MERS structure

Gytis Dudas[1], Luiz Max Carvalho[2], Allison Black[1], Trevor Bedford[1] & Andrew Rambaut[2,4,5]

[1]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA, [2]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK, [4]Fogarty International Center, National Institutes of Health, Bethesda, MD, USA, [5]Centre for Immunology, Infection and Evolution at the University of Edinburgh, Edinburgh, UK

March 3, 2017

Abstract

Middle East Respiratory Syndrome coronavirus (MERS-CoV) is a zoonotic virus causing significant mortality and morbidity largely in the Arabian Peninsula. The epidemiology of the virus is relatively poorly understood with most MERS patients being older males with comorbidities, although large nosocomial have occurred, notably in Jeddah, Riyadh and South Korea. Our understanding of MERS-CoV epidemiology is further complicated by some patients not recalling contact with camels, the accepted reservoir for MERS-CoV, or any other livestock, suggesting significant cryptic MERS-CoV circulation in communities. Seroepidemiology has been employed extensively during the time since the discovery of MERS-CoV to understand the exposure and therefore risk among camels and humans alike, but various sequencing efforts have not been fully leveraged against this zoonotic virus. Here we use existing MERS-CoV sequencing data to estimate the number of times the virus has been introduced into humans and use the distribution of sequence clusters to estimate the reproductive number for the virus from sequence data alone. We confirm that MERS-CoV has relatively poor capacity to spread in humans and arrive at an estimate of at least 65 zoonotic introductions of the virus into humans.

Introduction

Middle East Respiratory Syndrome Coronavirus (MERS-CoV) is a zoonotic infection of humans in the Arabian Peninsula originating from camels. The virus, first discovered in 2012 **?**, has gone on to cause more than 1800 infections with at least 600 associated deaths. Its epidemiology remains obscure, largely because outbreaks are observed among the most severely affected patients, such as older males with comorbidities. Whilst contact with camels is often reported by patients, many do not recall contact with any livestock, suggesting a significant but unobserved community contribution to the outbreak. Studies into MERS-CoV epidemiology often rely on serology to identify factors associated with MERS CoV exposure in potential risk groups or hosts. The new era of genomic epidemiology, however, has repeatedly shown the utility of sequence data in outbreak scenarios. Often only sequence data can pinpoint sources of pathogens and discriminate between multiple and single source scenarios, which are a fundamental part of quantifying risk. Sequencing MERS-CoV has been performed as part of initial attempts to link human infections with the camel reservoir, nosocomial outbreak investigations and routine surveillance.

It is accepted that human MERS-CoV infections are a result of multiple introductions of the virus into humans, but no study has attempted to go beyond educated guesses about this number. Here we use existing MERS-CoV sequence data to investigate the population structure of the virus between two of its known hosts: humans and camels. By explicitly modelling the evolution of MERS-CoV between these two distinct host populations we show that human MERS infections are the result of frequent and unidirectional spillover events from camels into humans. Inter-outbreak evolution of the virus occurs exclusively in camels, highlighting the need for genomic surveillance of MERS-CoV in the reservoir as well. We also use the sequence clusters recovered from our population structure analyses to estimate the reproductive number for the virus (R_0), which is in agreement with previous findings that MERS-CoV is, on average, poor at spreading in human populations. Our analyses also capture seasonal variation in zoonotic transmission patterns, with indications that particularly large outbreaks are the result of zoonotic transmissions that take place at the beginning of each new year.

Results

Cross species transmission of MERS-CoV

It is generally accepted that MERS-CoV is a continuously re-emerging human pathogen, however the exact number of cross-species transmission events remains unknown. We confirm this by identifying the camel reservoir as the host population where most of MERS-CoV evolution takes place (see figure 1). The human side of MERS outbreaks, on the other hand, are caused by repetitive zoonotic transmissions from camels, which we estimated by using an explicit model of population structure. We infer that across 216 MERS-CoV sequences collected from humans, we detect a median of 67 individual camel to human cross-species transmissions (95% highest posterior density: 60–72), with little support for back-transmission scenarios (median 3, 95% HPDs: 0–13). There is also

apparent heterogeneity in resulting human cluster sizes, which is not unexpected. Large clusters of MERS are apparent and mostly correspond to known large transmission chains, often in a nosocomial setting, such as the outbreak in South Korea and Riyadh. We also note a general predisposition for large MERS-CoV clusters to have crossed the species barrier into humans early in each year.
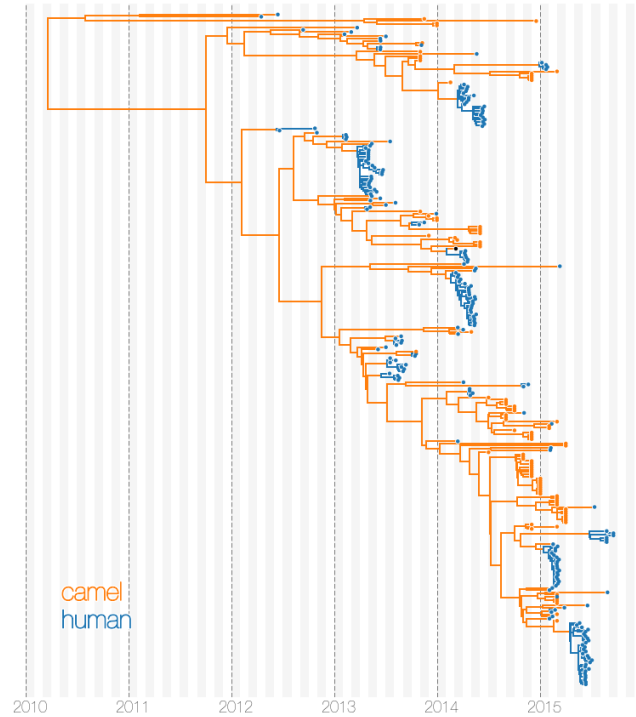

Figure 1. Typed maximum clade credibility tree of MERS-CoV sequences from humans and camels. Maximum clade credibility (MCC) tree showing inferred ancestral hosts for MERS-CoV recovered using the structured coalescent. Vast majority of MERS-CoV evolution is inferred to occur in camels (orange) with human outbreaks (blue) representing evolutionary dead-ends for the virus. Whilst large clusters of human cases are apparent in the tree, significant contributions to human outbreaks are made by singleton sequences, likely representing recent cross-species transmissions that were caught early.

Estimated epidemiological parameters

Removing the confounding effects of shared ancestry for MERS-CoV sequences has allowed us to treat every viral introduction into humans as an independent replicate of the transmission process in humans.

Discussion

Methods

Sequence data

All MERS-CoV sequences were downloaded from GenBank. Fragments of some strains submitted to GenBank as separate accessions were assembled into a single sequence. Protein coding sequences were extracted and concatenated. Sequences were annotated with available collection dates and hosts, designated as camel or human. For partitioned analyses the alignment was split into two fragments, the first containing all nucleotides up to nucleotide 21000, and the rest of the genome being assigned to the second fragment. Despite their divergence both Severe Acute Respiratory Syndrome (SARS) and MERS betacoronaviruses appear to recombine across position 21000, which may indicate the existence of a recombinogenic secondary RNA structure that remains to be described in its vicinity.

Structured coalescent analyses

MultiTypeTree module was used in BEAUti v2.4.3 to specify a structured coalescent model with two demes - humans and camels - based on GenBank records. Analyses were run on codon position partitioned data with two separate HKY+Γ_4 nucleotide substitution models specified for codon positions 1+2 and 3. A relaxed molecular clock with branch rates drawn from a lognormal distribution was used to infer the evolutionary rate from date calibrated tips. Default priors were used for all parameters except for migration rates between demes for which an exponential prior with mean 1.0 was used. An identical set up was used for analysing alignments split into two fragments, with independent clocks, trees and migration rates, but shared substitution models and deme population sizes. Eight independent whole-genome MCMC analyses were run, for 200 million states each subsampling every 20000 states. For two-fragment MCMC analyses ten independent runs were set up. Due to the increased complexity of multitype tree parameter space 50% of states from every analysis were discarded as burnin.

Epidemiological analyses

$$r_j = \frac{\Gamma(kj + j - 1)}{\Gamma(kj)\Gamma(j+1)} \frac{(\frac{R_0}{k})^{j-1}}{(1 + \frac{R_0}{k})^{kj+j-1}} \tag{1}$$

Data availability

Acknowledgements

References