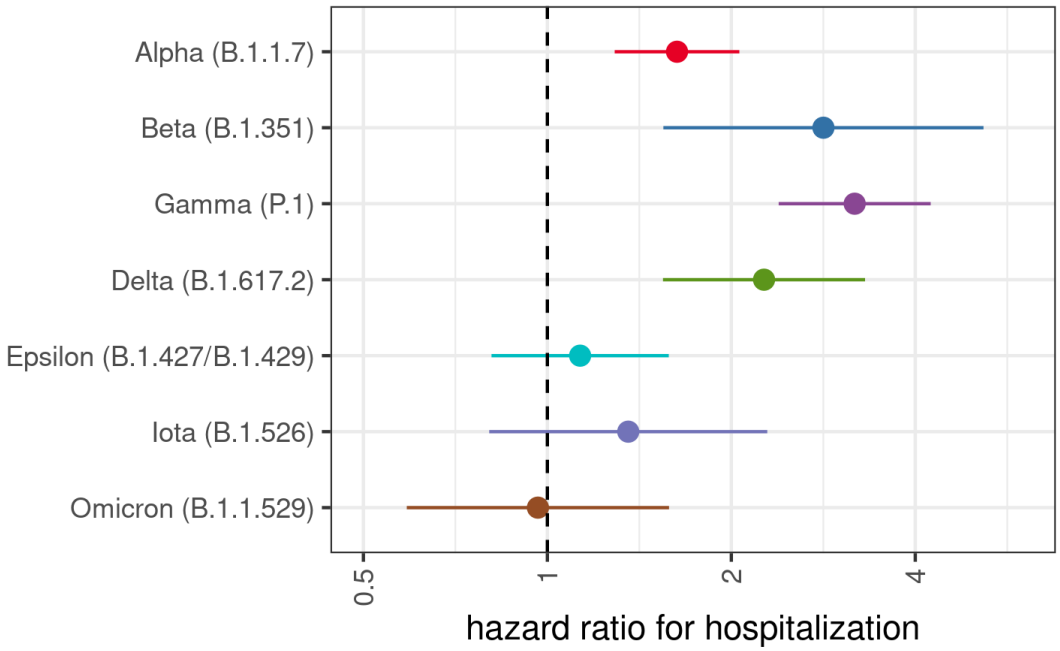

Figure 2.2: Relative Risk of Hospitalization by Variant Lineage. Risk of hospitalization is compared to individuals infected with an ancestral lineage. Error bars represent 95% CI. Estimates are adjusted for age, sex assigned at birth, calendar week, and vaccination status.

The association between variant lineage and hospitalization risk stratified by vaccination is shown in Figure 2.3 with unvaccinated individuals (unvaccinated or < 21 days post dose one) infected with ancestral lineages as the reference category. When compared to the reference, our model shows a higher risk of hospitalization for those unvaccinated individuals infected with Gamma, Delta, or Alpha, while those infected with Omicron showed no significant difference (Table 2.3). In the strata of individuals with an active vaccination but

Characteristics	Hospitalization	
	HR	95% CI
WHO Lineage		
Ancestral	REF	
Alpha	1.64	(1.29-2.07)
Beta	2.85	(1.56-5.23)
Gamma	3.20	(2.40-4.26)
Delta	2.28	(1.56-3.34)
Epsilon	1.13	(0.67-1.90)
Iota	1.34	(0.80-2.30)
Omicron	0.92	(0.56-1.52)
Vaccination		
Unvaccinated to <21 days post dose one	REF	
≥21 days post dose one to <21 days post booster	0.40	(0.35-0.45)
≥21 days post booster	0.31	(0.19-0.51)

Table 2.2: Adjusted Cox Proportional Hazard Estimates for Risk of Hospitalization. *Additional model covariates include: sex, age (in 10 year bins), calendar week

no active booster, no significant difference was observed in the risk of hospital admittance following infection with Gamma, Delta, or Alpha, but a lower risk of hospitalization was found following infection with Omicron (HR 0.49 95% CI 0.29-0.83), all when compared to unvaccinated, ancestral lineage cases. For those variant categories who had at least 4 hospitalizations following active booster vaccination (Table A.1), we find a significantly lower risk of hospitalization for cases infected with Omicron (HR 0.44 95% CI 0.21-0.93) but no significant difference for those infected with Delta, both compared to the unvaccinated, ancestral reference. Without stratification by variant lineage, we find that when compared to the unvaccinated group, cases with a record of an active vaccination but no booster and those with an active booster vaccination both have a lower risk of hospitalization ($\geq$ 21

days post dose one but $<$ 21 days post booster: HR 0.34, 95% CI 0.23–0.50; $\geq$ 21 days post booster: HR 0.31 95% CI 0.19-0.51).

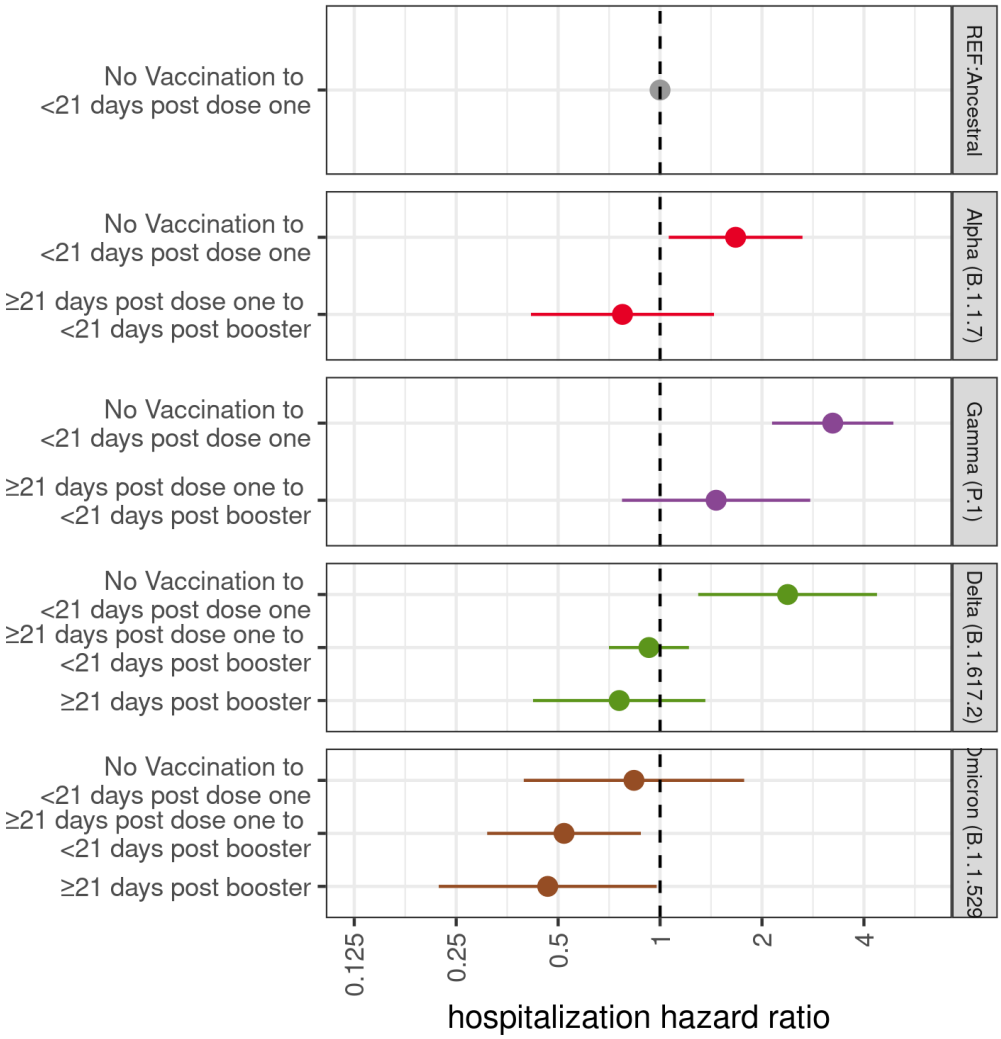

Figure 2.3: HR for risk of hospitalization following infection with a VOC (excluding Beta due to small sample size) stratified by vaccination status. Unvaccinated individuals infected with ancestral lineages serve as the reference category for each VOC HR. Error bars represent 95% CI. Estimates are adjusted for calendar week, age and sex assigned at birth. Categories with less than 4 hospitalizations are censored.

In a secondary analysis comparing hospitalization risk following infection with Omicron

	Hospitalization	
Characteristics	HR	95% CI
Vaccine*Variant		
Alpha		
Unvaccinated to <21 days post dose one	1.67	(0.91- 3.07)
≥21 days post dose one to <21 days post booster	0.78	(0.42-1.42)
≥21 days post booster	–	–
Gamma		
Unvaccinated to <21 days post dose one	3.24	(2.15 -4.89)
≥21 days post dose one to <21 days post booster	1.46	(0.77-2.78)
≥21 days post booster	—	—
Delta		
Unvaccinated to <21 days post dose one	2.39	(1.32-4.32)
≥21 days post dose one to <21 days post booster	0.93	(0.71-1.22)
≥21 days post booster	0.75	(0.41-1.34)
Omicron		
Unvaccinated to <21 days post dose one	0.79	(0.37-1.67)
≥21 days post dose one to <21 days post booster	0.49	(0.29-0.83)
≥21 days post booster	0.44	(0.21-0.93)
Ancestral		
Unvaccinated to <21 days post dose one	REF	

Table 2.3: Adjusted Cox Proportional Hazards Estimates for Risk of Hospitalization for Variant-Vaccine Interaction *Additional model covariates include: sex, age (in 10 year bins), calendar. Each variant lineage category risk estimate uses the "Unvaccinated to ¡21 days post dose one" vaccination group in cases infected with an ancestral lineage as the reference group. Categories with less than 4 hospitalizations are censored

to infection with Delta as the reference (Fig 2.4, Table A.2), we find a lower risk of hospitalization associated with Omicron infection (HR 0.34, 95% CI 0.23-0.50). When stratified by vaccination status, we find progressively lower risks of hospitalization for cases infected with Omicron for those unvaccinated (HR 0.37, 95% CI 0.21-0.66), vaccinated without a

booster (HR 0.23, 95% CI 0.14-0.39), and those $\geq$ 21 days post booster (HR 0.19, 95% CI 0.09-0.41), all when compared to unvaccinated cases with Delta infections.

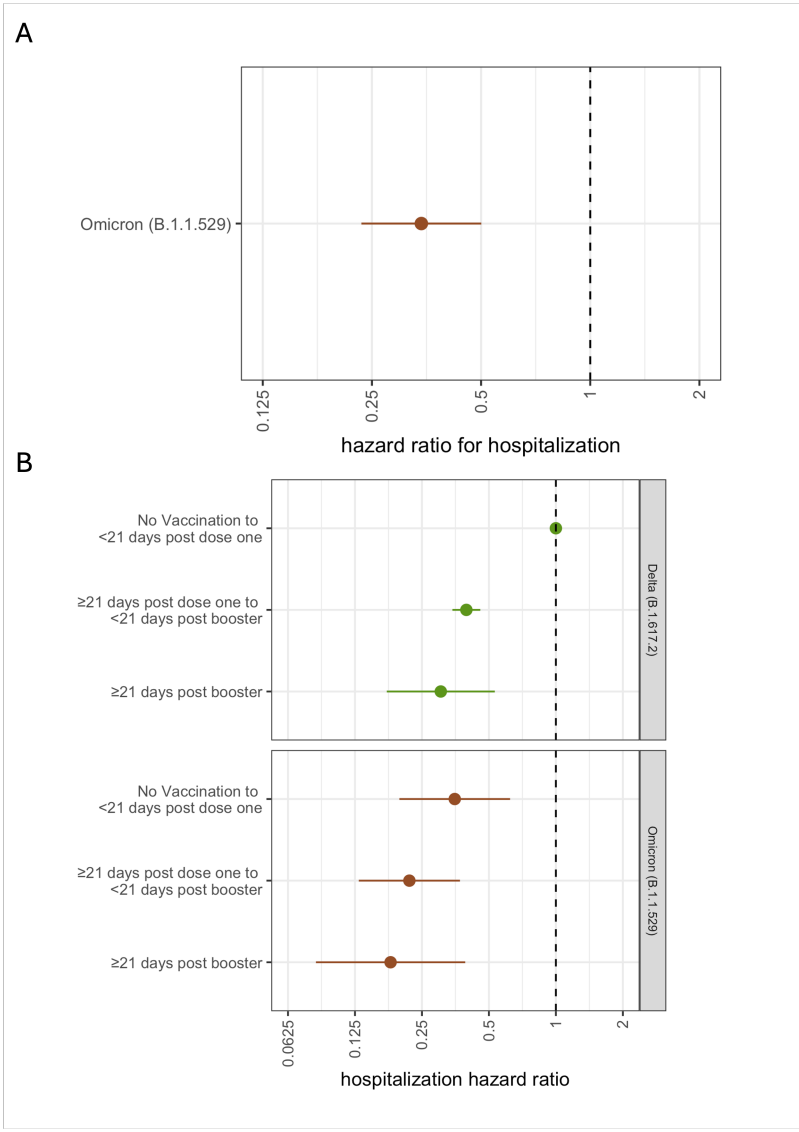

Figure 2.4: Risk of Hospitalization following Infection with Omicron vs Delta. A. Risk of hospitalization is compared to individuals infected with Delta. B.Unvaccinated individuals infected with Delta serve as the reference category for each VOC HR. Error bars represent 95% CI. Estimates are adjusted for calendar week, age and sex assigned at birth.

Estimates of the HR of the risk of hospitalization for cases infected with variants are robust to both model selection and inclusion of all sequences in our original database (Supp. Fig. A.3,A.4).

2.4 Discussion

In this study, we use SARS-CoV-2 cases in WA that were sequenced as part of sentinel surveillance to evaluate the differential risk of hospitalization following infection with a variant. We find that in our study period, cases infected with Alpha, Beta, Gamma, or Delta have a higher hospitalization risk compared to cases infected with an ancestral lineage, after adjusting for relevant covariates. We find similar estimates of higher hospitalization risk in the subset of unvaccinated individuals and no significant difference in hospitalization risk in individuals with an active vaccination, following infection with Gamma, Delta, or Alpha, while individuals infected with Omicron show a lower hospitalization risk in all vaccination categories, all compared to unvaccinated individuals infected with ancestral lineages.

Our findings are consistent with studies from around the world that have examined hospitalization risk following infection with SARS-CoV-2 variants [14]. Our estimates of hospitalization risk following infection with Delta (HR 2.28 95% CI 1.56-3.34) are similar to those from Scotland (HR 1.85 95% CI 1.39-2.47) [142] and Public Health England (HR 2.61, 95% CI 1.56-4.36) [133]. To our knowledge, few studies outside of ours have examined the hospitalization risk of Omicron compared to infection with ancestral lineages, but our estimates of hospitalization risk of Omicron vs Delta are highly similar to those calculated using S-gene target failure (SGTF) data [95, 13, 129]. Unlike studies using only SGTF data to identify probable Omicron cases, our study uses genomic sequencing to confirm the variant identity of each case, reducing the risk of misclassification. Verification with genomic sequencing is crucial for estimating the severity of Omicron, especially given the global rise of the BA.2 sublineage which does not cause S-gene dropout in TaqPath assay.

We also evaluated hospitalization risk following infection by Alpha, Gamma, Delta, or Omicron stratified by vaccination status. Following infection with Alpha, Gamma, or Delta, we saw a higher hospitalization risk for unvaccinated individuals and no significant difference in risk for vaccinated individuals without a booster when compared to those unvaccinated individuals infected with an ancestral lineage. Vaccinated individuals without a booster infected with any of these three variants all showed similar estimates of hospitalization risk with overlapping confidence intervals. The similar, overlapping estimates of risk in vaccinated individuals following infection with a VOC is supported by studies in the United Kingdom and Denmark showing no significant difference in hospitalization risk for vaccinated individuals infected with Delta when compared to those infected with Alpha, together suggesting that vaccination exerts a similar effect across these three variants [14, 149]. Unvaccinated individuals infected with Omicron showed no significant difference in risk of hospitalization, but individuals with any vaccination were found to have a lower hospitalization risk, all when compared to the unvaccinated, ancestral reference. When comparing hospitalization risk of Omicron vs Delta stratified by vaccination status, we find that any active vaccination is associated with a lower risk of hospitalization regardless of lineage when compared to unvaccinated cases with Delta infections, with estimates of risk similarly observed in Denmark [13].

Our sample sizes in some stratum are small (Table S1) limiting our ability to make conclusions. Additionally, cases were selected into our study based on test positivity; if vaccinated individuals are less likely to seek testing and severe illness leads to increased testing, conditioning study enrollment on testing can lead to collider stratification bias, or a distorted association between vaccination and disease severity[70]. Prior to July 27, 2021, CDC guidance stated that fully vaccinated individuals without symptoms did not need to get laboratory tested for SARS-CoV-2 following an exposure, meaning that cases in our sample with a vaccination record, which are conditional on being tested, are almost

certainly biased towards a subset of the population with a more severe clinical presentation than the population at-large, potentially underestimating estimates of vaccine protection on hospitalization risk. However, Delta and Omicron estimates largely derive from cases and hospitalizations after July 27, 2021.

Although our findings are consistent with previous studies, they are not without limitations. Variant classification is conditional on whole genome sequencing and a Ct threshold ¡30, meaning that our sequenced cohort may have been different from the general population of cases in WA. Sample sizes were determined by variant-specific circulation in WA and thus some variant categories have as few as 11 hospitalizations (Table 1: Beta); these respective estimates should be interpreted accordingly. Vaccination data in IIS is not comprehensive of federal vaccination efforts; vaccination status may therefore be misclassified for some cases. Sentinel specimens included in this study were randomly selected for sequencing within laboratories, but laboratories were not randomly sampled for inclusion in the sentinel surveillance program. The implementation of this program set proportions for sampling from specific laboratories to gather a geographically representative sample. It is possible that laboratory-level association with patient populations with differential risk of hospitalization over time may bias the study findings. The study is observational in nature, meaning that despite adjusting for potential confounders, there might be other confounders such as use of monoclonal therapy, social deprivation, etc., that might affect the association between SARS-CoV-2 variant and hospitalization risk. While previous studies have included comorbid conditions, race/ethnicity, or region of residence, their association with the risk of infection with a variant vs an ancestral strain in WA is unclear and thus were not included in our *a priori* set of confounders. Including these variables in an exploratory model did not affect estimates (Delta adjusting for race and county: HR 2.21 95% CI 1.50-3.30; Delta without race and county: HR 2.28 95% CI 1.56-3.34).

In conclusion, our retrospective cohort study found a higher hospitalization risk in cases

infected with Alpha, Beta, Gamma, and Delta, but not Omicron. Our study supports hospital preparedness in areas with uncontrolled viral spread as well as promoting vaccination. This study also highlights the importance of ongoing genomic surveillance at the state and federal level to monitor variant outcomes. Building a robust public health workforce as well as collaborations between public health and academia is critical to using genomic epidemiology to answer crucial questions about emerging SARS-CoV-2 variants.

2.5 Methods

2.5.1 Study Design

For this retrospective cohort study, we included cases with SARS-CoV-2 positive RT-PCR results in the Washington Disease Reporting System (WDRS) that contained linking information to corresponding sequences in the GISAID EpiCoV database [143, 45] with specimen collection dates between December 1, 2020 and January 14, 2022. Sequence quality was determined using Nextclade version 1.0.1 (https://clades.nextstrain.org/). Lineage was assigned using the Pangolin COVID-19 Lineage Assigner version 3.1.20 (https://pangolin.cog-uk.io/); only cases with an assigned PANGO lineage were included. The primary exposure of interest was SARS-CoV-2 variant, corresponding to all variant viruses that were given a Greek letter variant label by the WHO. These were all assigned a Nextstrain clade making the distinction clear [84]. Variants with less than ten hospitalization events were excluded, leaving Alpha, Beta, Gamma, Delta, Iota, Epsilon, and Omicron for analysis as well as ancestral viruses for reference. Vaccination data was collected from the WA IIS repository that is maintained by the Office of Immunizations at WADOH.

Cases without a known age, variant or vaccine manufacturer, cases with multiple lineages identified for the same infection, and cases where the linked viral sequence had ¿10% sequencing ambiguity, were excluded from the study. For cases with multiple specimens sequenced of the same virus, only the first sequenced specimen was used for analysis. The

main analysis was limited to cases with specimens sequenced as part of sentinel surveillance.

2.5.2 Sentinel Surveillance

As part of an initiative to monitor the genomic epidemiology of SARS-CoV-2, WADOH established a sentinel surveillance program with partner laboratories around the state. Laboratories and the percentage of randomly selected positive specimens they submit for sequencing were designated to optimize representation across WA [126]. Only PCR positive samples with a cycle threshold (Ct) of 30 or less are selected for sequencing. In addition to these designated sentinel laboratories, specimens were classified as sentinel surveillance if the sequencing laboratory indicated that they were conducting sequencing on randomly selected specimens. Specimens selected for sequencing as part of outbreak investigations, targeted due to travel history, targeted due to known vaccine breakthrough status, or targeted as part of investigations of S-gene target failures were not considered sentinel surveillance.

2.5.3 Hospitalizations

The primary outcome of interest was COVID-19 hospitalization. COVID-19 hospitalization is defined as a WA resident with confirmed COVID-19-positive lab who is identified as being hospitalized through hospital records, self-report of hospitalization, or linkage with syndromic surveillance hospitalization records (RHINO). If RHINO hospitalization records differ from the hospital record or self-report, the data is manually reviewed to adjudicate. Cases known to be hospitalized for a condition other than COVID-19 (e.g. labor and delivery) are not counted. In addition to the above data curation by WADOH, we additionally exclude cases where a positive viral collection date is more than 14 days after hospitalization in order to prevent misclassification of hospitalizations not attributable to COVID-19. Cases with a record of hospitalization but without an admission date were excluded from the study.

2.5.4 Covariates

We identified *a priori* confounders that were suspected to be associated with both risk of hospitalization following a COVID-19 infection and the epidemiological risk of acquiring a variant. These included age at sampling (categorized into 10-year increments), calendar week of collection, sex assigned at birth, and vaccination. Vaccination status was made into a three tier variable of 1) "Unvaccinated to $<$ 21 days post dose one", 2) " $\geq$ 21 days post dose one to $<$ 21 days post booster"', 3) "$\geq$ 21 days post booster" due to a low number of hospitalized cases having a record of vaccination. We consider active vaccination only after 21 days due to CDC guidance regarding active protection from symptomatic infection only after 14 days [7] and then allowing for an additional 7 days to allow for the development of protection from hospitalization, given that the mean time from symptom onset to hospitalization was found to be about 7 days [47]. Our vaccination covariate includes cases with a history of vaccination with BNT162b2 , mRNA-1273, and Ad26.COV2. Additionally, cases with a repeat positive test (defined as a case where the specimen collection date was more than 21 days after the first positive test date) were also excluded from the study to reduce confounding from previous immunity.

2.5.5 Statistical Analysis

We used descriptive statistics to explore characteristics of our sample stratified by SARS-CoV-2 lineage. For all descriptive analyses, we summarized categorical variables as frequencies and percentages.

We estimated the associations between SARS-CoV-2 variants and the risk of COVID-19 hospitalization by calculating hazard ratios (HRs) for the time to hospital admission through a Cox proportional hazard model with mixed effects using ancestral lineages as the reference group. We adjusted the HRs for the covariates of age, sex assigned at birth, calendar week (continuous), and vaccination status. Sex and vaccination status were added as random

effects to regularize adjustments for under-represented categories. A likelihood ratio test was used to examine the global effect of variant lineages on hospitalization risk.

In a secondary analysis to analyze how vaccination affected the risk of hospitalization by variant lineage, an interaction term of vaccination*lineage was introduced into the model and reran for those variants found to have the largest sample size and effect magnitude: Alpha, Gamma, Delta, and Omicron. Stratified risk of hospitalization by vaccination status was conditioned on the "Unvaccinated to <21 days post dose one" group for cases infected with an ancestral lineage.

The above analysis was repeated with a subset of the data only including cases infected with Delta (as the reference) or Omicron with a collection date after September 1st, 2021.

In order to account for differences in both model selection and case inclusion, sensitivity analyses were performed using a Cox proportional hazard model with fixed effects and a Poisson regression model for both the subsetted sentinel surveillance-only dataset as well as for the entire case dataset found in WDRS for the same study period. Statistical analyses were performed using R version 3.6.2 (R Project for Statistical Computing).

Analytic code can be found at https://github.com/blab/ncov-wa-variant-severity

2.6 Acknowledgments

Clinical and sentinel laboratories who forwarded specimens for sequencing, and sequencing laboratories that reported data to WADOH. The WADOH Data Science Support Unit for integrating sequencing data with epidemiologic case data. The WADOH IIS team and DIQA teams for maintaining and linking vaccination data. The WADOH surveillance team for case and hospitalization data. Natasha Close and the WADOH RHINO team for hospitalization data. James S. Miller (Epidemic Intelligence Service Officer, Centers for Disease Control and Prevention) for manuscript review and insightful feedback. The team at Altius Institute for Biomedical Sciences: Rebecca L. Bruders, Amanda C. Gale, Clementine B.M Green,

Suman Grewal, Muhammad H. Halimun, Kneshay N. Harper, Jessica M. Halow, William A. Isner, Audra K. Johnson, Jessica N. Kunder, Lauren E. Mitchell, Jemma S. Nelson, Alex S. Nguyen, Sofia E. Olsson, Sadie L. Patraw, Tobias F.C. Ragoczy, Ashly M. Senske, Julia Wald.

Chapter 3

LOCAL-SCALE PHYLODYNAMICS REVEAL DIFFERENTIAL COMMUNITY IMPACT OF SARS-COV-2 IN A METROPOLITAN US COUNTY

This work was originally published in *PLOS Pathogens* [124]

3.1 Abstract

SARS-CoV-2 transmission is largely driven by heterogeneous dynamics at a local scale, leaving local health departments to design interventions with limited information. We analyzed SARS-CoV-2 genomes sampled between February 2020 and March 2022 jointly with epidemiological and cell phone mobility data to investigate fine scale spatiotemporal SARS-CoV-2 transmission dynamics in King County, Washington, a diverse, metropolitan US county. We applied an approximate structured coalescent approach to model transmission within and between North King County and South King County alongside the rate of outside introductions into the county. Our phylodynamic analyses reveal that following stay-at-home orders, the epidemic trajectories of North and South King County began to diverge. We find that South King County consistently had more reported and estimated cases, COVID-19 hospitalizations, and longer persistence of local viral transmission when compared to North King County, where viral importations from outside drove a larger proportion of new cases. Using mobility and demographic data, we also find that South King County experienced a more modest and less sustained reduction in mobility following stay-at-home orders than North King County, while also bearing more socioeconomic inequities that might contribute to a disproportionate burden of SARS-CoV-2 transmission. Overall, our findings suggest a role

for local-scale phylodynamics in understanding the heterogeneous transmission landscape.

3.2 Author Summary

State- or county-level data collected as part of routine surveillance often mask significant local differences in SARS-CoV-2 transmission due to their lack of granularity. This leaves local public health departments with incomplete information for resource allocation. Using King County, Washington as an example of a diverse, metropolitan US county, we leveraged genomic epidemiology to understand differences in transmission between North and South King County, two adjacent regions within the same county with stark socioeconomic differences. By combining epidemiological, mobility, and demographic data, we found that these two regions had divergent SARS-CoV-2 epidemic trajectories following the start of statewide stay-at-home orders in March 2020. Our approach also revealed important differences in the role of viral importations and persistence of local viral transmission on changing SARS-CoV-2 incidence in the background of large-scale non-pharmaceutical interventions. Our work shows that we can use genomic epidemiology to reveal differences in transmission at a local scale, which can inform equitable resource allocation at a local level to reduce the burden of infectious diseases.

3.3 Introduction

The first confirmed SARS-CoV-2 infection in the United States was detected in Washington State (WA) on January 19, 2020. Since initial detection of the virus, genomic epidemiology has played a crucial role in identifying and estimating new introductions and community transmission in WA [18, 114, 147] and throughout the US [38, 94] and has motivated rapid public health interventions. While international introductions continue to seed new viral lineages into the US, the majority of transmission is driven by infections and movement at a local scale, wherein neighboring states, regions, counties, or even zip codes can have vastly

different epidemic dynamics [147, 108, 37].

In WA, genomic epidemiology has aided in understanding the spatiotemporal variation of the SARS-CoV-2 epidemic. At a statewide level, previous studies have examined changes in the relative frequency of variant viruses and the impact of non-pharmaceutical interventions on the estimated effective population size of the virus [114]. Phylodynamic analyses have estimated the role of introductions in promoting community spread in the state at large and revealed an asymmetrical interplay between the eastern and western regions of the state, wherein intra-state transmission accounts for more than half of the introductions into the eastern region of WA but only for less than 30% of the introductions into western WA [147].

Even a regional view fails to capture the nuance of epidemic dynamics needed to effectively curb transmission in the state because neighboring counties and even intra-county areas are affected by epidemic and demographic heterogeneity. King County, WA is a demographically diverse, metropolitan US county that has been proactive in promoting testing and vaccination throughout the SARS-CoV-2 epidemic. Despite these efforts, studies have revealed a large degree of variation in SARS-CoV-2 infection probability and hospitalization, with communities of color disproportionately impacted [79].

Previous studies have attributed differences in local case counts to unequal reductions in mobility [145, 83]. When compared to a baseline average from 2019, King County, WA as a whole shows a large decrease in mobility following the implementation of stay-at-home orders in March 2020 but differences between within-county regions are salient: North King County experienced a 60% reduction in mobility compared to the 40% reduction in South King County (Fig 3.1 A) . While South King County eventually returned to baseline levels of mobility by the end of 2020, North King County was able to maintain reduced levels through March 2022. The ability to significantly reduce and maintain mobility changes has been previously attributed to socioeconomic inequities, including geographical differences in income [161] and percentage of the community that contributes as an essential worker [145].

We see a similar pattern in King County: South King County has a lower median household income, a larger percentage of essential workers in the active workforce, and a higher average household size than North King County (Fig 3.2B-D), despite a smaller population size (Fig. 3.2E).

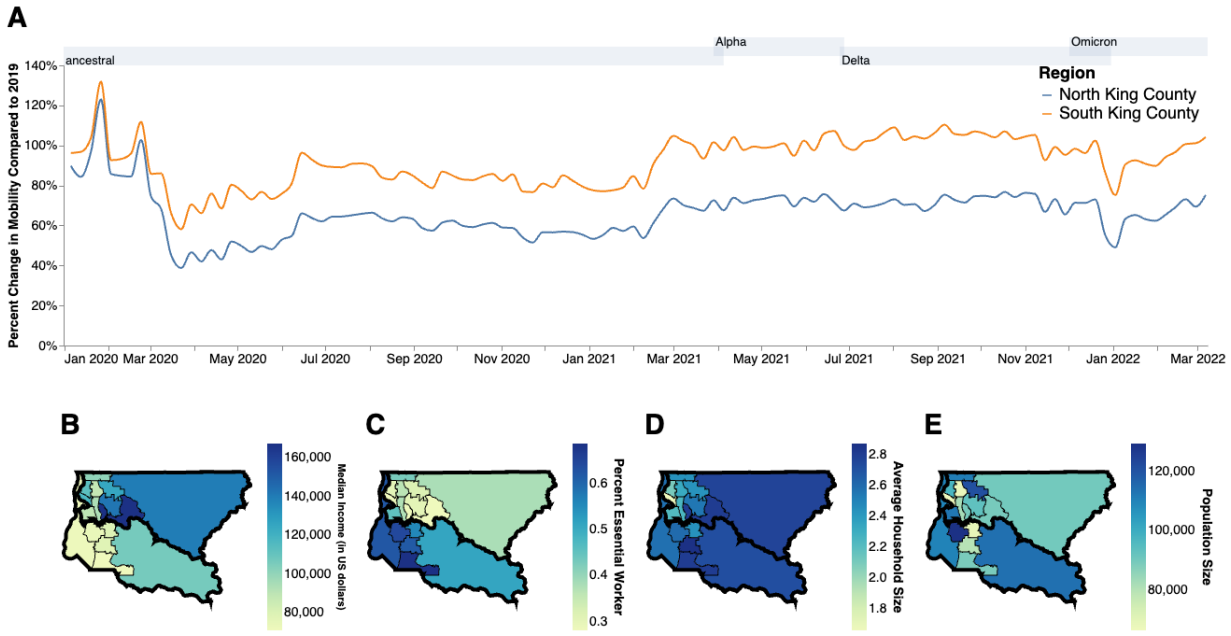

Figure 3.1: Socioeconomic Characteristics of King County A. Percent change in mobility from Feb 2020 to March 2022 over time using average mobility in 2019 as baseline for North (blue line) and South (orange line) King County. Dashed line denotes no change compared to baseline. B,C. Median household income in 2020. (**B**) Percentage of the active workforce whose occupation is defined as "essential" from 2015-2020 (**C**) average household size from 2015-2020 (**D**) and population size (**E**) in King County by Public Use Microdata Area (PUMA). Gray shaded regions above each figure show the time periods during which ancestral virus, Alpha, Delta, and Omicron respectively represented greater than 30% of sequenced case. Geojsons for King County PUMAs were made using shapefiles from the US Census Bureau [150] and can be found here: https://github.com/seattleflu/seattle-geojson/tree/master/seattle_geojsons

While some studies have used genomic epidemiology to examine transmission between US counties or boroughs [94, 108, 37], here we employ phylodynamic tools to understand the fine scale spatial and temporal dynamics of SARS-CoV-2 viral transmission both within and

between regions of King County, WA, as a case study of a demographically and socioeconomically diverse US metropolitan county. Using 11,737 viral sequences sampled from individuals in King County between January 2020 and March 2022, we examined the role of introductions in promoting community spread and the impact of non-pharmaceutical interventions on viral transmission dynamics.

3.4 Results

The COVID-19 epidemic in King County, WA shows distinct spatial and temporal patterns that persisted throughout our study from February 2020 to March 2022. At the PUMA level, confirmed COVID-19 cases and hospitalizations in King County are disproportionately higher in more southern PUMAs than in northern PUMAs (Fig 3.2A, B) during almost every time period analyzed. During the last time period encompassing the BA.1 Omicron wave, from December 2021 to March 2022, we observed a more equal geographic distribution of confirmed COVID-19 cases, but COVID-19 hospitalizations continue to disproportionately affect southern regions.

Due to the salient differences between northern and southern PUMAs, we then divided King County into two regions, North and South, and analyzed COVID-19 cases and hospitalizations continuously over time (Fig 3.2C,D). From January 2020 to the end of March 2020, during the beginning of the epidemic, we see that cases and hospitalizations are slightly higher in North King County. However, starting in April 2020 soon after a stay-at-home order on March 23, South King County consistently had higher confirmed cases and hospitalizations per capita than North King County, a trend that mostly persisted throughout the time period studied, except during the Omicron wave when cases were similar in both regions. Time series of cases and hospitalizations replicated the geographical trends seen in Fig.3.1A and B: while the difference in the number of confirmed cases seemed to contract in during the BA.1 Omicron wave (Dec 2021–March 2022), the magnitude of the difference

in hospitalizations remains roughly constant, with South King County disproportionately burdened.

To investigate transmission dynamics between and within these two King County regions, we analyzed 11,602 sequenced King County viruses alongside contextual sequences from around the world. Following the creation of time-resolved phylogenies using Nextstrain [74], we split the sequences into local outbreak clusters using parsimony-based clustering to identify groups of sequences whose ancestral states were inferred to be in King County (see Methods, Supp Fig. B.1). We identify 5964 clusters and find that the number of clusters increases over the time in both regions (Fig. 3.3A), most likely due to an increase in the number of cases being sequenced in WA. Additionally, we find that the majority of clusters are single introductions (n = 5,095), with larger clusters increasingly rare (Fig 3.3B, clusters with more than 10 sequences were excluded for clarity). South King County has a greater mean cluster size (South: 1.87; North: 1.61; two-sample t-test p-value: 0.048) as well as a larger maximum cluster size (max South cluster size of 280 vs max North cluster size of 150). Figure 3.3C shows the phylogenetic tree of all clusters with 5 or more sequences with inferred geographic location as coloring.

We also analyzed the inferred ancestral location for all clusters over time divided out by the dominant variant waves (Supp. Fig B.2). We found that Alpha and Delta arrived first into King County mainly from other US states before spreading into the larger WA region, with Alpha also arriving from the UK where it originated. As time progressed, the source of introductions switched from mainly North America (excluding WA) to predominantly from within Washington (excluding Omicron which was introduced into King County primarily from WA). Additionally, we saw that North King County has a larger proportion of viral introductions coming from outside WA, while the majority of introductions into South King County come from within the state.

We then employed phylodynamic inference methods on the identified outbreak clusters

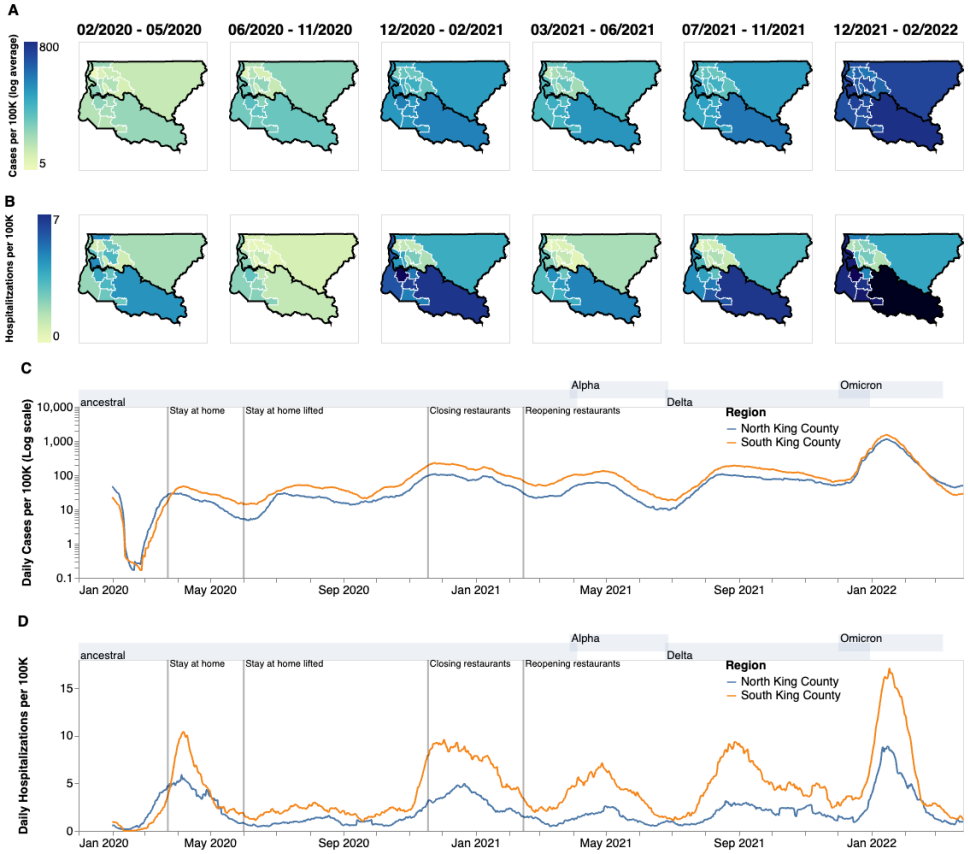

Figure 3.2: Descriptive Epidemiology of SARS-CoV-2 Epidemic in King County, WA (A, B) Confirmed positive cases (A) and hospitalizations (B) per 100,000 individuals of SARS-CoV-2 in King County by Public Use Microdata Area (PUMA) averaged for each of the six waves of the epidemic up until March 2022. Dark borders denote geographical boundaries between North and South King County (**C, D**) Daily positive cases and hospitalizations of SARS-CoV-2 from February 2020 to March 2022 by region of King County smoothed with a 14 day rolling average. Blue denotes North King County; Orange denotes South King County. Gray shaded regions above each figure show the time periods during which ancestral virus, Alpha, Delta, and Omicron respectively represented greater than 30% of sequenced case. Geojsons for King County PUMAs were made using shapefiles from the US Census Bureau [150] and can be found here: https://github.com/seattleflu/seattle-geojson/tree/master/seattle_geojsons

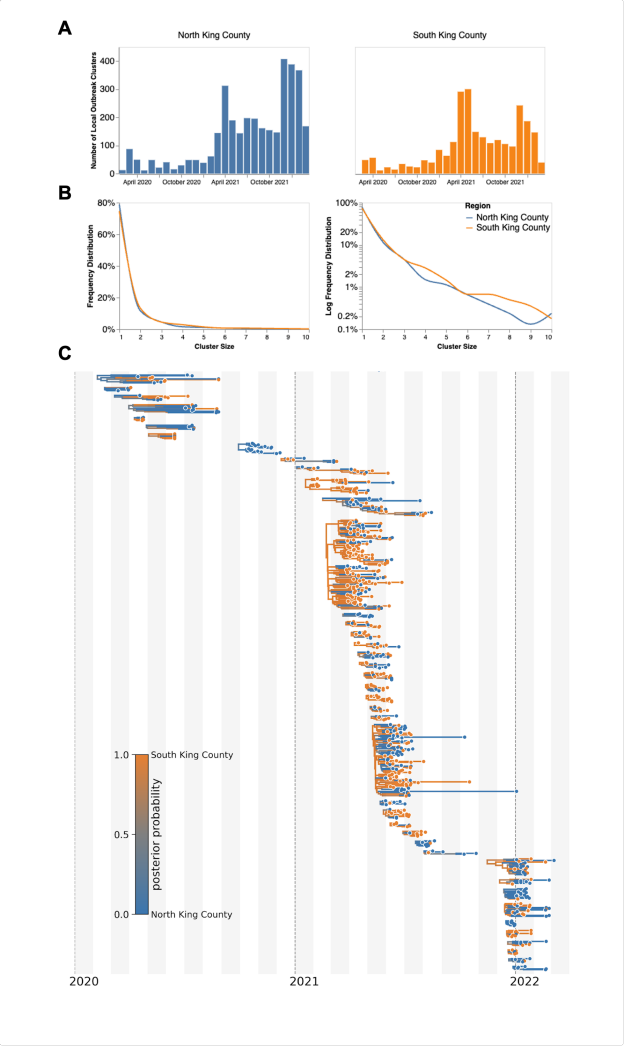

Figure 3.3: Representative SARS-CoV-2 Clusters by Region in King County. We combined more than 11,500 SARS-CoV-2 genomes from King County with more than 45,000 contextual sequences from around the world and built a time-resolved phylogeny. King County outbreak clusters were then extracted using a parsimony based clustering approach. We inferred geographic transmission history between each region using MASCOT-GLM. Here, we display the number of clusters over time by King County Region (**A**), the frequency of cluster size by region on a linear (**B left**) and log (**B right**) scale (up to a cluster size of 10. Larger clusters exist but were excluded from the graph for clarity), and the maximum clade credibility tree of all clusters with five or more sequences (**C**) where color represents posterior probability of being in South King County. The x-axis represents the collection date (for tips) or the inferred time to the most recent common ancestor (for internal nodes). Blue denotes North King County, Orange denotes South King County.

to analyze SARS-CoV-2 spread in the county. Following subsampling, we used a MASCOT-GLM approach with relevant predictors on a random subsample of 3000 sequences from our dataset of local outbreak clusters to reconstruct SARS-CoV-2 transmission dynamics (Supp. Fig. B.3). Phylodynamic estimates of the effective population size (N_e) of the virus in both King County regions over time mirror patterns seen in both confirmed COVID-19 hospitalizations and cases: while the *Ne* in North King County is initially greater until the end of March 2020, following WA stay-at-home orders, we find a consistently greater *Ne* in South King County throughout the study period (Fig. 3.4A). We also find that hospitalizations one week in the future was the most informative predictor for effective population size in our model (Fig 3.4B), while the migration rate predictors were not significantly informative (Fig 3.4C).

We next analyzed the posterior set of phylogenies produced by the MASCOT-GLM analysis to understand viral circulation within and between the two regions. Given the higher estimated *Ne* in South King County, we quantified the average persistence time of viral transmission chains in each region (Fig 3.5A, see Methods). While the average monthly persistence time remained relatively equal between the two regions during the early stages of the epidemic, following May 2020 up until 2022, we see that transmission chains in South King County consistently have significantly higher persistence times than in North King County, with the mean local transmission length averaged over the entire time period of 21.5 days in South King County and 13.5 days in North King County. We see an increase in average persistence times in both regions during large waves of COVID-19 cases attributable to the introduction of a new variant with a transmissibility advantage (such as in late 2020- early 2021 with the introduction of Alpha) and the relaxation of stay-at-home order, with South King County consistently having longer persistence times.

To understand if these longer transmission chains in South King County could be due to a higher number of viral introductions from outside the county, we reconstructed the ancestral

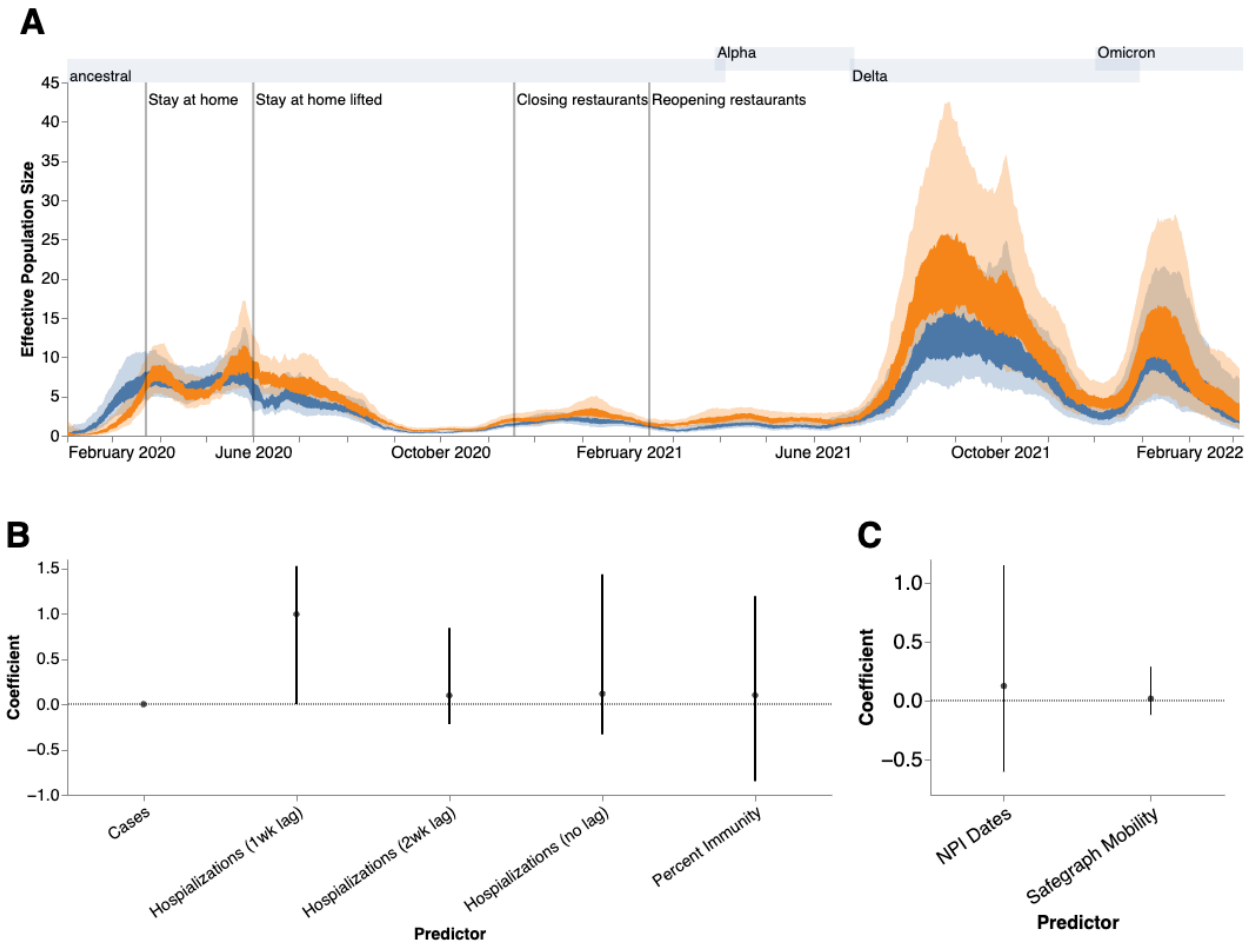

Figure 3.4: Phylodynamic Analysis via MASCOT-GLM (A)Estimates of effective population sizes from Feb 2020 to March 2022 in North (blue) and South (orange) King County using 3000 randomly subsampled sequences. The inner band denotes the 50% highest posterior density (HPD) interva,l and the outer band denotes the 95% HPD interval. Vertical gray lines denote dates of non-pharmaceutical interventions in Washington State. **(B)** Estimates of model predictor coefficients for *Ne* estimation and **(C)** for migration rate estimation. All of the predictors displayed on the x-axis were included in the analytic model. Dark line represents median estimates, light bands represent 95% HPD. Gray shaded regions above each figure show the time periods during which ancestral virus, Alpha, Delta, and Omicron, respectively represented greater than 30% of sequenced case

states of each *a priori* defined King County transmission cluster to quantify the relative number of introductions into each region (Fig 3.5B). While greater than 50% of introductions

prior to May 2020 were into South King County, the majority of the time period studied was characterized by a greater relative proportion of introductions from outside into North King County.

These fine scale phylodynamic analyses also allow us to investigate the interplay between local regions. Introductions from outside regions have been shown to play a driving force in maintaining local outbreaks [115] but often these introductions are focused on interstate or international travel. Here we quantify the interplay between two inner-county regions, examining the number of transmission events that occur between North and South King County (Fig 3.5C). By quantifying the number of migration jumps between the two regions, we see a clear pattern emerge in which prior to June 2020 when WA lifted emergency stay at home orders, there was little difference in the number of transmission events between regions. Following the elimination of the stay-at-home orders however, transmission events become asymmetrical, where we consistently see disproportionally more transmission from South King County to North King County than in the opposite direction, with the largest differences occurring in the beginning months of 2021.

Given the higher number of introductions into North King County but the larger *Ne* and longer transmission chain length in South King County, we sought to estimate the relative contribution of introductions versus local community spread in driving the epidemic in both King County regions. To do so, we calculated the percentage of new cases from introductions in each region using the estimated changes in *Ne* over time as well as the estimated rates of introduction both from outside King County and from the neighboring inner-county region. We estimated a relatively higher percentage of cases due to introductions in South vs North King County prior to emergency stay-at-home order in WA on March 23, 2020 (Fig 3.6A). Following the stay-at-home order, the pattern switched and was largely constant throughout the epidemic, with North King County averaging about 35% of new cases from introductions versus local spread while only about an average of 25% of new cases were estimated to be

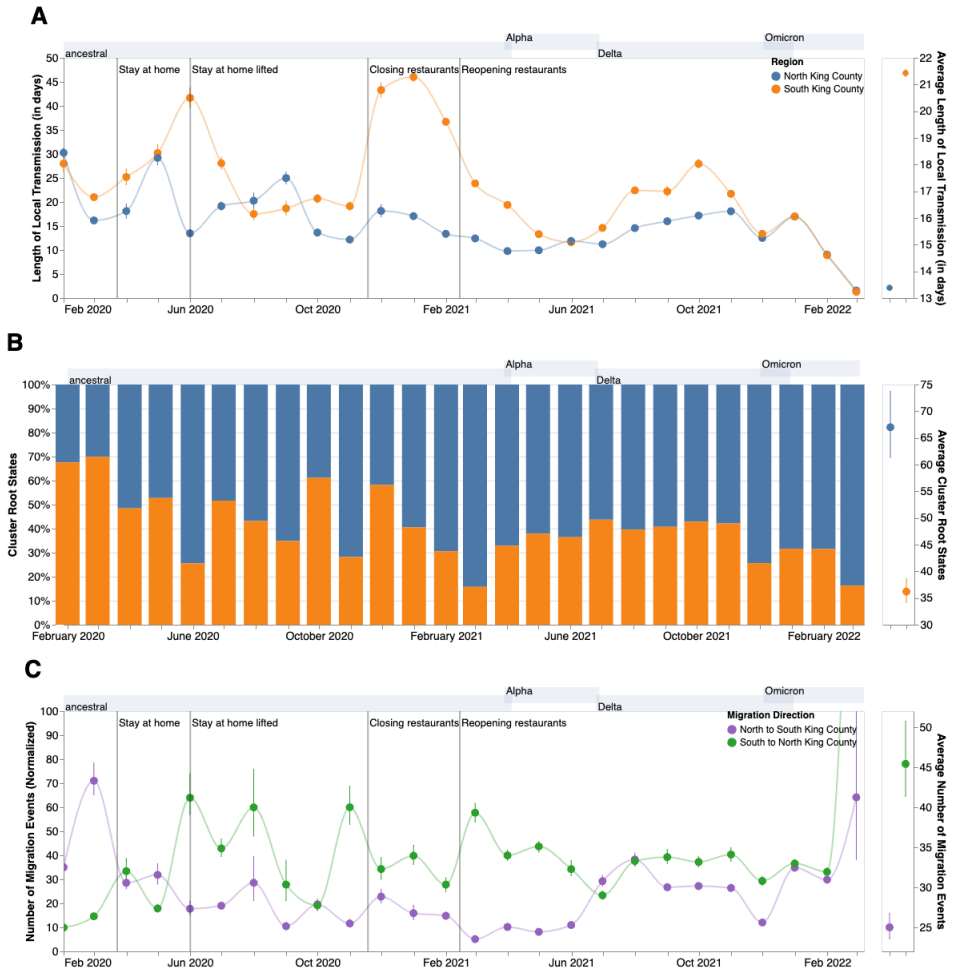

Figure 3.5: Within and Inter-Regional Dynamics in King County inferred from pathogen genomes and relevant covariates **A.** Persistence time (in days) of local transmission chains over time in both regions of King County. Accompanying graph showing persistence times averaged over the entire time period for both regions with error bars denoting 95% CIs. **B.** Inferred reconstruction of ancestral state for each transmission cluster over time. Blue denotes initial introduction in North King County and orange denotes initial introduction in South King County. Average values are normalized to 100% over time. The Accompanying graph showing inferred introductions averaged over the entire time period for both regions with error bars denoting 95% CIs. **C.** Number of migration events from North to South King County (purple) and from South to North King County (green) over time. Bands denote 95% CI. The accompanying figure shows the number of migration events between the two regions averaged over the entire time period with error bars denoting 95% CIs. Gray shaded regions above each figure show the time periods during which ancestral virus, Alpha, Delta, and Omicron respectively represented greater than 30% of sequenced cases.

from introductions in South King County. To further support this estimate, we calculated the percentage of visits to POIs in North and South King County for devices having an outside home location using SafeGraph mobility data. We find similar estimates ranging from about 25%-40% throughout time (Fig. 3.6A, black lines).

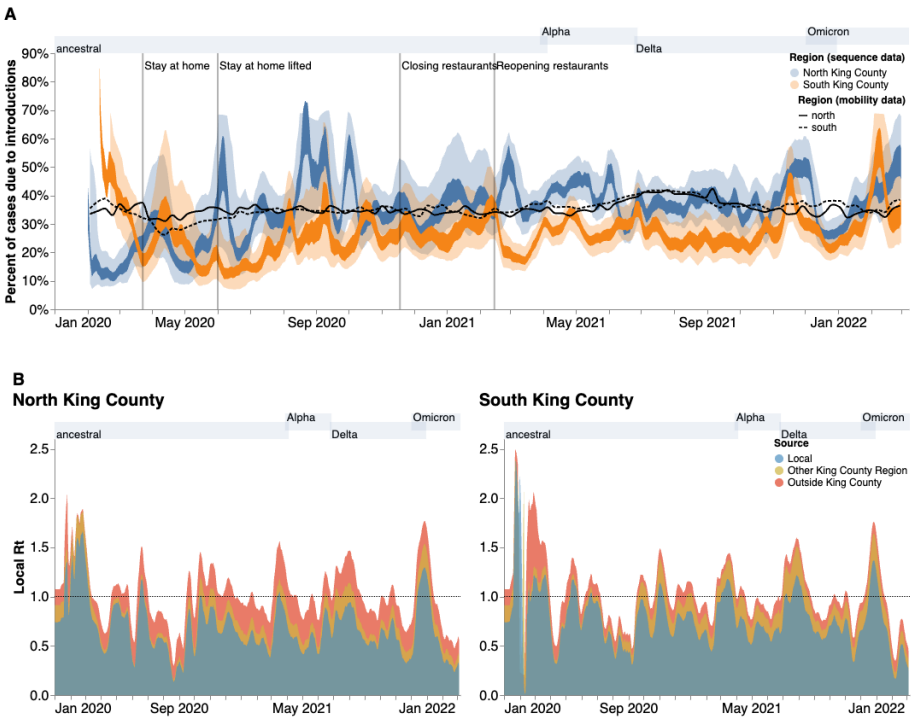

Figure 3.6: Phylodynamic estimates of the differential impact of introductions and local spread on transmission dynamics of SARS-CoV-2 by region in King County. (A) Percentages of new cases due to introductions were estimated as the relative contribution of introductions to the overall number of infections in the region. The inner area denotes the 50% HPD interval and the outer area denotes the 95% HPD interval. Blue = North King County; Orange = South King County. Black lines represent the same calculation using SafeGraph mobility data as parameter approximations. Solid black line is for North King County; Dashed black line is for South King County. (**B)** Estimates of local *Rt* highlighting the contribution of introductions from outside King County (red) and from the neighboring King County region (gold) on local transmission in each King County region. Dashed line denotes an Rt of 1. Estimates were smoothed using a 7 day rolling average. Estimates higher than 1 suggest an exponentially growing epidemic. Gray shaded regions above each figure show the time periods during which ancestral virus, Alpha, Delta, and Omicron respectively represented greater than 30% of sequenced cases.

To better compare transmission dynamics between the two regions, we next used the effective population size dynamics to compute R_t, the time-varying effective reproductive number (Fig. 3.6B, Supp. Fig. B.4). Additionally, we also employed our estimates of the percentage of new cases that are due to introductions to separate out the effects of local transmission and introductions on R_t. We find that the Rt for both regions closely follows variant waves, with an Rt above 1, which implies increasing transmission, matching with dates of increased case counts. Additionally, by separating out contributions into being from local transmission, introductions from the neighboring King County region, or introductions from outside King County, we find that local transmission is the main contributor to Rt in both regions but that introductions have a differential impact. We see that introductions as a whole play a much larger role in promoting and maintaining transmission in North King County, with outside regions being the main contributor of introductions. In South King County, Rt is more driven by local within-region spread, with introductions from North King County being more influential than introductions from outside the county.

Phylodynamic estimates of epidemic dynamics were similar regardless of subsampling strategy used (Supp. Figs B.5, B.6).

3.5 Discussion

The surge of whole genome sequencing has enabled large-scale investigation into key COVID-19 epidemiological dynamics. Yet, genomic epidemiology can also be employed to analyze transmission patterns at a local scale to aid in policy making and intervention evaluation. Here, we examined fine-scale SARS-CoV-2 transmission dynamics at a sub-county level for King County, WA, a large metropolitan area with a demographically diverse population.

We used novel phylodynamic methods to reconstruct the epidemic in King County from January 2020 to March 2022 and examine within-region dynamics and their interplay from pre-identified local outbreak clusters. We divide King County into North and South, informed

by the clear differences in outcomes (cases and hospitalizations) at the PUMA level, in which South King County has been disproportionately affected despite having a smaller population size (673,548 in South versus 1,400,211 in North King County in 2020 [1]). We estimated that for the majority of the time period studied, introductions accounted for a larger percentage of new cases in North than in South King County (Fig 3.5). While a higher proportion of introductions among new cases can be attributed to either a higher rate of introduction or a lower local transmission rate, we find evidence of a greater number of viral introductions into North King County over time, from both outside and within the county, but longer chains of local transmission in South King County (Fig 3.5). Together, our data suggest a larger impact of introductions in North King County and a larger role of local community spread in South King County in driving the respective regional epidemics. This conclusion is supported via our *Rt* estimates, or the time-varying estimate of secondary infections, which show that outside introductions play a significant role in transmission in North King County while local spread is more contributory in South King County (Fig 3.6. Importantly, cases being driven by a higher percentage of introductions can be due to either an increase in introductions from outside, a decrease in local spread, or a combination of both.

Given the smaller population size in South King County, one potential explanation for higher local spread in that region is reduced access to social and economic capital and health care resources needed to curb community transmission. Previous studies looking at SARS-CoV-2 test positivity in King County at a census tract level have found that a higher test positivity was associated with various socioeconomic indicators including lower educational attainment, higher rates of poverty, and high transportation costs [141, 75]. Additionally, they found that communities with a higher proportion of people of color, which are more likely to be located in South King County, were also associated with higher test positivity in 2020. Hansen et al. [75], specifically found that having a place of residence in South King County was associated with SARS-CoV-2 test positivity. The stark contrast in health

outcomes between North and South King County has been previously attributed to historical redlining and systemic racism, whereby decades of racial segregation prevented communities of color from residing in northern areas of Seattle and were forced into the south into present day South King County [24, 4].

The associations between test positivity and socioeconomic status are not a unique King County phenomenon; they have been found in various metropolitan areas around the US [145, 83]. Similarly, a previous study that used phylodynamics to analyze differences in SARS-CoV-2 spread in two Wisconsin counties found that the county with the highest basic reproductive number, an approximate measure of local spread in a naive population, was also the county with the higher proportion of people in poverty and lower access to health as well as with the highest proportion of communities of color, which mimics the transmission dynamics and demographic differences seen at a within-county level in King County [108]. While we are unable to ascribe causality, our work adds to the growing body of literature showing a correlation between geographic differences in SARS-CoV-2 transmission and socioeconomic inequities potentially related to the ability to reduce mobility following non-pharmaceutical interventions.

Our results are not without limitations. Whole genome sequencing in WA is conditional on laboratory-confirmed testing in which sample quality must meet minimum requirements in terms of PCR cycle threshold, potentially biasing our dataset towards more symptomatic cases, although previous studies have found no significant difference in viral load between symptomatic and asymptomatic individuals [85, 89, 134]. Additionally, the changing availability of genomic sequencing, as well as of at-home testing, is impacting the chance a case shows up in our data through the period studied (see Figure 3.4b). In order to limit the impact of the increased use of at-home antigen testing, we limited our analysis to only include sequences from before April 2022. Multiple subsampling strategies were considered and implemented in an effort to account for this variation (Supp. Figs 5, 6).

Our phylodynamic analyses are conditioned on inferred King County sequence clusters that are found through the incorporation of contextual sequences from around the world into a temporally-resolved phylogeny. As such, it is possible that differential sampling from other locations could impact our identified clusters. Limited SARS-CoV-2 sequence diversity, especially during periods of rapid transmission, could impact our ability to break up larger clusters [107], which might lead to collapsing multiple introductions into King County into shared clusters. Prior studies have used GLM approaches to ameliorate this bias [93], similar to our use of MASCOT-GLM. Optimally, we would like to avoid having to a priori define local outbreak clusters entirely by, for example, explicitly accounting for locations outside of King County in the model. This is currently not possible due to the additional computational cost of explicitly considering an outside deme. Additionally, Bayesian coalescent models assume random sampling of infected individuals, meaning that targeted sampling, such as super spreader events or contact tracing, could bias our phylodynamic estimations. Such sampling from outbreak analyses may also not be constant through time, complicating *Ne* inferences. Lastly, our *Rt* calculations assume that the change in *Ne* over time is proportional to the change in the number of infected individuals over time.

The transmission dynamics of the SARS-CoV-2 pandemic have been highly heterogeneous across countries. Here we show that even different areas of the same metropolitan region can have different trajectories. Changes in incidence throughout the course of an epidemic can be driven by changes in local transmission, importations, or both. Common methods to estimate incidence and changes in incidence via Rt often ignore or are unable to quantify these differences [31, 32], leading to situations where local health departments have limited information with which to tackle growing case counts. Our local scale genomic epidemiology approach can reveal these differences by quantifying the contribution of importations and local transmission on Rt (Fig 6B) through the joint integration of genomic and epidemiological information. Quantifying changes and differences in contribution to incidence can

directly lead to tailored interventions. For example, in an area where incidence is driven mostly by outside viral introductions, interventions could focus on limiting their impact by implementing testing at the airport or quarantine for recent travelers. Meanwhile, ramping up testing, vaccination, and masking as well as providing medical and economic aid to promote quarantine and isolation without furthering income inequities could be more impactful for areas where local community transmission is the main driver of epidemic growth.

3.6 Methods

3.6.1 Experimental Design and Data Sources

For this retrospective phylodynamic study, we aimed to understand local SARS-CoV-2 transmission dynamics in a diverse, metropolitan county. We analyzed 11,602 whole genome SARS-CoV-2 sequences from King County, WA and 69,588 genome sequences from around the world downloaded from GISAID [143] with sample collection dates between February 1 2020 and March 6 2022. In order to analyze local scale phylodynamics, ZIP code information for our primary dataset from King County was obtained from the Washington State Department of Health (WADOH) on March 22, 2022. 7289 (62%) of genomes from King County were sequenced by UW Virology and 2631 (22%) of genomes from King County were sequenced by Seattle Flu Study / Brotman Baty Institute for Precision Medicine. Three other laboratories (Altius, CDC and WA PHL) sequenced the remaining 1,917 (16%) of genomes collectively.

Time series of zip code-aggregated cases and hospitalizations were found on WADOH and Public Health Seattle King County's (PHSKC) Covid Data Dashboard [2]. Publicly available demographic information by ZIP code was obtained through the U.S. Census Bureau's American Community Survey (ACS). This study utilized both ACS 2015-2019 (5-Year Estimates) and ACS 2020 [1].

Additionally, we obtained mobile device location data from SafeGraph (https://safegraph.com/),

a data company that aggregates anonymized location data from 40 million devices, or approximately 10% of the United States population, to measure foot traffic to over 6 million physical places (points of interest) in the US [81]. We estimated population mobility within and between North and South King County and the in-flow of visitors residing outside of King County from January 2019 to March 2022, using SafeGraph's "Weekly Patterns" dataset, which provides weekly counts for the total number of unique devices visiting a point of interest (POI) from a particular home location. Points of interest (POIs) are fixed locations, such as businesses or attractions. A "visit" indicates that a device entered a building or the spatial perimeter designated as a POI. A "home location" of a device is defined as its common nighttime (18:00-7:00) census block group (CBG) for the past 6 consecutive weeks.

3.6.2 Geographic Scales

To understand local-scale dynamics, most of this study was focused on geographic scales finer than the county level. We divided King County into both Public Use Microdata Areas (PUMAs), which are non-overlapping, statistical geographic areas containing no fewer than 100,000 people each, and general regions, North and South. Information as to how we aggregate ZIP codes into PUMAs and PUMAs into North and South can be found in Supplementary Table 1.

3.6.3 Maximum likelihood tree generation

A temporally-resolve phylogeny was created using the Nextstrain (13) SARS-CoV-2 workflow (https://github.com/nextstrain/ncov), which aligns sequences against the Wuhan Hu-1 reference using nextalign (https://github.com/nextstrain/nextclade), infers a maximum-likelihood phylogeny using IQ-TREE [105] with a GTR nucleotide substitution model, and estimates molecular clock branch lengths using TreeTime [139]. All sequences were downloaded from the GISAID EpiCoV database on May 26 2022 [143].

In order to capture the SARS-CoV-2 epidemic in King County with high resolution and computational efficiency, we created four separate temporally resolved phylogenies that span from February 2020 to March 2022. To do so, we created specific phylogenies for Omicron (Nextstrain clades 21K, 21L, 21M comprising 2856 King County Sequences and 18,817 contextual sequences from around the world), Delta (Nextstrain clades 21A, 21I, 21J comprising 2955 King County Sequences and 19,197 contextual sequences from around the world), Alpha (Nextstrain clade 20I comprising 2941 King County Sequences and 15,406 contextual sequences from around the world), and all other SARS-CoV-2 lineages (2850 King County Sequences, 16,168 contextual sequences from around the world). These builds provided higher resolution during epidemic waves while also being mutually exclusive to sequences found in the alternative builds.

Contextual sequences are needed in order to investigate how King County samples relate to regional and global viral diversity, and to identify local outbreak clusters specific to King County. Given that cluster identification is conditional on the number of background sequences that interdigitate large clades on the phylogeny, we attempted to maximize the number of contextual sequences within the bounds of reasonable computational efficiency. We prioritized sequences from WA and North America in order to optimize regional context. For each variant, we specified contextual data sampling to include up to 10,000 genomes per time-period from WA, sampled from all counties and months, up to 7000 genomes per month from other US states, and up to 5000 genomes per month from the rest of the world. In each variant-specific phylogeny, contextual sequences comprise 83-86% of the total number of sequences. While we expect the number of the clusters to increase with an increasing number of contextual sequences, prior work has shown that changes in the proportion of background sequences that make up the analytical dataset above a proportion of 50% have a limited impact on the number of clusters identified and mean cluster size (Fig. S13 in [114]), and downstream phylodynamic analyses.

Phylogeographic reconstruction of spread around King County was conducted using the same Nextstrain workflow via ancestral trait reconstruction of PUMAs and North and South region geographic attributes. Metadata on ZIP code, PUMA, and region was manually added to the GISAID metadata using the ZIP code information obtained from WADOH as described above.

3.6.4 Clustering

To identify local outbreak groups in King County, we clustered all King County sequences based on inferred internal node location. Following Müller et al [114], we used a parsimony-based approach to reconstruct the locations of internal nodes. Briefly, using the Fitch parsimony algorithm, we inferred internal node locations by considering only two sequence locations: King County and then anywhere else. We then identified local outbreak clusters by selecting groups of sequences in which all their ancestral nodes were inferred to be from King County, up until there was a change in location.

After identifying relevant King County clusters from each of the four variant Nextstrain builds, we then annotated the clusters in a combined dataset.

3.6.5 Subsampling

To reduce computation times in subsequent MCMC analyses, we utilized three different subsampling schemes. Three thousand sequences from King County, WA from identified clusters were chosen either at random, through equal temporal subsampling for every year-week in the studied time period, or via weighted subsampling informed by daily hospitalization counts smoothed using a 14-day rolling average. The random subsampling scheme with 3000 sequences was chosen for the main result as it allowed for better resolution during variant waves.

3.6.6 MASCOT GLM on multiple local outbreak clusters

To analyze the transmission dynamics within and between South and North King County, we used an adapted version of MASCOT [112] on the 3000 subsampled King County clusters and sequences. MASCOT is an approximate structured coalescent approach [113] that models how lineages coalesce (share a common ancestor) within the same locations or migrate between them. In order to distinguish between local transmission and transmission occurring outside of King County, we extended MASCOT to jointly infer coalescent and migration rates from local outbreak clusters [114]. In short, we model the transmission dynamics in King County as a structured coalescent model. We then model the introduction of lineages into King County (independent of whether it is North or South King County) as a backwards in time process of lineages having originated from outside King County. This backwards in time process is assumed to be independent of the transmission dynamics in King County and occurs at a rate given by the introduction rate [114]. The rate of introduction that is estimated as part of the MCMC is allowed to vary over time.

We used generalized log-linear models [111] to estimate whether COVID-19 hospitalizations, cases, seroprevalence, NPIs, and mobility are predictive of SARS-CoV-2 effective population sizes and migration rates over time. The model included error terms to account for observation noise and omitted predictor variables. We implemented a MASCOT-GLM [111] analysis on King County transmission clusters with BEAST2 [23] software, allowing the effective population sizes and the rates of introduction to change every day and every 14 days, respectively. We performed effective population size and migration rate inference using an adaptive multivariate Gaussian operator [12] and ran the analyses using an adaptive Metropolis-coupled MCMC [110].

3.6.7 Empirical Predictors

We chose several predictors to inform estimates of the migration and effective population size of SARS-CoV-2 in King County regions. To inform the effective population size, we used daily COVID hospitalizations (lagged 1-3 weeks), daily confirmed SARS-CoV-2 cases, and percent immunity against SARS-CoV-2 in Western Washington.

Percent immunity for Western Washington was found via the Nationwide COVID-19 Infection- and Vaccination-Induced Antibody Seroprevalence from the Centers for Disease Control (CDC) (39). To include daily values, the monthly seroprevalence surveys estimates were plotted, fit to a spline and daily percent immunity values based on the fitted spline were extrapolated for the time period studied to include as a predictor.

We also used dates of non-pharmaceutical interventions (NPIs) in WA and between-region mobility to inform migration rates between North and South King County. Dates of NPIs were found as part of the COVID-19 US State Policy Database [135]. NPIs included are start and end of emergency stay at home orders as well as closing and reopening of bars and restaurants. We chose not to include the opening and closing of public schools due to a high degree of overlap with the NPIs already included. Washington State closed down public schools on March 16th, 2020, which was only a week before the statewide shelter in place was issued on March 23rd, 2020. Similarly, public schools returned to in-person instruction on April 5th, 2021, which is near to the date of restaurant reopening at the end of February 2021.

To measure movement between North and South King County, we extracted the home CBG of devices visiting either North or South points of interest (POIs) and limited our dataset to devices with home locations in South King County visiting North King County POIs, or vice versa, and to POIs that had been recorded in SafeGraph's dataset since January 2019. For each POI in each week, we excluded home census block groups with fewer than five visitors to that POI. To adjust for variation in SafeGraph's panel size over time, we

divided Washington's census population size by the number of devices in SafeGraph's panel with home locations in Washington state each month and multiplied the number of weekly visitors by that value. To estimate the total number of *visits* from each home CBG each week, we multiplied the number of weekly visitors by the total number of visits divided by the total number of unique visitors in Washington state each week. For each direction of movement, we summed these adjusted weekly visits across POIs and measured the percent change in movement from North to South or South to North over time relative to the average movement observed in all of 2019.

3.6.8 Posterior processing

Parameter traces were visually evaluated for convergence using Tracer (v1.7.1) [136], and 10% burn-in was applied for all phylodynamic analyses. All tree plotting was performed with baltic (https://github.com/evogytis/baltic) and data visualizations were done using Altair [151]. We summarized trees as maximum clade credibility trees using TreeAnnotator and visually inspected posterior tree distributions using IcyTree [152].

Transmission between regions was calculated by measuring the number of migration jumps from North to South King County and vice versa walking from tips to root in the posterior set of trees. In order to account for unequal sampling between the two regions, the rate of migration was estimated as the total number of migration jumps per month in each region divided by the average branch lengths for that region for the same month.

Persistence time was measured by calculating the average number of days for a tip to leave its sampled location (North vs South), walking backwards up the phylogeny from the tip up until node location was different from tip location (following Bedford et al. [16].

3.6.9 Estimating percentage of new cases due to introductions

We estimated the percentage of new cases due to introductions for both North and South King County by adapting the methods previously described in Müller et al [114]. The percentage of cases due to introductions at time t can be calculated by dividing the number of introductions at time t by the total number of new cases at time t. We first represented the total number of new cases in a region as the sum of the number of introductions and the number of new local infections due to local transmission, resulting in the following equation:

$$\pi(t) = \frac{\#\ of\ introductions(t)}{\#\ of\ new\ local\ cases(t)\ +\#\ of\ introductions}$$

We estimated the number of new local cases at time t by assuming the local epidemic in each King County region follows a simple transmission model, in which we estimate the number of new cases at time t as the product of the transmission rate (new infections per day per individual) multiplied by the number of people already infected in that region I. For the number of introductions, we similarly assumed that the number of introductions equals the product of the rate of introduction (introductions per day, which we refer to as migration rate m) and the number of people already infected in that region I. We use the number of infected individuals in the destination region rather than the origin region for calculating the number of introductions since the approximate structured coalescent approach models epidemic processes as backwards-in-time, resulting in the equation containing only information about the number of infected individuals in the destination region. We then rewrote the above equation as

$$\pi(t) = \frac{m(t)I(t)}{\beta(t)I(t) + m(t)I(t)}$$

where *I(t)* denotes the number of infected people in that region at time t. Given the presence of *I(t)* in every element, we factored out *I(t)* to arrive at

$$\pi(t) = \frac{m(t)}{\beta(t) + m(t)}$$

For each region in King County, we considered introductions at time t to be the sum of the introductions coming into the region from outside of King County and introductions coming from the neighboring King County region. Splitting up the introductions by source of contribution, we ultimately defined the percentage of new cases due to introductions at time t for region y as

$$\pi(t) = \frac{m^b_{zy}(t) + m_{out}(t)}{\beta(t) + m^b_{zy}(t) + m_{out}(t)}$$

where $m^b_{zy}(t)$ denotes the backwards migration rate per day from the neighboring King County region z into region y, and m_{out} refers to the migration rate per day into region y from outside of King County.

In a transmission modeling framework, the transmission rate is equal to the sum of the growth rate r and the per-day uninfectious rate where

$$\beta = r + \delta$$

To compute the growth rate in region y, we assume that differences in effective population size between adjacent time intervals can approximate the growth rate r and thus $\frac{d(log(Ne_y))}{dt} \approx r$. In addition, we assumed that $\frac{d(log(Ne))}{dt}$ is independent from the rate of introduction. We calculated the growth rate of the effective population size as

$$\frac{d(log(Ne_y))}{dt} = \frac{log(Ne(t + \Delta t)) - log(Ne(t))}{\Delta t}$$

where $Ne(t)$ denotes the effective population size of a region at time t. We ran our MASCOT-GLM analysis using daily time intervals but calculated Ne using a rolling weekly average in order to smooth our estimates.

By also assuming an expected time until becoming uninfectious for each individual of 7 days [51], we calculated the transmission rate at time t in region y as

$$\beta_y(t) = \frac{d(log(Ne))}{dt} + \delta$$

The rate of introduction per day from outside of King County $m_{out}(t)$ into a King County region y is a parameter that was directly inferred by MASCOT-GLM for each daily time interval by modeling everything outside of King County as a separate third deme.

To compute the backwards migration rate, we first calculate the forward-in-time varying migration rate $m^f_{yz}(t)$ for region y into region z over a linear combination of c different predictors:

$$m^f_{yz}(t) = \beta e^{(\sum_{i=1}^{c} \omega^i \sigma^i p^i(t))}$$

where the forward migration rate $m^f(t)$ is computed via MASCOT-GLM coefficients ω^i, indicators i, log-standardized predictor values p^i for predictor i and the respective error parameter e. The variable β outside the summation refers to the overall migration rate scaler while, ω^i refers to the migration rate scalar for each of the individual c predictors.

From the forward-in-time migration rate $m^f_{yz}(t)$, we can then calculate the backwards-in-time migration rate from state z to state y, $m^b_{zy}(t)$, as the product of the ratio of effective population sizes $\frac{Ne_y(t)}{Ne_z(t)}$ and the calculated forward migration rates:

$$m^b_{zy}(t) = \frac{Ne_y(t)}{Ne_z(t)} m^f_{yz}(t)$$

Where $Ne_y(t)$ refers to the effective population size in region y at time t and $Ne_z(t)$ refers to the effective population size in the neighboring King County region z at time t.

In addition to the calculation of percentage of new cases due to introductions, we repeated the above calculation using only SafeGraph mobility data. We used the in-flow of visitors from outside of King County and movement between each region of King County as approximations for the number of introductions and within-region mobility as an approximation for the transmission rate, following the same equation presented above. When estimating in-flows from outside King County and within-region movement, we applied the same filtering and normalization methods used when estimating between-region movement.

3.6.10 Estimating the effective reproductive number Rt

We calculated the effective reproductive number Rt , the time-varying average of secondary infections, in both regions, using both the daily time-varying transmission rate and the becoming uninfectious rate where $Rt = \frac{\beta}{\delta}$. Additionally, we sought to separate out the contributions of introductions versus local transmission to the Rt_t of each region. To do so, we modified the Rt equation to include the percent of new cases from introductions as an estimate of local community spread only:

$$Rt = \frac{\beta(1-\pi)}{\delta}$$

where refers to the percentage of new cases due to introductions as described above.

To estimate the contribution of introductions from outside of King County separately from that of the neighboring King County region, we calculated Rt using the above equation and the percent of cases from introductions as previously described but omitting introductions from outside King County. Briefly:

$$\pi_{yz}(t) = \frac{m_{yz}(t)}{\beta(t) + m_{yz}(t)}$$

where $\pi_{yz}(t)$ refers to the percentage of cases in region z due to introductions from region y into region z at time t, and m_{yz} refers to the per-day migration rate from region y to z as derived above.

3.6.11 Data Availability

Nextstrain builds, BEAST XMLS, scripts, sequence information, and de-identified data can be found at https://github.com/blab/ncov-king-county.

Acknowledgments

We would like to thank Mike Famulare for assembling the geojsons of King County PUMAs from the US Census Bureau that were used in this study. Clinical and sentinel

laboratories who forwarded specimens for sequencing, and sequencing laboratories that reported data to WADOH. We gratefully acknowledge all data contributors, ie the Authors and their Originating laboratories responsible for obtaining the specimens, and their Submitting laboratories for generating the genetic sequence and metadata and sharing via the GISAID Initiative, on which this research is based. The WADOH Data Science Support Unit for integrating sequencing data with epidemiologic case data. We also thank SafeGraph for providing foot traffic data.

Chapter 4

UNDERDETECTED DISPERSAL AND EXTENSIVE LOCAL TRANSMISSION DROVE THE 2022 MPOX EPIDEMIC

This work was originally published in *Cell* [122]

4.1 Abstract:

The World Health Organization declared mpox a public health emergency of international concern in July 2022. To investigate global mpox transmission and population-level changes associated with controlling spread, we built phylogeographic and phylodynamic models to analyze MPXV genomes from five global regions together with air traffic and epidemiological data. Our models reveal community transmission prior to detection, changes in case-reporting throughout the epidemic, and a large degree of transmission heterogeneity. We find that viral introductions played a limited role in prolonging spread after initial dissemination, suggesting that travel bans would have had only a minor impact. We find that mpox transmission in North America began declining before more than 10% of high-risk individuals in the USA had vaccine-induced immunity. Our findings highlight the importance of broader routine specimen screening surveillance for emerging infectious diseases and of joint integration of genomic and epidemiological information for early outbreak control.

4.2 Introduction

Mpox is a viral zoonotic disease caused by the mpox virus (MPXV), previously referred to as monkeypox virus, that is endemic to West and Central Africa [77, 98]. Prior to 2022, most cases of mpox outside of endemic regions occurred in individuals with either a recent

travel history to Nigeria or with an exposure to live animals from endemic areas. On May 7, 2022, an individual with a travel history to Nigeria was diagnosed with mpox in the United Kingdom (UK) [146]. Following this initial detection, the number of mpox cases without a travel history to endemic countries began to increase rapidly in various regions around the globe consistent with epidemic human-to-human spread [146]. As of July 19, 2023, the Centers for Disease Control and Prevention (CDC) reported 88,549 cases of mpox worldwide since Jan 2022 [29].

The 2022 mpox epidemic was characterized by human-to-human spread outside of endemic areas, mostly in men who have sex with men (MSM) [98, 146, 130]. The long incubation period of 5-21 days [146, 73], as well as the atypical and less severe illness presentation suggests that mpox may have spread undetected prior to initial case discovery. Presymptomatic transmission of mpox has also been documented, suggesting that the epidemic was at least partially fueled by transmission occurring prior to symptom onset [35, 56, 72].

The WHO declared mpox to be a public health emergency of international concern on July 23, 2022, promoting investigations into disease spread, the use of vaccines to control transmission, and potential guidelines for international travel [5]. Individual countries began vaccination efforts in an attempt to curb mpox spread but have been criticized for long delays in starting effective vaccination campaigns in high-risk areas [64]. To date, it is still unclear to what extent continued international travel contributed to the explosive spread of mpox in various global regions and whether or not national vaccination campaigns were wholly responsible for controlling the epidemic.

Genomic epidemiology is uniquely poised to explore global and regional transmission dynamics through the joint integration of viral genomic information and epidemiological metadata. This approach augments traditional public health surveillance, especially when case-based surveillance is limited [155]. While a few studies have looked into the regional spread of mpox at various stages of the 2022 epidemic [125, 59, 61, 80], most relied on very

few pathogen genomes. Overall, the extent of undetected mpox spread and the effectiveness of proposed interventions have yet to be examined. Here we employ recent advances in phylogeographic and phylodynamic methods to estimate changes in case detection rate, the impact of underdetection on transmission, and the role of introductions in promoting local community spread in various global regions. We also examine the impact of vaccination campaigns on epidemic growth and decay in North America as well as estimate the degree of transmission heterogeneity in the declining phase of the epidemic.

4.3 Results

4.3.1 Early mpox spread in Western Europe sparks prolonged outbreaks in Southern Europe, North America and South America

Following initial detection in the UK on May 7, 2022, the number of mpox cases reported worldwide grew rapidly (Figure 4.1). In early May, reported cases were found mainly in Western and Southern, and then Central, Europe where the epidemic peaked around mid-July (Figures 4.1A, B). Beginning in mid-May, however, cases began to be reported in North America, which ultimately led to the largest number of reported cases of any global region studied, peaking at the beginning of August. Around the same time as the North American peak, cases were detected and started rising in South America, which substantially contributed to the later tail of the 2022 mpox epidemic. Similarly, the number of sequences collected increased as more cases were detected, with heterogeneity between regions and North America (primarily the US) submitting the largest number of sequences to GenBank. (Figures 4.1C, D).

To investigate the spread of mpox throughout the course of the epidemic across global regions, we employed a phylogeographic approach with an asymmetrical discrete trait model on 1004 publicly available MPXV sequences subsampled based on confirmed case counts (Figure 4.2, Figure C.1A) in order to infer the global region of origin for all internal ancestral

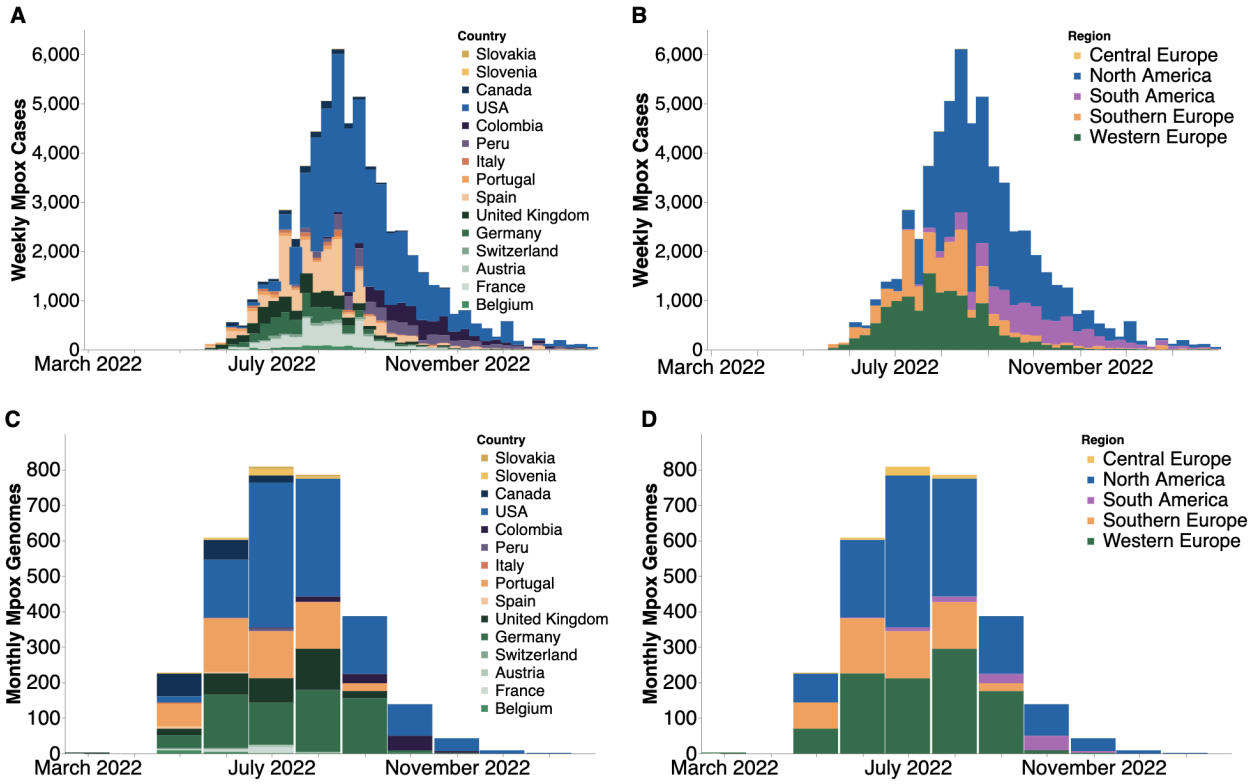

Figure 4.1: Case counts and publicly available sequences by geographic region(**A, B**) Confirmed positive weekly mpox cases by country (**A**) and global region (**B**) smoothed using a 7 day rolling average on daily data and then aggregating into weekly counts. Only countries with greater than 5 sequences on GenBank were included. (**C, D**) Monthly count of publicly-available MPXV genomes found on GenBank by country (**C**) and global region (**D**).

nodes. We chose a case count weighted subsampling scheme since discrete trait analysis assumes that sample sizes across subpopulations are proportional to their relative population prevalence [101]. Due to the low number of recorded cases in Central Europe, no sequences from that region were included in the final subset (Figure C.1B). We infer that the most recent common ancestor (MRCA) of the epidemic existed between March 9th and March 27th, 2022 (95% HPD) and phylogeographic estimation assigns this ancestor to Western Europe. We infer the evolutionary clock rate to be 8.41 × 10-5 (95% HPD 7.71 × 10-5 to 9.10 × 10-5) substitutions per site per year or approximately 16.8 substitutions per genome per year.

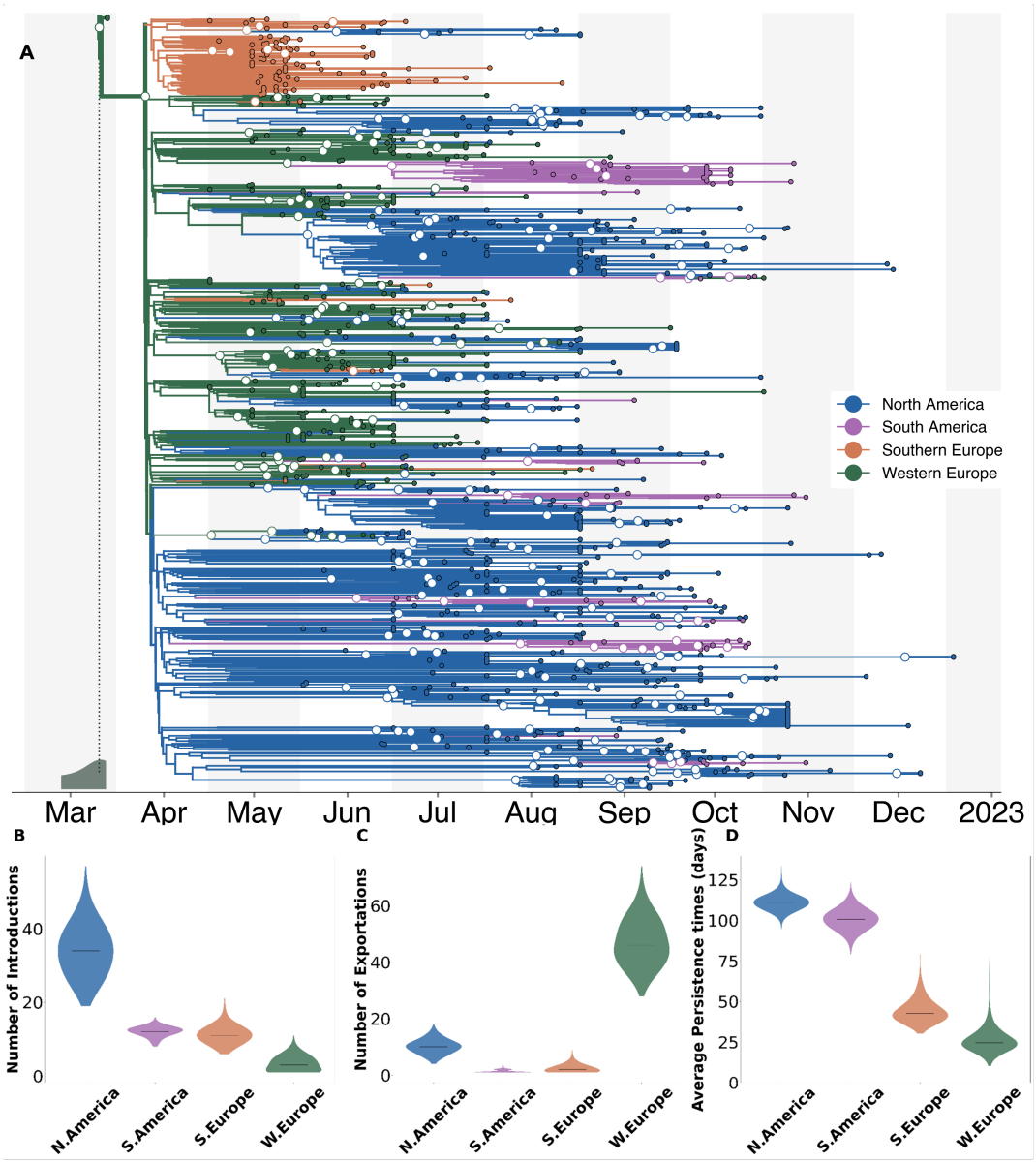

Figure 4.2: Phylogeographical estimates of MPXV spread in 4 global regions. (**A**) The maximum clade credibility tree summary of the Bayesian inference conducted using asymmetric discrete trait analysis and Skygrid prior on 1004 sequences. Colors correspond to the regions in the legend. Ancestral nodes with greater than 50% posterior support are highlighted with a white circle overlaid. Inset histogram on bottom left corner shows 95% interval for the time to most recent common ancestor (TMRCA)(**B-D**) Estimated number of introductions (**B**), exports (**C**), and average time of local persistence in days (**D**) for each global region. Horizontal black line denotes median estimates.

Alternative phylodynamic models place the MRCA in November or December 2021, but show similar estimates of clock rate (Table C.2).

We observe strong population structure where single introductions often result in large local clades. These large local clades suggest that local spread played a considerable role in their respective regional outbreaks. We find rapid early spread in Western Europe lead to a high number of introductions to other global regions (46 introduction events, IQR: 41-53), seeding regional outbreaks (Figure 4.2C). Our findings also show evidence of repeated dissemination into North America and subsequent sustained community transmission as North America had the highest median number of viral importations and longest median viral persistence time (111 days, IQR: 108-114) (Figure 2B,D).

To test the appropriateness and accuracy of the phylogeographic inference, we repeated the analysis in which 10% (100 tips in total) of the sequence locations were masked. We then inferred these locations via the same phylogeographic approach and found that the model correctly inferred 93% of the masked tip locations, suggesting a strong genomic signal (Figure C.2). Additionally, we also repeated our analysis using an equal spatiotemporal subsampling scheme for every year-week in the studied time period as well as by subsampling directly from each region rather than from individual countries and found highly similar estimates of the MRCA and of patterns of introductions, exportations, and persistence regardless of the subsampling scheme used (Figure C.3)

4.3.2 Rapid early spread characterized by significant underdetection of cases

In order to analyze within-region transmission dynamics, improve robustness to sampling bias, and enhance inference via the joint integration of genomic and epidemiological meta-data, we then employed an approximate structured coalescent (MASCOT) with a generalized linear model (GLM) approach with estimated prevalence and air passenger data as empirical predictors on 587 sequences in order to infer the effective population size and migration rates

within and between each region, respectively (Figure C.4A). We also included a predictor for each month within the time period studied to account for potential changes in case detection over time. The included sequences were subsampled with equal temporal weighting to increase representation of undersampled regions such as Central Europe (Figure C.1A, C, see Methods for more information). Despite the improved computational efficiency of MASCOT over standard structured coalescent approaches [112], parameter inference under MASCOT-GLM is still computationally demanding compared to discrete trait analysis. For reference, the runtime for our main DTA analysis with 1004 sequences was about 12.26 hours/million states while for MASCOT-GLM with only 587 sequences it was about 16.45 hours/million states. Using a minimum of $5 * 10^7$ MCMC steps to promote convergence, these runtimes translate to 25.5 days of computational demand for our main DTA analysis and 34.3 days for our main MASCOT-GLM analysis. As such, we reduced the number of sequences to 587 for MASCOT-GLM to allow for inference within actionable timescales (Figure C.1C). Additionally, the MASCOT-GLM subsampling scheme is different from the subsampling for the DTA analysis as the structured coalescent is more robust to differences in sampling across regions and is subsequently informed by regional prevalence [101, 88]. We used a GLM approach in order to draw inferential power from relevant predictors and reduce uncertainty relative to inferences using the coalescent alone. After separating out each introduction and its inferred descendents from the maximum clade credibility tree and comparing them to confirmed case counts, we see strong evidence of viral circulation before initial detection in each global region (Figure 4.3A). Additionally, we revealed that the largest downstream outbreak clusters arise from introductions prior to detection from public health surveillance while introductions after detection are more likely to be a single case and extinguish quickly (Figure C.4B).

We sought to investigate the extent of underdetection in each region by comparing the MASCOT-GLM estimates of effective population size *Ne* (Figure 4.3B) with the prevalence

estimated solely from case counts (Figure 4.3C), which we assume to approximate the census population size. Of note, the MASCOT-GLM estimates are informed by prevalence as an empirical predictor, allowing us to assume that differences between the coalescent-derived *Ne* and case-based prevalence estimates could be due to differential case reporting. While both estimates show regional peaks at similar points in time, we find a divergence between the two estimates in the early months of the outbreak – May, June, July 2022 – where our coalescent-derived *Ne* show continuous viral epidemic growth before case-based prevalence counts report any cases detected by local public health authorities, suggesting significant underdetection of cases in these months. This observation is supported by the estimated coefficients of the monthly predictors that show the direction and magnitude of each predictor's effect on the inference of regional *Ne.* Figure 4.3D shows that the predictors for the months of May, June, and July 2022 had a strong positive effect on predicting regional *Ne.* By August 2022, however, when a substantial number of cases had been detected in all five regions, we see that our model no longer finds the monthly predictors to be required, implying that prevalence estimates are sufficient to describe *Ne.* The strong positive effect of the monthly predictors from May through July, even in the presence of competing information from the prevalence predictor, suggests significant underreporting of cases in these first few months. For comparison with our predictor-informed MASCOT-GLM model, a strictly coalescent-based model without predictors (Figure C.5) shows similar trends in *Ne* but displays a larger degree of uncertainty, supporting the use of empirical predictors to inform our inference.

4.3.3 After initial dissemination, viral importations had limited impact on local spread

When analyzing transmission chains resulting from introductions (Figure 4.3A), we identified a bimodal pattern in each region, where most viral introductions resulted in a single imported case while a very small number of introductions spark explosive and widespread local transmission. Upon identifying the regional introductions with the highest posterior

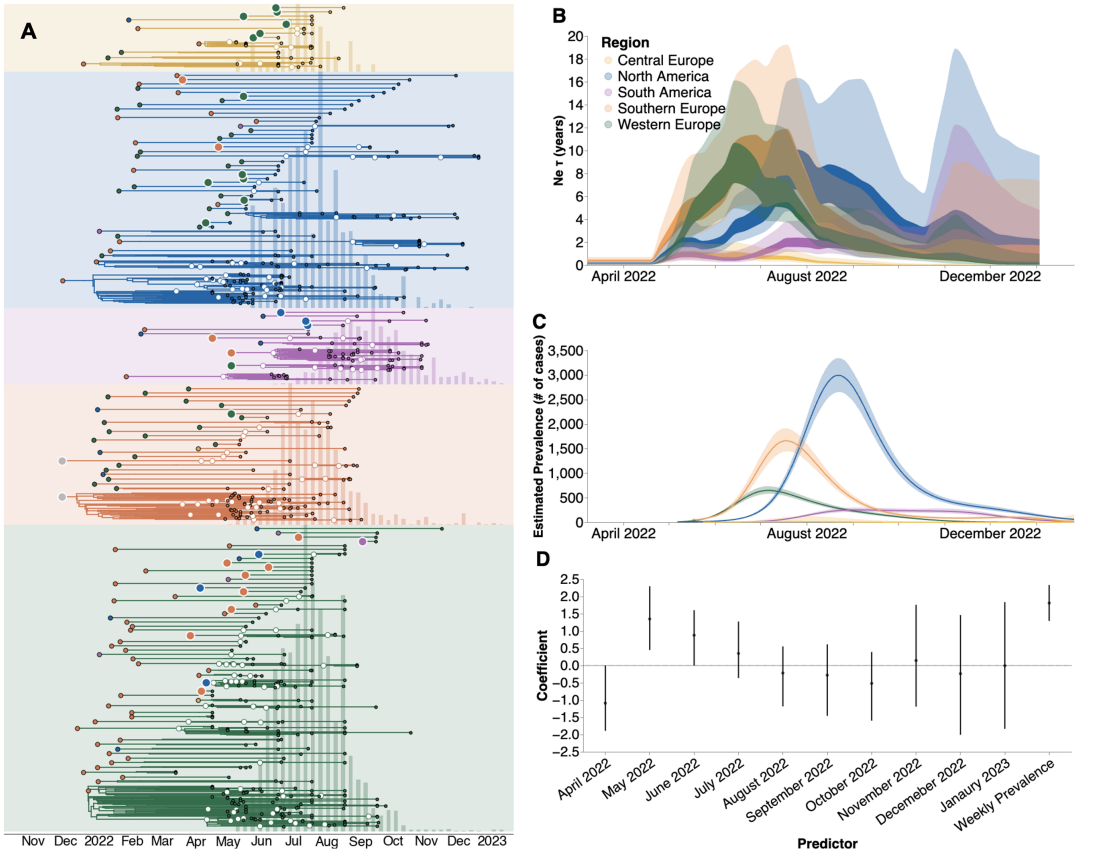

Figure 4.3: Phylodynamic investigation reveals underdetection of mpox. (**A**) Regional specific introductions and the resulting outbreak clusters extracted from the maximum clade credibility tree summary of the Bayesian inference conducted using MASCOT-GLM on 587 sequences. Colors correspond to the regions in the legend. Ancestral nodes with greater than 50% posterior support are highlighted with a white circle overlaid. (**B**) Estimates of effective population sizes ($Ne\tau$ in years) from April 2022 through December 2024 using 550 sequences subsampled equally throughout time. The coalescent time scale depends on both effective population size Ne (number of effective individuals) and on generation time τ (years per generation), resulting in $Ne\tau$ being a measure of coalescent time scale in years [17]. (**C**) Regional prevalence (in number of cases, interpreted as census population size N) estimated independently using publicly-available case counts, and (**D**) Estimates of model predictor coefficients for Ne estimation. All of the predictors displayed on the x-axis were included in the analytic model. Dark line represents median estimates, light bands represent 95% HPD.

support in our MCC tree, we find that introductions that occurred early in the global outbreak lead to larger and more persistent transmission chains, while those introductions that occurred after initial public health detection in each region resulted in smaller outbreaks that extinguished faster (Figure 4.4A, Figure C.4B). The clear negative correlation between time of introduction and persistence of downstream transmission chains remains even without the influence of the two large transmission chains following the first two inferred introductions (Figure C.4C). We also found that air passenger volumes between each global region were a significant positive predictor of viral migration between each region, highlighting the importance of regional connectivity in promoting international viral spread (Figure 4.4B).

We sought to estimate the relative contribution of introductions versus local community spread in driving the epidemic in each global region via inferred parameters from MASCOT-GLM. To quantify the impact of those introductions, we calculated the percentage of new cases from introductions in each region using the estimated changes in *Ne* over time, the rate of viral migration between regions, and the incubation and infectious periods distributions for mpox. We found that introductions played a relatively small role in each regional epidemic, with introductions resulting in an average of 1.5-10% of new cases over the time period studied (Figure 4.4C). We see the percentage of new cases due to viral introductions in North and South America peaking at the start of their respective epidemics and then quickly drop down once the epidemic begins to peak. This finding suggests that following the initial viral seeding from importation events, local transmission dominates and that viral introductions play a very limited role in the later stages of regional epidemics. We also see large variability in the contribution of introductions on local spread during the later months which could be driven by lack of genomic and case-based information at those time periods (Figure 4.4C).

To better understand transmission dynamics locally within each region, we computed *Rt*, the time-varying effective reproductive number, using the estimated growth rate derived from changes in effective population size (Figure 4.5). We also employed our estimates of the

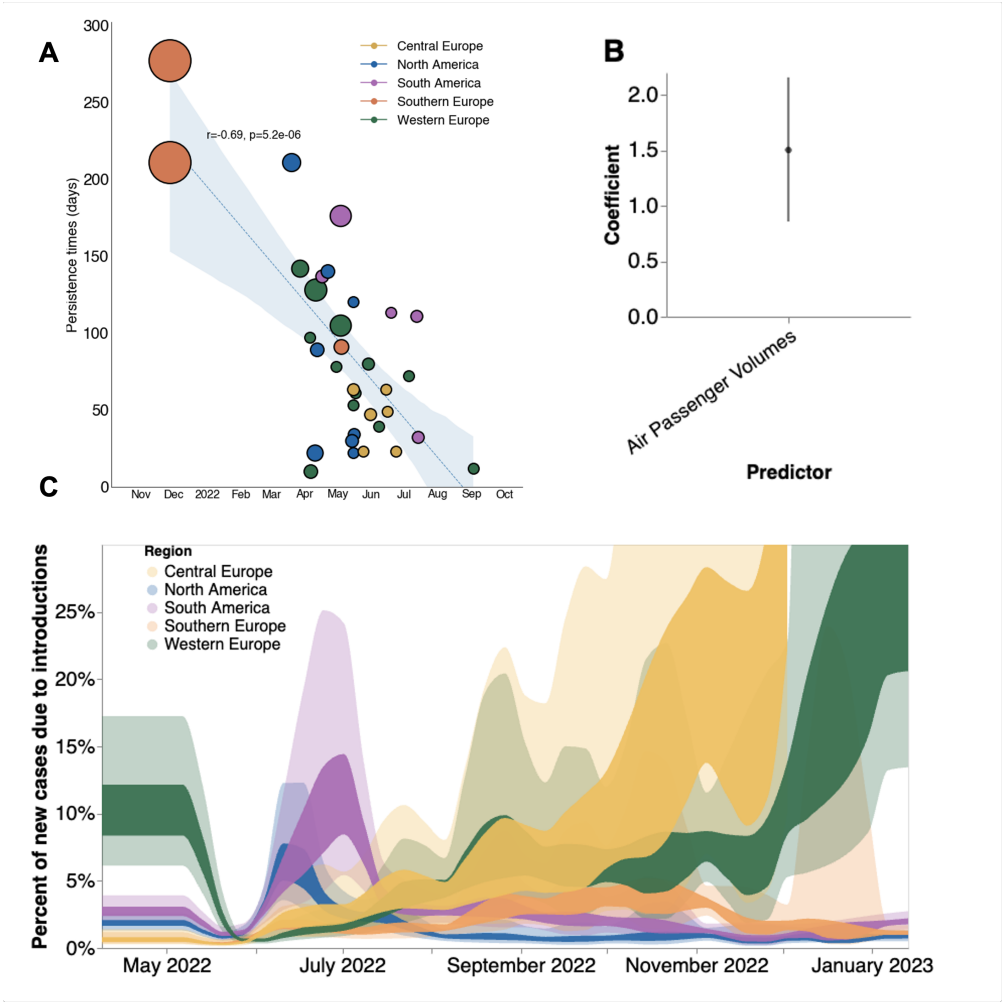

Figure 4.4: Phylodynamic estimates of MPXV transmission dynamics in 5 global regions (**A**) Relationship between estimated date of introduction and persistence time**.** Each circle represents a single viral introduction with greater than 50% posterior support into the region denoted by the color (i.e. a green point represents an introduction into Western Europe). The size of each point is proportional to the size of the outbreak cluster resulting from each introduction with larger circles representing more resulting downstream tips. Blue dashed line represents the linear best fit line using Pearson's correlation. Blue shaded region denotes the variability of the line and the resulting estimates from Pearson's correlation are shown in text above the shaded region. (**B)** Estimates of model predictor coefficients for migration rate estimation. Error bars denote 95% HPD interval for the magnitude of predictor coefficient (**C**) Percentages of new cases due to introductions were estimated as the relative contribution of introductions to the overall number of infections in the region. The inner area denotes the 50% HPD interval and the outer area denotes the 95% HPD interval. Estimates were smoothed using a 14 day rolling average.

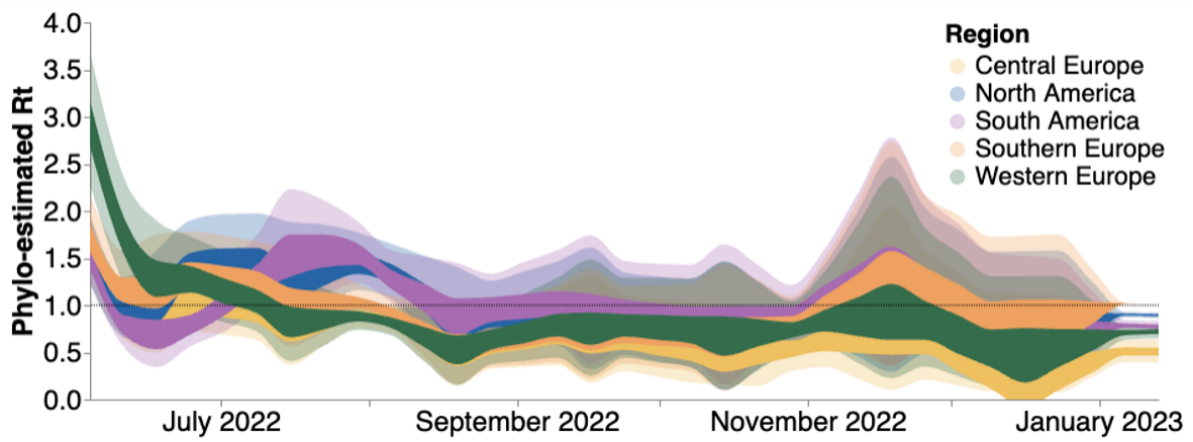

Figure 4.5: Estimates of time-varying reproductive number (Rt) in five global regions. Estimates of *Rt* from April 2022 through December 2023 via MASCOT-GLM using 587 sequences subsampled equally throughout time. The inner area denotes the 50% HPD interval and the outer area denotes the 95% HPD interval. Dashed line highlights an *Rt* value of 1 above which denotes an exponentially growing viral epidemic. Estimates were smoothed using a 14 day rolling average.

percentage of new cases that are due to introductions to calculate *Rt* without the influence of introductions (Figure C.6A). Initially, we observe high *Rt* with viral establishment in each respective region followed by a subsequent rapid decrease in which most regions achieve *Rt* < 1 (signaling an declining epidemic) by September 2022. Initial Rt of 1.5–3 corresponds to an epidemic doubling every 5.6–22.1 days. Removing the contribution of introductions, however, has a very small effect on regional *Rt,* showing the limited impact of introductions on local viral spread after initial regional establishment (Figure C.6A).

Additionally, we calculated *Rt* using only case counts (See Methods) to highlight the impact of accounting for underreporting on *Rt* estimation (Figure C.6B). Compared to our phylodynamic estimates that take into account underreporting in the first months of the epidemic, renewal based methods for *Rt* estimation from only case counts significantly over-estimate the transmissibility of mpox for every region.

4.3.4 US vaccine campaigns had limited impact on curbing the North American outbreak

Given that North America bore the highest burden of mpox cases throughout the epidemic, we focused on this region to explore the role introductions had on prolonging the local epidemic as well as the impact of mpox vaccination on *Rt*. We find that introductions accounted for only an average of about 5-15% of local spread. By focusing on the declining half of the North American epidemic (dates later than June 15, 2022), we additionally found that preventing introductions following the initial seeding event would have caused the *Rt* to fall below one only less than a week earlier (Figure C.6A), highlighting the relatively low importance of introductions.

When we overlaid North American *Rt* estimates alongside the cumulative percentage of high-risk individuals in the USA with mpox vaccine-derived immunity (for description of the data and definitions, see Methods**,** under *Data Sources*), we found that *Rt* began declining prior to initiation of vaccination in the US (Figure 4.6A). Vaccine-induced immunity was estimated via a two week lag since the date of vaccination. North American *Rt* estimates fell below one near the middle of August 2022, when the cumulative percentage of high-risk individuals with vaccine-derived immunity was less than 8%. Under an SIR model of infectious disease dynamics, vaccine-derived immunity impacts disease transmission by removing individuals from the susceptible population in a linear fashion. Before there was any mpox vaccine-derived immunity in the US, North American *Rt* peaked at 1.49. Assuming a linear decrease in *Rt* as cumulative vaccine-derived immunity increased, we would expect *Rt* to fall below 1 only after greater than 33% of the high-risk population of the US developed immunity against mpox (Figure 4.6B, dashed gray line). When we compare the actual decay of *Rt* in North America, we find that *Rt* falls below one before about 10% of the high risk population developed immunity (Figure 4.6B, blue scatter points and red spline), implying that vaccination is not primarily responsible for the drop of *Rt* below 1. The decay of *Rt* in North America before a substantial percentage of high-risk individuals developed vaccine-

related immunity remains clear even when assuming for no lag or a one week lag after vaccination for the development of immunity (Figure C.6C-D). Of note, we were only able to publicly access vaccination information for the US via the CDC but our regional *Rt* analysis for North America includes viral dynamics for both the US and Canada.

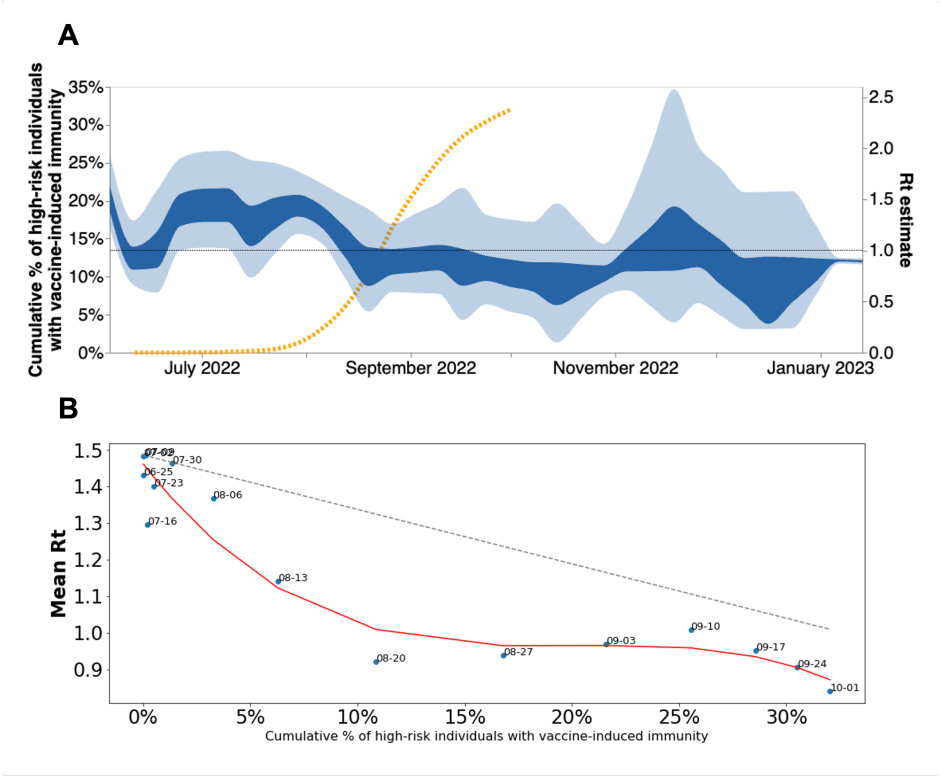

Figure 4.6: North American MPXV local transmission dynamics. (**A**) North American *Rt* estimated via phylodynamics (solid bands). Dashed orange line indicates the cumulative percentage of high-risk individuals with vaccine-induced immunity in the US. (**B**) Scatter plot comparing mean *Rt* calculated via MASCOT-GLM for North America vs cumulative percentage of high risk individuals with vaccine-induced immunity in the United States. Vaccine-induced immunity was estimated via a two week lag since the date of vaccination. Red line indicates the best fit spline for scattered points. Dashed gray line indicates expected linear decrease in *Rt* with increasing vaccine-immunity assuming SIR dynamics. Over each point are the dates that correspond to the mean *Rt* and percent of immunity at that moment.

4.3.5 High degree of transmission heterogeneity observed in the declining phase of the mpox epidemic

Upon separating out each introduction and its inferred descendents from the maximum clade credibility tree (Figure 4.3A), we noticed that a small number of introductions resulted in a sustained expansion of local transmission while the remaining majority of introductions produced few downstream infections. The extent to which some individuals tend to contribute disproportionately to infection events is measured by the dispersion parameter k that quantifies transmission heterogeneity [97]. Lower values of the dispersion parameter correspond to a higher degree of heterogeneity in transmission. When transmission heterogeneity is high, interventions targeting the most infectious individuals can have a considerable impact on epidemic burden. Quantifying transmission heterogeneity is hence important to guide control efforts. We thus sought to quantify mpox transmission heterogeneity using a method relying on the analysis of the size distribution of clusters of identical sequences [148].

We observed that the mean size of clusters of identical sequences decreased over the course of the epidemic (Figure 4.7A). We found that the timing of the decrease across locations was consistent with our estimates of Rt obtained from the analysis of case and sequence data (Figure 4.5), with larger cluster sizes observed in the US than in Europe during June 2022. Globally, the size of clusters of identical sequences ranged from 1 to 118 with 61% of sequences belonging to a cluster of size greater than 1 (Figure C.7). The probability to observe a cluster of a given size is determined by the effective reproduction number R across the period, the degree of transmission heterogeneity measured by the dispersion parameter k and the fraction of infections sequenced (see Methods). Figure 4.7B depicts how the probability to observe a cluster of size 118 (knowing we observed 2624 clusters) is impacted by R and k assuming that 5.5% of infections were sequenced (average proportion of cases sequenced throughout the epidemic). We find that for values of the reproduction number R greater than 1.5, observing a cluster of identical sequences of size 118 is not unlikely

regardless of the value of the dispersion parameter k. This is consistent with the fact that in this parameter regime, the expected mean number of offspring with identical genomes is greater than 1 so that we expect some clusters of identical sequences to not go extinct [148]. For a value of the dispersion parameter similar to what has been estimated during previous mpox outbreaks (e.g. 0.36 in [21]), the reproduction number would need to be greater than 1.31 for this probability to reach 5%. Considering a lower dispersion parameter value (0.1 which is on the lower range of what has been estimated across different pathogens [97]) would still require the reproduction number to be greater than 1.21 for this probability to reach 5%. This suggests that transmission heterogeneity alone (without a reproduction number greater than 1) is unlikely to explain the size of the large polytomy observed at the beginning of the epidemic. Overall, the large first polytomy is highly consistent with a reproduction number greater than 1 at the beginning of the mpox outbreak. This aligns with reproduction number estimates obtained from our phylodynamic analysis (Figure 4.5), which is indicative of mpox spread within the community.

We then estimated R and k during the decreasing phase of the epidemic in different geographical regions (Figure 4.7C-D, Tables C.3 C.4). Assuming that half of infections were detected, we estimated k across locations at 0.30 (95% CI: 0.18-0.54) and reproduction numbers below unity across locations (Table C.3). This corresponds to heterogeneity in transmission with 65% - 72% of infected individuals producing 0 offspring (and hence the remainder responsible for all transmission events). Assuming a greater fraction of infections were detected lead to lower estimates of R and greater estimates of k. This had however little impact on the fraction of individuals producing 0 offspring (Table C.3). Allowing the dispersion parameter k to vary between locations resulted in similar estimates, though with considerably more uncertainty (Table C.4). Our results suggest considerable transmission heterogeneity which could be explained by the structure of the sexual contact network in MSM [137]. Our estimate is consistent with those previously obtained for sexually-transmitted infections spread

between MSM (e.g. dispersion parameter of 0.257 estimated during a gonorrhea outbreak in MSM [162]).

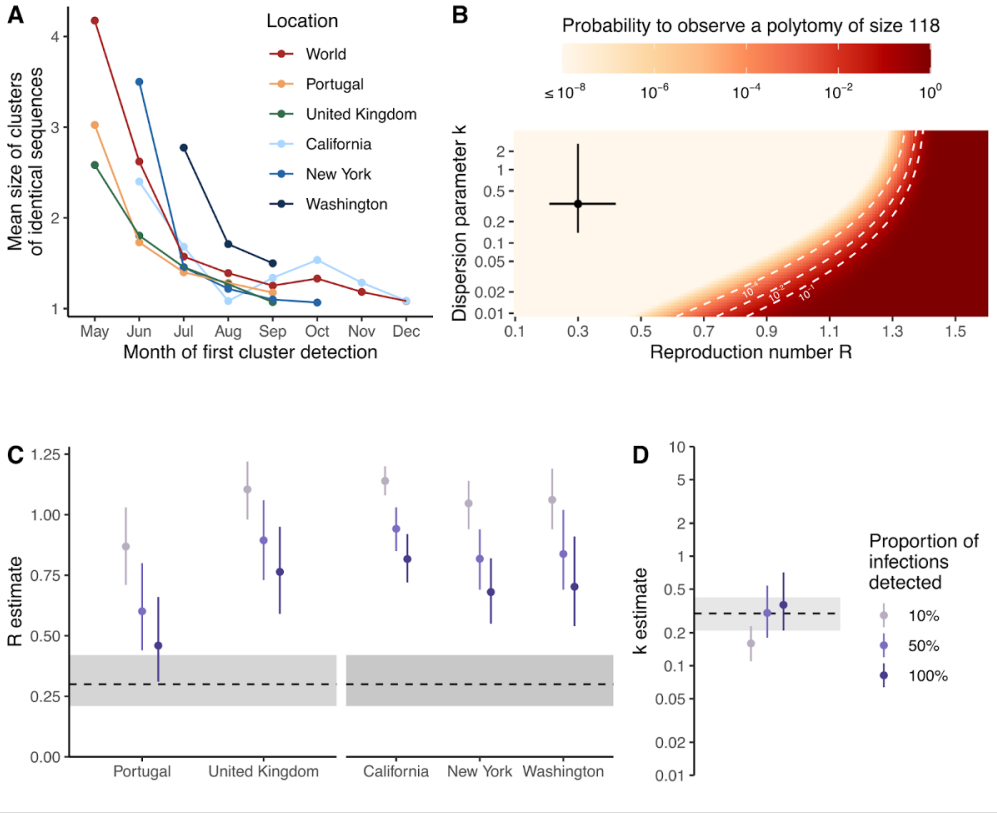

Figure 4.7: Transmission heterogeneity estimates obtained from clusters of identical mpox sequences (A) Mean size of clusters of identical sequences for different geographical regions by month of first cluster detection. **(B)** Probability to observe a cluster of size 118 among 2624 clusters as a function of the reproduction number R and the dispersion parameter k assuming 5.5% of infections are sequenced. Estimates of **(C)** the reproduction number R by geographical unit and **(D)** the dispersion parameter k across geographical units from August 2022 exploring different assumptions regarding the proportion of infections detected. In B, the point corresponds to estimates obtained by Blumberg and Lloyd-Smith (43) from the analysis of epidemiological clusters during previous outbreaks. The segments correspond to the associated 95% confidence intervals. In C-D, points correspond to maximum likelihood estimates and vertical segments to 95% likelihood profile confidence intervals. The horizontal dotted line and the shaded area correspond to estimates obtained by Blumberg and Lloyd-Smith [21] from the analysis of epidemiological clusters during previous outbreaks. In B, the dotted white lines correspond to contour lines for probabilities of 10^{-4}, 10^{-2} and 10^{-1}.

4.4 Discussion

Despite the heightened focus on public health surveillance of emerging infections since the start of the SARS-CoV-2 pandemic, MPXV sparked regional epidemics around the world, contributing to a high degree of morbidity among those affected [137, 53, 106]. In this study, we present both a global and regional view of mpox detection, expansion, and containment by jointly analyzing genomic, mobility, and epidemiological data. We find evidence of rapid spread following initial regional viral seeding events, community transmission prior to detection by local public health surveillance, differential changes in case-detection throughout the epidemic, a limited role of viral introductions in prolonging regional epidemics, a large degree of transmission heterogeneity, and limited impact of vaccination campaigns during the early phases of the North American epidemic.

Despite double stranded DNA viruses typically exhibiting a slower evolutionary rate than RNA viruses [128], clade IIb of MPXV has been found to have a significantly faster evolutionary rate since transitioning to sustained human-to-human transmission driven by APOBEC3 editing [120]. While the evolutionary rate of the variola virus (a closely related poxvirus to MPXV) has been previously estimated to be about 9 $\times$ 10-6 substitutions per site per year [54], we infer the evolutionary rate of the B.1 lineage of MPXV to be 8.41 $\times$ 10-5 (95% HPD 7.71 $\times$ 10-5 to 9.10 $\times$ 10-5) substitutions per site per year or approximately 16.6 substitutions per genome per year (compared to the 1-2 substitutions per genome per year for variola virus). This increased evolutionary rate approaches the rate of many RNA viruses [140] and allows for a strong phylogenetic signal (Figures S2 and S5) to analyze epidemic spread and dynamics.

While prior studies have analyzed the global spread of MPXV via phylogenetic methods [125, 59, 61], they were often limited by small sample sizes and a superficial description of regional trends. Recent advances in phylodynamic and phylogenetic methods have been developed to tackle issues of low genetic diversity and biased sampling where phylodynamic

uncertainty is reduced by the joint inference of genomic information alongside relevant predictors, such as epidemiological and mobility information [93, 111, 114, 102]. In the present study, we leverage these recent advances through the use of MASCOT-GLM, an approximate structured coalescent approach found to be more robust to sampling bias than traditional phylogeographic methods that allows for the integration of important predictors, notably estimated prevalence and air passenger volumes, to inform estimates of local transmission dynamics and regional viral migration.

These phylodynamic estimates, in addition to untangling global dispersion, allow us to explore changes in case detection and the impact of viral introductions on local spread on a regional level, highlighting global differences in epidemic outcomes. Despite the heightened interest in public health surveillance, we found evidence of early undetected spread in each region (Figure 3). These early undetected transmission events were often associated with the largest downstream clusters, while later viral introductions were quickly contained. Additionally, we found a strong influence of monthly predictors for the beginning months of the epidemic – May, June, July 2022 – with regards to estimating regional effective population size. The strong effect of the early monthly predictors implies the presence of significant case underreporting as the prevalence predictor in our model was not solely sufficient to inform inference of *Ne.* Despite worldwide attempts to improve public health surveillance, our study shows the limitations of current surveillance systems, promoting the need for broader routine specimen screening for a wide range of pathogens with outbreak potential.

Rt is a measure of transmissibility and has been widely used for monitoring changes in transmission dynamics and evaluating the impact of interventions [131]. The most common methods for estimating *Rt* rely on a time series of case counts and the distribution of the generation time, and rely on the assumption of constant detection rates [65, 31]. As our results suggest that case detection for mpox varied significantly in the early stages of the epidemic (Fig 3), such methods will result in biased estimates of the reproduction number

and the impact of control measures for the mpox epidemic [65]. In contrast to *Rt* estimates obtained solely from case counts (Figure S6A or [86]), we obtained estimates that are smaller in magnitude by relying on phylodynamic models informed by prevalence estimates and monthly predictors that account for changes in case detection. Case-based *Rt* calculations will be overestimated if case detection is increasing as the estimates capture both the true rise in infections and the rise in detection of infections. For mpox, case-detection stabilizes in August 2022 (Fig. 3D), after which time we expect case-based Rt estimates to be more accurate. This suggests that our approach of integrating multiple data sources would provide a more accurate estimation of mpox transmissibility and of the impact of interventions, especially in the beginning stages of an epidemic where accurate knowledge of *Rt* can have high impact in informing public health action.

An outstanding question raised during the beginning of the mpox epidemic that remains unclear is the potential impact of interventions in preventing and controlling spread [43]. Similar to the early phases of the SARS-CoV-2 pandemic, the MPXV epidemic prompted considerations around travel bans and restrictions in an attempt to curb transmission to previously unaffected areas. While travel bans were ultimately not implemented, the CDC issued a series of travel recommendations and warnings for both individuals exposed to MPXV and for those traveling to areas with a high number of mpox cases on June 6, 2022 [29]. Despite these travel recommendations, our models show that there were already many introduced lineages circulating in North America before June 6th (Figure 3), limiting the impact and effectiveness of these recommendations on curbing disease spread. Our results show that following initial viral seeding, viral introductions played a limited role in promoting local transmission, accounting for less than 15% of new cases in any given region studied (Figure 4). We also found that removing the influence of introductions also would have had limited impact in the timing of North American *Rt* dropping below one (Figure S6). Together this suggests little potential impact of travel restrictions after mid-May 2022 once MPXV had

already been established in the region. Our estimates of transmission heterogeneity, where we found that only 28-35% of infected individuals were responsible for all transmission events observed during the decreasing phase of the epidemic (Figure 7), promotes tailoring public health interventions to high risk groups rather than population-wide policies.

We also examined the impact of vaccination campaigns on controlling the mpox epidemic in North America by comparing changes in local transmission as measured by *Rt* to the cumulative percentage of high-risk individuals in the US with vaccine-derived immunity (Figure 6). While even a half vaccination dose has been found to be effective at providing robust immunity against mpox [58, 163], there was concern over the delayed start of vaccination campaigns in the US. We find that local transmission in North America decreased below one in mid-August 2022 before 10% of the population had any vaccine-induced immunity. In the present analysis, we only accounted for vaccine-derived immunity. However, we can attempt to account for immunity derived from natural infection by comparing mpox cases to the size of the at-risk population. Doing so, we find that less than 2% of the high risk MSM population in the US had reported cases of mpox as of Nov 22, 2023 (see *Methods* under *Data Sources)*. Converting this crude cumulative incidence into an estimate of the total proportion of the population infected requires knowing the reporting rate of mpox infections. If we assume complete reporting then we expect just 2% cumulative incidence, which should have negligible impact on lowering epidemic Rt. However, if the reporting rate was 10% then we expect approximately 20% cumulative incidence, which starts to have an impact on Rt. While we were unable to find a precise estimate of reporting rate in the US, prior studies in Portugal [22] and in North Carolina [132] estimate the rate of detection to be 62% (95% CI, 43%-83%) and 66% (95% CI, 44%-91%). If we assume that the US reporting rate falls on the lower bound of those estimates, we expect 4.7% cumulative incidence. Thus, in general we expect that natural immunity will have played a minor role in reducing epidemic Rt compared to behavioral modification and vaccine-derived immunity.

The incongruent population definitions (high-risk population in the US for vaccine-derived immunity and a single *Rt* estimate for the US and Canada combined) could conceivably bias our conclusions regarding the relationship between Rt and vaccine coverage. Despite the lack of publicly available vaccination information for the whole of Canada, vaccination data for Montreal [118] show that pre-exposure vaccination began at a similar time, if not later, than the vaccination efforts in the US (immunization in Montreal began on May 27, 2022 and in Ontario on June 9, 2022 compared to May 22, 2022 in the US [4]). We believe that the similar, if not later, start of vaccination campaigns in Canada biases our results in a conservative and limited fashion compared to what would be expected if the Canadian vaccination efforts began earlier than those in the US. In addition, the Canadian MSM population is estimated to be 10% of the US population [69, 66] further suggesting that vaccination in Canada should have a limited role in reducing mpox *Rt* in North America. More broadly, our estimates of *Rt* and vaccine-derived immunity aggregate across large spatial regions. Further spatially resolved analyses could provide additional information about the relationship between *Rt* and vaccine coverage.

Mpox in the US and Canada spread predominantly among high-risk MSM populations [19], suggesting that the majority of the North American sequences in our study were derived from a similar (but not identical) population as used to estimate vaccine coverage. Our conclusions are concordant with those from the CDC which also found that *Rt* fell below one in August 2022 when only about 1.3% of the high risk population in the US had any vaccine-induced immunity [28]. Similarly, modeling of mpox in Washington D.C. suggests that behavioral modifications within the MSM community were the main contributing factor to slowing initial mpox spread, but that vaccination campaigns were ultimately needed to definitively curb the local epidemic and prevent future outbreaks [30, 164]. A UK-based modeling study focusing on men who have sex with men found that vaccination could not explain the drop in mpox incidence in the region, but rather attribute the declining incidence

to changes in behavior within the same community [25]. Together, these findings highlight the significant effect of behavioral change among men who have sex with men in curbing the epidemic as well as emphasize the need for prompt public health response in order to maximize the population-level effectiveness of vaccination campaigns.

In conclusion, our study integrates diverse data sources to provide crucial insights on the spread and control of mpox. Despite global efforts to improve molecular surveillance, our study shows that early unrecognized spread was critical to driving the initial epidemic. Once the mpox epidemic was recognized, behavioral modification in the MSM community resulted in a sharp decline in Rt in North America ahead of vaccination rollout in the US. Our findings are relevant for policymakers in promoting broader routine specimen screening as a core tenant of pandemic preparedness. Recent emerging disease outbreaks – Zika, Ebola, SARS-CoV-2, and now mpox – have been characterized by late public health detection and cryptic local transmission as a result [41, 18, 48] Our work shows that rapid pathogen detection and subsequent behavioral change could be sufficient to curb epidemic spread. Additionally, our work prompts swift public health investments and interventions to protect marginalized and vulnerable populations from mpox and other emerging infections [64, 106].

4.4.1 Limitations of the study

Our study has noteworthy limitations. Our genomic data from GenBank only cover a small selection of countries and regions, suggesting that we are missing transmission events that involve unsampled countries especially from regions such as Asia, Oceania, and Africa, although mpox cases in these areas in summer 2022 were limited and unlikely to significantly impact our results. Additionally, the changing availability of genomic sequencing, as well as unequal sampling across the regions study affect the probability that a case shows up as a sequence in our dataset through the period studied. If viruses migrate frequently between our study countries and countries that lack genomic sampling, the lack of samples that might

interdigitate with samples from the study country may affect our ability to distinguish separate introductions. Despite this potential bias, the 2022 mpox epidemic mainly affected Europe and the Americas, which are regions that are well represented in our study, limiting the effect of this bias. Additionally, we attempted to account for this variation by weighting the subsampling for our phylogeographic (DTA) analysis according to confirmed case counts, and by oversampling undersampled regions (and downsampling overrepresented regions) in our MASCOT-GLM analysis (Figure S1) as well as by adding in estimated prevalence as a predictor in the model in an effort to account for this variation.

Bayesian coalescent models assume random sampling of infected individuals, meaning that targeted sampling of superspreader events, or via contact tracing, could bias our phylodynamic estimations. We attempt to quantify the extent of transmission heterogeneity via our estimates of overdispersion (Figure 7). In the analysis of transmission heterogeneity, we explicitly accounted for the fraction of cases sequenced and explored several assumptions regarding the proportion of infections detected by the surveillance system. This was done assuming that all infections had the same probability of being detected as cases and sequenced. Active surveillance targeting larger clusters could lead to underestimating the extent of transmission heterogeneity [21, 20].

We see a discrepancy in the TMRCA between various models (Table C.2) and find our estimates to be highly dependent on the tree prior and thus should be interpreted with caution. Inference of TMRCA is dependent on the estimate of effective population size in early 2022. Different tree priors assume different parametric forms of effective population size and so differ in TMRCA estimates. The rapid exponential growth observed in early 2022 suggests that effective population size should be low in January-March 2022. This information is used by the DTA skyline and skygrid models, as well as the MASCOT-skyline model, resulting in TMRCA estimates close to the earliest March sequences. Consistently, the MASCOT-GLM model estimates the coefficient of the monthly predictor for April 2022

and earlier at -1.09 (95% HPD: -1.89 – 0.00, Fig. 3D), again supporting a small effective population size in this time period. We suggest a conservative interpretation of these results supporting a TMRCA between September 2021 and March 2022.

4.5 Methods

4.5.1 Genomic data and maximum likelihood tree generation

All available MPXV sequences were downloaded from GenBank while excluding sequences from countries with five or fewer sequences, leaving Austria, Belgium, Canada, Colombia, France, Germany, Italy, Peru, Portugal, Slovakia, Slovenia, Spain, Switzerland, United Kingdom, and the USA. Sequences with ambiguous date of collection in the month column, with a sample collection earlier than January 2022, and flagged as being low quality by Nextclade https://docs.nextstrain.org/projects/nextclade/en/stable/user/algorithm/07-quality-control.html) [10] were excluded. Given that the 2022 epidemic was found to be driven by MPXV clade II, lineage B.1 [59, 120], any sequences not part of lineage B.1 were also excluded, resulting in 3013 genome sequences included in our analysis.

A temporally-resolved phylogeny was created using a modified version of the Nextstrain [74] monkeypox workflow (https://github.com/nextstrain/monkeypox), which aligns sequences against the MPXV_USA_2021_MD (accession ON918611) reference using nextalign [10], infers a maximum-likelihood phylogeny using IQ-TREE [105] with a GTR nucleotide substitution model, and estimates molecular clock branch lengths using TreeTime [139]. The resulting phylogeny specific to this dataset can be found at:
https://nextstrain.org/groups/blab/monkeypox/hmpxv1.

4.5.2 Regional geographic scales

Due to the low number of sequences from various countries, we analyzed mpox spread at the scale of global regions. We focused on five regions with the highest number of publicly

available sequences on Genbank: Central Europe, North America, South America, Southern Europe, and Western Europe. Country to region mapping can be found in Table B.1.

4.5.3 Data Sources

Data on the number of reported mpox cases per region per month were downloaded from OWID (https://ourworldindata.org/; last accessed on February 13 2023).

Population sizes for each country were downloaded from the World Bank [6] and aggregated based on respective countries and then regions as described in the previous section.

To compare vaccination rates with changes in Rt, we accessed publicly available vaccination counts from the CDC [8] as well as the cumulative percentage of high risk individuals vaccinated [28]. The CDC defined "high-risk individuals" as MSM for whom preexposure prophylaxis against infection for HIV is clinically indicated as well as MSM who are living with HIV. In order to account for the development of immunity, we followed the CDC method of assuming the development of immunity took two weeks following vaccination [28] and thus only considered individuals as "having vaccine-induced immunity" after reaching two weeks from the date of first vaccination.

We additionally estimated that 1.88% of the high risk population in the US was infected with mpox (calculated by dividing the total number of confirmed mpox infections in the US, which is 31,010 as of November 22, 2023, by 1,647,121, which is the total number of US individuals estimated by the CDC to be at high risk for mpox infection.)

We used air travel data from the International Air Transport Association (IATA) quantifying the monthly number of passengers on origin-destination itineraries between airports in the 15 included countries [62].

4.5.4 Site masking

We found that fewer than 1% of nucleotide positions out of 197,209 total sites in the MPXV sequence alignment were phylogenetically informative, ie. polymorphic. To reduce computational runtime for phylogeographic reconstruction (discrete trait analysis), we masked 90% of invariant positions from the MPXV alignment prior to further analysis. The Nextstrain monkeypox workflow produces a BED file containing phylogenetically uninformative or misleading alignment positions to be masked. A VCF file was generated from the alignment using SNP-sites v2.5.1 [121]. We identified variable positions from the VCF using Pysam v0.20.0 [33]. Next, we selected a random subset of 90% of all invariant positions to remove and appended the remaining nucleotides to the BED file. A new alignment of 19,721 positions was generated with the modified BED file using the Nextstrain workflow. Clock rate estimates inferred with a masked alignment were adjusted by a magnitude of 10 in order to account for the degree of masking.

4.5.5 Phylogeographic analysis

To investigate the dispersal history of MPXV among five global regions, we first conducted an asymmetric discrete trait phylogeographic analysis [92] using the Bayesian stochastic search variable selection (BSSVS) model implemented in BEAST 1.10 [144]. For this analysis, we considered each global region as a discrete location and employed subsampling weighted by mpox case counts for each region using a random seed, resulting in a final subset of 1004 sequences (distribution across countries and regions shown in Supplementary Table C.1). We masked the alignment as described above. We employed a strict molecular clock with a uniform distribution from 0 to 1 and an initial value of 610-5 and a GTR+Γ nucleotide substitution model. We used a Skygrid coalescent tree prior allowing grid points to change every two weeks [63]. Two independent Markov chain Monte Carlo (MCMC) procedures were run for 5108 iterations and sampled every 1000 iterations. Resulting posterior distributions

were combined after discarding initial 20% of sampled trees as burn-in from each of them. We used Tracer 1.7 [136] to assess convergence and to estimate effective sampling size (ESS). These values were all ¿150. We then used TreeAnnotator 1.10 to obtain a maximum clade credibility (MCC) tree removing the first 20% of iterations for burn-in.

The number of viral imports and exports between regions was estimated by calculating the number of regional transitions walking from tips to root in the posterior set of trees and calculating the median as well as the 50% and 95% highest posterior density estimates (HPD). Following Bedford et al. [16], persistence time was measured by calculating the average number of days for a lineage to leave its sampled region, walking backwards up the phylogeny from the tip up until the node location was different from the tip region.

In a secondary analysis, in order to check the accuracy of ancestral state reconstructions as well as the strength of genomic signal, out of the same 1004 sequences, 10% had their locations masked and then reconstructed [40] via the same discrete trait analysis described above. Reconstruction accuracy was assessed by comparing the most likely reconstructed location with the true location.

4.5.6 Estimation of mpox incidence, prevalence, and effective reproduction number via case counts

To jointly estimate mpox case incidence, prevalence, and effective reproduction number, we used the renewal equation framework from Figgins and Bedford [52]. The time-varying effective reproduction number (i.e. the average number of secondary cases infected by a single primary case) was modeled using a 4th order spline with 5 evenly spaced knots assuming a discretized gamma-distributed generation time with mean 12.6 days and standard deviation 5.7 days [73]. Case counts were modeled using a zero-inflated negative binomial distribution. This model produces posterior estimates of daily incidence (defined as the number of newly infected individuals in absolute counts) and effective reproduction number. We then used

this incidence and an assumed gamma-distributed infectious period with a mean of 4.5 days to compute the prevalence, which we define as the number of actively infected individuals in absolute counts [73].

Models were fit to aggregated case counts for each region using full-rank stochastic variational inference. Optimization was performed using the ADAM optimizer with learning rate 4e-3 and for 50,000 iterations and 500 samples were drawn from the approximate posterior.

4.5.7 MASCOT-GLM

To analyze the transmission dynamics within and between each global region, we used an adapted version of MASCOT [112]. MASCOT is an approximate structured coalescent approach [113] that models how lineages coalesce (share a common ancestor) within the same locations and migrate between locations. We used generalized log-linear models [111] to estimate whether estimated regional mpox prevalence and air passenger volumes are predictive of MPXV effective population sizes and migration rates over time, respectively. Additionally, in order to account for differential underreporting by month, ten additional effective population size predictors were added, one for every month of the time period studied from April 2022 through January 2023. Empirical predictors were obtained via data sources described above. The model included error terms to account for observation noise and omitted predictor variables. We implemented a MASCOT-GLM [111] analysis with BEAST2 [23] software, allowing the effective population sizes and the migration rates to change every week. We performed effective population size and migration rate inference using an adaptive multivariate Gaussian operator [12] and ran the analyses using an adaptive Metropolis-coupled MCMC[110] using four chains with a length of 2.5108. For this analysis, we employed equal temporal subsampling to enrich for undersampled regions by randomly choosing a max of 11 sequences per region per calendar month via Augur filter [78]. The unmasked alignment was used for all MASCOT-GLM analyses.

4.5.8 MASCOT- Skyline

In order to investigate the degree of genomic signal and influence of empirical predictors on tree reconstruction, we reran our MASCOT analysis without empirical predictors using a MASCOT-Skyline approach. To allow for population sizes to change over time, we modeled the effective population sizes similar to the Skygrid approach for unstructured populations [63]. We estimated the effective population size for each location between time t=0×tree height, ..., t=1×tree height. Between each time point where we estimated the Ne, we assumed exponential growth. We assume the prior on the effective population size over time to be a Gaussian Markov random field (GMRF) and estimate the variance of the GMRF prior on the effective population size over time. We assume the GMRF prior for each state to have the same variance. We assumed the migration rate to be constant forward-in-time, m^f_{zy}, between states y and z . As the structured coalescent assumes backwards-in-time migration rates, we assumed that the backwards-in-time rate of migration between state y and z, $m^b_{yz} = m^f_{zy} \times \frac{Ne(t)_z}{Ne(t)_y}$. To infer effective population sizes and migration rates over time, we employed an adaptable multivariate gaussian operator [12].

4.5.9 Posterior processing

Parameter traces were visually evaluated for convergence using Tracer, tree distributions were visually inspected using IcyTree [152], and 20% burn-in was applied for all phylodynamic analyses. All tree plotting was performed with baltic (https://github.com/evogytis/baltic) and data plotting was done using Altair [151].

The number of migration events between regions was estimated by calculating the number of regional transitions walking from tips to root in the posterior set of trees and calculating the median as well as the 50% and 95% highest posterior density estimates (HPD). In order to calculate the migration rate for each model, we divided the total migration count for each

tree in the posterior set by the tree length (the sum of all branch lengths) and then calculated the mean and 95% HPD.

4.5.10 Estimating percentage of new cases due to introductions

We estimated the percentage of new cases due to introductions for each global region by adapting the methods previously described [114, 124]. The percentage of cases due to introductions at time t can be calculated by dividing the number of introductions at time t by the total number of new cases at time t. We first represented the total number of new cases in a region as the sum of the number of introductions and the number of new local infections due to local transmission, resulting in the following equation:

$$\pi(t) = \frac{\#\ of\ introductions(t)}{\#\ of\ new\ local\ cases(t)\ +\#\ of\ introductions}$$

We estimated the number of new local cases at time t by assuming the local epidemic in each global region follows a simple transmission model, in which we derived the number of new cases at time t as the product of the transmission rate (new infections per day per individual) multiplied by the number of people already infected in that region I. For the number of introductions, we similarly assumed that the number of introductions equals the product of the rate of introduction (introductions per day per infectious individual, which we refer to as migration rate m) and the number of people already infected in that region I. We use the number of infected individuals in the destination region rather than the origin region for calculating the number of introductions since the approximate structured coalescent approach models epidemic processes as backwards-in-time, resulting in the equation containing only information about the number of infected individuals in the destination region (more information on backwards migration rates below). We then rewrote the above equation as

$$\pi(t) = \frac{m(t)I(t)}{\beta(t)I(t) + m(t)I(t)}$$

where $I(t)$ denotes the number of infected people in that region at time t. Given the presence of $I(t)$ in every element, we factored out $I(t)$ to arrive at

$$\pi(t) = \frac{m(t)}{\beta(t) + m(t)}$$

For each region, we considered introductions at time t to be the sum of the introductions coming into the region from each other global region, assuming a negligible number of introductions from unincluded regions. We define the percentage of new cases due to introductions at time t for region y as

$$\pi_y(t) = \frac{\sum_{i \neq y} m^b_{y \to i}(t)}{\beta_y(t) + \sum_{i \neq y} m^b_{y \to i}(t)}$$

where $m_{i \to y}$ denotes the migration rate per lineage per day into region y from every other region.

In a SEIR transmission modeling framework (employed due to the incubation period of MPXV), the transmission rate is a function of the infectious period , the incubation period , and the exponential growth rate r *(*as adapted from Example 4 in Ma 2020 [99]):

$$\beta = \frac{(2r + \gamma + \sigma)^2 - (\sigma - \gamma)^2}{4\sigma}$$

To compute the growth rate in region y, we assume that differences in effective population size between adjacent time intervals can approximate the growth rate r and thus $\frac{d(log(Ne_y))}{dt} \approx r$. In addition, we assumed that $\frac{d(log(N_e))}{dt}$ is independent from the rate of introduction. We calculated the growth rate of the effective population size as

$$\frac{d(log(Ne_y))}{dt} = \frac{log(Ne(t + \Delta t)) - log(Ne(t))}{\Delta t}$$

where $Ne(t)$ denotes the effective population size of a region at time t. We ran our MASCOT-GLM analysis using weekly time intervals but averaged over three week intervals ($\Delta t = 3$) for the growth rate in order to reduce noise and account for the long generation time for mpox.

By also assuming an expected time until becoming uninfectious for each individual of 4.5 days and an incubation period of 8 days [6], we calculated the transmission rate at time t in region y as

$$\beta = \frac{(2\frac{d(log(Ne_y))}{dt} + \gamma + \sigma)^2 - (\sigma - \gamma)^2}{4\sigma}$$

Since the coalescent, which MASCOT approximates, works backward-in-time, we calculated the rate of introductions into each global region $m_y(t)$ as the backwards migration rate $m_y^b(t)$ from inferred MASCOT parameters. To compute the backwards migration rate, we extract the forward-in-time migration rate $m_{yi}^f(t)$, where i refers to a different region in a combination of global regions c, that is inferred via MASCOT-GLM, and then calculate the backwards-in-time migration rate into region y, as the sum of the products of the ratio of effective population sizes $\frac{Ne_y(t)}{Ne_z(t)}$ and the forward migration rates:

$$m_y^b(t) = \sum_{i=1}^{c} \frac{Ne_y(t)}{Ne_i(t)} \times m_{yi}^f(t)$$

where $Ne_y(t)$ refers to the effective population size in region y at time t and $Ne_i(t)$ refers to the effective population size in a different region i from a combination of global regions c at time t.

4.5.11 Estimating the effective reproductive number Rt from pathogen genomes

We calculated the effective reproductive number Rt , the time-varying average of secondary infections from a primary infected individuals, in each region, assuming an exponentially distributed infectious and incubation period of mean respectively $\frac{1}{\gamma}$ and $\frac{1}{\sigma}$, yielding $Rt = (1+\frac{r}{\gamma})(1+\frac{r}{\sigma})$ [157] Additionally, we sought to separate out the contributions of introductions versus local transmission to Rt_t in each region. To do so, we modified the Rt equation to include the percent of new cases from introductions as an estimate of local community

spread so that $Rt = (1 + \frac{r}{\gamma})(1 + \frac{r}{\sigma})(1 - \pi)$, where refers to the percentage of new cases due to introductions as described above.

4.5.12 Estimating transmission heterogeneity

We analyzed the size distribution of clusters of identical mpox sequences to characterize the disease's offspring distribution [148]. We assumed that the offspring distribution follows a negative binomial distribution characterized by its reproduction number R and its dispersion parameter k [97]. The probability r_j that a cluster of identical sequences of size j can be derived as

$$r_j = \frac{\Gamma(kj + j - 1)}{\Gamma(kj) \times \Gamma(j + 1)} \times \frac{(\frac{pR}{k})^{j-1}}{(1 + \frac{pR}{k})^{j-1}},$$

where p denotes the probability that a transmission event occurs before a mutation event.

In practice, only a fraction of infections are sequenced. The probability r_j to observe a cluster of size j was thus derived as:

$$\tilde{r_j} = \frac{\sum_{l \geq j} r_l \binom{l}{j} (p_{detect})^j (1 - p_{detect})^{l-j}}{1 - \sum_{l \geq 0} r_l (1 - p_{detect})^l}$$

where p_{detect} denotes the fraction of infections sequenced.

The probability for an observed cluster of identical sequences to be of size at least J can then be computed as $P_J = 1 - \sum_{j=1}^{J-1} \tilde{r_j}$. The probability to observe at least a cluster of size J among n_{clust} clusters is thus equal to $1 - (1 - P_J)^{n_{clust}}$.

Former work has shown that the size distribution of clusters of identical sequences can be used to infer the reproduction number and the dispersion parameter when the mean number of offspring with identical sequences lies below 1 [148]. For the 2022 mpox epidemic, this would correspond to values of the reproduction number lying below 1.5 [148]. To ensure this criterion was met, we analyzed the size distribution of clusters of identical mpox sequences for different geographical units (Portugal, the United Kingdom and US states California,

New York and Washington) from August 2022, which corresponds to the decreasing phase of the epidemic (Figure S7). We generated the size distribution of clusters of identical mpox sequences for these different geographical units and defined clusters temporally based on their date of first detection. We estimated the fraction of cases sequenced in these different regions from August 2022 by computing the ratio between the number of sequences used and the number of cases publicly reported. We first inferred the parameters of the offspring distribution assuming that the dispersion parameter was the same across these geographical units and estimating a reproduction number for each of these geographical units. This was done by considering different assumptions regarding the fraction of infections detected (10%, 50% and 100%) and assuming a probability that transmission occurs before mutation equal to 66% [157].We also ran a location-specific model and estimated the reproduction number and the dispersion parameter for these each region. We assumed that clusters of identical sequences stemmed from local transmission dynamics. This hypothesis is supported by the small contribution played by introductions estimated from the phylogeographic analysis.

We also generated the distribution of cluster sizes worldwide. We explored how different assumptions regarding R and k impacted the probability to observe a cluster of size 118 (the largest cluster observed) among the 2624 clusters of identical sequences observed. This was done assuming that 5.5% of infections were sequenced (which corresponds to the fraction of cases sequenced since the beginning of the epidemic).

4.5.13 Data Availability

Nextstrain builds, BEAST XMLS, scripts, sequence information, and de-identified data can be found at https://github.com/blab/mpox-dynamics. All sequences are available on GenBank with accession numbers found in the Github repository.

4.6 Acknowledgments

We would like to thank Allison Black for constructive feedback, discussions, and edits. We gratefully acknowledge all data contributors, i.e. the Authors and their Originating laboratories responsible for obtaining the specimens, and their Submitting laboratories for generating the genetic sequence and metadata and sharing via GenBank. The laboratories and institutions that contributed more than ten sequences for this study are as follows: Los Angeles County Public Health Laboratories, Los Angeles County Department of Public Health, UW Virology, Laboratory Medicine, UKHSA, Research and Evaluation, CDC, DHCPP-PRB, Robert Koch Institute, Centre for Biological Threats, Highly Pathogenic Viruses, National Institute of Health Doutor Ricardo Jorge, Portugal (INSA), Department of Infectious Diseases, Laboratorio Departamental de Salud Publica de Antioquia, Antioquia, Instituto Nacional de Salud, Direccion de Investigacion en Salud Publica, National Microbiology Laboratory, Public Health Agency of Canada, Institute National de Saude Doutor Ricardo Jorge (INSA), Portugal, CDPH, VRDL, IHU - Mediterranee Infection, MEPHI, Centre for Biological Threats, Highly Pathogenic Viruses, Robert Koch Institute, Germany, New Jersey Department of Health, Public Health and Environmental Laboratories, Rush University Meical Center, Regional Innovative Public Health Laboratory (RIPHL), University of Nebraska Medical Center, Environmental, Agricultural, and Occupational Health, Universidad Tecnologica de Pereira, Laboratorio de Biologia Molecular y Biotecnologia / Facultad de ciencias de la salud, Royal Infirmary of Edinburgh, Viral Genotyping Reference Laboratory, Institute of Microbiology and Immunology, Faculty of Medicine, University of Ljubljana, Laboratory for Diagnostics of Zoonoses and WHO Centre, Institute of Tropica Medicine, Department of Clinical Sciences, Medical University of Vienna, Center for Virology, Instituto Nacional de Salud Peru, Laboratorio de Biotecnologia y Biologia Molecular.

Chapter 5

CONCLUSION

The preceding chapters demonstrate the potential for pathogen genomics to augment epidemiological investigations into understanding disease spread in heterogeneous contexts as well as in evaluating the impact of infection control measures (which include vaccination, travel bans, and stay-at-home orders). In Chapter 2, I discuss how pathogen genomes linked to individual-level metadata allowed for the examination of SARS-CoV-2 variant specific effects on the risk of hospitalization following infection. When previous studies examined the risk of hospitalization following infection with a SARS-CoV-2 variant [15, 95], classification of the variant was often done via PCR testing for those variants that exhibited an S-gene target failure (SGTF). Unlike studies using only SGTF data to identify probable Omicron cases, our study used genomic sequencing to confirm the variant identity of each case, reducing the risk of misclassification. Additionally, genomic sequencing allowed us to analyze differences between variants that could not be differentiated via SGTF (SGTF is primarily a signature of Alpha and Omicron BA.1 [104]), highlighting the broad utility of our approach not only to future variants without SGTF but also to future emerging infections.

It is important to note that our study was possible due to the robust sentinel surveillance system and the detailed and careful matching of genomic information to epidemiological and case data that is available in Washington State [116]. Sentinel surveillance systems that focus on representative sequencing of communicable diseases are instrumental for monitoring of the rise and spread of emerging and endemic pathogens and their respective genetic variants. In Chapter 4 for example, I showed that despite the general advocacy for improved surveillance following the COVID-19 pandemic, we found cryptic local transmission and nu-

merous introductions of mpox in the Americas and three European subregions prior to public health detection of the first mpox case. These conclusions should constitute a warning signal for public health jurisdictions to swiftly build and improve broad pathogen surveillance infrastructures in order to prepare for future outbreaks.

While Chapter 2 demonstrated how to use matched epidemiological and genomic data to examine variant-specific effects on individual-level outcomes, Chapters 3 and 4 represent how the integration of diverse data sources alongside pathogen genomics can improve our understanding of infection spread at a population-level, especially with regards to infections marked by asymptomatic and presymptomatic spread [100, 158]. Chapter 3 showed how enhancing our phylodynamic methods with empirical epidemiological predictors drove exploration into the relative contributions of viral importations or local transmission in driving differing epidemics at a local level. Methods that rely only on epidemiological case count data often ignore or are unable to disentangle the interplay between introductions and local transmission in driving disease incidence [31, 32], limiting the ability of public health to tailor interventions and effectively distribute often-limited resources.

Chapter 4 builds upon this work to show that the use of empirical predictors in phylodynamic models not only results in more precise inference of within and between region transmission dynamics, but also allows us to test hypotheses regarding the drivers of disease spread. For example, the inclusion of monthly predictor variables alongside estimated prevalence in our MASCOT-GLM model examining transmission dynamics for the 2022 mpox epidemic allowed us to test the hypothesis that there was underdetected mpox spread that was occurring before public health detection. Additionally, we were able to quantify the effect size of each monthly predictor, giving us insight into the extent of underreporting on a monthly basis and a relative quantification of how case-reporting was changing over time. Given that traditional epidemiological methods that rely solely on case counts often assume a constant case reporting rate [31], our inclusion of monthly predictors while also adjusting for

estimated prevalence allowed us to estimate important epidemiological parameters, such as the time-varying effective reproductive number (Rt), in a less biased manner. This suggests that our approach of integrating multiple data sources provides a more accurate estimation of measures of transmissibility and of the impact of interventions, especially in the beginning stages of an epidemic where accurate knowledge of Rt can have high impact on informing public health action.

Finally, Chapters 3 and 4 also highlight that emerging infectious diseases do not occur in a vacuum; they exploit inequitable systems affecting marginalized communities. In Chapter 3, I demonstrated how even two adjacent regions within the same US county can have such starkly different SARS-CoV-2 epidemic dynamics. I found that South King County bore a heavier burden of disease throughout the entire epidemic when compared to North King County, especially following the implementation of stay-at-home orders. In an attempt to understand the drivers of this difference, I explored different markers of socioeconomic inequities, both past and present.

The stark contrast in health outcomes between North and South King County has been previously attributed to historical redlining and systemic racism, whereby decades of racial segregation prevented communities of color from residing in northern areas of Seattle, limiting their residence to present day South King County [4]. Previous work has demonstrated that redlining and structural racism are major contributors to poor health outcomes through the suppression of economic opportunity and human capital [44]. We can see this effect in King County, WA: South King County has a lower median household income, a larger percentage of essential workers that cannot work from home, a larger population of people of color, and a higher average household size than North King County, despite a smaller overall population size. The legacy of structural racism resulted in conditions where individuals living in South King County were at higher risk for SARS-CoV-2 infection when compared to their counterparts in South King County. This is salient in the fact that in response to

stay-at-home orders residents in North King County not only reduced their movement more than those in South King County but were also able to stay at home through 2022, more than two years after South King County residents had to return to their typical levels of movement (See 3.1). Given that previous studies have attributed differences in local case counts to unequal reductions in mobility [83, 145], Chapter 3 presents a clear picture of how emerging infectious diseases can highlight and exploit existing, historically-determined systems of inequity, disproportional affecting already-marginalized communities.

In Chapter 4, I describe the transmission dynamics of the 2022 mpox epidemic that spread though queer sexual networks, primarily affecting men who have sex with men (MSM). There was and continues to be a concern that anti-LGBTQ+ sentiments around the world worsened the epidemic. März et al. [109], argue that socio-political factors surrounding anti-LGBTQ+ sentiments were poorly considered when crafting a global public health response. They outline two main factors. The first is that outbreak control measures and public health action might be stunted due to regional anti-LGBTQ+ sentiments, given the historical reality that occurred during the rise of HIV in the 1980s and the still-present criminalization of queerness and same-sex acts as well as the ongoing discrimination against queer personhood [64]. The second factor being that the rise of mpox within queer sexual networks could further discrimination against LGBTQ+ individuals. The authors document instances of mpox being used by anti-LGBTQ+ politicians in Latvia, Iran, and the US that might have led the queer community to be less likely to seek healthcare and vaccination, not only increasing morbidity within the community but also reducing the effectiveness of public health interventions. Indeed, in the US, there were several documented cases of affected individuals having difficulty receiving appropriate care for mpox [57].

Promoting LGBTQ+ health equity early on could also have ameliorated the impact of mpox. März et al. [109] outline policy recommendations that could promote LGBTQ+ equality in mpox response. The most salient example that could have been enacted early

is community involvement in the processes to develop mpox outbreak response measures, including development of evidence-based and nondiscriminatory information to the general public and healthcare providers, and tailored information to high risk groups, as well as funding and institutional support for community based outbreak response strategies.

Studies have consistently shown that once the queer community was aware of the risk of mpox, they took steps to modify their behavior. From August 5-15, 2022, the CDC conducted an mpox-specific survey among MSM. Their results showed that 48% of respondents reported reducing their number of sex partners, 50% reported reducing one-time sexual encounters, and 50% reported reducing sex with partners met on dating apps or at sex venues since learning about the mpox outbreak [36]. A number of modeling studies have also found that, given the delayed vaccine rollout in many countries, behavior change, in combination with infection-acquired immunity, was a key driver of the decline of the 2022 mpox epidemic[25, 46, 166]. The documented willingness of the MSM community to change their behavior upon learning about the mpox outbreak, and the subsequent studies showing that behavior change strongly promoted the decline of mpox cases, shows that improving targeted and tailored communication to vulnerable populations delivered early could have resulted in faster outbreak control.

One possible model for how public health could have worked in tandem with the queer community can be found in New York. By working together with queer activists, the New York City Department of Health assembled a fleet of vans with privacy windows that not only provided information and testing but also acted as mobile vaccination units, bringing vaccines to bars, bathhouses, and local sex parties [96, 119]. Additionally, the Department of Health contracted local activists to convince sex parties and bathhouses to pause activity, especially before the arrival of vaccines. A similar template, tailored to the community in various cities, could have been adopted to help promote the dissemination of information and vaccination in a non-stigmatizing manner.

It is important to note that the documented behavior change in queer communities in response to the 2022 mpox epidemic is linked to perceived risk. This means that in order to promote behavior change we must first be aware of the existence and spread of an emerging disease. Chapter 4 shows that the 2022 mpox epidemic, similar to recent emerging disease outbreaks such as Zika, Ebola, SARS-CoV-2, was characterized by late public health detection and cryptic local transmission as a result. In order for behavioral change to be an effective tool for slowing the spread of emerging infections, it must be coupled with robust public health surveillance systems that is augmented by routine pathogen sequencing [116].

The above statements are in no way meant to imply that the responsibility for controlling mpox falls on the queer community, but rather are meant to highlight public health interventions that could have been implemented early in the outbreak in the setting of current governmental structures and global public health policies. Increased global vaccine and medication equity, stronger queer, racial and ethnic inclusion policies, improved public health surveillance systems, faster and more coordinated public health responses could have all worked to prevent the emergence of mpox, reduce preventable suffering in endemic areas, and contributed to faster control of the 2022 epidemic [9, 64, 138].

Collectively, the body of work presented herein represents a push towards the joint integration of genomic and epidemiological metadata into the study of the epidemiology of emerging pathogens. As shown throughout the text, pathogen genomes augment traditional epidemiological study designs, allowing for estimation of lineage specific effects on human health. Additionally, incorporating epidemiological and mobility data into our phylodynamic analyses also allowed for more precise inference of epidemic dynamics and for the ability to test hypotheses on the drivers of disease spread. The unification of epidemiological and genomic data, alongside efforts to create and sustain robust genomic surveillance systems, has the potential to strengthen our investigations on emerging pathogens, especially in the aid of marginalized populations that bear the disproportionate burden of infectious diseases.

Appendix A

SUPPLEMENTARY MATERIAL FOR CHAPTER 2

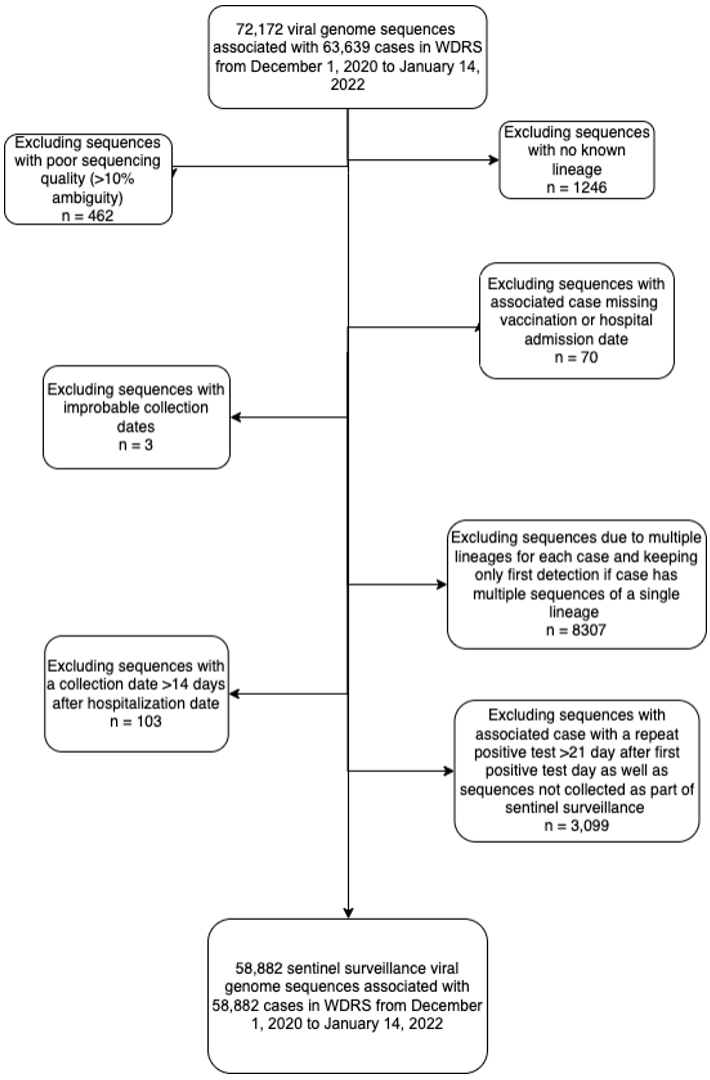

Figure A.1: Inclusion flow diagram for study population

.

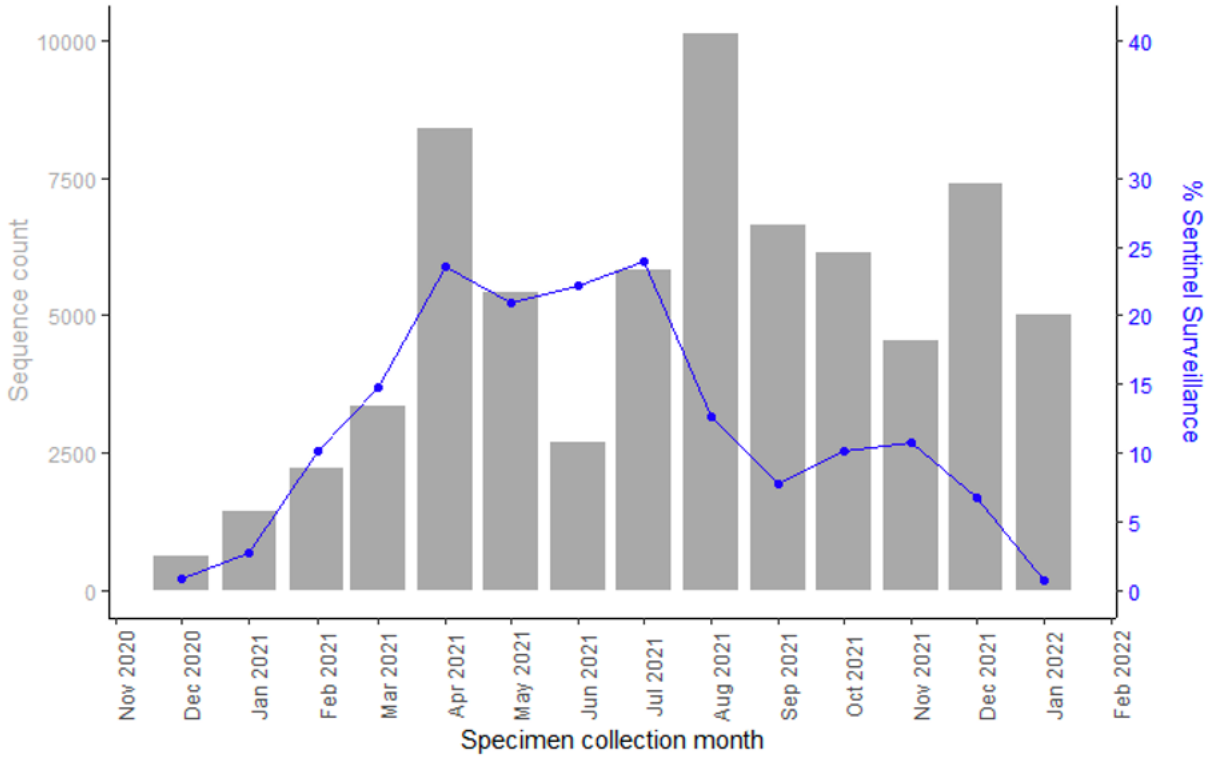

Figure A.2: Proportion of total SARS-CoV-2 cases in Washington sequenced over time as part of sentinel surveillance. Bars represent total sequence count while blue line represents percentage of total SARS-CoV-2 positive cases in Washington that sequenced as part of sentinel surveillance for each time period. Specimens submitted from sentinel labs have decreased in January 2022 due to lab capacity during the omicron peak.

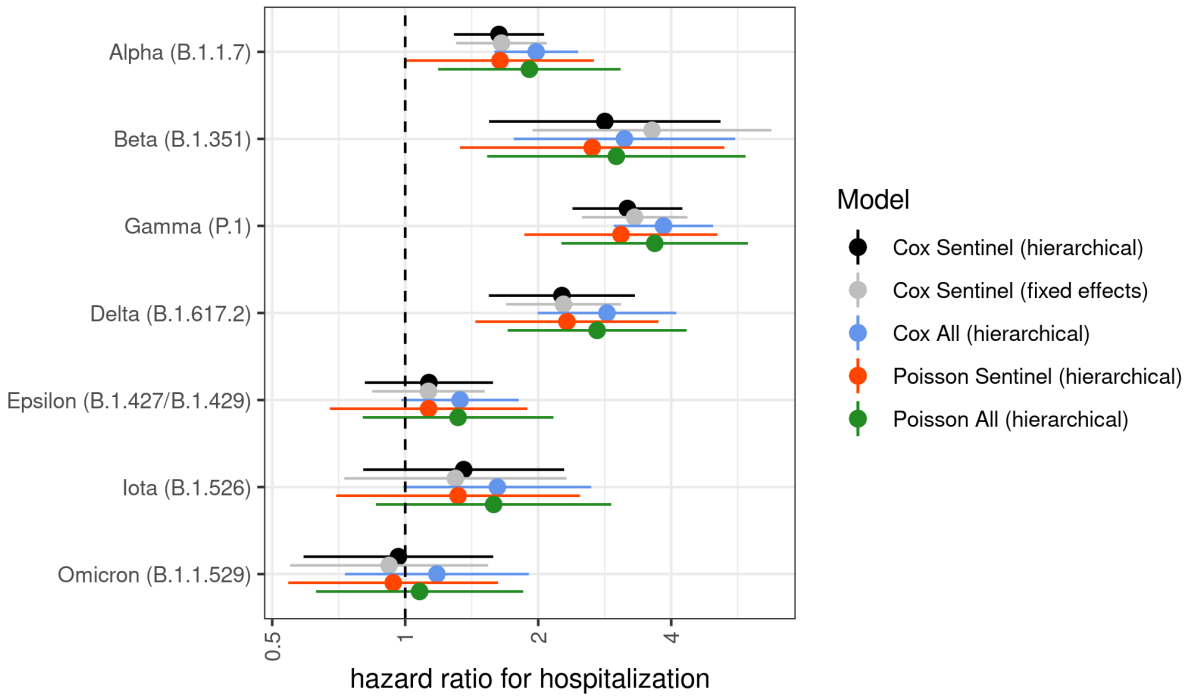

Figure A.3: Risk of Hospitalization by Variant Lineage via differing model selection. Error bars represent 95% CI. "Hierarchical" refers to a model with mixed effects and "Sentinel" refers to the sample restricted to only to cases collected through sentinel surveillance. "Fixed effects" describes all model covariates being treated as fixed effects and "All" refers to the entire dataset from Dec 1, 2020 to Jan 14, 2022, irrespective of participation in sentinel surveillance.

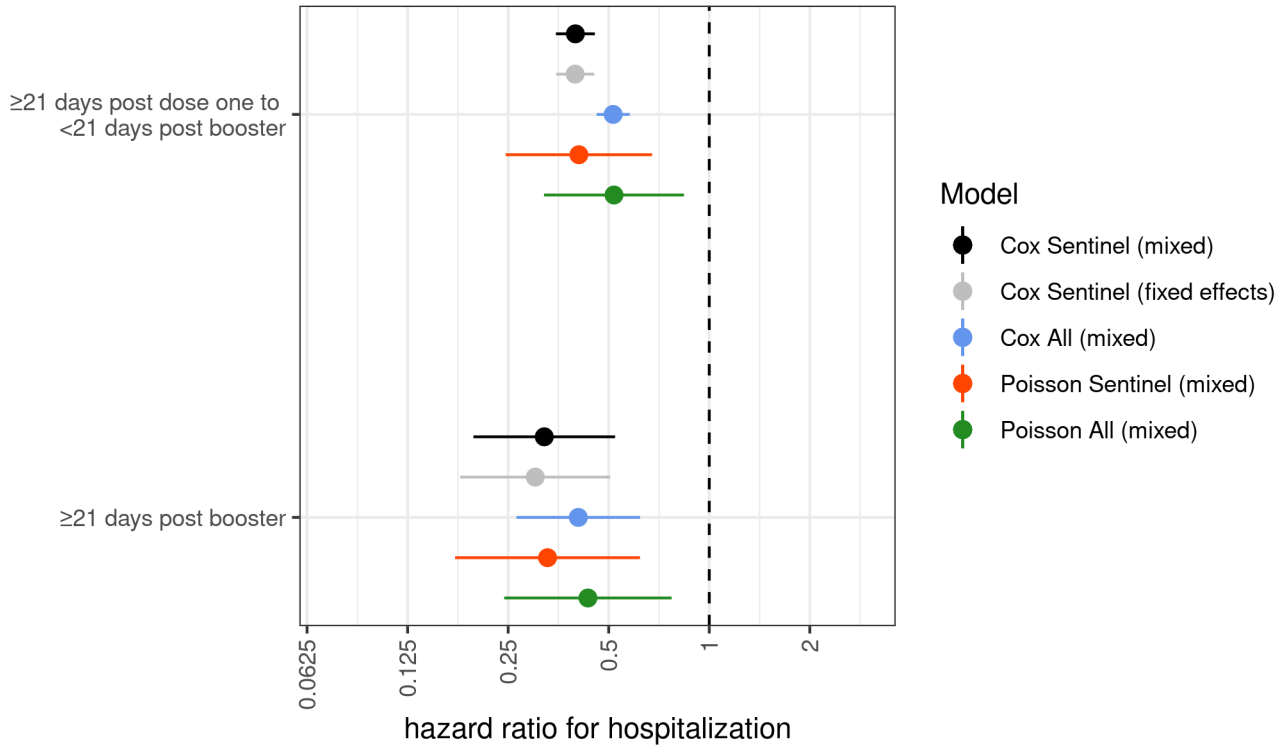

Figure A.4: Risk of Hospitalization by vaccination dosage and technology via differing model selection. Error bars represent 95% CI. "Mixed" refers to a model with mixed effects and "Sentinel" refers to the sample restricted to only to cases collected through sentinel surveillance. "Fixed effects" describes all model covariates being treated as fixed effects and "All" refers to the entire dataset from Dec 1, 2020 to Jan 14, 2022 , irrespective of participation in sentinel surveillance and includes targeted sequencing of suspected breakthrough infections.

	Other		Alpha		Delta		Gamma		Omicron	
Hospitalized	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Unvaccinated to <21 days post dose one	114	4975	224	8188	837	22271	102	1844	16	1984
≥21 days post dose one to <21 days post booster	–	–	9	300	261	9173	10	144	16	2793
≥21 days post booster	–	–	–	–	11	554	–	–	4	549

Table A.1: Hospitalizations by Vaccination Status

Characteristics	Hospitalization	
	HR	95% CI
Lineage		
Delta	REF	
Omicron	0.34	(0.23-0.50)
Vaccination* Lineage		
Delta		
Unvaccinated to < 21 days post dose one	REF	
≥21 days post dose one to <21 days post booster	0.40	(0.34-46)
≥21 days post booster	0.30	(0.17-0.52)
Omicron		
Unvaccinated to < 21 days post dose one post dose one	0.35	(0.20-0.62)
≥21 days post dose one to <21 days post booster	0.22	(0.13-0.37)
≥21 days post booster	0.18	(0.08-0.39)

Table A.2: Adjusted Cox Proportional Hazards Estimates for Risk of Hospitalization for Omicron vs DeltaAdditional model covariates include: sex, age (in 10 year bins), calendar week. Each variant lineage category risk estimate uses the "Unvaccinated to ¡21 days post dose one" vaccination group in cases infected with Delta as the reference group.

Appendix B

SUPPLEMENTARY MATERIAL FOR CHAPTER 3

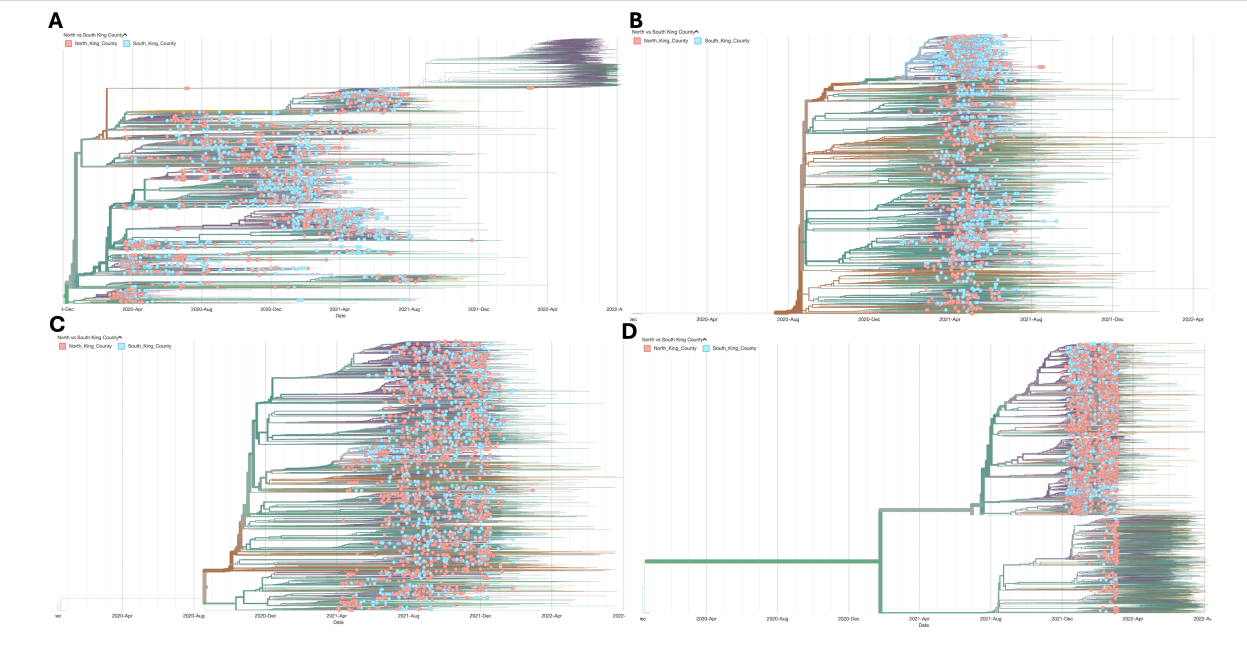

Figure B.1: Time-resolved maximum likelihood phylogenies for King County, WA by dominant variant wave with sample collection dates between February 1 2020 and March 6 2022. Trees are filtered to highlight genomes from King County among contextual sequences from around the globe. Tip color represents the region within King County, with pink corresponding to North King County and blue representing South King County. Branches are colored based on inferred ancestry. Panel **A** represents all variant clades excluding Alpha, Delta, and Omicron (the full tree can be explored interactively at https://nextstrain.org/groups/blab/ncov-king-county/other), the other panels represent Alpha (**B**, https://nextstrain.org/groups/blab/ncov-king-county/alpha), Delta (**C**, https://nextstrain.org/groups/blab/ncov-king-county/delta), and Omicron (**D**, https://nextstrain.org/groups/blab/ncov-king-county/omicron

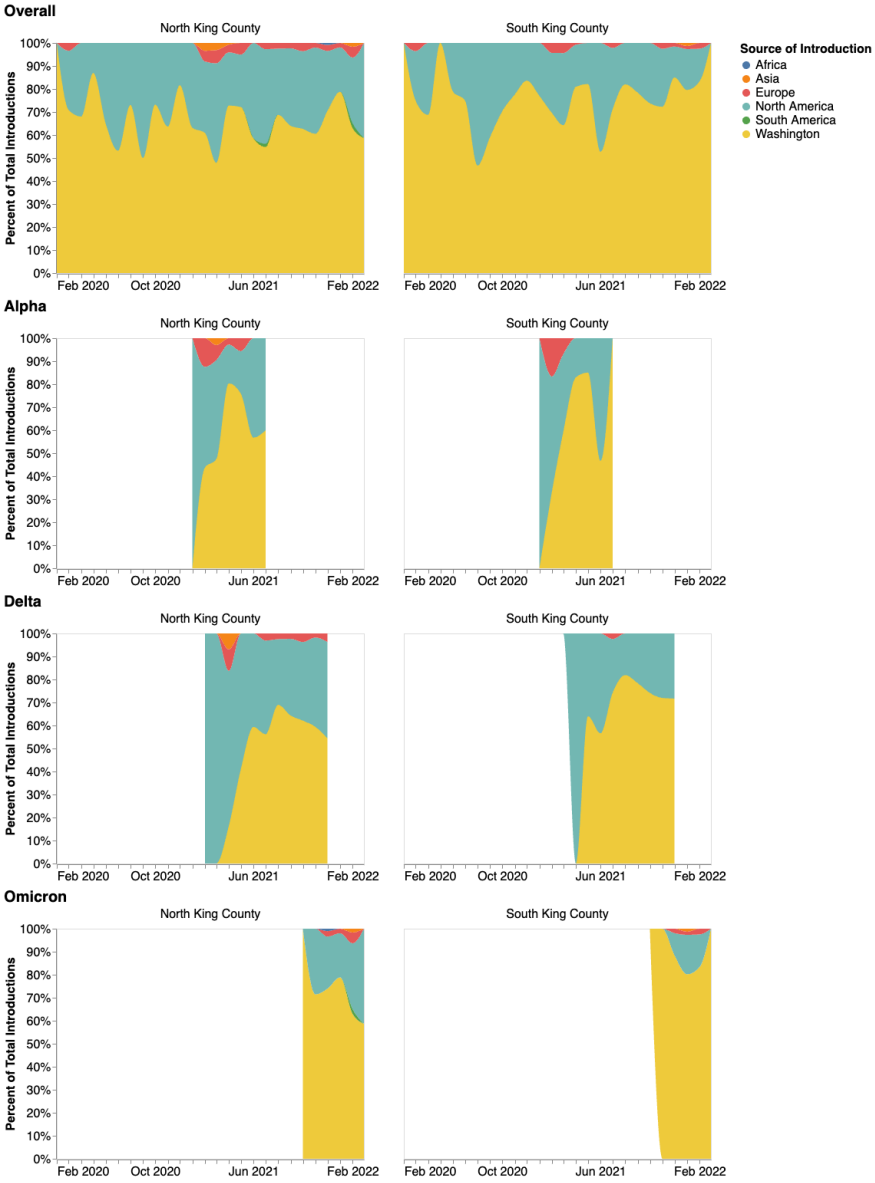

Figure B.2: Source of introduction for each identified King County cluster. The left column is introductions into North King County, the right into South King County. The panels show how the inferred geographical source of each introduction changes over time as a percentage of all introductions into the regions for that time period. The top row contains all the introductions among the four different time-resolved phylogenies. Each subsequent row represents a different variant studied and is labeled accordingly.

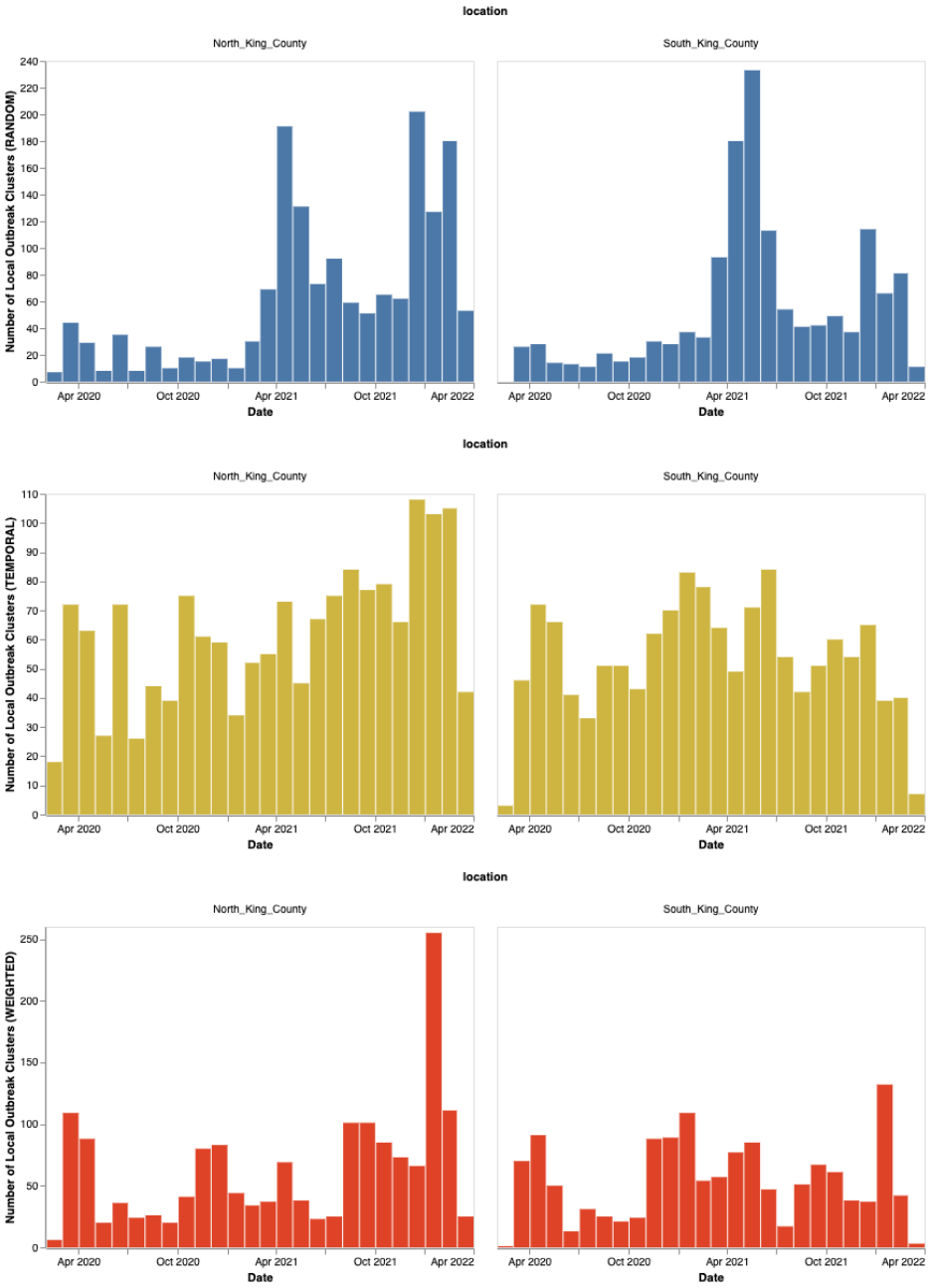

Figure B.3: Number of local outbreak clusters over time by subsampling scheme: random (A, Blue), equal temporal weighting by year-week (B, Gold), and subsampling weighted by daily hospitalizations calculated using a 14 day moving average (C, Red)

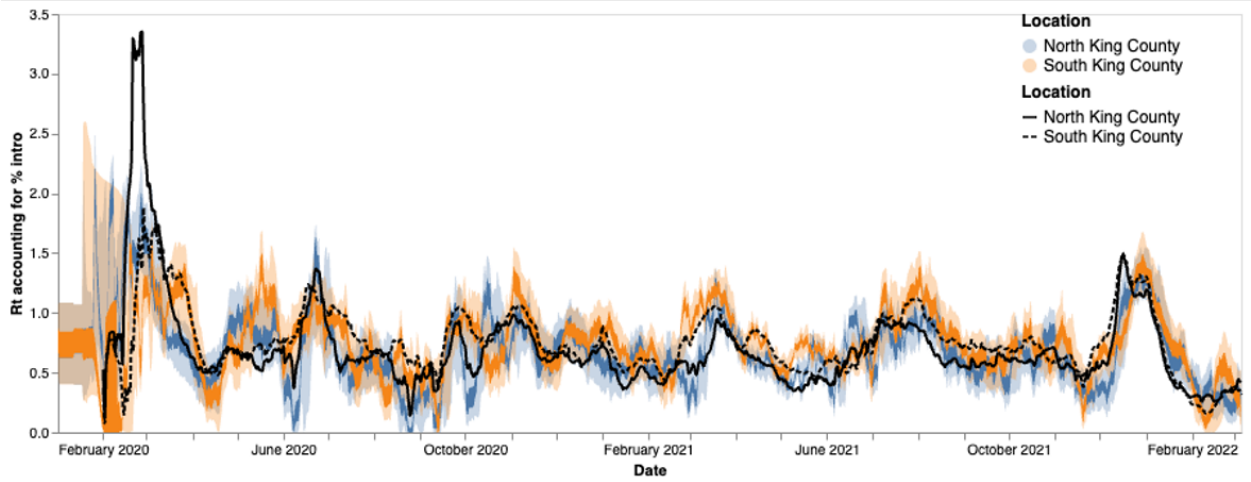

Figure B.4: Rt estimation using phylodynamic estimates (Blue North King County; Orange = South King County) and case data (Black lines, solid = North King County, dashed = South King County) The inner area denotes the 50% HPD interval and the outer area denotes the 95% HPD interval.

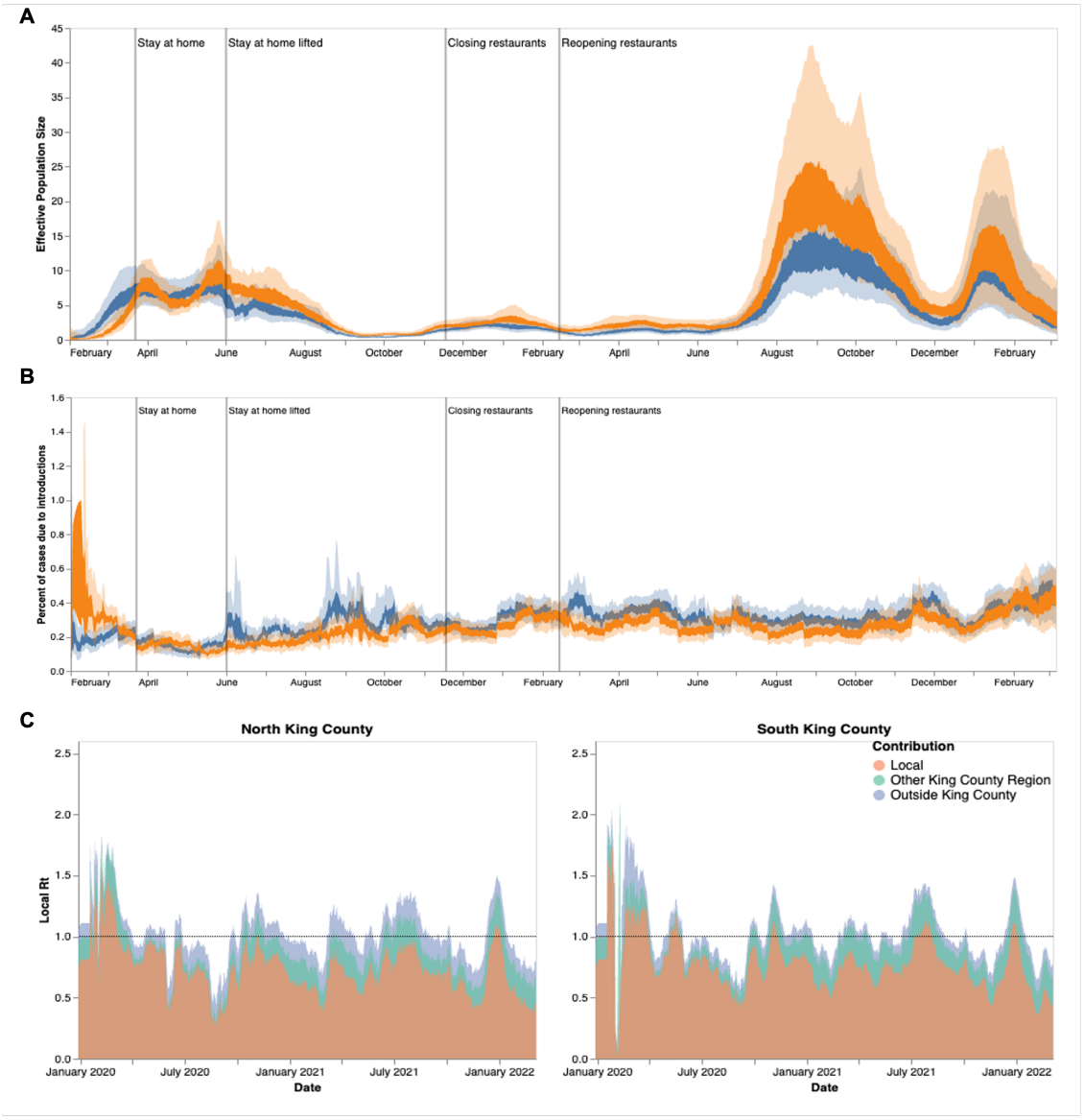

Figure B.5: Phylodynamic estimates of SARS-CoV-2 transmission in King County with equal temporal subsampling. Results presented above were inferred using 3000 sequences subsampled using equal temporal weighting by year-week. Analyses presented, as defined previously, are: effective population size over time (A), percent of cases due to introductions (B), and local Rt estimations divided by region and source of contribution (C). Orange denotes South King County; blue denotes North King County.

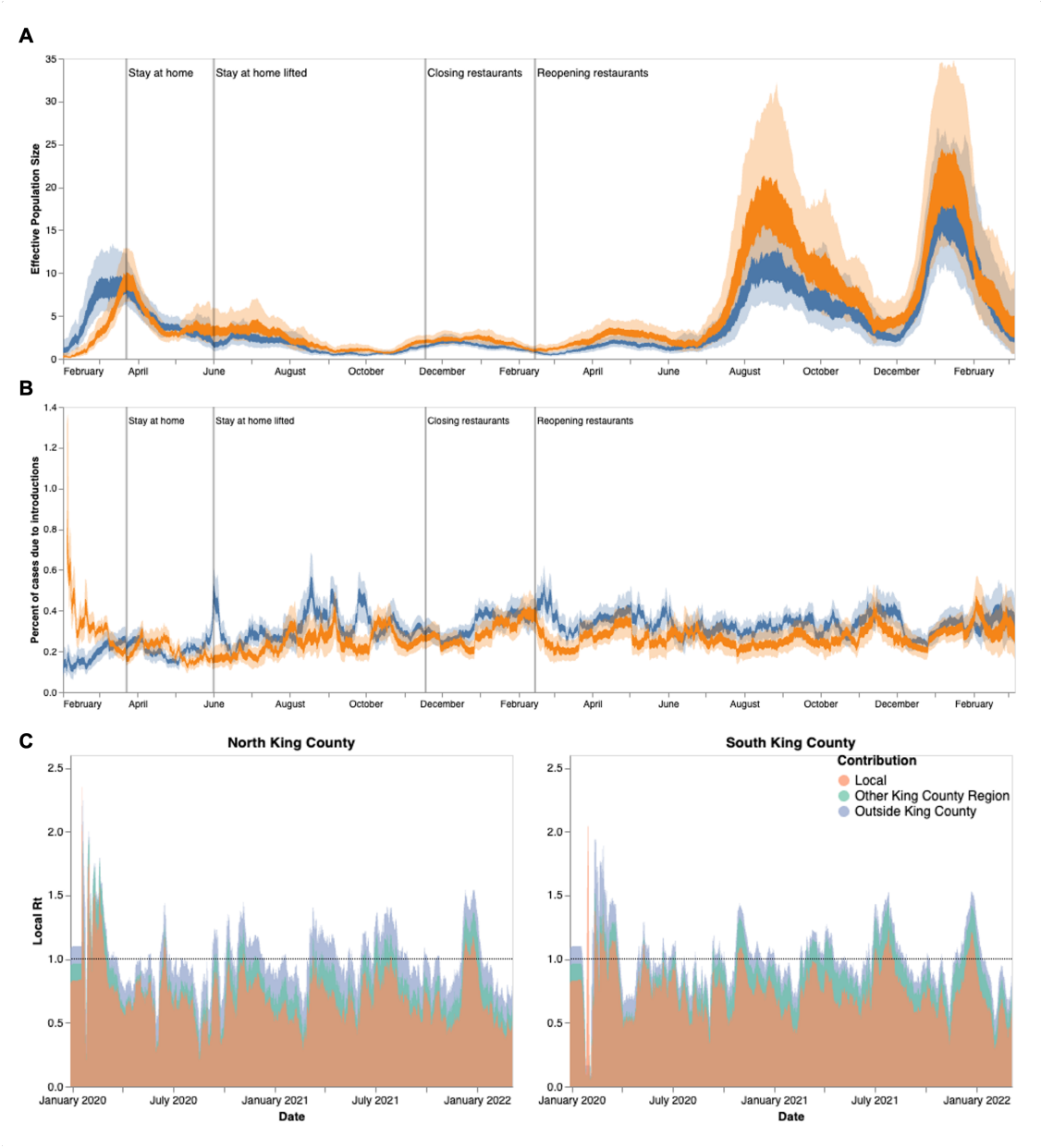

Figure B.6: Phylodynamic estimates of SARS-CoV-2 transmission in King County with subsampling weighted by hospitalizations. Results presented above were inferred using 3000 sequences subsampled using weighting by hospitalizations over time using a 14 day rolling average. Analyses presented, as defined previously, are: effective population size over time (A), percent of cases due to introductions (B), and local Rt estimations divided by region and source of contribution (C). Orange denotes South King County; blue denotes North King County.

Region	PUMA	ZIPCODE
North King County	11601	98103
		98107
		98117
	11602	98105
		98115
		98125
		98195
	11603	98101
		98102
		98104
		98109
		98119
		98121
		98154
		98164
		98199
	11604	98112
		98118
		98122
		98144
	11605	98106
		98108
		98116
		98126
		98134
		98136
	11606	98133
		98155
		98177
		98011
		98028
	11607	98033
		98034
		98052
	11608	98004
		98005
		98006
		98007
		98008
		98039
	11609	98040
		98029
		98076
		98075
	11616	98045
		98065
		98014
		98077
		98053
		98024
		98072
		98019
South King County	11610	98055
		98057
		98056
		98178
	11611	98146
		98148
		98166
		98168
		98188
	11612	98003
		98023
		98198
		98070
	11613	98030
		98031
		98032
		98092
	11614	98001
		98002
		98047
	11615	98010
		98022
		98038
		98051
		98027
		98042
		98059
		98058

Table B.1: Geocoding for different geographical scales in King County, WA

Appendix C

SUPPLEMENTARY MATERIAL FOR CHAPTER 4

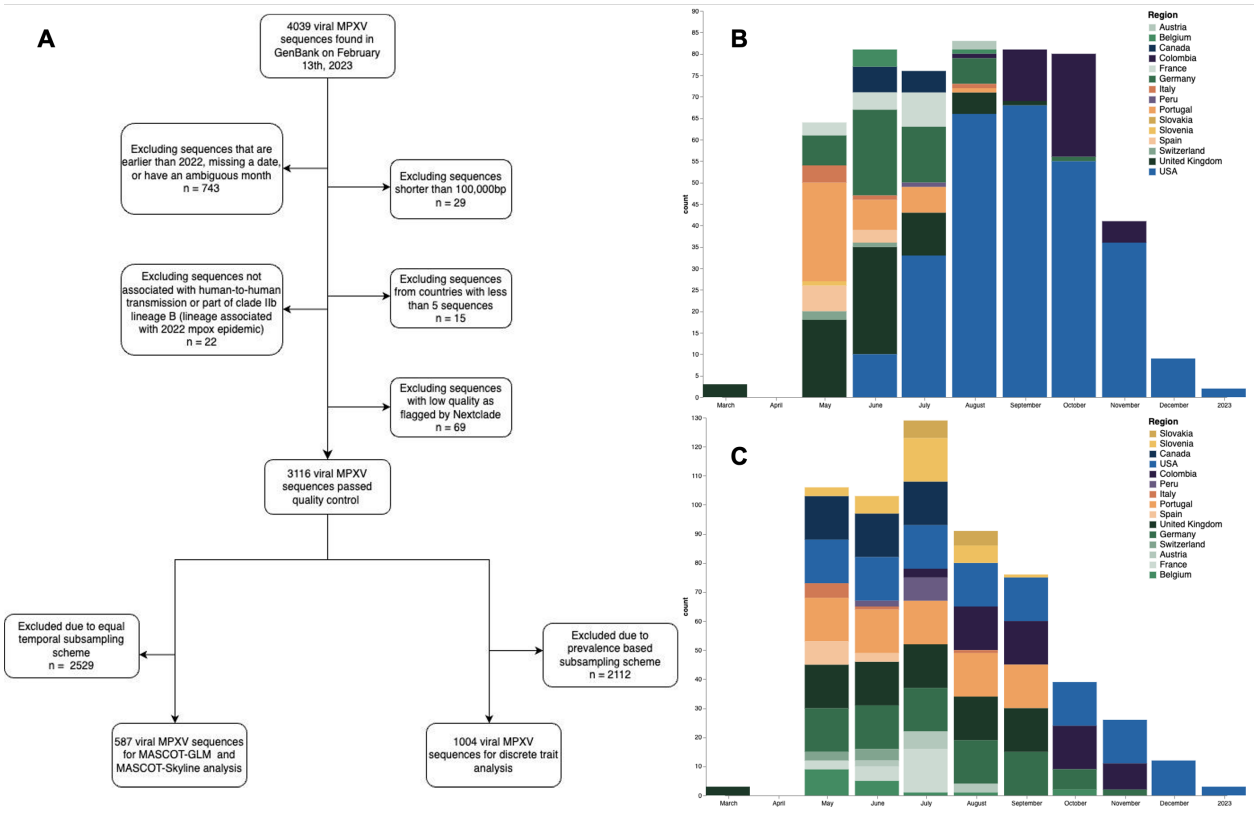

Figure C.1: Subsampling for phylogeographic and phylodynamic inference, related to Figure 1. (A) Flow diagram displaying the inclusion and exclusion criteria for the final two analytic samples **(B)** Temporal Distribution of 1004 genomes used for phylogeographic analysis. Genomes were subsampled using confirmed case counts as weights. **(C)** Temporal distribution of 587 genomes used for MASCOT-GLM analysis. Subsampling was done to promote an equal number of samples from each deme for each month in order to oversample underrepresented countries.

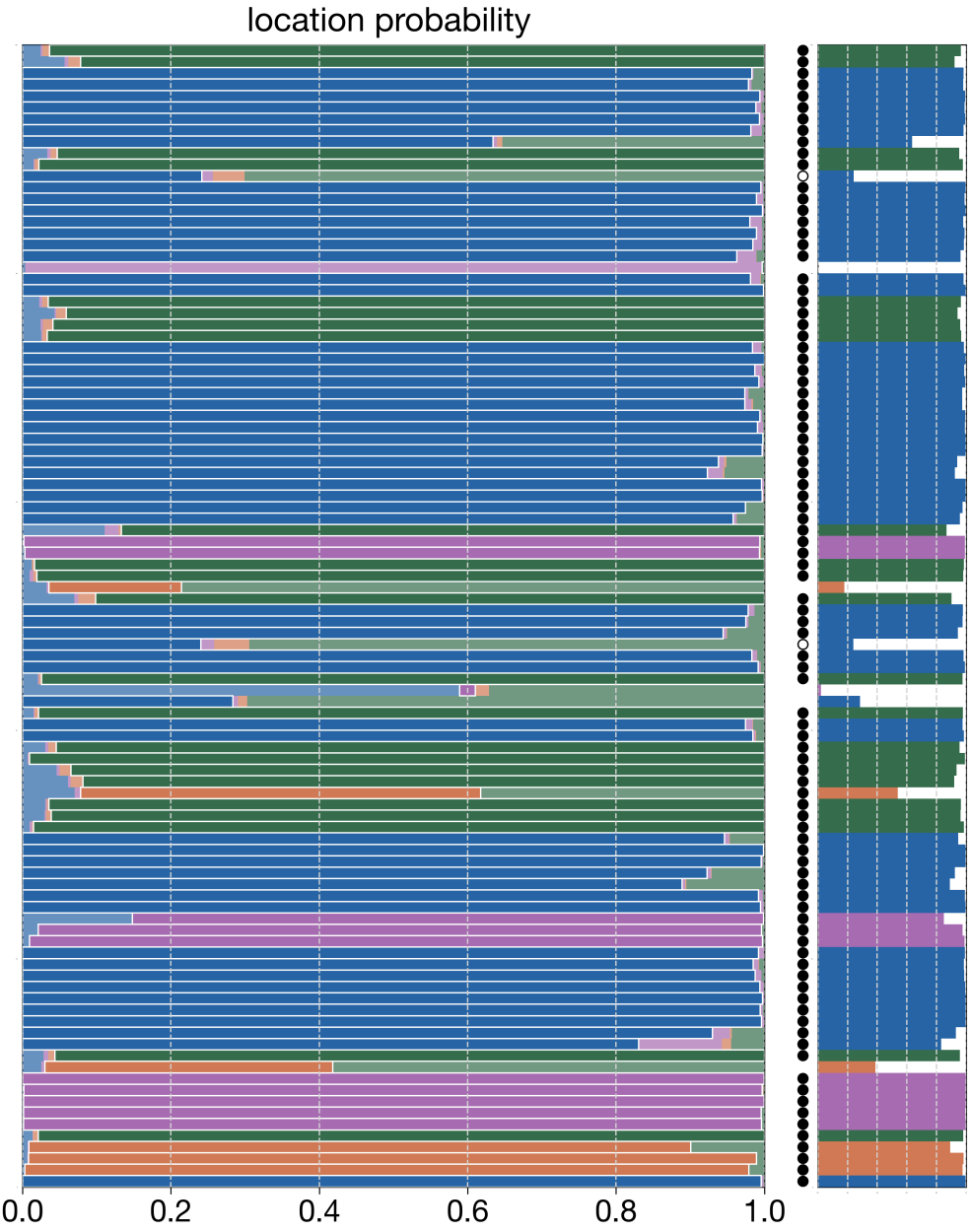

Figure C.2: Masked tip location inference, related to Figure 2.Horizontal bars indicate the posterior distribution of masked tip locations, coloured by region.. The correct location of each tip is outlined in white with the smaller plot to the right showing only the posterior probability of the correct location. Bars marked with an open circle indicate cases where the correct location is within the 95% credible set and solid circles indicate cases where the location with the most probability mass is also the correct location.

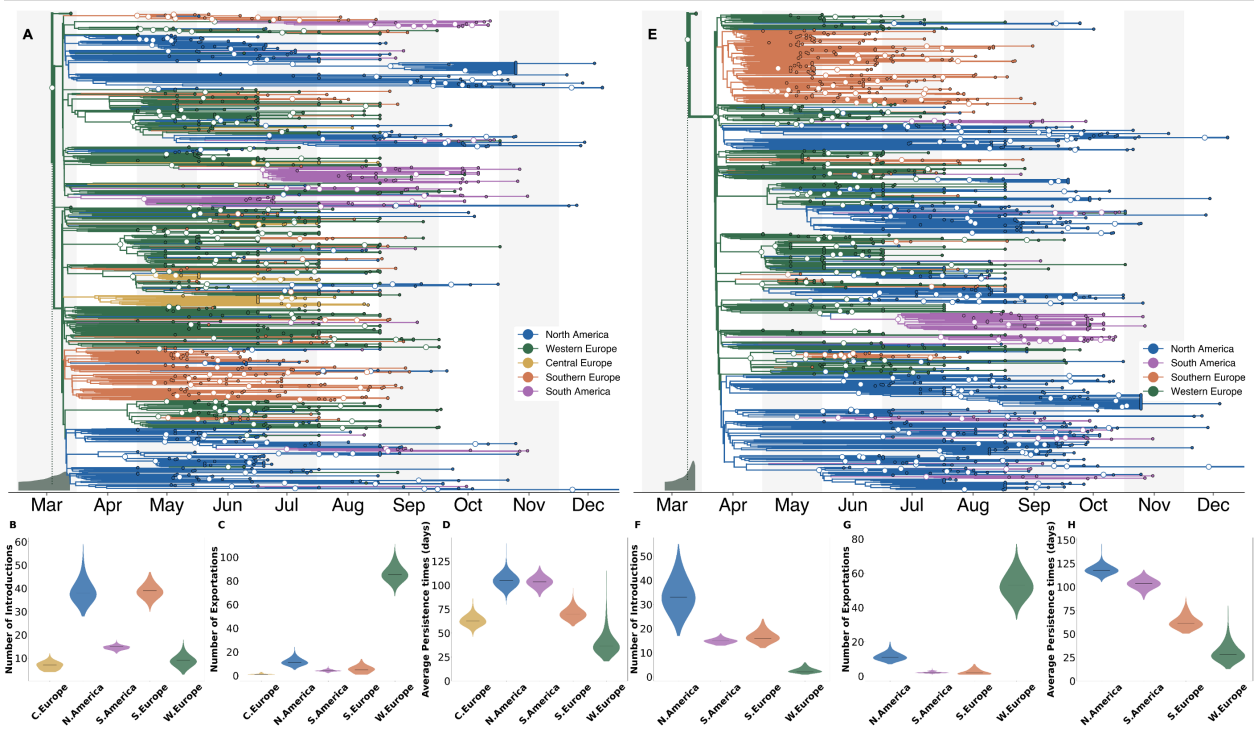

Figure C.3: Phylogeographic analysis using alternative subsampling schemes, related to Figure 2. In comparison with main Figure 2 which uses a case-count-based subsampling scheme, **A-D** used an equal spatiotemporal subsampling scheme which attempted to sample an equal number of sequences from each region for each year-month, which is the same strategy used for the MASCOT-GLM analyses. **E-H** repeat the analyses but sampled sequences directly from global regions irrespective of country of origin. (**A & E**) The maximum clade credibility tree summary of the Bayesian inference conducted using asymmetric discrete trait analysis and Skygrid prior on 991 (**A**) and 1019 (**E**) sequences. Colors correspond to the regions in the legend. Ancestral nodes with greater than 50% posterior support are highlighted with a white circle overlaid. Inset histogram on bottom left corner shows 95% interval for the time to most recent common ancestor (TMRCA)(**B-D & F-H**) Estimated number of introductions (**B & F**), exports (**C & G**), and average time of local persistence in days (**D & H**) for each global region. Horizontal black line denotes median estimates.

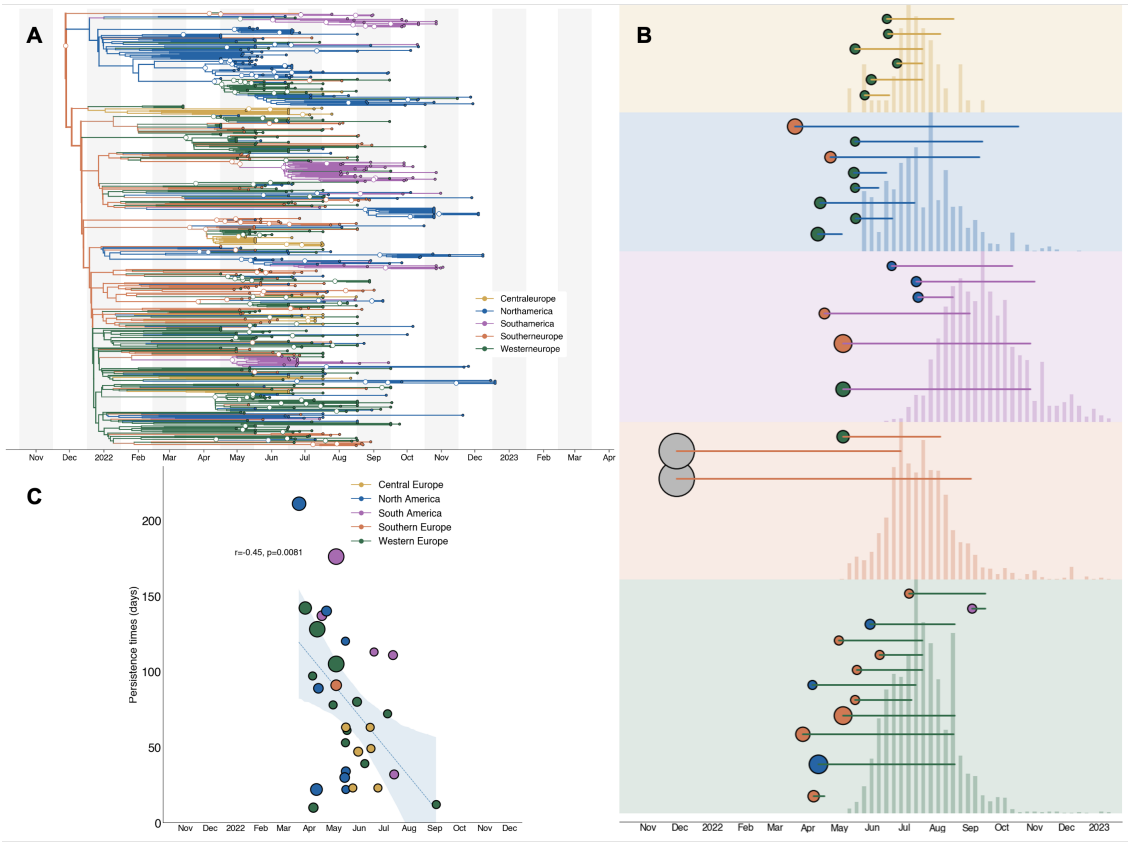

Figure C.4: Analysis of introductions inferred via MASCOT-GLM, related to Figure 3 and 4 (A) The maximum clade credibility tree summary of the Bayesian inference conducted using MASCOT-GLM on 587 sequences. Colors correspond to the regions in the legend. Ancestral nodes with greater than 50% posterior support are highlighted with a white circle overlaid. **(B)** Exploded subtrees for each region with only the introductions with greater than 50% posterior support showing that early underdetected introductions lead to longer transmission chains. Color at introduction origin represents inferred source region and size of the circle at the origin is proportional to the number of downstream tips. Length of line coming out of each introduction origin represents the length of the transmission chain. Case counts are overlaid for each region .**(C)** Relationship between estimated date of introduction and persistence time with the first two large introductions removed. Each circle represents a single viral introduction with greater than 50% posterior support into the region denoted by the color (i.e. a green point represents an introduction into Western Europe). The size of each point is proportional to the size of the outbreak cluster resulting from each introduction with larger circles representing more resulting downstream tips. Blue dashed line represents the linear best fit line using Pearson's correlation. Blue shaded region denotes the variability of the line and the resulting estimates from Pearson's correlation are shown in text above the shaded region.

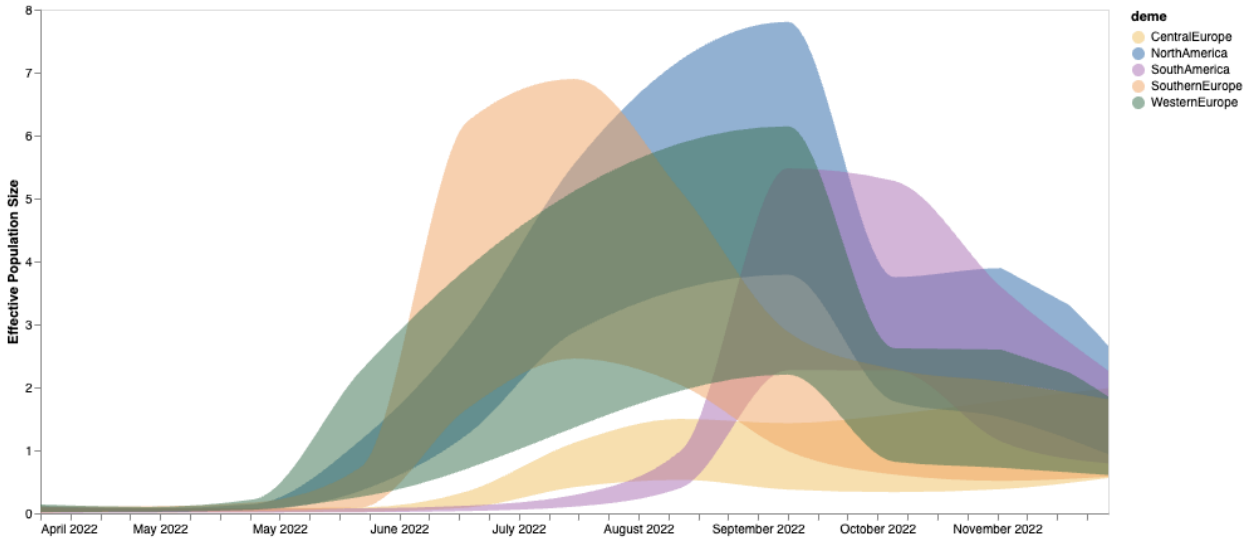

Figure C.5: Effective population size estimated via MASCOT-Skyline, related to Figure 3.Estimates of effective population sizes (NeTao in years) from April 2022 through December 2024 using 587 sequences subsampled equally throughout time. In contrast to the main MASCOT-GLM analysis, no empirical predictors were used, showing the extent of phylogenetic signal and uncertainty when using only genomes.

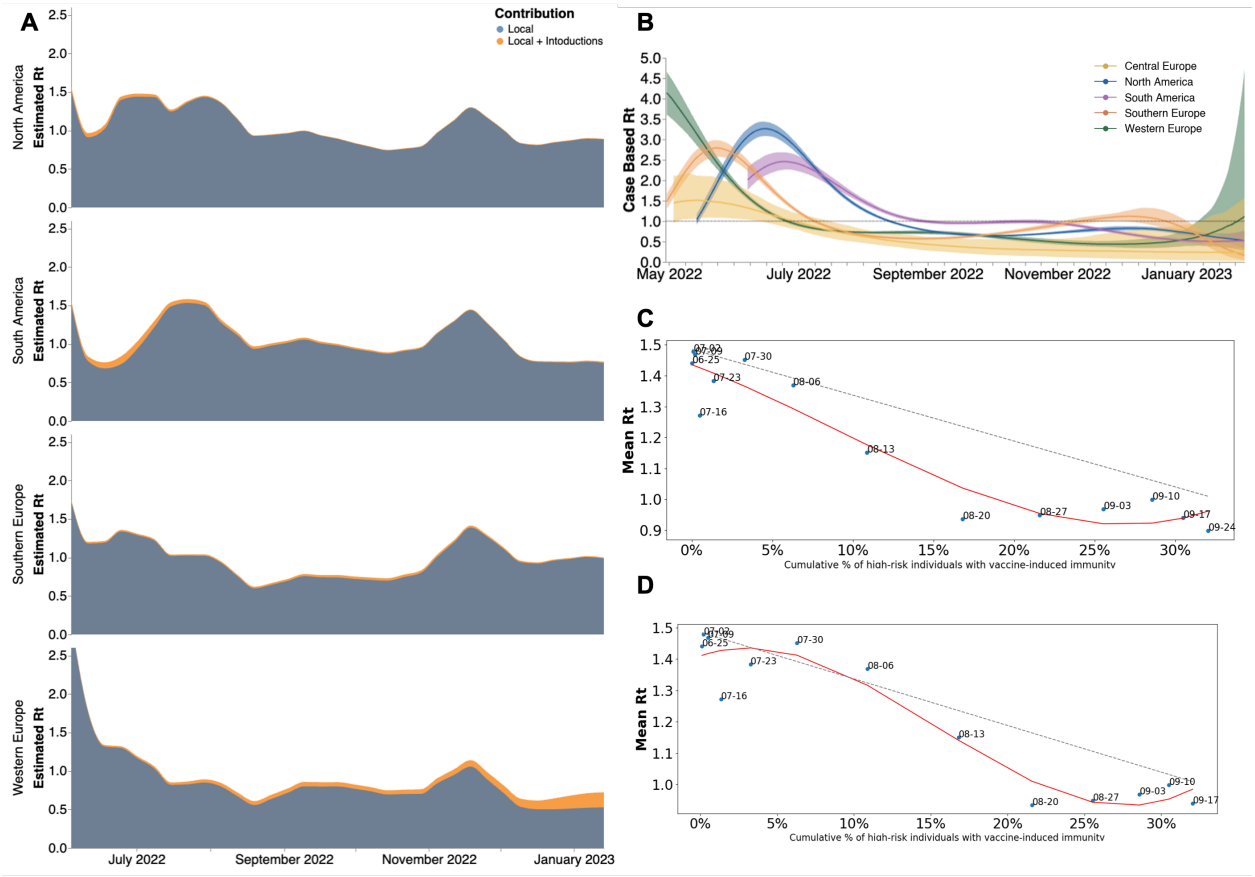

Figure C.6: Estimates of time-varying reproductive number (Rt), related to figures 5 and 6. (A) Estimates of Rt from April 2022 through December 2023 via MASCOT-GLM for four global regions separated by source of contribution. Blue denotes local Rt without the influence of outside viral introductions while orange shows the added contribution of introductions. Central Europe was removed due to limited data on introductions. **(B)** Estimates of time-varying reproductive number (Rt) in five global regions Estimates of Rt from April 2022 through December 2022 using renewal model framework from case counts only. The inner area denotes the 50% HPD interval and the outer area denotes the 95% HPD interval. Dashed line highlights an Rt value of 1 above which denotes an exponentially growing viral epidemic. (**C-D**) Scatter plot comparing mean *Rt* calculated via MASCOT-GLM for North America vs cumulative percentage of high risk individuals with vaccine-induced immunity in the United States with a one week lag to account account for the development of immunity **(C)** and no lag following date of vaccination **(D).** Red line indicates the best fit spline for scattered points. Dashed gray line indicates expected linear decrease in *Rt* with increasing vaccine-immunity assuming SIR dynamics. Over each point are the dates that correspond to the mean *Rt* and percent of immunity at that moment.

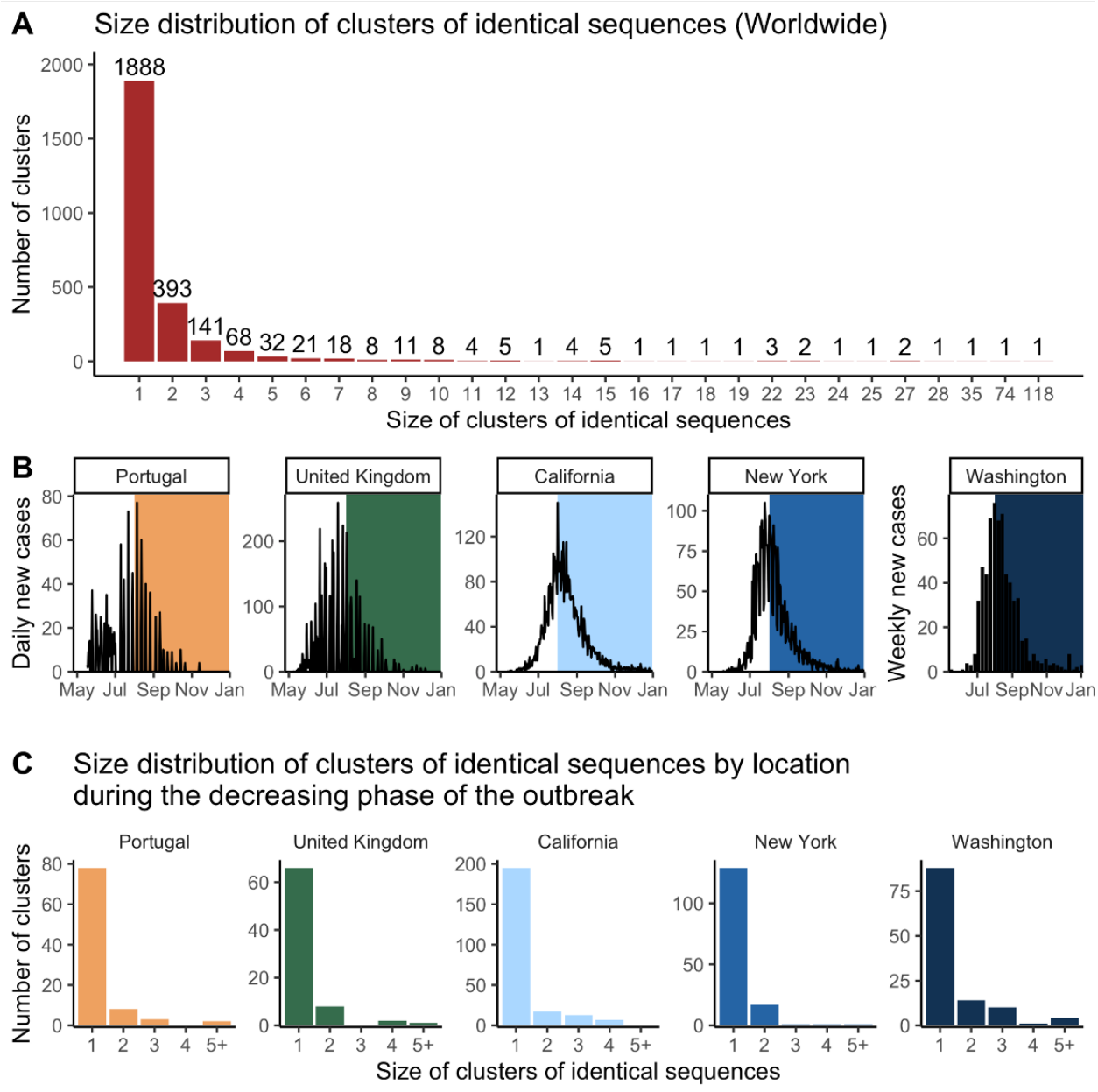

Figure C.7: Size distribution of clusters of identical mpox sequences, related to Figure 7 (A) Size distribution of clusters of identical mpox sequences worldwide. **(B)** Dynamics of mpox cases in the location of study. The coloured rectangles correspond to the study period. **(C)** Size distribution of clusters of identical mpox sequences by location during the decreasing phase of the outbreak (study period defined in B).

Region	Country	DTA	MASCOT-GLM
North America	Canada	27	45
	United States	533	120
Western Europe	Austria	3	11
	Germany	116	84
	Switzerland	7	7
	United Kingdom	94	78
	France	25	23
	Belgium	15	18
Central Europe	Slovakia	0	11
	Slovenia	0	31
Southern Europe	Italy	6	7
	Portugal	88	75
	Spain	9	11
South America	Colombia	77	57
	Peru	3	10

Table C.1: Geocoding for various country and regional scales used in this study, related to Figures 2 and 3. DTA denotes the samples for the phylogeographic analysis which was subsampled using confirmed case counts as weights. MASCOT-GLM column denotes the sample for the phylodynamic inference which was subsampled by enforcing equal temporal sampling per country per month.

Analysis	Tree prior	TMRCA		Migration rate		Clock rate	
		Mean	95% HPD	Mean	95% HPD	Mean	95% HPD
With three retrospectively-collected March 2022 sequences from the UK							
MASCOT-GLM	Approximate Structured Coalescent	2021-12-03	(2021-09-21 to 2022-02-01)	1.77	(1.45 -2.13)	6.27E-05	(5.62e-5 – 6.96e-5)
MASCOT-Skyline	Approximate Structured Coalescent	2022-01-29	(2021-12-07 to 2022-03-12)	1.83	(1.38-3.00)	5.71E-05	(5.06e-5 – 6.39e-5)
DTA	SkyGrid	2022-03-24	(2022-03-09 to 2022-03-27)	0.72	(0.57-0.89)	8.41E-05	(7.71e-5 –9.10e-5)
DTA	Skyline	2022-03-18	(2022-02-25 to 2022-03-27)	1.34	(1.03-1.66)	5.64E-05	(5.00e-5 – 6.39e-5)
DTA	Constant	2022-01-23	(2021-12-21 to 2022-02-24)	0.81	(0.68-0.89)	9.44E-05	(8.64e-5 –1.05e-4)
Without March 2022 sequences from the UK							
MASCOT-GLM	Approximate Structured Coalescent	2021-11-29	(2021-09-19 to 2022-01-28)				
DTA	SkyGrid	2022-03-30	(2022-03-05 to 2022-04-19)				

Table C.2: Comparison of time to most recent common ancestor (TMRCA), migration rate (migration events per year), and clock rate (substitutions per site per year) by method, related to Figures 2 and 3. First five rows denote the comparison of key summary statistics from main and alternative models used (with varying tree priors and inclusion of empirical predictors) which include three sequences from March 2022 which were found retrospectively in the UK. The last two rows represent the main analyses but without the three retrospective march 2022 samples.

Assumption regarding the proportion of infections detected	Location	*R estimate (maximum likelihood estimate with 95% CI)*	*k estimate (maximum likelihood estimate with 95% CI)*	Proportion of infected individuals with 0 offspring (obtained from R and k MLE)
10%	Portugal	0.87 (0.71-1.03)	0.16 (0.11-0.23)	74%
	United Kingdom	1.10 (0.98-1.22)		72%
	California (USA)	1.14 (1.08-1.20)		72%
	New York (USA)	1.05 (0.94-1.14)		72%
	Washington (USA)	1.06 (0.94-1.19)		72%
50%	Portugal	0.60 (0.44-0.80)	0.30 (0.18-0.54)	72%
	United Kingdom	0.89 (0.73-1.06)		66%
	California (USA)	0.94 (0.85-1.03)		65%
	New York (USA)	0.82 (0.69-0.94)		67%
	Washington (USA)	0.84 (0.69-1.02)		67%
100%	Portugal	0.46 (0.31-0.66)	0.36 (0.21-0.71)	74%
	United Kingdom	0.76 (0.59-0.95)		66%
	California (USA)	0.82 (0.72-0.92)		65%
	New York (USA)	0.68 (0.55-0.82)		68%
	Washington (USA)	0.70 (0.54-0.91)		68%

Table C.3: Reproduction numbers and dispersion parameter estimates from the analysis of the size distribution of clusters of identical sequences using a joint-likelihood, related to Figure 7. For each location, we report maximum likelihood estimates (MLE) along 95% likelihood profile confidence intervals. Different assumptions regarding the proportion of infections sequenced are explored. These estimates were obtained by allowing the reproduction numbers to vary between regions but assuming a similar value of the dispersion parameter k across locations.

Location	Assumption regarding the proportion of infections detected	*R estimate (maximum likelihood estimate with 95% CI)*	*k estimate (maximum likelihood estimate with 95% CI)*	Proportion of infected individuals with 0 offspring (obtained from R and k MLE)
Portugal	10%	0.8 (0.64-0.98)	0.09 (0.02-0.3)	81%
	50%	0.57 (0.40-0.79)	0.19 (0.04-1.8)	77%
	100%	0.45 (0.29-0.67)	0.23 (0.06-6.5)	78%
United Kingdom	10%	0.97 (0.81-1.12)	0.11 (0.03-0.34)	78%
	50%	0.76 (0.58-0.95)	0.23 (0.04-2.6)	72%
	100%	0.63 (0.45-0.84)	0.28 (0.05- >10)	72%
California (USA)	10%	1.03 (0.95-1.11)	0.07 (0.03-0.11)	83%
	50%	0.89 (0.8-0.98)	0.19 (0.08-0.42)	72%
	100%	0.79 (0.68-0.89)	0.25 (0.1-0.75)	70%
New York (USA)	10%	1.21 (1.14-1.28)	1.26 (0.37 - >10)	42%
	50%	0.92 (0.79-1.03)	1.04 (0.24- >10)	52%
	100%	0.75 (0.61-0.88)	0.99 (0.19- >10)	58%
Washington (USA)	10%	1.1 (0.99-1.21)	0.23 (0.07-0.72)	67%
	50%	0.85 (0.71-1.02)	0.38 (0.13-2.3)	64%
	100%	0.71 (0.55-0.9)	0.43 (0.16-2.8)	66%

Table C.4: Location-specific reproduction number and dispersion parameter estimates from the analysis of the size distribution of clusters of identical sequences, related to Figure 7. For each location, we report maximum likelihood estimates (MLE) along 95% likelihood profile confidence intervals. Different assumptions regarding the proportion of infections sequenced are explored.

BIBLIOGRAPHY

[1] Census Bureau Data.

[2] Current COVID-19 metrics - King County.

[3] NERVTAG Note on B.1.1.7 Severity.

[4] Seattle's Race and Segregation Story in Maps 1920-2020 - Seattle Civil Rights and Labor History Project.

[5] WHO Director-General's statement at the press conference following IHR Emergency Committee regarding the multi-country outbreak of monkeypox - 23 July 2022.

[6] World Bank Open Data.

[7] Interim Clinical Considerations for Use of COVID-19 Vaccines | CDC, September 2021.

[8] Mpox Vaccine Administration in the U.S. | Mpox | Poxvirus | CDC, January 2024.

[9] Ifedayo Adetifa, Jean-Jacques Muyembe, Daniel G Bausch, and David L Heymann. Mpox neglect and the smallpox niche: a problem for Africa, a problem for the world. *The Lancet*, 401(10390):1822–1824, May 2023.

[10] Ivan Aksamentov, Cornelius Roemer, Emma B. Hodcroft, and Richard A. Neher. Nextclade: clade assignment, mutation calling and quality control for viral genomes. *Journal of Open Source Software*, 6(67):3773, November 2021.

[11] Stephen W. Attwood, Sarah C. Hill, David M. Aanensen, Thomas R. Connor, and Oliver G. Pybus. Phylogenetic and phylodynamic approaches to understanding and combating the early SARS-CoV-2 pandemic. *Nature Reviews Genetics*, 23(9):547–562, September 2022. Publisher: Nature Publishing Group.

[12] Guy Baele, Philippe Lemey, Andrew Rambaut, and Marc A. Suchard. Adaptive MCMC in Bayesian phylogenetics: an application to analyzing partitioned data in BEAST. *Bioinformatics (Oxford, England)*, 33(12):1798–1805, June 2017.

[13] Peter Bager, Jan Wohlfahrt, Samir Bhatt, Sofie Marie Edslev, Raphael Niklaus Sieber, Anna Cäcilia Ingham, Marc Stegger, Rebecca Legarth, Camilla Holten Møller, Robert Leo Skov, Palle Valentiner-Branth, Maria Overvad, Mie Agermose Gram, Frederikke Kristensen Lomholt, Louise Hallundbæk, Caroline Hjorth Espensen, Sophie Madeleine Gubbels, Marianne Voldstedlund, Marianne Karakis, Karina Lauenborg Møller, Stefan Schytte Olsen, Thea K. Fischer, Zitta Barrella Harboe, Caroline Klint Johannesen, Maarten van Wijhe, Jon Gitz Holler, Lone Simonsen, Ram Benny Christian Dessau, Martin Barfred Friis, David Fuglsang-Damgaard, Mette Pinholt, Nikolai Søren Kirkby, Marianne Kragh Thomsen, Thomas Vognbjerg Sydenham, John Eugenio Coia, Ea Sofie Marmolin, Anders Fomsgaard, Jannik Fonager, Morten Rasmussen, Katja Spiess, Ellinor Marving, Arieh Cohen, Nicolai Balle Larsen, Troels Lillebaek, Henrik Ullum, Kåre Mølbak, and Tyra Grove Krause. Reduced Risk of Hospitalisation Associated With Infection With SARS-CoV-2 Omicron Relative to Delta: A Danish Cohort Study. SSRN Scholarly Paper ID 4008930, Social Science Research Network, Rochester, NY, January 2022.

[14] Peter Bager, Jan Wohlfahrt, Jannik Fonager, Morten Rasmussen, Mads Albertsen, Thomas Yssing Michaelsen, Camilla Holten Møller, Steen Ethelberg, Rebecca Legarth, Mia Sarah Fischer Button, Sophie Gubbels, Marianne Voldstedlund, Kåre Mølbak, Robert Leo Skov, Anders Fomsgaard, and Tyra Grove Krause. Risk of hospitalisation associated with infection with SARS-CoV-2 lineage B.1.1.7 in Denmark: an observational cohort study. *The Lancet Infectious Diseases*, June 2021.

[15] Peter Bager, Jan Wohlfahrt, Morten Rasmussen, Mads Albertsen, and Tyra Grove Krause. Hospitalisation associated with SARS-CoV-2 delta variant in Denmark. *The Lancet Infectious Diseases*, 0(0), September 2021. Publisher: Elsevier.

[16] Trevor Bedford, Sarah Cobey, Peter Beerli, and Mercedes Pascual. Global Migration Dynamics Underlie Evolution and Persistence of Human Influenza A (H3N2). *PLOS Pathogens*, 6(5):e1000918, May 2010. Publisher: Public Library of Science.

[17] Trevor Bedford, Sarah Cobey, and Mercedes Pascual. Strength and tempo of selection revealed in viral gene genealogies. *BMC Evolutionary Biology*, 11(1):220, July 2011.

[18] Trevor Bedford, Alexander L. Greninger, Pavitra Roychoudhury, Lea M. Starita, Michael Famulare, Meei-Li Huang, Arun Nalla, Gregory Pepper, Adam Reinhardt, Hong Xie, Lasata Shrestha, Truong N. Nguyen, Amanda Adler, Elisabeth Brandstetter, Shari Cho, Danielle Giroux, Peter D. Han, Kairsten Fay, Chris D. Frazar, Misja Ilcisin, Kirsten Lacombe, Jover Lee, Anahita Kiavand, Matthew Richardson, Thomas R. Sibley, Melissa Truong, Caitlin R. Wolf, Deborah A. Nickerson, Mark J. Rieder, Janet A.

Englund, The Seattle Flu Study Investigators‡, James Hadfield, Emma B. Hodcroft, John Huddleston, Louise H. Moncla, Nicola F. Müller, Richard A. Neher, Xianding Deng, Wei Gu, Scot Federman, Charles Chiu, Jeffrey S. Duchin, Romesh Gautom, Geoff Melly, Brian Hiatt, Philip Dykema, Scott Lindquist, Krista Queen, Ying Tao, Anna Uehara, Suxiang Tong, Duncan MacCannell, Gregory L. Armstrong, Geoffrey S. Baird, Helen Y. Chu, Jay Shendure, and Keith R. Jerome. Cryptic transmission of SARS-CoV-2 in Washington state. *Science*, 370(6516):571–575, October 2020. Publisher: American Association for the Advancement of Science Section: Report.

[19] Dawn Blackburn, Nicole M. Roth, Jeremy A.W. Gold, Leah Zilversmit Pao, Evelyn Olansky, Elizabeth A. Torrone, R. Paul McClung, Sascha R. Ellington, Kevin P. Delaney, Neal Carnes, and Patrick Dawson. Epidemiologic and Clinical Features of Mpox in Transgender and Gender-Diverse Adults — United States, May–November 2022. *Morbidity and Mortality Weekly Report*, 71(5152):1605–1609, December 2022.

[20] S. Blumberg and J. O. Lloyd-Smith. Comparing methods for estimating R0 from the size distribution of subcritical transmission chains. *Epidemics*, 5(3):131–145, September 2013.

[21] Seth Blumberg and James O. Lloyd-Smith. Inference of R0 and Transmission Heterogeneity from the Size Distribution of Stuttering Chains. *PLOS Computational Biology*, 9(5):e1002993, May 2013. Publisher: Public Library of Science.

[22] Vítor Borges, Mariana Perez Duque, João Vieira Martins, Paula Vasconcelos, Rita Ferreira, Daniel Sobral, Ana Pelerito, Isabel Lopes de Carvalho, Maria Sofia Núncio, Maria José Borrego, Cornelius Roemer, Richard A. Neher, Megan O'Driscoll, Raquel Rocha, Sílvia Lopo, Raquel Neves, Paula Palminha, Luís Coelho, Alexandra Nunes, Joana Isidro, Miguel Pinto, João Dourado Santos, Verónica Mixão, Daniela Santos, Silvia Duarte, Luís Vieira, Fátima Martins, Jorge Machado, Vítor Cabral Veríssimo, Berta Grau, André Peralta-Santos, José Neves, Margarida Caldeira, Mafalda Pestana, Cândida Fernandes, João Caria, Raquel Pinto, Diana Póvoas, Fernando Maltez, Ana Isabel Sá, Mafalda Brito Salvador, Eugénio Teófilo, Miguel Rocha, Virginia Moneti, Luis Miguel Duque, Francisco Ferreira e Silva, Teresa Baptista, Joana Vasconcelos, Sara Casanova, Kamal Mansinho, João Vaz Alves, João Alves, António Silva, Miguel Alpalhão, Cláudia Brazão, Diogo Sousa, Paulo Filipe, Patrícia Pacheco, Francesca Peruzzu, Rita Patrocínio de Jesus, Luís Ferreira, Josefina Mendez, Sofia Jordão, Frederico Duarte, Maria João Gonçalves, Eduarda Pena, Claúdio Nunes Silva, André Rodrigues Guimarães, Margarida Tavares, Graça Freitas, Rita Cordeiro, and João Paulo Gomes. Viral genetic clustering and transmission dynamics of the 2022 mpox outbreak in Por-

tugal. *Nature Medicine*, 29(10):2509–2517, October 2023. Number: 10 Publisher: Nature Publishing Group.

[23] Remco Bouckaert, Timothy G. Vaughan, Joëlle Barido-Sottani, Sebastián Duchêne, Mathieu Fourment, Alexandra Gavryushkina, Joseph Heled, Graham Jones, Denise Kühnert, Nicola De Maio, Michael Matschiner, Fábio K. Mendes, Nicola F. Müller, Huw A. Ogilvie, Louis du Plessis, Alex Popinga, Andrew Rambaut, David Rasmussen, Igor Siveroni, Marc A. Suchard, Chieh-Hsi Wu, Dong Xie, Chi Zhang, Tanja Stadler, and Alexei J. Drummond. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Computational Biology*, 15(4):e1006650, April 2019. Publisher: Public Library of Science.

[24] Kaya Bramble, Magali N. Blanco, Annie Doubleday, Amanda J. Gassett, Anjum Hajat, Julian D. Marshall, and Lianne Sheppard. Exposure Disparities by Income, Race and Ethnicity, and Historic Redlining Grade in the Greater Seattle Area for Ultrafine Particles and Other Air Pollutants. *Environmental Health Perspectives*, 131(7):077004, July 2023. Publisher: Environmental Health Perspectives.

[25] Samuel P. C. Brand, Massimo Cavallaro, Fergus Cumming, Charlie Turner, Isaac Florence, Paula Blomquist, Joe Hilton, Laura M. Guzman-Rincon, Thomas House, D. James Nokes, and Matt J. Keeling. The role of vaccination and public awareness in forecasts of Mpox incidence in the United Kingdom. *Nature Communications*, 14(1):4100, July 2023. Number: 1 Publisher: Nature Publishing Group.

[26] Thiago Carvalho, Florian Krammer, and Akiko Iwasaki. The first 12 months of COVID-19: a timeline of immunological insights. *Nature Reviews. Immunology*, 21(4):245–256, 2021.

[27] CDC. Coronavirus Disease 2019 (COVID-19), February 2020.

[28] CDC. Monkeypox Technical Reports, January 2023.

[29] CDC. Mpox in the U.S., February 2023.

[30] Patrick A. Clay, Jason M. Asher, Neal Carnes, Casey E. Copen, Kevin P. Delaney, Daniel C. Payne, Emily D. Pollock, Jonathan Mermin, Yoshinori Nakazawa, William Still, Anil T. Mangla, and Ian H. Spicknall. Modelling the impact of vaccination and sexual behavior change on reported cases of mpox in Washington D.C, February 2023. Pages: 2023.02.10.23285772.

[31] Anne Cori, Neil M. Ferguson, Christophe Fraser, and Simon Cauchemez. A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics. *American Journal of Epidemiology*, 178(9):1505–1512, November 2013.

[32] R. Creswell, D. Augustin, I. Bouros, H. J. Farm, S. Miao, A. Ahern, M. Robinson, A. Lemenuel-Diot, D. J. Gavaghan, B. C. Lambert, and R. N. Thompson. Heterogeneity in the onwards transmission risk between local and imported cases affects practical estimates of the time-dependent reproduction number. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 380(2233):20210308, August 2022. Publisher: Royal Society.

[33] Petr Danecek, James K Bonfield, Jennifer Liddle, John Marshall, Valeriu Ohan, Martin O Pollard, Andrew Whitwham, Thomas Keane, Shane A McCarthy, Robert M Davies, and Heng Li. Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2):giab008, February 2021.

[34] Nicholas G. Davies, Sam Abbott, Rosanna C. Barnard, Christopher I. Jarvis, Adam J. Kucharski, James D. Munday, Carl A. B. Pearson, Timothy W. Russell, Damien C. Tully, Alex D. Washburne, Tom Wenseleers, Amy Gimma, William Waites, Kerry L. M. Wong, Kevin van Zandvoort, Justin D. Silverman, CMMID COVID-19 Working Group1‡, COVID-19 Genomics UK (COG-UK) Consortium‡, Karla Diaz-Ordaz, Ruth Keogh, Rosalind M. Eggo, Sebastian Funk, Mark Jit, Katherine E. Atkins, and W. John Edmunds. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science*, 372(6538), April 2021. Publisher: American Association for the Advancement of Science Section: Research Article.

[35] Irith De Baetselier, Christophe Van Dijck, Chris Kenyon, Jasmine Coppens, Johan Michiels, Tessa de Block, Hilde Smet, Sandra Coppens, Fien Vanroye, Joachim Jakob Bugert, Philipp Girl, Sabine Zange, Laurens Liesenborghs, Isabel Brosius, Johan van Griensven, Philippe Selhorst, Eric Florence, Dorien Van den Bossche, Kevin K. Ariën, Antonio Mauro Rezende, Koen Vercauteren, and Marjan Van Esbroeck. Retrospective detection of asymptomatic monkeypox virus infections among male sexual health clinic attendees in Belgium. *Nature Medicine*, 28(11):2288–2292, November 2022. Number: 11 Publisher: Nature Publishing Group.

[36] Kevin P. Delaney. Strategies Adopted by Gay, Bisexual, and Other Men Who Have Sex with Men to Prevent Monkeypox virus Transmission — United States, August 2022. *MMWR. Morbidity and Mortality Weekly Report*, 71, 2022.

[37] Simon Dellicour, Samuel L. Hong, Bram Vrancken, Antoine Chaillon, Mandev S. Gill, Matthew T. Maurano, Sitharam Ramaswami, Paul Zappile, Christian Marier, Gordon W. Harkins, Guy Baele, Ralf Duerr, and Adriana Heguy. Dispersal dynamics of SARS-CoV-2 lineages during the first epidemic wave in New York City. *PLOS Pathogens*, 17(5):e1009571, May 2021. Publisher: Public Library of Science.

[38] Xianding Deng, Wei Gu, Scot Federman, Louis du Plessis, Oliver G. Pybus, Nuno R. Faria, Candace Wang, Guixia Yu, Brian Bushnell, Chao-Yang Pan, Hugo Guevara, Alicia Sotomayor-Gonzalez, Kelsey Zorn, Allan Gopez, Venice Servellita, Elaine Hsu, Steve Miller, Trevor Bedford, Alexander L. Greninger, Pavitra Roychoudhury, Lea M. Starita, Michael Famulare, Helen Y. Chu, Jay Shendure, Keith R. Jerome, Catie Anderson, Karthik Gangavarapu, Mark Zeller, Emily Spencer, Kristian G. Andersen, Duncan MacCannell, Clinton R. Paden, Yan Li, Jing Zhang, Suxiang Tong, Gregory Armstrong, Scott Morrow, Matthew Willis, Bela T. Matyas, Sundari Mase, Olivia Kasirye, Maggie Park, Godfred Masinde, Curtis Chan, Alexander T. Yu, Shua J. Chai, Elsa Villarino, Brandon Bonin, Debra A. Wadford, and Charles Y. Chiu. Genomic surveillance reveals multiple introductions of SARS-CoV-2 into Northern California. *Science*, 369(6503):582–587, July 2020. Publisher: American Association for the Advancement of Science.

[39] Allison Black and Gytis Dudas. *An applied genomic epidemiological handbook.*

[40] Gytis Dudas and Trevor Bedford. The ability of single genes vs full genomes to resolve time and space in outbreak analysis. *BMC Evolutionary Biology*, 19(1):232, December 2019.

[41] Gytis Dudas, Luiz Max Carvalho, Trevor Bedford, Andrew J. Tatem, Guy Baele, Nuno R. Faria, Daniel J. Park, Jason T. Ladner, Armando Arias, Danny Asogun, Filip Bielejec, Sarah L. Caddy, Matthew Cotten, Jonathan D'Ambrozio, Simon Dellicour, Antonino Di Caro, Joseph W. Diclaro, Sophie Duraffour, Michael J. Elmore, Lawrence S. Fakoli, Ousmane Faye, Merle L. Gilbert, Sahr M. Gevao, Stephen Gire, Adrianne Gladden-Young, Andreas Gnirke, Augustine Goba, Donald S. Grant, Bart L. Haagmans, Julian A. Hiscox, Umaru Jah, Jeffrey R. Kugelman, Di Liu, Jia Lu, Christine M. Malboeuf, Suzanne Mate, David A. Matthews, Christian B. Matranga, Luke W. Meredith, James Qu, Joshua Quick, Suzan D. Pas, My V. T. Phan, Georgios Pollakis, Chantal B. Reusken, Mariano Sanchez-Lockhart, Stephen F. Schaffner, John S. Schieffelin, Rachel S. Sealfon, Etienne Simon-Loriere, Saskia L. Smits, Kilian Stoecker, Lucy Thorne, Ekaete Alice Tobin, Mohamed A. Vandi, Simon J. Watson, Kendra West, Shannon Whitmer, Michael R. Wiley, Sarah M. Winnicki, Shirlee Wohl, Roman Wölfel, Nathan L. Yozwiak, Kristian G. Andersen, Sylvia O. Blyden, Fatorma Bolay,

Miles W. Carroll, Bernice Dahn, Boubacar Diallo, Pierre Formenty, Christophe Fraser, George F. Gao, Robert F. Garry, Ian Goodfellow, Stephan Günther, Christian T. Happi, Edward C. Holmes, Brima Kargbo, Sakoba Keïta, Paul Kellam, Marion P. G. Koopmans, Jens H. Kuhn, Nicholas J. Loman, N'Faly Magassouba, Dhamari Naidoo, Stuart T. Nichol, Tolbert Nyenswah, Gustavo Palacios, Oliver G. Pybus, Pardis C. Sabeti, Amadou Sall, Ute Ströher, Isatta Wurie, Marc A. Suchard, Philippe Lemey, and Andrew Rambaut. Virus genomes reveal factors that spread and sustained the Ebola epidemic. *Nature*, 544(7650):309–315, April 2017. Publisher: Nature Publishing Group.

[42] Gytis Dudas, Luiz Max Carvalho, Andrew Rambaut, and Trevor Bedford. MERS-CoV spillover at the camel-human interface. *eLife*, 7:e31257, January 2018. Publisher: eLife Sciences Publications, Ltd.

[43] Christopher Dye and Moritz U. G. Kraemer. Investigating the monkeypox outbreak. *BMJ*, 377:o1314, May 2022. Publisher: British Medical Journal Publishing Group Section: Editorial.

[44] Leonard E. Egede, Rebekah J. Walker, Jennifer A. Campbell, Sebastian Linde, Laura C. Hawks, and Kaylin M. Burgess. Modern Day Consequences of Historic Redlining: Finding a Path Forward. *Journal of General Internal Medicine*, 38(6):1534–1537, May 2023.

[45] Stefan Elbe and Gemma Buckland-Merrett. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Global challenges (Hoboken, NJ)*, 1(1):33–46, January 2017.

[46] Akira Endo, Hiroaki Murayama, Sam Abbott, Ruwan Ratnayake, Carl A. B. Pearson, W. John Edmunds, Elizabeth Fearon, and Sebastian Funk. Heavy-tailed sexual contact networks and monkeypox epidemiology in the global outbreak, 2022. *Science*, 378(6615):90–94, October 2022. Publisher: American Association for the Advancement of Science.

[47] Christel Faes, Steven Abrams, Dominique Van Beckhoven, Geert Meyfroidt, Erika Vlieghe, Niel Hens, and Belgian Collaborative Group on COVID-19 Hospital Surveillance. Time between Symptom Onset, Hospitalisation and Recovery or Death: Statistical Analysis of Belgian COVID-19 Patients. *International Journal of Environmental Research and Public Health*, 17(20):7560, January 2020. Number: 20 Publisher: Multidisciplinary Digital Publishing Institute.

[48] N. R. Faria, J. Quick, I. M. Claro, J. Thézé, J. G. de Jesus, M. Giovanetti, M. U. G. Kraemer, S. C. Hill, A. Black, A. C. da Costa, L. C. Franco, S. P. Silva, C.-H. Wu, J. Raghwani, S. Cauchemez, L. du Plessis, M. P. Verotti, W. K. de Oliveira, E. H. Carmo, G. E. Coelho, A. C. F. S. Santelli, L. C. Vinhal, C. M. Henriques, J. T. Simpson, M. Loose, K. G. Andersen, N. D. Grubaugh, S. Somasekar, C. Y. Chiu, J. E. Muñoz-Medina, C. R. Gonzalez-Bonilla, C. F. Arias, L. L. Lewis-Ximenez, S. A. Baylis, A. O. Chieppe, S. F. Aguiar, C. A. Fernandes, P. S. Lemos, B. L. S. Nascimento, H. a. O. Monteiro, I. C. Siqueira, M. G. de Queiroz, T. R. de Souza, J. F. Bezerra, M. R. Lemos, G. F. Pereira, D. Loudal, L. C. Moura, R. Dhalia, R. F. França, T. Magalhães, E. T. Marques, T. Jaenisch, G. L. Wallau, M. C. de Lima, V. Nascimento, E. M. de Cerqueira, M. M. de Lima, D. L. Mascarenhas, J. P. Moura Neto, A. S. Levin, T. R. Tozetto-Mendoza, S. N. Fonseca, M. C. Mendes-Correa, F. P. Milagres, A. Segurado, E. C. Holmes, A. Rambaut, T. Bedford, M. R. T. Nunes, E. C. Sabino, L. C. J. Alcantara, N. J. Loman, and O. G. Pybus. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. *Nature*, 546(7658):406–410, June 2017. Number: 7658 Publisher: Nature Publishing Group.

[49] P. Farmer. Social inequalities and emerging infectious diseases. *Emerging Infectious Diseases*, 2(4):259–269, 1996.

[50] Joseph Felsenstein. Accuracy of Coalescent Likelihood Estimates: Do We Need More Sites, More Sequences, or More Loci? *Molecular Biology and Evolution*, 23(3):691–700, March 2006.

[51] Luca Ferretti, Chris Wymant, Michelle Kendall, Lele Zhao, Anel Nurtay, Lucie Abeler-Dörner, Michael Parker, David Bonsall, and Christophe Fraser. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science*, 368(6491):eabb6936, May 2020. Publisher: American Association for the Advancement of Science.

[52] Marlin D. Figgins and Trevor Bedford. SARS-CoV-2 variant dynamics across US states show consistent differences in effective reproduction numbers, August 2022. Pages: 2021.12.09.21267544.

[53] Douglas L. Fink, Helen Callaby, Akish Luintel, William Beynon, Helena Bond, Eleanor Y. Lim, Effrossyni Gkrania-Klotsas, Jospeh Heskin, Margherita Bracchi, Balram Rathish, Iain Milligan, Geraldine O'Hara, Stephanie Rimmer, Joanna R. Peters, Lara Payne, Nisha Mody, Bethany Hodgson, Penny Lewthwaite, Rebecca Lester, Stephen D. Woolley, Ann Sturdy, Ashley Whittington, Leann Johnson, Nathan Jacobs, John Quartey, Brendan AI Payne, Stewart Crowe, Ivo AM Elliott, Thomas

Harrison, Joby Cole, Katie Beard, Tomas-Paul Cusack, Imogen Jones, Rishi Banerjee, Tommy Rampling, Jake Dunning, Douglas L. Fink, Iain D. Milligan, Akish Luintel, Alison J. Rodger, Sanjay R. Bhagani, Lucy E. Lamb, Rachel C. Moores, Simon F. K. Lee, Colin S. Brown, Susan Hopkins, Stephen Mepham, Simon Warren, Aoife Molloy, Ian Cropley, Alex Kew, Natasha Karunaharan, Antonia Scobie, Jennifer Hart, Dianne Irish, Tanzina Haque, Hamid Jalal, Robin Smith, Damien Mack, Tristan Barber, Fiona Burns, Robert Miller, Eleanor Hamlyn, Pedro Simoes, Breda Athan, Jennifer Abrahamsen, Jessica Joyce, Caroline Taylor, Sally Reddecliffe, Chloe Miller, Brooke Reeve, Hugh Kingston, Tim Crocker-Buque, Nicolas Massie, Ankush Dhariwal, Angelina Jayakumar, Robert Hammond, Alexandra Bramley, Tanmay Kanitkar, Laura Maynard-Smith, Eliza Gil, Cavan O'Connor, Derek Cocker, Wendy Spicer, Marisa Lanzman, Meera Thacker, Zoe O. Anorson, Dharmesh Patel, Alan Williams, Catherine F. Houlihan, Dominic Wakerley, Claire N. Gordon, Helen Callaby, Daniel J. Bailey, Jenna Furneaux, Abbie M. Bown, Elizabeth J. Truelove, Marian J. Killip, David Jackson, Tracy L. B. Beetar-King, Ulrike M. V. Arnold, Rhea M. Strachan, Jones Matthew, Hannah J. Matthew, Jane C. Osborne, Tommy Rampling, Richard Vipond, Barry Gibney, Jodie Owen, Helena Bond, Will Beynon, Michael Hunter, Louise McCorry, Carol Emerson, Say Quah, Suzanne Todd, Emma McCarty, Eoin Walker, Susan Feeney, Tanya Curran, Kathy Li, J. D. Mullan, Kate Jackson, Peter Nelson, Kevin Lewis, Mark McNicol, Marcus Pratt, Anna Smith, Erin Vos, Fahad Alsalemee, Daniel O. Leary, John Canny, Katherine McGinnity, Carly Culbert, Conor McDowell, Cathy McQuillan, Eunjin Jeong, Lynsey Glass, Jessica Dyche, Paula McClean, Rebecca Stewart, Harold Ursolino, Melissa Perry, Hannah McCormick, Eleanor Y. Lim, Effrossyni Gkrania-Klotsas, Margherita Bracchi, Joseph Heskin, Nicklas Brown, Thomas Juniper, Borja Mora-Peris, Alessia Dalla-Pria, Nicola Mackie, Lucy Garvey, Alan Winston, Graham Cooke, Mark Nelson, Emer Kilbride, Ala Elbishi, William Kerrigan, Joshua Silva, Jesal Gohil, Sasha Payagala, Yasmin Walters, Joanna Smith, Jonathan Goodfellow, Kitty Lyons, Hsiu Tung, Kinjal Patel, Merle Henderson, Michael Butler, Edu Peres, Taiana Silva Carvalho, Antoine Joly, Molly Dickinson, Luke S. P. Moore, Nabeela Mughal, Stephen Hughes, Shrada Chitlangia, Priyanka Viramgana, Ruth Byrne, Paul Randell, Luigi Strangis, Nicola Poveda, Deborah Bovey, Poppy Richardson, Vivian Heaslip, Christopher Higgs, Marta Boffito, Nicolo Girometti, Gary Whitlock, Victoria Tittle, Rachel Jones, Michael Rayment, Christopher Scott, David Asboe, Marcus Pond, David Muir, Balram Rathish, Geraldine O'Hara, Movin Abeywickrema, Sarah-Lou Bailey, Sara E. Boyd, Dayana Da Silva Fontoura, Anna Daunt, Claire Y. Mason, Jamie Murphy, Vasanth V. Naidu, Aatish Patel, Caitlin Pley, Ethan Redmore, Katherine Sharrocks, Luke B. Snell, Rohan Sundramoorthi, Jerry C. H. Tam, Aisling Brown, Sam Douthwaite, Anna Goodman, Gaia Nebbia, William Newsholme, Nicholas Price, Emily Shaw, Alex Salam, Claire van Nispen tot Panner-

den, Helen Winslow, Julia Bilinska, Sarah Keegan, Harry Coleman, Jessica Doctor, Nasreen Moini, Daniella Chilton, Golaleh Haidari, Rebecca Simons, Rajababu Kulasegaram, Nick Larbalestier, Achyuta Nori, Jack R. Potter, Cecilia Tuudah, Paul Wade, Alexandra Travers, Sarah Dunford, Joshua Greenwood, Georgina Oledimmah, Lesley Gyampo, Pedro SA Pinto, AbdulKadir Muse, Zoe Parker, Charlotte Alexander, Alexander Khan, Medinat Ajayi, Abigail Baltazar, Davis Sharella, Nasra Hersi, Thuy Nguyen, Rugiatu Timbo, Ismail Jalloh, Susan Bryan, Patricia Clarke, Marcia Kerr, Fidelis Amedu, Maria BohoBonaba, Sarah Haque, Michelle Howson, Norbai Tambilawan, Soledad Yupanqui Estay, Hawanatu Bangura, Tseday Gideon, Damilola Jerome-oboh, Linda Tetteh, Chioma Nwagu, Viwoalo Agbaglah, Nona Narag, Mahima Zaveri, Maedhbh Ni Luanaigh, Peggy Keane, Joanna R. Peters, Stephanie Rimmer, Aula Abbara, Borja Mora-Peris, Olamide Dosekun, Lucy Garvey, Mhairi Bolland, Adam Stafford, Dina Saleh, Rhianna Sheridan, Ella Davies, Kristi Sun, Mark Gilchrist, Priti Kukadia, Muhammed Embrahimsa, Christopher Chiu, Michael Butler, Kinjal Patel, Lauren Taylor, Charlotte Short, Jasmini Alagratnam, Iresh Jayaweera, Kavitha Gundugola, Lara V. S. Payne, Nisha Mody, Killian Quinn, Caoimhe Nic Fhogartaigh, Nivenjit Kaur, Salmaan Bholah, Kajann Kantha, Jonathan Youngs, Temi Lampejo, Nicholas Pitto, David S. Lawrence, Holly Middleditch, Lourdes Dominguez-Dominguez, Ayoma Ratnappuli, Sara Al-Hashimi, Amelia Oliveira, Zoe Ottaway, Larissa Mulka, Bethany Hodgson, Penny Lewthwaite, Anne M. Neary, Michael R. Downey, Danielle C. Lucy, Craig I. McCallum, Michael Beadsworth, Libuse Ratcliffe, Tom E. Fletcher, Gerry Davies, Nicholas Wong, Stephen Aston, Thomas E. Wingfield, Thomas Blanchard, Paul Hine, Rebecca Lester, Stephen D. Woolley, Susie Gould, Christopher Smith, Michael Abouyannis, Abolaji Atomode, James Cruise, Merna Samual, Nicola Scott, Vino Srirathan, Joseph Lewis, Lauren Richards, Mary-Ann Cummings, Emily Gillan, Rebecca Peers, Amy Tickle, Grace Keating, Tendi Chinyanda, Mav Sanchez, Daniel Harrison, Hoyle, Ben Metcalfe, Jennifer Taylor, Nicky Johnson, Neil Kelle, Kirsty McDowell, Ian Richardson, Monette Saguidan, Nicky Farmer, Angella Gillespie, Shay Willoughby, Samantha Parker, Shamseena Avulan, Shazia Arif, Suzanne Marshall, David Carlisle, Mohsen Rezaei, Angela Booth, Joanne Watts, Lauren Tremarco, Priyanga Jeyanayagam, Odinaka Ubochi, Daniel Vagianos, Mark Richardson, Anthony Jarvis, Kyra Gow, Jade Walmsley, Adam O'keefe, Anna Smielewska, Mark Hopkins, Fatima Balane, Sarah Bradley, Tumena Corrah, Venus Daquiz, Christopher Dugan, Joshua Elliot, Fiona Foley, Dawn Friday, May Gamit, David Garner, Karishma Gokani, Laurence John, Deepa Joseph, Nuzhath Khan, Cherifer Mamuyac, Alastair McGregor, John McSorley, Victoria Parris, Luciana Rubinstein, Julian Rycroft, Kelcy Salinas, Jason Salinas, Jency Sebatian, Melanie Smith, Marina Tejero Garcia, Uchenna Ume, Margarete Vicentine, Gabriel Wallis, Ann Sturdy, Ashley Whittington, Nathan Jacobs, Leann Johnson, Alec Bonington, Alison Uriel, Andrew Ustianowski, Balazs

Dancso, Celia Hogan, Clare van Halsema, F. Javier Vilar, Karen Devine, Katherine Ajdukiewicz, Rajesh Rajendran, Samit Ghosh, Michael Riste, Nicholas Machin, Chitra Babu, Shazaad Ahmad, Dorcas Obeng, Farnaz Dave, Gavin Conolley, Joseph Thompson, Maya Tickell-Painter, Prasun Chakravorty, Rachel Pringle, Mohammad R. Zafar, Sarah Lawrence, Amada Sanchez-Gonzalez, Cristina Fernandez, Lynsey Goodwin, David Carey, Molly Howarth-Maddison, Samuel Moody, Rebecca Upton, Christina Apthorp, Charlotte Murray, Kirstie Salthouse, Sabah Nadeem, Grant Ridley, Francesca White, Andrew Brown, Michael Lawless, Mohamed Mohamed, Robert Mulligan, Amy Belfield, Jacob Brolly, Maria Calderon, James Cheveau, Milo Cullinan, Sophie Garrad, Will Griffiths, Aidan Ireland, Peter Ireland, Charlotte Milne, Paul Nwajiugo, John Quartey, Bijan Ghavami-Kia, Chris Duncan, Adam Evans, Ewan Hunter, Ashley Price, Matthias Schmid, Uli Schwab, Yusri Taha, Brendan Payne, Ivo A. M. Elliott, Stewart Crowe, Charles J. Woodrow, Drosos E. Karageorgopoulos, Peter J. Davis, Emily Lord, Oliver J. Bannister, Andrew B. Dagens, Thomas Harrison, Joby Cole, Anne Tunbridge, Saher Choudry, Adam Telfer, Ihsan Jhibril, Syed N. Atta, Ben Stone, Cariad Evans, Mike Ankcorn, Suha Akili, Mehmet Yavuz, Vicky Goodall, Sam Farrow, Georgina Mountford, Tomas-Paul Cusack, Kate Beard, Julian Sutton, Tristan Clark, Annette Mason, Mike Vickers, Derek Macallan, Tihana Bicanic, Angela Houston, Cassie Pope, NgeeKeong Tan, Christopher Ward, Imogen Jones, Rishi Banerjee, Jonathan Cohen, Marieke Emonts-le Clercq, David Porter, Andrew Riordan, Ruchi Sinha, Elizabeth Whittaker, and Jake Dunning. Clinical features and management of individuals admitted to hospital with monkeypox and associated complications across the UK: a retrospective cohort study. *The Lancet Infectious Diseases*, 23(5):589–597, May 2023. Publisher: Elsevier.

[54] Cadhla Firth, Andrew Kitchen, Beth Shapiro, Marc A. Suchard, Edward C. Holmes, and Andrew Rambaut. Using Time-Structured Data to Estimate Evolutionary Rates of Double-Stranded DNA Viruses. *Molecular Biology and Evolution*, 27(9):2038–2051, September 2010.

[55] David N. Fisman and Ashleigh R. Tuite. Progressive Increase in Virulence of Novel SARS-CoV-2 Variants in Ontario, Canada. preprint, Infectious Diseases (except HIV/AIDS), July 2021.

[56] Aaron T. Fleischauer, James C. Kile, Molly Davidson, Marc Fischer, Kevin L. Karem, Robert Teclaw, Hans Messersmith, Pamela Pontones, Bradley A. Beard, Zachary H. Braden, Joanne Cono, James J. Sejvar, Ali S. Khan, Inger Damon, and Matthew J. Kuehnert. Evaluation of Human-to-Human Transmission of Monkeypox from Infected Patients to Health Care Workers. *Clinical Infectious Diseases*, 40(5):689–694, March 2005.

[57] Jackie Fortier. It took him multiple trips to the ER to start receiving treatment for monkeypox. *NPR*, July 2022.

[58] Sharon E. Frey, Anna Wald, Srilatha Edupuganti, Lisa A. Jackson, Jack T. Stapleton, Hana El Sahly, Samer S. El-Kamary, Kathryn Edwards, Harry Keyserling, Patricia Winokur, Wendy Keitel, Heather Hill, Johannes B. Goll, Edwin L. Anderson, Irene L. Graham, Christine Johnston, Mark Mulligan, Nadine Rouphael, Robert Atmar, Shital Patel, Wilbur Chen, Karen Kotloff, C. Buddy Creech, Paul Chaplin, and Robert B. Belshe. Comparison of lyophilized versus liquid modified vaccinia Ankara (MVA) formulations and subcutaneous versus intradermal routes of administration in healthy vaccinia-naïve subjects. *Vaccine*, 33(39):5225–5234, September 2015.

[59] Liping Gao, Qi Shi, Xiaoping Dong, Miao Wang, Zhiguo Liu, and Zhenjun Li. Mpox, Caused by the MPXV of the Clade IIb Lineage, Goes Global. *Tropical Medicine and Infectious Disease*, 8(2):76, February 2023. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.

[60] Jennifer L. Gardy and Nicholas J. Loman. Towards a genomics-informed, real-time, global pathogen surveillance system. *Nature Reviews Genetics*, 19(1):9–20, January 2018. Publisher: Nature Publishing Group.

[61] Crystal M. Gigante, Bette Korber, Matthew H. Seabolt, Kimberly Wilkins, Whitni Davidson, Agam K. Rao, Hui Zhao, Todd G. Smith, Christine M. Hughes, Faisal Minhaj, Michelle A. Waltenburg, James Theiler, Sandra Smole, Glen R. Gallagher, David Blythe, Robert Myers, Joann Schulte, Joey Stringer, Philip Lee, Rafael M. Mendoza, LaToya A. Griffin-Thomas, Jenny Crain, Jade Murray, Annette Atkinson, Anthony H. Gonzalez, June Nash, Dhwani Batra, Inger Damon, Jennifer McQuiston, Christina L. Hutson, Andrea M. McCollum, and Yu Li. Multiple lineages of monkeypox virus detected in the United States, 2021–2022. *Science*, 378(6619):560–565, November 2022. Publisher: American Association for the Advancement of Science.

[62] Marius Gilbert, Giulia Pullano, Francesco Pinotti, Eugenio Valdano, Chiara Poletto, Pierre-Yves Boëlle, Eric D'Ortenzio, Yazdan Yazdanpanah, Serge Paul Eholie, Mathias Altmann, Bernardo Gutierrez, Moritz U. G. Kraemer, and Vittoria Colizza. Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study. *The Lancet*, 395(10227):871–877, March 2020. Publisher: Elsevier.

[63] Mandev S. Gill, Philippe Lemey, Nuno R. Faria, Andrew Rambaut, Beth Shapiro, and Marc A. Suchard. Improving Bayesian Population Dynamics Inference: A Coalescent-

Based Model for Multiple Loci. *Molecular Biology and Evolution*, 30(3):713–724, March 2013.

[64] Gregg S. Gonsalves, Kenneth Mayer, and Chris Beyrer. Déjà vu All Over Again? Emergent Monkeypox, Delayed Responses, and Stigmatized Populations. *Journal of Urban Health*, 99(4):603–606, August 2022.

[65] Katelyn M. Gostic, Lauren McGough, Edward B. Baskerville, Sam Abbott, Keya Joshi, Christine Tedijanto, Rebecca Kahn, Rene Niehus, James A. Hay, Pablo M. De Salazar, Joel Hellewell, Sophie Meakin, James D. Munday, Nikos I. Bosse, Katharine Sherrat, Robin N. Thompson, Laura F. White, Jana S. Huisman, Jérémie Scire, Sebastian Bonhoeffer, Tanja Stadler, Jacco Wallinga, Sebastian Funk, Marc Lipsitch, and Sarah Cobey. Practical considerations for measuring the effective reproductive number, Rt. *PLoS Computational Biology*, 16(12):e1008409, December 2020.

[66] Statistics Canada Government of Canada. The Daily — A statistical portrait of Canada's diverse LGBTQ2+ communities, June 2021. Last Modified: 2021-06-15.

[67] Allison J. Greaney, Andrea N. Loes, Katharine H. D. Crawford, Tyler N. Starr, Keara D. Malone, Helen Y. Chu, and Jesse D. Bloom. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host & Microbe*, 29(3):463–476.e6, March 2021.

[68] Bryan T. Grenfell, Oliver G. Pybus, Julia R. Gog, James L. N. Wood, Janet M. Daly, Jenny A. Mumford, and Edward C. Holmes. Unifying the Epidemiological and Evolutionary Dynamics of Pathogens. *Science*, 303(5656):327–332, January 2004. Publisher: American Association for the Advancement of Science.

[69] Jeremy A Grey, Kyle T Bernstein, Patrick S Sullivan, David W Purcell, Harrell W Chesson, Thomas L Gift, and Eli S Rosenberg. Estimating the Population Sizes of Men Who Have Sex With Men in US States and Counties Using Data From the American Community Survey. *JMIR Public Health and Surveillance*, 2(1):e14, April 2016.

[70] Gareth J. Griffith, Tim T. Morris, Matthew J. Tudball, Annie Herbert, Giulia Mancano, Lindsey Pike, Gemma C. Sharp, Jonathan Sterne, Tom M. Palmer, George Davey Smith, Kate Tilling, Luisa Zuccolo, Neil M. Davies, and Gibran Hemani. Collider bias undermines our understanding of COVID-19 disease risk and severity. *Nature Communications*, 11(1):5749, November 2020. Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Epidemiology;Infectious

diseases;Risk factors;Statistical methods Subject_term_id: epidemiology;infectious-diseases;risk-factors;statistical-methods.

[71] Nathan D. Grubaugh, Jason T. Ladner, Philippe Lemey, Oliver G. Pybus, Andrew Rambaut, Edward C. Holmes, and Kristian G. Andersen. Tracking virus outbreaks in the twenty-first century. *Nature Microbiology*, 4(1):10–19, January 2019. Publisher: Nature Publishing Group.

[72] Sarah Anne J. Guagliardo, Benjamin Monroe, Christian Moundjoa, Ateba Athanase, Gordon Okpu, Jillybeth Burgado, Michael B. Townsend, Panayampalli S. Satheshkumar, Scott Epperson, Jeffrey B. Doty, Mary G. Reynolds, Elisabeth Dibongue, Georges Alain Etoundi, Els Mathieu, and Andrea M. McCollum. Asymptomatic Orthopoxvirus Circulation in Humans in the Wake of a Monkeypox Outbreak among Chimpanzees in Cameroon. *The American Journal of Tropical Medicine and Hygiene*, 102(1):206–212, November 2019. Publisher: The American Society of Tropical Medicine and Hygiene Section: The American Journal of Tropical Medicine and Hygiene.

[73] Giorgio Guzzetta, Alessia Mammone, Federica Ferraro, Anna Caraglia, Alessia Rapiti, Valentina Marziano, Piero Poletti, Danilo Cereda, Francesco Vairo, Giovanna Mattei, Francesco Maraglino, Giovanni Rezza, and Stefano Merler. Early Estimates of Monkeypox Incubation Period, Generation Time, and Reproduction Number, Italy, May–June 2022 - Volume 28, Number 10—October 2022 - Emerging Infectious Diseases journal - CDC.

[74] James Hadfield, Colin Megill, Sidney M Bell, John Huddleston, Barney Potter, Charlton Callender, Pavel Sagulenko, Trevor Bedford, and Richard A Neher. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*, 34(23):4121–4123, December 2018.

[75] Chelsea L. Hansen, Amanda Perofsky, Roy Burstein, Michael Famulare, Shanda Boyle, Robin Prentice, Cooper Marshall, Benjamin JJ McCormick, David Reinhart, Ben Capodanno, Melissa Truong, Kristen Schwabe-Fry, Kayla Kuchta, Brian Pfau, Zack Acker, Jover Lee, Thomas R. Sibley, Evan McDermot, Leslie Rodriguez-Salas, Jeremy Stone, Luis Gamboa, Peter D. Han, Jeffery S. Duchin, Alpana Waghmare, Janet A. Englund, Jay Shendure, Trevor Bedford, Helen Y. Chu, Lea M. Starita, and Cécile Viboud. Trends in risk factors and symptoms associated with SARS-CoV-2 and Rhinovirus test positivity in King County, Washington: A Test-Negative Design Study of the Greater Seattle Coronavirus Assessment Network, August 2022. Pages: 2022.08.12.22278203.

[76] Verity Hill, Christopher Ruis, Sumali Bajaj, Oliver G. Pybus, and Moritz U. G. Kraemer. Progress and challenges in virus genomic epidemiology. *Trends in Parasitology*, 37(12):1038–1049, December 2021. Publisher: Elsevier.

[77] Yong Huang, Li Mu, and Wei Wang. Monkeypox: epidemiology, pathogenesis, treatment and prevention. *Signal Transduction and Targeted Therapy*, 7(1):1–22, November 2022. Number: 1 Publisher: Nature Publishing Group.

[78] John Huddleston, James Hadfield, Thomas R. Sibley, Jover Lee, Kairsten Fay, Misja Ilcisin, Elias Harkins, Trevor Bedford, Richard A. Neher, and Emma B. Hodcroft. Augur: a bioinformatics toolkit for phylogenetic analyses of human pathogens. *Journal of Open Source Software*, 6(57):2906, 2021.

[79] Carolyn Ingram, Esther Min, Edmund Seto, BJ Cummings, and Stephanie Farquhar. Cumulative Impacts and COVID-19: Implications for Low-Income, Minoritized, and Health-Compromised Communities in King County, WA. *Journal of Racial and Ethnic Health Disparities*, 9(4):1210–1224, August 2022.

[80] Joana Isidro, Vítor Borges, Miguel Pinto, Daniel Sobral, João Dourado Santos, Alexandra Nunes, Verónica Mixão, Rita Ferreira, Daniela Santos, Silvia Duarte, Luís Vieira, Maria José Borrego, Sofia Núncio, Isabel Lopes de Carvalho, Ana Pelerito, Rita Cordeiro, and João Paulo Gomes. Phylogenomic characterization and signs of microevolution in the 2022 multi-country outbreak of monkeypox virus. *Nature Medicine*, 28(8):1569–1572, August 2022. Number: 8 Publisher: Nature Publishing Group.

[81] Levente Juhasz and Hartwig Hochmair. Studying Spatial and Temporal Visitation Patterns of Points of Interest Using SafeGraph Data in Florida. *GIS Center*, June 2020.

[82] J. F. C. Kingman. The coalescent. *Stochastic Processes and their Applications*, 13(3):235–248, September 1982.

[83] Stephen M. Kissler, Nishant Kishore, Malavika Prabhu, Dena Goffman, Yaakov Beilin, Ruth Landau, Cynthia Gyamfi-Bannerman, Brian T. Bateman, Jon Snyder, Armin S. Razavi, Daniel Katz, Jonathan Gal, Angela Bianco, Joanne Stone, Daniel Larremore, Caroline O. Buckee, and Yonatan H. Grad. Reductions in commuting mobility correlate with geographic differences in SARS-CoV-2 prevalence in New York City. *Nature Communications*, 11(1):4674, September 2020. Number: 1 Publisher: Nature Publishing Group.

[84] Frank Konings, Mark D. Perkins, Jens H. Kuhn, Mark J. Pallen, Erik J. Alm, Brett N. Archer, Amal Barakat, Trevor Bedford, Jinal N. Bhiman, Leon Caly, Lisa L. Carter, Anne Cullinane, Tulio de Oliveira, Julian Druce, Ihab El Masry, Roger Evans, George F. Gao, Alexander E. Gorbalenya, Esther Hamblion, Belinda L. Herring, Emma Hodcroft, Edward C. Holmes, Manish Kakkar, Shagun Khare, Marion P. G. Koopmans, Bette Korber, Juliana Leite, Duncan MacCannell, Marco Marklewitz, Sebastian Maurer-Stroh, Jairo Andres Mendez Rico, Vincent J. Munster, Richard Neher, Bas Oude Munnink, Boris I. Pavlin, Malik Peiris, Leo Poon, Oliver Pybus, Andrew Rambaut, Paola Resende, Lorenzo Subissi, Volker Thiel, Suxiang Tong, Sylvie van der Werf, Anne von Gottberg, John Ziebuhr, and Maria D. Van Kerkhove. SARS-CoV-2 Variants of Interest and Concern naming scheme conducive for global discourse. *Nature Microbiology*, 6(7):821–823, July 2021. Bandiera_abtest: a Cg_type: Nature Research Journals Number: 7 Primary_atype: Comments & Opinion Publisher: Nature Publishing Group Subject_term: Policy and public health in microbiology;SARS-CoV-2 Subject_term_id: policy-and-public-health-in-microbiology;sars-cov-2.

[85] Andrew M. Laitman, Joshua A. Lieberman, Noah G. Hoffman, Pavitra Roychoudhury, Patrick C. Mathias, and Alexander L. Greninger. The SARS-CoV-2 Omicron Variant Does Not Have Higher Nasal Viral Loads Compared to the Delta Variant in Symptomatic and Asymptomatic Individuals. *Journal of Clinical Microbiology*, 60(4):e00139–22, March 2022. Publisher: American Society for Microbiology.

[86] Henry Laurenson-Schafer, Nikola Sklenovská, Ana Hoxha, Steven M. Kerr, Patricia Ndumbi, Julia Fitzner, Maria Almiron, Luis Alves de Sousa, Sylvie Briand, Orlando Cenciarelli, Soledad Colombe, Meg Doherty, Ibrahima Soce Fall, Christian García-Calavaro, Joana M. Haussig, Masaya Kato, Abdi Rahman Mahamud, Oliver W. Morgan, Pierre Nabeth, Jeremias Domingos Naiene, Wildo Araujo Navegantes, Opeayo Ogundiran, Charles Okot, Richard Pebody, Tamano Matsui, Hugo López-Gatell Ramírez, Catherine Smallwood, Raúl Francisco Pérez Tasigchana, Aisling M. Vaughan, George Sie Williams, Basma Abdelgawad, Amarnath Babu, Evans Buliva, Finlay Campbell, Daniel Cardoso Portela Câmara, Zainab Eleiba, Blanche Johanna Greene-Cramer, Esther Hamblion, Mahmoud Hassan, Kaja Kaasik-Aaslav, Basant Mohamed, Victoria Ndarukwa, James Richard Otieno, Jeffrey Pires, Jukka Pukkila, Felix Sanni, Craig Schultz, Tika Sedai, Laila Skrowny, Manilay Phengxay, Ariuntuya Ochirpurev, Jozica Skufca, Laura Goddard, Viema Biaukula, Peter Omondi Mala, Rosamund F. Lewis, Boris I. Pavlin, and Olivier le Polain de Waroux. Description of the first global outbreak of mpox: an analysis of global surveillance data. *The Lancet Global Health*, 11(7):e1012–e1023, July 2023. Publisher: Elsevier.

[87] Adam S. Lauring, Judith Frydman, and Raul Andino. The role of mutational robust-

ness in RNA virus evolution. *Nature Reviews Microbiology*, 11(5):327–336, May 2013. Publisher: Nature Publishing Group.

[88] Maylis Layan, Nicola F Müller, Simon Dellicour, Nicola De Maio, Hervé Bourhy, Simon Cauchemez, and Guy Baele. Impact and mitigation of sampling bias to determine viral spread: Evaluating discrete phylogeography through CTMC modeling and structured coalescent model approximations. *Virus Evolution*, 9(1):vead010, January 2023.

[89] Seungjae Lee, Tark Kim, Eunjung Lee, Cheolgu Lee, Hojung Kim, Heejeong Rhee, Se Yoon Park, Hyo-Ju Son, Shinae Yu, Jung Wan Park, Eun Ju Choo, Suyeon Park, Mark Loeb, and Tae Hyong Kim. Clinical Course and Molecular Viral Shedding Among Asymptomatic and Symptomatic Patients With SARS-CoV-2 Infection in a Community Treatment Center in the Republic of Korea. *JAMA Internal Medicine*, 180(11):1447–1452, November 2020.

[90] Sonja Lehtinen, Peter Ashcroft, and Sebastian Bonhoeffer. On the relationship between serial interval, infectiousness profile and generation time. *Journal of The Royal Society Interface*, 18(174):20200756, January 2021. Publisher: Royal Society.

[91] Philippe Lemey, Andrew Rambaut, Trevor Bedford, Nuno Faria, Filip Bielejec, Guy Baele, Colin A. Russell, Derek J. Smith, Oliver G. Pybus, Dirk Brockmann, and Marc A. Suchard. Unifying Viral Genetics and Human Transportation Data to Predict the Global Transmission Dynamics of Human Influenza H3N2. *PLOS Pathogens*, 10(2):e1003932, February 2014. Publisher: Public Library of Science.

[92] Philippe Lemey, Andrew Rambaut, Alexei J. Drummond, and Marc A. Suchard. Bayesian Phylogeography Finds Its Roots. *PLOS Computational Biology*, 5(9):e1000520, September 2009. Publisher: Public Library of Science.

[93] Philippe Lemey, Nick Ruktanonchai, Samuel L. Hong, Vittoria Colizza, Chiara Poletto, Frederik Van den Broeck, Mandev S. Gill, Xiang Ji, Anthony Levasseur, Bas B. Oude Munnink, Marion Koopmans, Adam Sadilek, Shengjie Lai, Andrew J. Tatem, Guy Baele, Marc A. Suchard, and Simon Dellicour. Untangling introductions and persistence in COVID-19 resurgence in Europe. *Nature*, 595(7869):713–717, July 2021. Number: 7869 Publisher: Nature Publishing Group.

[94] Jacob E. Lemieux, Katherine J. Siddle, Bennett M. Shaw, Christine Loreth, Stephen F. Schaffner, Adrianne Gladden-Young, Gordon Adams, Timelia Fink, Christopher H. Tomkins-Tinch, Lydia A. Krasilnikova, Katherine C. DeRuff, Melissa

Rudy, Matthew R. Bauer, Kim A. Lagerborg, Erica Normandin, Sinéad B. Chapman, Steven K. Reilly, Melis N. Anahtar, Aaron E. Lin, Amber Carter, Cameron Myhrvold, Molly E. Kemball, Sushma Chaluvadi, Caroline Cusick, Katelyn Flowers, Anna Neumann, Felecia Cerrato, Maha Farhat, Damien Slater, Jason B. Harris, John A. Branda, David Hooper, Jessie M. Gaeta, Travis P. Baggett, James O'Connell, Andreas Gnirke, Tami D. Lieberman, Anthony Philippakis, Meagan Burns, Catherine M. Brown, Jeremy Luban, Edward T. Ryan, Sarah E. Turbett, Regina C. LaRocque, William P. Hanage, Glen R. Gallagher, Lawrence C. Madoff, Sandra Smole, Virginia M. Pierce, Eric Rosenberg, Pardis C. Sabeti, Daniel J. Park, and Bronwyn L. MacInnis. Phylogenetic analysis of SARS-CoV-2 in Boston highlights the impact of superspreading events. *Science*, 371(6529):eabe3261, February 2021. Publisher: American Association for the Advancement of Science.

[95] Joseph A. Lewnard, Vennis X. Hong, Manish M. Patel, Rebecca Kahn, Marc Lipsitch, and Sara Y. Tartof. Clinical outcomes associated with SARS-CoV-2 Omicron (B.1.1.529) variant and BA.1/BA.1.1 or BA.2 subvariant infection in Southern California. *Nature Medicine*, 28(9):1933–1943, September 2022. Publisher: Nature Publishing Group.

[96] Gideon Lichfield. The Bittersweet Defeat of Mpox. *Wired*. Section: tags.

[97] J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, and W. M. Getz. Superspreading and the effect of individual variation on disease emergence. *Nature*, 438(7066):355–359, November 2005. Number: 7066 Publisher: Nature Publishing Group.

[98] Fok-Moon Lum, Anthony Torres-Ruesta, Matthew Z. Tay, Raymond T. P. Lin, David C. Lye, Laurent Rénia, and Lisa F. P. Ng. Monkeypox: disease epidemiology, host immunity and clinical interventions. *Nature Reviews Immunology*, 22(10):597–613, October 2022. Number: 10 Publisher: Nature Publishing Group.

[99] Junling Ma. Estimating epidemic exponential growth rate and basic reproduction number. *Infectious Disease Modelling*, 5:129–141, January 2020.

[100] Qiuyue Ma, Jue Liu, Qiao Liu, Liangyu Kang, Runqing Liu, Wenzhan Jing, Yu Wu, and Min Liu. Global Percentage of Asymptomatic SARS-CoV-2 Infections Among the Tested Population and Individuals With Confirmed COVID-19 Diagnosis: A Systematic Review and Meta-analysis. *JAMA Network Open*, 4(12):e2137257, December 2021.

[101] Nicola De Maio, Chieh-Hsi Wu, Kathleen M. O'Reilly, and Daniel Wilson. New Routes to Phylogeography: A Bayesian Structured Coalescent Approximation. *PLOS Genetics*, 11(8):e1005421, August 2015. Publisher: Public Library of Science.

[102] John T. McCrone, Verity Hill, Sumali Bajaj, Rosario Evans Pena, Ben C. Lambert, Rhys Inward, Samir Bhatt, Erik Volz, Christopher Ruis, Simon Dellicour, Guy Baele, Alexander E. Zarebski, Adam Sadilek, Neo Wu, Aaron Schneider, Xiang Ji, Jayna Raghwani, Ben Jackson, Rachel Colquhoun, Áine O'Toole, Thomas P. Peacock, Kate Twohig, Simon Thelwall, Gavin Dabrera, Richard Myers, Nuno R. Faria, Carmen Huber, Isaac I. Bogoch, Kamran Khan, Louis du Plessis, Jeffrey C. Barrett, David M. Aanensen, Wendy S. Barclay, Meera Chand, Thomas Connor, Nicholas J. Loman, Marc A. Suchard, Oliver G. Pybus, Andrew Rambaut, and Moritz U. G. Kraemer. Context-specific emergence and growth of the SARS-CoV-2 Delta variant. *Nature*, 610(7930):154–160, October 2022. Number: 7930 Publisher: Nature Publishing Group.

[103] John T McCrone, Robert J Woods, Emily T Martin, Ryan E Malosh, Arnold S Monto, and Adam S Lauring. Stochastic processes constrain the within and between host evolution of influenza virus. *eLife*, 7:e35962, April 2018. Publisher: eLife Sciences Publications, Ltd.

[104] Tracy McMillen, Krupa Jani, Elizabeth V. Robilotti, Mini Kamboj, and N. Esther Babady. The spike gene target failure (SGTF) genomic signature is highly accurate for the identification of Alpha and Omicron SARS-CoV-2 variants. *Scientific Reports*, 12(1):18968, November 2022. Publisher: Nature Publishing Group.

[105] Bui Quang Minh, Heiko A Schmidt, Olga Chernomor, Dominik Schrempf, Michael D Woodhams, Arndt von Haeseler, and Robert Lanfear. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, 37(5):1530–1534, May 2020.

[106] Oriol Mitjà, Andrea Alemany, Michael Marks, Jezer I. Lezama Mora, Juan Carlos Rodríguez-Aldama, Mayara Secco Torres Silva, Ever Arturo Corral Herrera, Brenda Crabtree-Ramirez, José Luis Blanco, Nicolo Girometti, Valentina Mazzotta, Aniruddha Hazra, Macarena Silva, Juan José Montenegro-Idrogo, Kelly Gebo, Jade Ghosn, María Fernanda Peña Vázquez, Eduardo Matos Prado, Uche Unigwe, Judit Villar-García, Noah Wald-Dickler, Jason Zucker, Roger Paredes, Alexandra Calmy, Laura Waters, Cristina Galvan-Casas, Sharon Walmsley, Chloe M. Orkin, Viviana Leiro, Lucila Marchetta, Patricia Fernandez Pardal, María Inés Figueroa, Pedro Cahn, Katharina Grabmeier-Pfistershammer, Agnes Libois, Laurens Liesenborghs, Beatriz Grinsztejn, Mauro Schechter, Alberto dos Santos de Lemos, Alvaro Furtado Costa, Si-

mone Queiroz Rocha, José Valdez Madruga, Darrell H. S. Tan, Sharmistha Mishra, Shreya Shah, Camila Jorquera, Alberto Castillo, Mauricio Carrión, Nelson Cevallos, Romain Palich, Valerie Pourcher, Emma Rubenstein, Pascal Migaud, Christoph Boesecke, Christian Hoffmann, Konstantinos Protopapas, Silvia Nozza, Anna Maria Cattelan, Cristina Mussini, Antonella d'Arminio Monforte, Raúl Adrian Cruz Flores, Edgar Pérez Barragán, Alma Leticia Rodríguez Guzmán, Dimie Ogoina, Nneka Marian Chika-Igwenyi, Onyeaghala Chizaram, Jenny Valverde López, Angelica García Tello, Maria Ubals, Martí Vall, Adrià Mendoza, Clara Suñer, Bonaventura Clotet, Jordi Bechini, Jose A. Lepe, M. Dolores Navarro-Amuedo, Jose Ignacio Bernadino, Alba Català, Eloy José Tarín Vicente, Borja González Rodríguez, Sergi Rodriguez-Mercader, Francisca Sánchez-Martinez, Esperanza Cañas-Ruano, Laura Parra-Navarro, Finn Filén, Carmen Tallón de Lara, Dominique Braun, Vanja Piezzi, Michael Burkhard, Helen Kovari, Anja Mönch, Jake Dunning, Pedro Simoes, Achyuta Nori, Sarah Keegan, John P. Thornhill, Vanessa Apea, Teymur Noori, Joyce L. Jones, Seth Judson, Elizabeth A. Gilliams, Matthew M. Hammill, Jeanne Keruly, Andrés F. Henao Martínez, Aung Lin, Jessica So, Kusha Davar, Diana Villareal, and Miguel Tapia Paredes. Mpox in people with advanced HIV infection: a global case series. *The Lancet*, 401(10380):939–949, March 2023. Publisher: Elsevier.

[107] Benoit Morel, Pierre Barbera, Lucas Czech, Ben Bettisworth, Lukas Hübner, Sarah Lutteropp, Dora Serdari, Evangelia-Georgia Kostaki, Ioannis Mamais, Alexey M Kozlov, Pavlos Pavlidis, Dimitrios Paraskevis, and Alexandros Stamatakis. Phylogenetic Analysis of SARS-CoV-2 Data Is Difficult. *Molecular Biology and Evolution*, 38(5):1777–1791, May 2021.

[108] Gage K. Moreno, Katarina M. Braun, Kasen K. Riemersma, Michael A. Martin, Peter J. Halfmann, Chelsea M. Crooks, Trent Prall, David Baker, John J. Baczenas, Anna S. Heffron, Mitchell Ramuta, Manjeet Khubbar, Andrea M. Weiler, Molly A. Accola, William M. Rehrauer, Shelby L. O'Connor, Nasia Safdar, Caitlin S. Pepperell, Trivikram Dasu, Sanjib Bhattacharyya, Yoshihiro Kawaoka, Katia Koelle, David H. O'Connor, and Thomas C. Friedrich. Revealing fine-scale spatiotemporal differences in SARS-CoV-2 introduction and spread. *Nature Communications*, 11(1):5558, November 2020. Number: 1 Publisher: Nature Publishing Group.

[109] Julian W. März, Søren Holm, and Nikola Biller-Andorno. Monkeypox, stigma and public health. *The Lancet Regional Health – Europe*, 23, December 2022. Publisher: Elsevier.

[110] Nicola F. Müller and Remco R. Bouckaert. Adaptive Metropolis-coupled MCMC for BEAST 2. *PeerJ*, 8:e9473, 2020.

[111] Nicola F Müller, Gytis Dudas, and Tanja Stadler. Inferring time-dependent migration and coalescence patterns from genetic sequence and predictor data in structured populations. *Virus Evolution*, 5(2):vez030, July 2019.

[112] Nicola F Müller, David Rasmussen, and Tanja Stadler. MASCOT: parameter and state inference under the marginal structured coalescent approximation. *Bioinformatics*, 34(22):3843–3848, November 2018.

[113] Nicola F. Müller, David A. Rasmussen, and Tanja Stadler. The Structured Coalescent and Its Approximations. *Molecular Biology and Evolution*, 34(11):2970–2981, November 2017.

[114] Nicola F. Müller, Cassia Wagner, Chris D. Frazar, Pavitra Roychoudhury, Jover Lee, Louise H. Moncla, Benjamin Pelle, Matthew Richardson, Erica Ryke, Hong Xie, Lasata Shrestha, Amin Addetia, Victoria M. Rachleff, Nicole A. P. Lieberman, Meei-Li Huang, Romesh Gautom, Geoff Melly, Brian Hiatt, Philip Dykema, Amanda Adler, Elisabeth Brandstetter, Peter D. Han, Kairsten Fay, Misja Ilcisin, Kirsten Lacombe, Thomas R. Sibley, Melissa Truong, Caitlin R. Wolf, Michael Boeckh, Janet A. Englund, Michael Famulare, Barry R. Lutz, Mark J. Rieder, Matthew Thompson, Jeffrey S. Duchin, Lea M. Starita, Helen Y. Chu, Jay Shendure, Keith R. Jerome, Scott Lindquist, Alexander L. Greninger, Deborah A. Nickerson, and Trevor Bedford. Viral genomes reveal patterns of the SARS-CoV-2 outbreak in Washington State. *Science Translational Medicine*, 13(595):eabf0202, May 2021. Publisher: American Association for the Advancement of Science.

[115] Nicola F. Müller, Daniel Wüthrich, Nina Goldman, Nadine Sailer, Claudia Saalfrank, Myrta Brunner, Noémi Augustin, Helena MB Seth-Smith, Yvonne Hollenstein, Mohammedyaseen Syedbasha, Daniela Lang, Richard A. Neher, Olivier Dubuis, Michael Naegele, Andreas Buser, Christian H. Nickel, Nicole Ritz, Andreas Zeller, Brian M. Lang, James Hadfield, Trevor Bedford, Manuel Battegay, Rita Schneider-Sliwa, Adrian Egli, and Tanja Stadler. Characterising the epidemic spread of influenza A/H3N2 within a city through phylogenetics. *PLOS Pathogens*, 16(11):e1008984, November 2020. Publisher: Public Library of Science.

[116] Hanna N. Oltean, Krisandra J. Allen, Lauren Frisbie, Stephanie M. Lunn, Laura Marcela Torres, Lillian Manahan, Ian Painter, Denny Russell, Avi Singh, JohnAric MoonDance Peterson, Kristin Grant, Cara Peter, Rebecca Cao, Katelynn Garcia, Drew Mackellar, Lisa Jones, Holly Halstead, Hannah Gray, Geoff Melly, Deborah Nickerson, Lea Starita, Chris Frazar, Alexander L. Greninger, Pavitra Roychoudhury, Patrick C. Mathias, Michael H. Kalnoski, Chao-Nan Ting, Marisa Lykken, Tana

Rice, Daniel Gonzalez-Robles, David Bina, Kelly Johnson, Carmen L. Wiley, Shaun C. Magnuson, Christopher M. Parsons, Eugene D. Chapman, C. Alexander Valencia, Ryan R. Fortna, Gregory Wolgamot, James P. Hughes, Janet G. Baseman, Trevor Bedford, and Scott Lindquist. Sentinel Surveillance System Implementation and Evaluation for SARS-CoV-2 Genomic Data, Washington, USA, 2020-2021. *Emerging Infectious Diseases*, 29(2):242–251, February 2023.

[117] Sean Wei Xiang Ong, Calvin J. Chiew, Li Wei Ang, Tze-Minn Mak, Lin Cui, Matthias Paul HS Toh, Yi Ding Lim, Pei Hua Lee, Tau Hong Lee, Po Ying Chia, Sebastian Maurer-Stroh, Raymond Tzer Pin Lin, Yee-Sin Leo, Vernon J. Lee, David Chien Lye, and Barnaby E. Young. Clinical and Virological Features of SARS-CoV-2 Variants of Concern: A Retrospective Cohort Study Comparing B.1.1.7 (Alpha), B.1.315 (Beta), and B.1.617.2 (Delta). SSRN Scholarly Paper ID 3861566, Social Science Research Network, Rochester, NY, June 2021.

[118] Ontario Agency for Health Protection and Promotion (Public Health Ontario). Mpox immunization and post-immunization cases in Ontario, March 2023.

[119] Sharon Otterman. Last Call for the Monkeypox Vaccine Van. *The New York Times*, November 2022.

[120] Áine O'Toole, Richard A. Neher, Nnaemeka Ndodo, Vitor Borges, Ben Gannon, João Paulo Gomes, Natalie Groves, David J. King, Daniel Maloney, Philippe Lemey, Kuiama Lewandowski, Nicholas Loman, Richard Myers, Ifeanyi F. Omah, Marc A. Suchard, Michael Worobey, Meera Chand, Chikwe Ihekweazu, David Ulaeto, Ifedayo Adetifa, and Andrew Rambaut. APOBEC3 deaminase editing in mpox virus as evidence for sustained human transmission since at least 2016. *Science*, 382(6670):595–600, November 2023. Publisher: American Association for the Advancement of Science.

[121] Andrew J. Page, Ben Taylor, Aidan J. Delaney, Jorge Soares, Torsten Seemann, Jacqueline A. Keane, and Simon R. Harris. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microbial Genomics*, 2(4):e000056, April 2016.

[122] Miguel I. Paredes, Nashwa Ahmed, Marlin Figgins, Vittoria Colizza, Philippe Lemey, John T. McCrone, Nicola Müller, Cécile Tran-Kiem, and Trevor Bedford. Underdetected dispersal and extensive local transmission drove the 2022 mpox epidemic. *Cell*, 187(6):1374–1386.e13, March 2024. Publisher: Elsevier.

[123] Miguel I. Paredes, Stephanie M. Lunn, Michael Famulare, Lauren A. Frisbie, Ian Painter, Roy Burstein, Pavitra Roychoudhury, Hong Xie, Shah A. Mohamed Bakhash,

Ricardo Perez, Maria Lukes, Sean Ellis, Saraswathi Sathees, Patrick C. Mathias, Alexander Greninger, Lea M. Starita, Chris D. Frazar, Erica Ryke, Weizhi Zhong, Luis Gamboa, Machiko Threlkeld, Jover Lee, Evan McDermot, Melissa Truong, Deborah A. Nickerson, Daniel L. Bates, Matthew E. Hartman, Eric Haugen, Truong N. Nguyen, Joshua D. Richards, Jacob L. Rodriguez, John A. Stamatoyannopoulos, Eric Thorland, Geoff Melly, Philip E. Dykema, Drew C. MacKellar, Hannah K. Gray, Avi Singh, JohnAric M. Peterson, Denny Russell, Laura Marcela Torres, Scott Lindquist, Trevor Bedford, Krisandra J. Allen, and Hanna N. Oltean. Associations between SARS-CoV-2 variants and risk of COVID-19 hospitalization among confirmed cases in Washington State: a retrospective cohort study. *Clinical Infectious Diseases*, page ciac279, April 2022.

[124] Miguel I. Paredes, Amanda C. Perofsky, Lauren Frisbie, Louise H. Moncla, Pavitra Roychoudhury, Hong Xie, Shah A. Mohamed Bakhash, Kevin Kong, Isabel Arnould, Tien V. Nguyen, Seffir T. Wendm, Pooneh Hajian, Sean Ellis, Patrick C. Mathias, Alexander L. Greninger, Lea M. Starita, Chris D. Frazar, Erica Ryke, Weizhi Zhong, Luis Gamboa, Machiko Threlkeld, Jover Lee, Jeremy Stone, Evan McDermot, Melissa Truong, Jay Shendure, Hanna N. Oltean, Cécile Viboud, Helen Chu, Nicola F. Müller, and Trevor Bedford. Local-scale phylodynamics reveal differential community impact of SARS-CoV-2 in a metropolitan US county. *PLOS Pathogens*, 20(3):e1012117, March 2024. Publisher: Public Library of Science.

[125] Luz H. Patiño, Susana Guerra, Marina Muñoz, Nicolas Luna, Keith Farrugia, Adriana van de Guchte, Zain Khalil, Ana Silvia Gonzalez-Reiche, Matthew M. Hernandez, Radhika Banu, Paras Shrestha, Bernadette Liggayu, Adolfo Firpo Betancourt, David Reich, Carlos Cordon-Cardo, Randy Albrecht, Rebecca Pearl, Viviana Simon, Aria Rooker, Emilia Mia Sordillo, Harm van Bakel, Adolfo García-Sastre, Dusan Bogunovic, Gustavo Palacios, Alberto Paniz Mondolfi, and Juan David Ramírez. Phylogenetic landscape of Monkeypox Virus (MPV) during the early outbreak in New York City, 2022. *Emerging Microbes & Infections*, 12(1):e2192830, December 2023. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/22221751.2023.2192830.

[126] Prabasaj Paul. Genomic Surveillance for SARS-CoV-2 Variants Circulating in the United States, December 2020–May 2021. *MMWR. Morbidity and Mortality Weekly Report*, 70, 2021.

[127] Pearson, Carl, Russell, Timothy W, Davies, Nicholas G., Kucharski, Adam J, CMMID COVID-19 working group, Edmunds, W John, and Eggo, Rosalind M. Estimates of severity and transmissibility of novel SARS-CoV-2 variant 501Y.V2 in South Africa, January 2021.

[128] Kayla M. Peck and Adam S. Lauring. Complexities of Viral Mutation Rates. *Journal of Virology*, 92(14):10.1128/jvi.01031–17, June 2018. Publisher: American Society for Microbiology.

[129] André Peralta-Santos, Eduardo Freire Rodrigues, Joana Moreno, Vasco Ricoca, Pedro Casaca, Eugenia Fernandes, João Paulo Gomes, Rita Ferreira, Joana Isidro, Miguel Pinto, Vítor Borges, Luís Vieira, Silvia Duarte, Carlos Sousa, José Pedro Almeida, Luís Menezes, Bibiana I. Ferreira, Ana Matias, Ana Pelerito, Samanta Freire, Teresa Grilo, Cláudia Medeiros Borges, Vera Moutinho, Andreia Leite, Irina Kislaya, Ana Paula Rodrigues, Pedro Pinto Leite, and Baltazar Nunes. Omicron (BA.1) SARS-CoV-2 Variant Is Associated With Reduced Risk of Hospitalization and Length of Stay Compared With Delta (B.1.617.2). SSRN Scholarly Paper ID 4017381, Social Science Research Network, Rochester, NY, January 2022.

[130] Eskild Petersen, Anu Kantele, Marion Koopmans, Danny Asogun, Adesola Yinka-Ogunleye, Chikwe Ihekweazu, and Alimuddin Zumla. Human Monkeypox: Epidemiologic and Clinical Characteristics, Diagnosis, and Prevention. *Infectious Disease Clinics of North America*, 33(4):1027–1043, December 2019.

[131] Virginia E Pitzer, Melanie Chitwood, Joshua Havumaki, Nicolas A Menzies, Stephanie Perniciaro, Joshua L Warren, Daniel M Weinberger, and Ted Cohen. The Impact of Changes in Diagnostic Testing Practices on Estimates of COVID-19 Transmission in the United States. *American Journal of Epidemiology*, 190(9):1908–1917, September 2021.

[132] Christopher Polk, Mindy Sampson, Robert T. Fairman, Michael E. DeWitt, Michael Leonard, Anupama Neelakanta, Lisa Davidson, Danya Roshdy, Chris Branner, Lewis McCurdy, Tom Ludden, Hazel Tapp, and Catherine Passaretti. Evaluation of a health system's implementation of a monkeypox care model under the RE-AIM framework. *Therapeutic Advances in Infectious Disease*, 10:20499361231158463, March 2023.

[133] Public Health England. SARS-CoV-2 variants of concern and variants under investigation. page 77, June 2021.

[134] Sang Hyun Ra, Joon Seo Lim, Gwang-un Kim, Min Jae Kim, Jiwon Jung, and Sung-Han Kim. Upper respiratory viral load in asymptomatic individuals and mildly symptomatic patients with SARS-CoV-2 infection. *Thorax*, 76(1):61–63, January 2021. Publisher: BMJ Publishing Group Ltd Section: Respiratory infection.

[135] J Raifman, K Nocka, D Jones, J Bor, S Lipson, J Jay, and P Chan. COVID-19 US state policy database, 2020.

[136] Andrew Rambaut, Alexei J Drummond, Dong Xie, Guy Baele, and Marc A Suchard. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology*, 67(5):901–904, September 2018.

[137] Aspen P. Riser. Epidemiologic and Clinical Features of Mpox-Associated Deaths — United States, May 10, 2022–March 7, 2023. *MMWR. Morbidity and Mortality Weekly Report*, 72, 2023.

[138] Stefan Rothenburg, Zhilong Yang, Pip Beard, Sara L. Sawyer, Boghuma Titanji, Gregg Gonsalves, and Jason Kindrachuk. Monkeypox emergency: Urgent questions and perspectives. *Cell*, 185(18):3279–3281, September 2022. Publisher: Elsevier.

[139] Pavel Sagulenko, Vadim Puller, and Richard A Neher. TreeTime: Maximum-likelihood phylodynamic analysis. *Virus Evolution*, 4(1):vex042, January 2018.

[140] Rafael Sanjuán. From Molecular Genetics to Phylodynamics: Evolutionary Relevance of Mutation Rates Across Viruses. *PLOS Pathogens*, 8(5):e1002685, May 2012. Publisher: Public Library of Science.

[141] Edmund Seto, Esther Min, Carolyn Ingram, BJ Cummings, and Stephanie A. Farquhar. Community-Level Factors Associated with COVID-19 Cases and Testing Equity in King County, Washington. *International Journal of Environmental Research and Public Health*, 17(24):9516, December 2020.

[142] Aziz Sheikh, Jim McMenamin, Bob Taylor, and Chris Robertson. SARS-CoV-2 Delta VOC in Scotland: demographics, risk of hospital admission, and vaccine effectiveness. *The Lancet*, 397(10293):2461–2462, June 2021. Publisher: Elsevier.

[143] Yuelong Shu and John McCauley. GISAID: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance*, 22(13):30494, March 2017.

[144] Marc A Suchard, Philippe Lemey, Guy Baele, Daniel L Ayres, Alexei J Drummond, and Andrew Rambaut. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution*, 4(1):vey016, January 2018.

[145] Karla Therese L Sy, Micaela E Martinez, Benjamin Rader, and Laura F White. Socioeconomic Disparities in Subway Use and COVID-19 Outcomes in New York City. *American Journal of Epidemiology*, 190(7):1234–1242, July 2021.

[146] John P. Thornhill, Sapha Barkati, Sharon Walmsley, Juergen Rockstroh, Andrea Antinori, Luke B. Harrison, Romain Palich, Achyuta Nori, Iain Reeves, Maximillian S. Habibi, Vanessa Apea, Christoph Boesecke, Linos Vandekerckhove, Michal Yakubovsky, Elena Sendagorta, Jose L. Blanco, Eric Florence, Davide Moschese, Fernando M. Maltez, Abraham Goorhuis, Valerie Pourcher, Pascal Migaud, Sebastian Noe, Claire Pintado, Fabrizio Maggi, Ann-Brit E. Hansen, Christian Hoffmann, Jezer I. Lezama, Cristina Mussini, AnnaMaria Cattelan, Keletso Makofane, Darrell Tan, Silvia Nozza, Johannes Nemeth, Marina B. Klein, and Chloe M. Orkin. Monkeypox Virus Infection in Humans across 16 Countries — April–June 2022. *New England Journal of Medicine*, 387(8):679–691, August 2022. Publisher: Massachusetts Medical Society.

[147] Diana M. Tordoff, Alexander L. Greninger, Pavitra Roychoudhury, Lasata Shrestha, Hong Xie, Keith R. Jerome, Nathan Breit, Meei-Li Huang, Mike Famulare, and Joshua T. Herbeck. Phylogenetic estimates of SARS-CoV-2 introductions into Washington State. *The Lancet Regional Health – Americas*, 1, September 2021. Publisher: Elsevier.

[148] Cécile Tran-Kiem and Trevor Bedford. Estimating the reproduction number and transmission heterogeneity from the size distribution of clusters of identical pathogen sequences. preprint, Epidemiology, April 2023.

[149] Katherine A. Twohig, Tommy Nyberg, Asad Zaidi, Simon Thelwall, Mary A. Sinnathamby, Shirin Aliabadi, Shaun R. Seaman, Ross J. Harris, Russell Hope, Jamie Lopez-Bernal, Eileen Gallagher, Andre Charlett, Daniela De Angelis, Anne M. Presanis, Gavin Dabrera, Cherian Koshy, Amy Ash, Emma Wise, Nathan Moore, Matilde Mori, Nick Cortes, Jessica Lynch, Stephen Kidd, Derek Fairley, Tanya Curran, James McKenna, Helen Adams, Christophe Fraser, Tanya Golubchik, David Bonsall, Mohammed Hassan-Ibrahim, Cassandra Malone, Benjamin Cogger, Michelle Wantoch, Nicola Reynolds, Ben Warne, Joshua Maksimovic, Karla Spellman, Kathryn McCluggage, Michaela John, Robert Beer, Safiah Afifi, Sian Morgan, Angela Marchbank, Anna Price, Christine Kitchen, Huw Gulliver, Ian Merrick, Joel Southgate, Martyn Guest, Robert Munn, Trudy Workman, Thomas Connor, William Fuller, Catherine Bresner, Luke Snell, Amita Patel, Themoula Charalampous, Gaia Nebbia, Rahul Batra, Jonathan Edgeworth, Samuel Robson, Angela Beckett, David Aanensen, Anthony Underwood, Corin Yeats, Khalil Abudahab, Ben Taylor, Mirko Menegazzo, Gemma Clark, Wendy Smith, Manjinder Khakh, Vicki Fleming, Michelle Lister, Hannah Howson-Wells, Louise Berry, Tim Boswell, Amelia Joseph, Iona Willingham, Carl Jones, Christopher Holmes, Paul Bird, Thomas Helmer, Karlie Fallon, Julian Tang, Veena Raviprakash, Sharon Campbell, Nicola Sheriff, Victoria Blakey, Lesley-Anne

Williams, Matthew Loose, Nadine Holmes, Christopher Moore, Matthew Carlile, Victoria Wright, Fei Sang, Johnny Debebe, Francesc Coll, Adrian Signell, Gilberto Betancor, Harry Wilson, Sahar Eldirdiri, Anita Kenyon, Thomas Davis, Oliver Pybus, Louis du Plessis, Alex Zarebski, Jayna Raghwani, Moritz Kraemer, Sarah Francois, Stephen Attwood, Tetyana Vasylyeva, Marina Escalera Zamudio, Bernardo Gutierrez, M. Estee Torok, William Hamilton, Ian Goodfellow, Grant Hall, Aminu Jahun, Yasmin Chaudhry, Myra Hosmillo, Malte Pinckert, Iliana Georgana, Samuel Moses, Hannah Lowe, Luke Bedford, Jonathan Moore, Susanne Stonehouse, Chloe Fisher, Ali Awan, John BoYes, Judith Breuer, Kathryn Harris, Julianne Brown, Divya Shah, Laura Atkinson, Jack Lee, Nathaniel Storey, Flavia Flaviani, Adela Alcolea-Medina, Rebecca Williams, Gabrielle Vernet, Michael Chapman, Lisa Levett, Judith Heaney, Wendy Chatterton, Monika Pusok, Li Xu-McCrae, Darren Smith, Matthew Bashton, Gregory Young, Alison Holmes, Paul Randell, Alison Cox, Pinglawathee Madona, Frances Bolt, James Price, Siddharth Mookerjee, Manon Ragonnet-Cronin, Fabricia F. Nascimento, David Jorgensen, Igor Siveroni, Rob Johnson, Olivia Boyd, Lily Geidelberg, Erik Volz, Aileen Rowan, Graham Taylor, Katherine Smollett, Nicholas Loman, Joshua Quick, Claire McMurray, Joanne Stockton, Sam Nicholls, Will Rowe, Radoslaw Poplawski, Alan McNally, Rocio Martinez Nunez, Jenifer Mason, Trevor Robinson, Elaine O'Toole, Joanne Watts, Cassie Breen, Angela Cowell, Graciela Sluga, Nicholas Machin, Shazaad Ahmad, Ryan George, Fenella Halstead, Venkat Sivaprakasam, Wendy Hogsden, Chris Illingworth, Chris Jackson, Emma Thomson, James Shepherd, Patawee Asamaphan, Marc Niebel, Kathy Li, Rajiv Shah, Natasha Jesudason, Lily Tong, Alice Broos, Daniel Mair, Jenna Nichols, Stephen Carmichael, Kyriaki Nomikou, Elihu Aranday-Cortes, Natasha Johnson, Igor Starinskij, Ana da Silva Filipe, David Robertson, Richard Orton, Joseph Hughes, Sreenu Vattipally, Joshua Singer, Seema Nickbakhsh, Antony Hale, Louissa Macfarlane-Smith, Katherine Harper, Holli Carden, Yusri Taha, Brendan Payne, Shirelle Burton-Fanning, Sheila Waugh, Jennifer Collins, Gary Eltringham, Steven Rushton, Sarah O'Brien, Amanda Bradley, Alasdair Maclean, Guy Mollett, Rachel Blacow, Kate Templeton, Martin McHugh, Rebecca Dewar, Elizabeth Wastenge, Samir Dervisevic, Rachael Stanley, Emma Meader, Lindsay Coupland, Louise Smith, Clive Graham, Edward Barton, Debra Padgett, Garren Scott, Emma Swindells, Jane Greenaway, Andrew Nelson, Clare McCann, Wen Yew, Monique Andersson, Timothy Peto, Anita Justice, David Eyre, Derrick Crook, Tim Sloan, Nichola Duckworth, Sarah Walsh, Anoop Chauhan, Sharon Glaysher, Kelly Bicknell, Sarah Wyllie, Scott Elliott, Allyson Lloyd, Robert Impey, Nick Levene, Lynn Monaghan, Declan Bradley, Tim Wyatt, Elias Allara, Clare Pearson, Husam Osman, Andrew Bosworth, Esther Robinson, Peter Muir, Ian Vipond, Richard Hopes, Hannah Pymont, Stephanie Hutchings, Martin Curran, Surendra Parmar, Angie Lackenby, Tamyo Mbisa, Steven Platt, Shahjahan Miah, David Bibby, Carmen Manso, Jonathan

Hubb, Meera Chand, Gavin Dabrera, Mary Ramsay, Daniel Bradshaw, Alicia Thornton, Richard Myers, Ulf Schaefer, Natalie Groves, Eileen Gallagher, David Lee, David Williams, Nicholas Ellaby, Ian Harrison, Hassan Hartman, Nikos Manesis, Vineet Patel, Chloe Bishop, Vicki Chalker, Juan Ledesma, Katherine Twohig, Matthew Holden, Sharif Shaaban, Alec Birchley, Alexander Adams, Alisha Davies, Amy Gaskin, Amy Plimmer, Bree Gatica-Wilcox, Caoimhe McKerr, Catherine Moore, Chris Williams, David Heyburn, Elen De Lacy, Ember Hilvers, Fatima Downing, Giri Shankar, Hannah Jones, Hibo Asad, Jason Coombes, Joanne Watkins, Johnathan Evans, Laia Fina, Laura Gifford, Lauren Gilbert, Lee Graham, Malorie Perry, Mari Morgan, Matthew Bull, Michelle Cronin, Nicole Pacchiarini, Noel Craine, Rachel Jones, Robin Howe, Sally Corden, Sara Rey, Sara Kumziene-SummerhaYes, Sarah Taylor, Simon Cottrell, Sophie Jones, Sue Edwards, Justin O'Grady, Andrew Page, Alison Mather, David Baker, Steven Rudder, Alp Aydin, Gemma Kay, Alexander Trotter, Nabil-Fareed Alikhan, Leonardo de Oliveira Martins, Thanh Le-Viet, Lizzie Meadows, Anna Casey, Liz Ratcliffe, David Simpson, Zoltan Molnar, Thomas Thompson, Erwan Acheson, Jane Masoli, Bridget Knight, Sian Ellard, Cressida Auckland, Christopher Jones, Tabitha Mahungu, Dianne Irish-Tavares, Tanzina Haque, Jennifer Hart, Eric Witele, Melisa Fenton, Ashok Dadrah, Amanda Symmonds, Tranprit Saluja, Yann Bourgeois, Garry Scarlett, Katie Loveson, Salman Goudarzi, Christopher Fearn, Kate Cook, Hannah Dent, Hannah Paul, David Partridge, Mohammad Raza, Cariad Evans, Kate Johnson, Steven Liggett, Paul Baker, Stephen Bonner, Sarah Essex, Ronan Lyons, Kordo Saeed, Adhyana Mahanama, Buddhini Samaraweera, Siona Silveira, Emanuela Pelosi, Eleri Wilson-Davies, Rachel Williams, Mark Kristiansen, Sunando Roy, Charlotte Williams, Marius Cotic, Nadua Bayzid, Adam Westhorpe, John Hartley, Riaz Jannoo, Helen Lowe, Angeliki Karamani, Leah Ensell, Jacqui Prieto, Sarah Jeremiah, Dimitris Grammatopoulos, Sarojini Pandey, Lisa Berry, Katie Jones, Alex Richter, Andrew Beggs, Angus Best, Benita Percival, Jeremy Mirza, Oliver Megram, Megan Mayhew, Liam Crawford, Fiona Ashcroft, Emma Moles-Garcia, Nicola Cumley, Colin Smith, Giselda Bucca, Andrew Hesketh, Beth Blane, Sophia Girgis, Danielle Leek, Sushmita Sridhar, Sally Forrest, Claire Cormie, Harmeet Gill, Joana Dias, Ellen Higginson, Mailis Maes, Jamie Young, Leanne Kermack, Ravi Gupta, Catherine Ludden, Sharon Peacock, Sophie Palmer, Carol Churcher, Nazreen Hadjirin, Alessandro Carabelli, Ellena Brooks, Kim Smith, Katerina Galai, Georgina McManus, Chris Ruis, Rose Davidson, Andrew Rambaut, Thomas Williams, Carlos Balcazar, Michael Gallagher, Áine O'Toole, Stefan Rooke, Verity Hill, Kathleen Williamson, Thomas Stanton, Stephen Michell, Claire Bewshea, Ben Temperton, Michelle Michelsen, Joanna Warwick-Dugdale, Robin Manley, Audrey Farbos, James Harrison, Christine Sambles, David Studholme, Aaron Jeffries, Leigh Jackson, Alistair Darby, Julian Hiscox, Steve Paterson, Miren Iturriza-Gomara, Kathryn Jackson, Anita Lucaci, Edith

Vamos, Margaret Hughes, Lucille Rainbow, Richard Eccles, Charlotte Nelson, Mark Whitehead, Lance Turtle, Sam Haldenby, Richard Gregory, Matthew Gemmell, Claudia Wierzbicki, Hermione Webster, Thushan de Silva, Nikki Smith, Adrienn Angyal, Benjamin Lindsey, Danielle Groves, Luke Green, Dennis Wang, Timothy Freeman, Matthew Parker, Alexander Keeley, Paul Parsons, Rachel Tucker, Rebecca Brown, Matthew Wyles, Max Whiteley, Peijun Zhang, Marta Gallis, Stavroula Louka, Chrystala Constantinidou, Meera Unnikrishnan, Sascha Ott, Jeffrey Cheng, Hannah Bridgewater, Lucy Frost, Grace Taylor-Joyce, Richard Stark, Laura Baxter, Mohammad Alam, Paul Brown, Dinesh Aggarwal, Alberto Cerda, Tammy Merrill, Rebekah Wilson, Patrick McClure, Joseph Chappell, Theocharis Tsoleridis, Jonathan Ball, David Buck, John Todd, Angie Green, Amy Trebes, George MacIntyre-Cockett, Mariateresa de Cesare, Alex Alderton, Roberto Amato, Cristina Ariani, Mathew Beale, Charlotte Beaver, Katherine Bellis, Emma Betteridge, James Bonfield, John Danesh, Matthew Dorman, Eleanor Drury, Ben Farr, Luke Foulser, Sonia Goncalves, Scott Goodwin, Marina Gourtovaia, Ewan Harrison, David Jackson, Dorota Jamrozy, Ian Johnston, Leanne Kane, Sally Kay, Jon-Paul Keatley, Dominic Kwiatkowski, Cordelia Langford, Mara Lawniczak, Laura Letchford, Rich Livett, Stephanie Lo, Inigo Martincorena, Samantha McGuigan, Rachel Nelson, Steve Palmer, Naomi Park, Minal Patel, Liam Prestwood, Christoph Puethe, Michael Quail, Shavanthi Rajatileka, Carol Scott, Lesley Shirley, John Sillitoe, Michael Spencer Chapman, Scott Thurston, Gerry Tonkin-Hill, Danni Weldon, Diana Rajan, Iraad Bronner, Louise Aigrain, Nicholas Redshaw, Stefanie Lensing, Robert Davies, Andrew Whitwham, Jennifier Liddle, Kevin Lewis, Jaime Tovar-Corona, Steven Leonard, Jillian Durham, Andrew Bassett, Shane McCarthy, Robin Moll, Keith James, Karen Oliver, Alex Makunin, Jeff Barrett, and Rory Gunson. Hospital admission and emergency care attendance risk for SARS-CoV-2 delta (B.1.617.2) compared with alpha (B.1.1.7) variants of concern: a cohort study. *The Lancet Infectious Diseases*, 0(0), August 2021. Publisher: Elsevier.

[150] US Census Bureau. 2010_census_tract_to_2010_puma.

[151] Jacob VanderPlas, Brian Granger, Jeffrey Heer, Dominik Moritz, Kanit Wongsuphasawat, Arvind Satyanarayan, Eitan Lees, Ilia Timofeev, Ben Welsh, and Scott Sievert. Altair: Interactive statistical visualizations for python. *Journal of open source software*, 3(32):1057, 2018.

[152] Timothy G Vaughan. IcyTree: rapid browser-based visualization for phylogenetic trees and networks. *Bioinformatics*, 33(15):2392–2394, August 2017.

[153] Timothy G. Vaughan, Denise Kühnert, Alex Popinga, David Welch, and Alexei J.

Drummond. Efficient Bayesian inference under the structured coalescent. *Bioinformatics*, 30(16):2272–2279, August 2014.

[154] Erik M. Volz, Katia Koelle, and Trevor Bedford. Viral Phylodynamics. *PLOS Computational Biology*, 9(3):e1002947, March 2013. Publisher: Public Library of Science.

[155] Erik M. Volz, Sergei L. Kosakovsky Pond, Melissa J. Ward, Andrew J. Leigh Brown, and Simon D. W. Frost. Phylodynamics of infectious disease epidemics. *Genetics*, 183(4):1421–1430, December 2009.

[156] John Wakely. *Coalescent Theory: An Introduction*. Macmillan Learning, April 2016. Google-Books-ID: x30RAgAACAAJ.

[157] J Wallinga and M Lipsitch. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences*, 274(1609):599–604, February 2007.

[158] Thomas Ward, Rachel Christie, Robert S. Paton, Fergus Cumming, and Christopher E. Overton. Transmission dynamics of monkeypox in the United Kingdom: contact tracing study. *BMJ*, 379:e073153, November 2022. Publisher: British Medical Journal Publishing Group Section: Research.

[159] Washington State Department of Health. First cases of U.K. COVID-19 strain found in Washington state, January 2021.

[160] Washington State Department of Health. SARS-CoV-2 Sequencing and Variants in Washington State. page 27, July 2021.

[161] Joakim A. Weill, Matthieu Stigler, Olivier Deschenes, and Michael R. Springborn. Social distancing responses to COVID-19 emergency declarations strongly differentiated by income. *Proceedings of the National Academy of Sciences*, 117(33):19658–19660, August 2020. Publisher: Proceedings of the National Academy of Sciences.

[162] Lilith K. Whittles, Peter J. White, and Xavier Didelot. A dynamic power-law sexual network model of gonorrhoea outbreaks. *PLOS Computational Biology*, 15(3):e1006748, March 2019. Publisher: Public Library of Science.

[163] Yael Wolff Sagy, Roy Zucker, Ariel Hammerman, Hila Markovits, Noa Gur Arieh, Wiessam Abu Ahmad, Erez Battat, Noga Ramot, Guy Carmeli, Avner Mark-Amir, Gal Wagner-Kolasko, Hadar Duskin-Bitan, Shlomit Yaron, Alon Peretz, Ronen Arbel,

Gil Lavie, and Doron Netzer. Real-world effectiveness of a single dose of mpox vaccine in males. *Nature Medicine*, 29(3):748–752, March 2023. Number: 3 Publisher: Nature Publishing Group.

[164] Shiting Yang, Xiaohao Guo, Zeyu Zhao, Buasiyamu Abudunaibi, Yunkang Zhao, Jia Rui, Yao Wang, Wentao Song, Hongjie Wei, and Tianmu Chen. Possibility of mpox viral transmission and control from high-risk to the general population: a modeling study. *BMC Infectious Diseases*, 23(1):119, February 2023.

[165] Ziheng Yang. *Computational Molecular Evolution*. Oxford Series in Ecology and Evolution. University Press, UK, Oxford, 2006.

[166] Xu-Sheng Zhang, Sema Mandal, Hamish Mohammed, Charlie Turner, Isaac Florence, Josephine Walker, Siwaporn Niyomsri, Gayatri Amirthalingam, Mary Ramsay, Andre Charlett, and Peter Vickerman. Transmission dynamics and effect of control measures on the 2022 outbreak of mpox among gay, bisexual, and other men who have sex with men in England: a mathematical modelling study. *The Lancet Infectious Diseases*, 24(1):65–74, January 2024. Publisher: Elsevier.