Population bottlenecks and founder effects: implications for mosquito-borne arboviral emergence

Scott C. Weaver1,2,3✉, Naomi L. Forrester4, Jianying Liu1,2,3 and Nikos Vasilakis1,2,5,6

Abstract | Transmission of arthropod-borne viruses (arboviruses) involves infection and replication in both arthropod vectors and vertebrate hosts. Nearly all arboviruses are RNA viruses with high mutation frequencies, which leaves them vulnerable to genetic drift and fitness losses owing to population bottlenecks during vector infection, dissemination from the midgut to the salivary glands and transmission to the vertebrate host. However, despite these bottlenecks, they seem to avoid fitness declines that can result from Muller’s ratchet. In addition, founder effects that occur during the geographic introductions of human-amplified arboviruses, including chikungunya virus and Zika virus, can affect epidemic and endemic circulation, as well as virulence. In this Review, we discuss the role of genetic drift following population bottlenecks and founder effects in arboviral evolution and spread, and the emergence of human disease.

Zoonotic infections that spill over from non-human animals to humans.

Arthropod-borne viruses (arboviruses) are transmitted to vertebrates by arthropod vectors (mainly mosquitoes and ticks)1 (FIG. 1). Most of the more than 500 known arboviruses are zoonotic agents that use wild animals as amplification and/or reservoir hosts [BOX 1]. A small fraction of these (about 100) are human-pathogenic and cause disease via ‘spillover’ infections (that is, enzootic or bridge vectors bite humans following infection from an enzootic vertebrate host). Notable examples include West Nile virus (WNV) and Japanese encephalitis virus, which use avian hosts but can be highly virulent when humans are infected by Culex spp. mosquito vectors. However, four arboviruses with non-human primate enzootic hosts have initiated human-amplified transmission cycles that last from months to centuries, which increases their ability to infect humans and to cause human disease. These include dengue virus (DENV), which infects nearly 400 million humans annually throughout many tropical and subtropical regions and causes disease in nearly 100 million individuals, thus making DENV the most important arbovirus2; and yellow fever virus (YFV), which despite the availability of an effective vaccine for nearly a century still infects tens of thousands of humans annually, with recent epidemics in Brazil3, Nigeria4 and Angola5. In addition, the past 15 years have brought unexpected outbreaks caused by two other human-amplified arboviruses, namely chikungunya virus (CHIKV) and Zika virus (ZIKV). Like DENV and YFV, these viruses are transmitted by the anthropophilic mosquitoes Aedes (Stegomyia) aegypti and sometimes Aedes (Stegomyia) albopictus. The expansion of these mosquitoes from Africa (A. aegypti) or Asia (A. albopictus) increased the disease burden throughout much of the tropics and permitted transmission in temperate regions (A. albopictus)6. However, our understanding of the emergence mechanisms remains incomplete. For example, why was CHIKV relatively inactive in the urban setting between the 1960s and 2005, and why were no ZIKV outbreaks documented before 2007? The ability of all arboviruses to survive and circulate despite adverse weather conditions (for example, droughts or winters when vectors and sometimes vertebrate hosts are inactive or limited in number) remains largely enigmatic, although current research is beginning to shed light on overwintering mechanisms7,8.

Nearly all arboviruses have RNA genomes that replicate in the absence of error correction mechanisms, and thus they exhibit high mutation frequencies in the range of 10−3 to 10−5 per nucleotide replicated9. Populations of RNA viruses exist as mutant swarms (FIG. 2), often selected at the population level as quasispecies. These swarms comprise a range of mutants (representing intra-host variation) that are dominated by a master sequence, which is typically characterized by high fitness in the current replicative environment. These mutant distributions are thought to enable RNA viruses to rapidly adapt to changing environments, which sometimes results in the emergence of epidemic and/or epizootic disease. Despite their potential for rapid evolution, many arboviruses exhibit high degrees of consensus genome sequence stability in nature, which may reflect, in part, the requirement to maintain fitness for infection
**Enzootic**
Ancestral, often continuous, transmission cycles of zoonotic arboviruses involving wild animals serving as amplification and/or reservoir hosts.

**Anthropophilic mosquitoes**
Mosquitoes with a preference for and typically a tendency to bite humans.

**Mutant swarms**
A population of RNA viruses with numerous, randomly derived mutations resulting from low-fidelity RNA replication (lack of proofreading). Mutant swarms are also termed quasispecies when they are selected and evolve as populations rather than as individual genetic variants.

**Quasispecies**
A diverse RNA (virus) population that contains many mutants closely related to a master sequence (usually the most abundant) and selected as a population, rather than as individuals, during its evolution.

**Population bottlenecks**
Major reductions in the population size of organisms, often used in the context of near-extinction events. Population bottlenecks result in a loss in genetic diversity and can also fix mutations at random. Population bottlenecks also represent a form of genetic drift.

---

**Fig. 1 | Infection and transmission of arthropod-borne viruses by mosquitoes and bottlenecks.**
Arthropod-borne viruses (arboviruses) undergo repeated bottlenecks during vector infection, dissemination from the midgut to the salivary glands and transmission to the vertebrate host. All of these population bottlenecks can present fitness challenges for the virus and may result in drift related to founder effects at each step outlined below. Infection of the mosquito midgut occurs upon ingestion of an infectious bloodmeal (steps 1, 2). Depending on the virus and the mosquito, the number of epithelial cells infected varies greatly (typically between 10 and most of the cells in the single layer of midgut epithelial cells). The severity of the midgut entry bottleneck seems to primarily depend on the titre of the initial infection and the genetic diversity of the virus population, as reductions in titre and diversity both result in severe reduction in the population to as little as 1–10 virus particles that found the next population following dissemination. Following infection and replication in the midgut cells, the arbovirus disseminates into the haemocoel (or open body cavity) (step 3). Then, access is gained to diverse secondary target organs and tissues such as the fat body, neurons and muscles, which support replication and may enhance infection of the salivary glands. Dissemination into the haemocoel is considered the major bottleneck during mosquito infection. The mechanism remains obscure, but generally the virus infects the midgut epithelium and then inefficiently disseminates into the haemocoel, either through microperforations in the basal lamina during basal lamina remodelling after blood feeding. Infection of the salivary glands is essential for transmission, and evidence suggests that there is a small bottleneck at this stage, probably because the arbovirus must pass through a basal lamina surrounding this organ (step 4). However, the impact and severity of this bottleneck remains to be determined. Following replication in the salivary gland acinar cells and shedding into the apical cavities, transmission can occur during a subsequent bloodmeal when saliva is injected during mosquito probing and feeding on blood vessels (step 5). The severity of this virus population bottleneck has been estimated to average between approximately 10 infectious units for some alphaviruses and over 10⁴ infectious units for West Nile virus. However, most of these estimates come from in vitro saliva collection, and few have been confirmed experimentally during in vivo transmission. Both dissemination from the midgut and infection of the salivary glands necessitate passage through basal laminae, predominantly consisting of collagen IV and laminin. The midgut basal lamina seems to become permissive for virus passage during bloodmeal digestion and stretching during a subsequent bloodmeal can create microperforations that seem to enhance dissemination, but passage through the salivary gland basal lamina remains poorly understood.

One critical challenge to arbovirus survival is represented by sequential anatomic barriers, especially in the vector (FIG. 1). These anatomic barriers to viral dissemination result in repeated population bottlenecks, which strongly reduce the number (to as little as one) of virus particles available to maintain infection and permit transmission. Population bottlenecks during the transmission cycle combined with the high mutation frequencies suggest that arboviruses undergo major challenges to maintain their fitness. Efficient natural selection (positive; or, more typically, purifying selection to remove random and potentially deleterious mutations) usually dominates virus evolution when populations are large. By contrast, reductions in virus populations reduce the efficiency of selection, which can result in stochastic changes in mutant frequencies and loss of genetic
Genetic drift
Random changes in the genetic make-up of a population due to chance, random sampling. Drift can dominate the evolution of a virus when population sizes remain small, reducing the efficiency of selection and genetic diversity.

Founder effects
The loss of genetic variation, sometimes resulting in the fixation of random mutations. Founder effects occur when a new population is established by a small number of individuals (founder population) randomly derived from a larger ancestral population. Founder effects represent a form of genetic drift.

Box 1 | Arbovirus transmission cycles and emergence

Arbovirus transmission involves two ecologically and genetically distinct cycles: the enzootic cycle (often called the sylvatic cycle) and the human-amplified cycle (often called the urban cycle) (see the figure). Zoonotic viruses, such as West Nile virus (WNV) and western equine encephalitis virus (WEEV) (see the figure, part a), use various wild animals as enzootic hosts, with spillover infections affecting humans. Some viruses, such as Venezuelan equine encephalitis virus (VEEV) (see the figure, part b), undergo secondary amplification involving domesticated animals, which increases spillover to humans in agricultural settings. Several of the most medically important viruses, including dengue virus (DENV), chikungunya virus (CHIKV), yellow fever virus (YFV) and Zika virus (ZIKV), use humans as direct amplification hosts (see the figure, part c), resulting in endemic and/or epidemic transmission and sometimes severe outbreaks.

For mosquito-borne viruses such as ZIKV, DENV and CHIKV, extant, ancestral virus strains cycle between various arboreal Aedes mosquito species and non-human primates (NHPs), and remain active in several parts of Africa and Southeast Asia (Asian enzootic transmission demonstrated only for DENV). During the human cycle, viruses circulate between humans (amplification hosts), and peri-domestic Aedes albopictus and domestic Aedes aegypti mosquitoes (see the figure, part c). Other Aedes mosquito species, mostly belonging to the subgenus Stegomyia, may vector these viruses in niche geographic regions within this endemic and/or epidemic cycle.

In rural areas of Africa and Asia, also known as the ‘zone of emergence’, these viruses can transferred between NHPs and humans through direct spillover, whereby an enzootic or a bridge vector transmits the virus from an enzootic host to a human. Under appropriate conditions, this can result in a transition from the enzootic cycle to a human–mosquito–human cycle (see the figure, part c).

Arboviruses with this potential for human-amplified transmission are among the most important for public health, such as ZIKV and CHIKV. They are likely to have emerged from the forest, associated with clearing of forests and development of human settlements. On the basis of phylogenetic analyses and detection in NHPs and A. aegypti mosquitoes, ZIKV spread to Asia at least decades ago126,132,133. By contrast, CHIKV is believed to have spread to Asia and the Americas centuries ago through the sailing ship trade2. The first ZIKV outbreak outside Africa was in Micronesia in 2007, followed by another outbreak in French Polynesia in 2013, before being introduced into South America and reaching pandemic levels (reviewed in Ref.113).

The historical emergence of CHIKV from Africa (from the East/Central South American (ECSA) enzootic lineage) has been traced through phylogenetics to South and Southeast Asia, where the emergent Asian strain continues to circulate and was the source of a recent introduction into the Americas124. Currently, a distinct ECSA strain introduced directly into Brazil co-circulates with Asian lineage strains in South America125. A second recent emergence from Africa (Indian Ocean lineage (IOL)) affected the islands of the Indian Ocean, India, Southeast Asia and Europe. Spread of all epidemic strains involved transmission mainly by A. aegypti, except for IOL strains that in some locations have adapted through a series of envelope glycoprotein substitutions for efficient transmission by A. albopictus126.

www.nature.com/nrmicro

R E V I E W S

© 2021 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
ratchet has been demonstrated to occur experimentally with several arboviruses\textsuperscript{27,28}, which underscores the risk they face if they are subjected to repeated bottlenecks during their transmission cycle (requiring replication and spread in divergent hosts)\textsuperscript{21}.

In this Review, we discuss viral population genetic factors that have recently been shown to be involved in arboviral spread and emergence of human disease. Adaptive evolution has been the focus of several recent, well-publicized examples of arboviral emergence\textsuperscript{25-28}. However, genetic drift has been far less investigated and has received less attention despite its potential to have an equally strong impact on arboviral spread and disease emergence. Therefore, we discuss new evidence — best exemplified by recent studies of CHIKV and ZIKV — that genetic drift following founder effects accompanying geographic introductions, and other population bottlenecks (FIG. 2), can also have dramatic influences on arboviral epidemics and disease. Although we focus on horizontal transmission, many arboviruses also undergo vertical, trans-ovarial transmission, which may also influence their evolution (but this is not discussed in this Review).

**Bottlenecks during transmission**

*Infection of and dissemination in the vector.* Infection of the mosquito vector is a driving force in arbovirus evolution, given the persistent nature of the infection as well as the length of time it takes for the virus to disseminate from the site of primary infection to the salivary glands. It has long been understood that the anatomy of the mosquito, in particular the basal laminae that surround the midgut, salivary glands and other organs, imposes bottlenecks on arboviral populations (FIG. 1). Thus, there has been great interest in the response of arboviruses to these bottlenecks and how they overcome them for transmission. We discuss three critical stages of infection with regard to the bottlenecks: infection of the mosquito midgut; dissemination into the haemocoel; and infection of the salivary glands.

Infection of the midgut represents the first bottleneck encountered by arboviruses (FIG. 1). This bottleneck was first thought to represent an anatomical barrier, a notion supported by early work on the transmission of WNV by *Culex pipiens quinquefasciatus*\textsuperscript{29} and Venezuelan equine encephalitis virus (VEEV) by *Aedes (Ochlerotatus) taeniopus*\textsuperscript{30}, showing infection of only a subset of enterocytes with common functions, or inefficient entry into the small number of these cells where infection is initiated, remains to be determined. However, recent work with the enzootic subtype IE VEEV vector *Culex taeniopus* suggests the latter. When infection of this highly susceptible vector occurs by a high-titre blood-meal, most cells become infected, which is a substantial departure from previous work\textsuperscript{29}. However, these findings may reflect the long-term co-evolution of enzootic subtype IE VEEV–*C. taeniopus* compared with more recent WNV–vector interactions in North America, and may therefore represent an outlier for human pathogenic...
arthropod to become infected

Vector competence
The intrinsic ability of an arthropod to become infected and transmit a pathogen.

The mechanism of arbovirus dissemination from the midgut into the haemocoel (FIG. 1) remains poorly understood. In addition to periodic basal lamina remodelling, which may permit virions to penetrate the otherwise exclusionary structure, the bloodmeal stretches the midgut basal lamina due to the approximately 20-fold increase in luminal volume. This results in basal lamina tears near the surrounding circular and longitudinal muscles. When midguts of CHIKV-infected mosquitoes were studied using transmission electron microscopy, virions were present in the basal lamina 24 h after feeding, and by 48 h the virus had disseminated from the midgut. Structural alterations and microperforations provide a likely escape mechanism for the virus from the midgut, especially in A. aegypti mosquitoes that feed more frequently than most other species. Recent studies examining the effect of repeated feeding suggest that a second, non-infectious bloodmeal expedites dissemination from the midgut of arboviruses infecting via an initial bloodmeal, probably due to similar mechanisms of microperforations.

Despite these physical mechanisms that promote arbovirus dissemination from the midgut, the dissemination barrier still presents a population bottleneck as demonstrated for CHIKV, ZIKV, VEEV, WNV and DENV. Depending on the virus–vector pairing, this bottleneck can be more or less stringent. For example, in subtype IE VEEV–C. tarsalis, the bottleneck has been estimated to reduce the population to approximately 2–50 virions, depending on the titre of the bloodmeal. However, 95% of infected mosquitoes support dissemination into the haemocoel. For WNV and CHIKV, the number of mosquitoes that develop disseminated infections is generally between 50 and 80%, again depending on the strains of mosquito and virus. Regardless of the severity of the bottleneck, arboviruses seem to recover genetic diversity quickly; for example, VEEV in approximately 8 days and DENV even more quickly. The ability of arboviruses to restore diversity following a bottleneck probably enables them to complete dissemination and, ultimately, transmission.

Prior to horizontal transmission, arboviruses must infect the salivary glands (FIG. 1). Regardless of how the virus passes through the surrounding basal lamina to infect the salivary glands, a population bottleneck occurs. Using intrathoracic infections that bypass the midgut bottleneck, several studies have demonstrated that a salivary gland infection bottleneck nevertheless persists. These founder effects have not been demonstrated to the same degree for alphaviruses as for WNV because the former are generally transmitted in much smaller amounts, which results in technical challenges to assessing the alphavirus populations in saliva.

For completion of the transmission cycle, an arbovirus must be present in the saliva of the vector during a bloodmeal subsequent to the infectious meal (FIG. 1). Some mosquitoes, especially A. aegypti, take frequent, partial bloodmeals and others only take one per reproductive cycle, which lasts approximately 1 week. The extrinsic incubation period, the minimum time between infection and transmission potential, can be as little as 2 days for CHIKV and a few other viruses. Given the
difficulties in establishing experimental transmission by infected mosquitoes in the laboratory, where in vitro collection of saliva can overestimate the amount of virus transmitted. It is unclear how stringent this transmission bottleneck is. This is especially true due to the artificial methods of expectoration typically used to assess virus titres in mosquito saliva. Flaviviruses such as YFV or DENV can transmit as large populations with titres up to $10^6$ infectious units, whereas alphaviruses such as CHIKV and VEEV are present in saliva at much lower titres, which suggests a more stringent transmission bottleneck.

**Replication and viraemia in the vertebrate host.** The early stages of arboviral replication and resulting viraemia in the vertebrate host are not well understood, in part because of limitations in the ability of animal models to reproduce natural hosts. For many arboviruses, critical sites of initial infection are immune cells near the site of the vector bite that precede dissemination to the lymphoid system and then transition to the circulatory system for viraemia. Studies with WNV suggest that the initial sites of replication are keratinocytes, which may represent a staging point from which the virus then spreads to the blood and other tissues and organs. Of the few arboviral population genetic studies in vertebrates, one showed that there was little to no bottleneck involved in generating viraemia for VEEV. A second study with ZIKV indicated few bottlenecks during the first 7 days of non-human primate infection, which suggests that vectors represent a more important bottleneck challenge than vertebrate hosts. This conclusion is also supported by studies with WNV that suggest limited bottlenecks and a strong purifying selection in birds. Thus, it is likely that the level of viraemia and the degree of viral diversity within the viraemia population are the major factors in transmission from vertebrates to arthropods, if only because this influences the degree to which mosquitoes are infected.

**Arbovirus evolution and emergence**

Overall, the above reviewed findings suggest that arboviruses regularly undergo major population bottlenecks at several stages of their transmission cycle. Therefore, they may face a major challenge of avoiding fitness declines due to drift when population sizes become too small for effective natural selection. Because total mutation frequencies of arboviruses are typically about $10^{-4}$ per nucleotide replicated, resulting in a per-genome frequency of about 1, direct reversion of the same mutation fixed in a population by drift occurs less frequently (once in about $10^4$ genomes), and repeated bottlenecks can rapidly reduce the fitness of a population due to the accumulation of random mutations that are typically deleterious. This phenomenon, Muller’s ratchet, occurs in the absence of efficient recombination, which can quickly and at higher frequency restore the wild-type sequence if two different genomes with deleterious mutations in different genome regions recombine. Some arboviruses have segmented genomes that may facilitate recombination via segment re-assortment, but most do not. Experimental studies of arboviruses including vesicular stomatitis virus and eastern equine encephalitis virus have shown that Muller’s ratchet reduces fitness when repeated plaque cloning is used to introduce bottlenecks to in vitro evolution systems.

However, considerable evidence suggests that many arboviruses avoid major fitness losses and continue to circulate efficiently in the face of both genetic and environmental challenges. In fact, their consensus sequences can be remarkably stable in nature, which suggests that they can avoid major drift and its potentially deleterious effects. One of the few exceptions of an arbovirus that stably circulates in nature is western equine encephalitis virus, an avian virus that during the mid-nineteenth century caused tens of thousands of annual spillover infections of equids and humans in North America, with high rates of mortality. However, western equine encephalitis virus underwent a dramatic reduction in its enzootic circulation, as well as human and equine disease, during the late twentieth century for unknown reasons. Recent reverse genetic studies provided no evidence of overall fitness declines, which suggests that ecological factors might have had a role in the decline of western equine encephalitis virus.

One potential explanation for the ability of arboviruses to avoid Muller’s ratchet is that, following bottlenecks during various stages of mosquito infection and transmission, populations of viable virus (for example, a plaque-purified recombinant strain indeed occurs, there is ‘leakage’ from the midgut, which restores viral diversity, including the most fit variants. Alternatively, it is possible that arboviruses have optimized their genome structure and mutation frequencies such that specific mutations, which confer some benefits during transmission of the recombinant strain indeed occurs, the following prerequisites should be met: the recombinant crossover should be demonstrated in a single PCR amplicon following cloning to ensure that it occurs in a single cDNA molecule; the recombination event should be demonstrated repeatedly in clonal populations of viable virus (for example, a plaque-purified or limited end point dilutions); and the recombinant
should maintain adequate sequence conservation during post-recombination evolution\textsuperscript{87}. These conditions have rarely been met in descriptions of arbovirus recombinants in the literature\textsuperscript{65}, although recombinants between closely related genomes would be technically challenging to detect.

Although drift would be frequently expected immediately following major bottlenecks regardless of mutant swarm restoration, its impacts may be minimized across multiple arboviral lineages. Discrete viral lineages within individual transmission chains may indeed undergo fitness decline in nature due to drift. However, very large numbers of individual transmission chains circulate in an active focus or region. This large number of lineages may ensure that, by chance, high-fitness genomes with master sequences remain propagated in nature to outcompete those that are affected by bottleneck-associated drift. Therefore, the ability of arboviruses to sustain huge numbers of transmission chains during efficient circulation in endemic or enzootic locations may enable them to avoid deleterious effects on their fitness, and thus permit them to persist with the ability to emerge via positive selection under favourable conditions.

**Geographic spread and founder effects**

In addition to the virus population bottlenecks encountered during the transmission cycle as described above, arboviruses frequently undergo bottlenecks during spread into new geographic regions or during reintroductions following dry seasons or winters that are not conducive to continuous circulation or persistence. In recent years, WNV\textsuperscript{68}, ZIKV\textsuperscript{69} and CHIKV\textsuperscript{70} spread for the first time in the modern scientific era to the New World. Although other arboviruses, especially DENV, also undergo frequent geographic introductions, their more widespread distribution for many decades may render the associated fitness effects less dramatic. Although the mechanism for the introduction of WNV into New York in 1999 remains enigmatic, modelling and phylogenetic studies suggest that most introductions of human-amplified ZIKV, CHIKV and DENV occurred via infected air travellers or individuals travelling by ship\textsuperscript{71–73} rather than through mosquitoes that have been inadvertently transported\textsuperscript{74} (Fig. 2). Although nearly all expansions of the geographic ranges of these arboviruses seem to represent point-source introductions\textsuperscript{65,75–79}, detailed studies of ZIKV in Florida in 2016 demonstrated multiple independent introductions, with only a few of these initiating prolonged transmission of weeks to a few months\textsuperscript{80}. For a stable geographic introduction, the arrival of an infected traveller must be followed by exposure to a competent vector (or vectors) that survives extrinsic incubation and then bites a susceptible person to initiate a transmission chain. This is presumably rare, considering the large numbers of infected travellers diagnosed with CHIKV, ZIKV and DENV in permisive regions such as the southern United States, and almost always without evidence of ensuing vector-borne circulation. However, when an infected traveller upon arrival is exposed to large populations of vectors such as A. aegypti, a major epidemic may occasionally ensue, as detected in multiple locations for all of these viruses.

The above geographic spread scenario includes several bottlenecks that provide the potential for founder effects (Fig. 2) to result in suboptimal transmission efficiency and altered amplification and/or virulence in humans. Based on the experimental evidence discussed above, in the case of many arboviruses, including CHIKV\textsuperscript{82} and ZIKV\textsuperscript{79,80}, the infected traveller is likely to be inoculated with a small number of infectious virions present in mosquito saliva (Fig. 1), with the potential for this population to contain random, sometimes fixed, deleterious mutations. When the infected traveller is subsequently bitten by a susceptible vector, the mosquito is likely to be infected by a small number of virions owing to the poor susceptibility of A. aegypti to these viruses and the midgut bottleneck described above\textsuperscript{65,85}. Thus, point-source introduction of these viruses by infected travellers probably involves at least two major arboviral bottlenecks during traveller infection and initial vector infection. Random mutations are far more likely by chance to be deleterious because non-synonymous mutations outnumber synonymous mutations, and random amino acid changes usually reduce protein function. Therefore, random mutations often result in less efficient vector transmission and/or host viraemia (sometimes with an accompanying reduction in virulence) than beneficial effects that increase transmission efficiency. Thus, these bottlenecks may influence emergence and human disease. Because large numbers of transmission chains, which may overcome drift as described above, do not initially exist following a new introduction, founder effects may have more dramatic impacts than the bottlenecks that all arboviruses undergo during the transmission cycle. Below, we discuss the evidence that these bottlenecks result in substantial fitness changes that affect the potential for arboviral outbreaks, further spread and disease severity.

**Chikungunya virus.** In the modern scientific era, several geographic CHIKV introductions have resulted in ongoing and amplified human–mosquito–human transmission chains that resulted in major disease outbreaks. In 2004, an epidemic was initiated by an East/Central/South African (ECSA) (Box 1) strain in coastal Kenya\textsuperscript{81}, followed by its spread into islands in the Indian Ocean\textsuperscript{82} and also, independently, from Kenya into South Asia\textsuperscript{83}, and subsequently to Southeast Asia\textsuperscript{84}. Spread of this ECSA-derived Indian Ocean lineage (IOL) was characterized by severe arthralgia with high rates of chronic disease, and the virus exhibited a series of envelope glycoprotein E2/E3 and E1 mutations that adapted some IOL strains for more efficient transmission by the recently invasive mosquito, A. albopictus\textsuperscript{12,13,96} (Fig. 3). Compared with all other ECSA strains examined, descendant IOL strains exhibit similar mosquito infection phenotypes and virulence in a mouse model\textsuperscript{9}, with no indication of fitness declines following the emergence in 2004. Thus, any founder effects and bottlenecks associated with the 2004–2005 CHIKV emergence did not reduce fitness, possibly because of multiple introductions from coastal Kenya. Although IOL strains were exported from Asia via infected travellers to initiate
the ability of CHIKV to efficiently transmit via and blue asterisks indicate ECSA strains with E2-211I; both of these amino acids decrease fitness. The red asterisk indicates the epistatic founder mutation (E1-A98T)...

Prior to the emergence of the IOL in 2004, the first phylogenetically delineated spread from Africa resulted from the earlier introduction of an ECSA enzootic strain in South and Southeast Asia approximately 60–100 years ago. This endemic and/or epidemic strain, called the Asian lineage (Fig. 3), has continued to circulate in Southeast Asia ever since the first outbreaks were reported in the 1950s. These outbreaks in Asia generally less severe than those caused by the IOL beginning in 2005, and it was initially unclear whether this reflected higher levels of human herd immunity during the 1950s compared with 2005, lower transmission efficiency during the earlier outbreaks or lower virulence of Asian lineage strains. Compared with African and IOL CHIKV strains, there is little or no difference in infectivity for A. aegypti, the principal urban CHIKV vector in most locations. However, based on epidemiologic and animal model studies, the Asian lineage strains seem to be less virulent in terms of apparent and human infection ratios and rates of chronic arthralgia. In 2013, this Asian CHIKV lineage was introduced into the Caribbean island of Saint Martin, followed by the extensive spread through much of the New World tropics and subtropics.

In contrast to the IOL that underwent adaptive evolution for more efficient mosquito transmission, evidence indicates that the Asian CHIKV lineage underwent founder effects and fitness losses following its introduction from Africa many decades ago. Asian strains possess a unique, low-fitness (in all hosts examined) 3′ untranslated region (UTR) that is much longer than that of African or IOL CHIKV strains, which strongly suggests a founder effect. The determinants of the lower virulence of the Asian lineage have not been thoroughly investigated, but phylogenetic reconstructions of the 3′ UTR indicate that the progenitor arrived from Africa with a major deletion in two direct sequence repeats, which reduced its fitness for replication in both mosquito and vertebrate cells. This fitness loss of the Asian lineage was partially restored over time by a series of point mutations and repeated sequence duplications. However, in every CHIKV sequence studied so far, the 3′ UTR of the Asian lineage remains debilitated (it results in reduced fitness for all hosts examined when placed using reverse genetics into the backbone of any major CHIKV lineage) compared with the ancestral ECSA and IOL 3′ UTR sequences. Furthermore, the 3′ UTR was also implicated in a probable founder effect following the 2013 introduction of the Asian lineage into the Americas. All Asian/American CHIKV strains contain a unique 177-nucleotide duplication in the repeated sequence region of the 3′ UTR. This insertion increases fitness for replication in mosquito cells, which suggests that CHIKV founder effects can also confer fitness advantages (as expected in a minority of introductions based on the majority of random mutations expected to be deleterious).

In addition to the direct 3′ UTR-mediated fitness effects resulting from the introductions of the Asian CHIKV lineage into Asia and the Americas, other founder effects have limited the ability of this strain to adapt for more efficient urban transmission. The detection of convergent evolution of the initial envelope glycoprotein 1 (E1)-A226V substitution in IOL strains as they spread into the Indian Ocean basin and South Asia suggested a strong selective advantage of this mutation for epidemic transmission. This hypothesis was supported by the major fitness advantage conferred by this mutation for infection of A. albopictus, which had spread from its ancestral range in Asia to many regions.
of Africa, southern Europe and the Americas since 1985 \cite{ref}. However, the lack of Asian lineage CHIKV strains with the E1-226V residue seemed paradoxical because this lineage had circulated in Asia, native territory for \textit{A. albopictus}, for many decades without evidence of the same adaptive mutation. This conundrum was resolved when subsequent reverse genetic studies demonstrated that the E1-A226V substitution does not appreciably increase infectivity of the Asian lineage CHIKV for \textit{A. albopictus} due to epistasis\cite{ref}. The generation of chimeric Asian–IOL viruses followed by mutagenesis of an Asian lineage clone showed that a single epistatic E1 glycoprotein threonine residue at position 98 is responsible for the inability of the Asian strain to benefit from the E1-A226V mutation (Fig. 3). Furthermore, the presence of the implicated E1-98T in all Asian strains and the lack of this residue in any other CHIKV strain, including those collected in Africa (all have E1-98A), along with the lack of a detectable phenotype for this mutation alone, strongly suggest that the threonine at position 98 in the Asian lineage resulted from a founder effect that accompanied the 3' UTR mutation upon introduction into Asia many decades ago.

In summary, major founder effects that accompanied the introduction of an ECSA CHIKV strain into Asia many decades ago probably limited its ability to cause major outbreaks due to 3' UTR-mediated fitness reductions that were difficult to restore, along with an amino acid substitution in the E1 envelope glycoprotein that prevented adaptation for transmission by \textit{A. albopictus}. The latter epistatic constraint was implicated in limiting the spread of the Asian and/or American CHIKV strains into temperate regions, a notion that is supported by the lack of the E1-226V residue in any sequenced CHIKV strain from the Americas, as well as by a lack of evidence for transmission in the Americas by \textit{A. albopictus}\cite{ref}.

\textbf{Zika virus.} The pathway and timing of the recent ZIKV introduction into the Americas is remarkably similar to that of CHIKV (Fig. 4). Following its evolution in sub-Saharan Africa in enzootic cycles involving arboreal mosquitoes and non-human primates, ZIKV was introduced many decades ago into Asia\cite{ref}. Following many decades without evidence of causing severe human disease, the first recognized outbreaks occurred during 2007 in Gabon\cite{ref} and Yap Island\cite{ref}. Then, in 2013, ZIKV spread into the South Pacific followed by Brazil, before causing a major epidemic that spread throughout the American tropics, subtropics and beyond\cite{ref}.

The reasons for the sudden emergence of ZIKV accompanied by Guillain–Barré syndrome and congenital Zika syndrome remain at least partially unresolved. Studies examining a mutation that occurred just prior to the spread of ZIKV into the South Pacific suggest that the virus adapted for more efficient urban transmission, similar to what was observed for the CHIKV IOL. An amino acid in the non-structural protein 1 (NS1-188) (Fig. 4) differs among ancestral African, some Asian and American ZIKV strains, with all African and American strains including valine and early Asian strains alanine. When the valine residue was engineered into a pre-epidemic 2010 Cambodian ZIKV strain, the resultant virus generated slightly increased infectivity for \textit{A. aegypti} mosquitoes\cite{ref}, the principal urban ZIKV vector\cite{ref,ref}. Furthermore, phylogenetic reconstruction of the evolution of this NS1-188 residue (Fig. 4) reveals that the pre-epidemic A188V substitution actually reflects the direct reversion of an earlier V188A substitution that occurred following the introduction of ZIKV into Asia, or soon thereafter. This strongly suggests that the V188A substitution confers reduced fitness for infection of \textit{A. aegypti}. This substitution therefore likely resulted from a founder effect, much like the CHIKV 3' UTR and E1-A98V mutations that accompanied the same geographic introduction from Africa to Asia. In addition, three additional substitutions in the capsid (C-A106T), precursor membrane (PrM-A1V) and non-structural protein 5 (NS5-V872M), also representing direct reversions of earlier mutations, confer enhanced transmission by \textit{A. aegypti} and/or
enhanced viraemia in human cell or mouse models of human viraemia. However, comparisons of ancestral African strains with Asian and American ZIKV strains suggest that the ability to infect A. aegypti has not been completely restored by these four mutations, as indicated by the greater transmissibility of African strains by A. aegypti as suggested by a 1966 isolation from this mosquito species in Malaysia, is surprising and underscores potential evolutionary constraints of the ability of arboviruses to adapt to new transmission opportunities due to their complex transmission cycles. Recently, another transmission-adaptive mutation that occurred shortly before the American outbreaks was identified. A substitution of methionine for valine in the envelope glycoprotein (E-V473M) enhances viraemia in mice and cynomolgus macaques, which suggests that it increased the epidemic potential via greater human viraemia. However, unlike the four revertant substitutions, E-V473M did not involve reversion of a prior mutation, which suggests classic Darwinian evolution.

The similarity in the geographic spread and fitness histories of ZIKV and CHIKV, with founder effects followed by direct reversions (ZIKV) or pseudo-reversions (CHIKV 3’ UTR) to only partially restore fitness, underscores the importance of founder effects in determining the emergence and epidemic potential of arboviruses. Both viruses reached the Americas only after many decades of circulation in Asia, with only partial restoration of the ancestral fitness exhibited by African strains. In the case of CHIKV, this limited vector usage (not by A. albopictus) and virulence in the Americas, whereas there was no direct evidence that ZIKV virulence was affected by founder effects or drift (NS1-A188V does not affect this phenotype). Thus, founder effects may limit both the timing and magnitude of emergence events, and their stochastic nature is likely to limit our ability to predict future epidemics.

Conclusions
There is compelling evidence that arboviruses face major population bottlenecks during several steps of their transmission cycles: infection of the mosquito midgut, dissemination from the midgut into the haemocoel, infection of the salivary glands and transmission to the vertebrate host. These bottlenecks can result in the stochastic fixation of random mutations that result from the inherently error-prone RNA virus replication, a form of genetic drift. Despite these bottlenecks and their potential for fitness losses via Muller’s ratchet, arboviruses are typically genetically stable in nature with no evidence of frequent fitness losses. These ubiquitous bottlenecks may be overcome through the rapid regeneration of the genetically diverse population following subsequent replication, along with large numbers of viral lineages in independent transmission chains from which lineages with master or high-fitness sequences can be selected. Founder effects, also a form of a bottleneck that can fix stochastic mutations via drift, can affect emergence trajectories in particularly unpredictable ways. They seem to have negatively affected the fitness of CHIKV and ZIKV following their introduction from Africa to Asia, possibly accompanied by drift from initially inefficient urban circulation in Asia. Although only one of several CHIKV lineages introduced from Africa to Asia or the Americas shows evidence for such a founder effect-related fitness decline, it is likely that this resulted in constraints on its vector usage and probably also affected its virulence. The emergence of ZIKV into the Americas followed the partial restoration of its fitness following founder effect-mediated drift and fitness loss. The lack of detected outbreaks caused by the ancestral African ZIKV lineage seems counter-intuitive, and contrasts with the observations of African CHIKV epidemics. However, the low symptomatic to asymptomatic ratio of ZIKV infections, compared with CHIKV, may result in greater under-reporting of the former virus. Furthermore, the greater experimental evidence of epidemic potential (A. aegypti infection and viraemia in animal models) of African ZIKV strains compared with Asian and American strains could also reflect limitations of the animal models.

It is important to recognize that predicting most arboviral emergence events remains nearly impossible due to the random nature of the mutations that accompany founder effects and bottleneck-mediated drift. Neither the emergence of the Asian lineage of CHIKV nor the emergence of ZIKV into the Americas could have been predicted without the retrospectively acquired information reviewed above. Nonetheless, some important inferences for the potential of a pathogen to emerge can be deduced through comprehensive surveillance, modelling and carefully controlled experimental studies.

Finally, drift and founder effects are only a few of many evolutionary mechanisms known to affect arboviral emergence. Other major drivers include changes in land use, increased international commerce and travel, all of which have been key for the spread of urban arboviruses. These factors have not been shown to directly impact bottlenecks that accompany arbovirus circulation, and more research is required. During past centuries, DENV, YFV and CHIKV (and possibly ZIKV) have spread regularly but more slowly from Africa to Asia and the Americas on ships, with only seasonal persistence in many temperate locations. However, the widespread dissemination of A. aegypti centuries ago, and that of A. albopictus since 1985 resulting from used tyre and other trades, along with urban population growth, especially in the tropics, has greatly accelerated global arbovirus spread. Huge populations of uncontrolled A. aegypti in large cities in the tropics have only compounded the risk for efficient transmission, along with resistance to insecticides by this and other arbovirus vectors. Finally, adaptation to vectors and hosts can also have substantial effects on arboviral spread, as underscored by both ZIKV and CHIKV. All of these emergence drivers require concerted, sustained study, and overcoming their effect on human health will necessitate innovative combinations of control approaches.
1. Karabatsos, N. International Catalog of Arboviruses Including Certain Other Viruses of Vertebrates 4th edn (American Society of Tropical Medicine and Hygiene, 1985).

2. Bilski, S. et al. The global distribution and burden of dengue. Nature 496, 504–507 (2013).

3. Paulis, C. I. & Fauci, A. S. Yellow fever — once again on the radar. Am. J. Engl. J. Med. 376, 1357–1359 (2017).

4. Nwaohuobi, W. E. et al. The response to re-emergence of avian influenza in Nigeria, 2017. Int. J. Infect. Dis. 92, 189–196 (2020).

5. Kraemer, M. U. G. et al. Spread of yellow fever outbreak in Angola and the Democratic Republic of the Congo, N. Engl. J. Med. 372, 561–567 (2020).

6. Weaver, S. C. et al. West Nile virus experimental infection of Aedes aegypti, Microbes and Infection 3, 113–118 (2001).

7. Deardorff, E. R. et al. West Nile virus experimental infection of Aedes aegypti, Microbes and Infection 3, 113–118 (2001).

8. Muller, H. J. The relation of recombination to constraints on molecular evolution. J. Mol. Evol. 18, 105–108 (1983).

9. Ko, H. Y. et al. Inter- and intra- host sequence diversity of yellow fever in Nigeria, 2017. N. Engl. J. Med. 376, 504–507 (2017).

10. Anishchenko, M. et al. Venezuelan encephalitis virus experimental infection of Aedes aegypti, Microbes and Infection 3, 496–499 (1990).

11. Karabatsos, N. Systematics and the Origin of Species (eds Huxley, J., Hard, A. C., & Ford E. B.) 157–180 (Princeton Univ. Press, 1954).

12. Hall, B. et al. The role of host alternation in the adaptive evolution of RNA viruses. Proc. Natl Acad. Sci. USA 105, 504–507 (2008).

13. Garsona, S. et al. Potential for co- infection of a sylvatic New World mosquito species, shows a low reproductive fitness in infected mosquitoes. Am. J. Trop. Med. Hyg. 78, 681–686 (2008).

14. Bergren, N. A. et al. Western equine encephalitis virus: evolutionary analysis of a declining alphavirus population. J. Virol. 86, 4348–4354 (2012).

15. Boylan, B. T., Moreira, F. R., Carlson, T. W. & Bernard, K. A. Mosquito cell-derived West Nile virus replicon particles mimic arbovirus inoculum and have reduced spread in mice. PLoS Negl. Trop. Dis. 11, e0003279 (2017).

16. Aloita, M. T. et al. Moleculare barcode Zika virus libraries to probe in vivo evolutionary dynamics. PLoS Pathog. 10, e1004884 (2014).

17. Grubau, L. et al. Mosquitoes transmit unique west nile virus populations during each feeding cycle. Cell Rep. 19, 1665–1673 (2017).

18. Weaver, S. C. et al. West Nile virus Experimental infection of Aedes aegypti, Microbes and Infection 3, 383–388 (2018).

19. Bergren, N. A. et al. “Submergence” of western equine encephalitis virus: evidence of positive selection argues against genetic shifts or recombination. PLoS Pathog. 16, e1008120 (2020).

20. This paper describes declines in an arbovirus fitness that were not attributed to drift or bottlenecks, and shows that fitness increased over time, suggesting ecologic explanations for the decline in enzootic transmission and human/equine cases.

21. Muratorio, A. C. & Vignali, R. Ultrastructural analysis of chikungunya virus dissemination from an insect vector. Cell Host Microbe 16, 1134–1139 (2014).

22. Tsetsarkin, K. A. et al. Multi-peak adaptive landscape for chikungunya virus evolution predicts continued fitness optimization in Aedes albopictus mosquitoes. Nat. Commun. 5, 4084 (2014).

23. This paper summarizes multiple A. albopictus- adaptive mutations that accompanied the spread of the 10th isolated strain of an Oceanic and Asia Scholle, F., Giraud, Y. A., Zhao, H., Oggs, S. & Mason, P. W. trans-Packaged West Nile virus-like particles infected and in infected mosquito vectors. J. Virol. 78, 11605–11614 (2004).

24. Smith, D. R., Adams, A. P., Kenney, J. L., Wang, E. & Weaver, S. C. Venezuelan encephalitis virus infection of the mosquito Aedes taeniorhynchus: infection initiated by a small number of susceptible epithelial cells and a population bottleneck. Virology 372, 176–186 (2008).

25. Deardorff, E. R. et al. West Nile virus experimental infection of Aedes aegypti, Microbes and Infection 3, 113–118 (2001).

26. Le Coupance, A. et al. Co-infection of mosquitoes with chikungunya and dengue viruses reveals modulation of the replication of both viruses in midguts and salivary glands of Aedes aegypti mosquitoes. J. Mol. Microbiol. 75, 217–220 (2017).

27. Domingo, E. & Holland, J. J. RNA virus mutations and the ratchet using an animal virus and an arbovirus as models. J. Mol. Microbiol. Biotecnol. 21, 71–81 (2011).

28. Anishchenko, M. et al. Venezuelan encephalitis virus experimental infection of Aedes aegypti, Microbes and Infection 3, 113–118 (2001).

29. Goenaga, S. et al. Potential for co-infection of a mosquito-specific flavivirus, Nuimuravirus, to block West Nile virus transmission in mosquitoes. Viruses 7, 5801–5812 (2015).

30. Rouany, C. M. et al. Viruses: a historical overview and recent developments. Adv. Virus Res. 98, 119–146 (2017).

31. Telzer, R. M. et al. The role of recombination in adaptive evolution. Nat. Rev. Genet. 15, 501–512 (2014).

32. Patterson, E. I. et al. Mosquito bottlenecks alter viral evolutionary dynamics. PLoS Pathog. 15, e1007955 (2019).

33. Karpavici, A. & Franz, A. W. E. Ultrastructural analysis of chikungunya virus infection of Aedes albopictus, Nanopore sequencing elucidates the rapid evolution of defective- interfering RNAs in Flock House virus. J. Virol. 88, 3260–3267 (2014).

34. Bergren, N. A. et al. “Submergence” of western equine encephalitis virus: evidence of positive selection argues against genetic shifts or recombination. PLoS Pathog. 16, e1008120 (2020).
66. Weaver, S. C. & Vasiliakos, N. Molecular evolution of dengue viruses: contributions of phylogenetics to understanding the history and epidemiology of the prearival arboviral disease. *Infect. Genet. Evol.* 9, 523–540 (2009).

67. Weaver, S. C. et al. Recombinational history and molecular evolution of western equine encephalomyelitis complex alphaviruses. *J. Virol.* 71, 613–625 (1997).

68. Lecointre, R. S. et al. Origin of the West Nile virus responsible for an arthropod-borne encephalitis in the northeastern United States. *Science* 286, 2533–2537 (1999).

69. Cardoso, C. W. et al. Outbreak of exanthematous illness associated with Zika, chikungunya, and dengue viruses, Salvador, Brazil. *Emerg. Infect. Dis.* 21, 2274–2276 (2015).

70. Leoparc-Goffart, I., Nougarede, A., Cassadou, S., Prat, C. & de Lamberlante, X. Chikungunya in the Americas. *Lancet* 383, 514 (2014).

71. Rezza, G. et al. Infection with chikungunya virus in Italy: an outbreak in a temperate region. *Lancet* 370, 1840–1846 (2007).

72. Venturi, G. et al. Detection of a chikungunya outbreak in central Italy, August to September 2017. *Euro Surveill.* 22, 17–00646 (2017).

73. Grubaugh, N. D. et al. Genomic epidemiology reveals multiple introductions of Zika virus into the United States. *Nature* 546, 401–405 (2017).

This paper describes multiple introductions of ZIKV into the United States, quickly becoming extinct and only a few persisting for weeks to months. Whether any of the extensions were related to the virus that was going to be determined.

74. Mier, Y. T. R., Latem, A. J. & Johansson, M. A. Mosquitoes on a plane: dissection will not stop the spread of vector-borne pathogens, a simulation study. *J. Virol.* 86, e005683 (2018).

75. Faria, N. R. et al. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. *Nature* 546, 406–410 (2017).

76. Faria, N. R. et al. Zika virus in the Americas: early epidemiological and genetic findings. *Science* (2016).

77. Nunes, M. R. et al. Emergence and potential for spread of chikungunya virus in Brazil. *BMC Med.* 13, 102 (2015).

78. Chen, R. et al. Comprehensive genome scale phylogenetic study provides new insights on the global expansion of chikungunya virus. *J. Virol.* 90, 10600–10610 (2016).

79. Azar, S. R. et al. Differential vector competency of *Aedes albopictus* populations from the Americas for Zika virus. *Am. J. Trop. Med. Hyg.* 97, 330–337 (2017).

80. Roundy, C. M. et al. Variation in *Aedes aegypti* mosquito competency for Zika virus transmission. *Emerg. Infect. Dis.* 23, 625–632 (2017).

81. Christen, J. P. et al. Drought-associated chikungunya emergence along coastal East Africa. *Am. J. Trop. Med. Hyg.* 86, 560–567 (2012).

82. Schuffenecker, I. et al. Genome microevolution of chikungunya viruses causing the Indian Ocean outbreak. *J. Virol.* 86, 5313–5320 (2012).

83. Haddox, A. D. et al. Genetic characterization of Zika virus strains: geographic expansion of the Asian lineage. *PLoS Negl. Trop. Dis.* 6, e1477 (2012).

84. Grand, G. et al. Zika virus in Gabon (Central Africa) — 2007: a new threat from *Aedes albopictus*? *PLoS Negl. Trop. Dis.* 8, e2681 (2014).

85. Lanciotti, R. S. et al. Genomic and ecologic properties of Zika virus associated with an epidemic, Yap State, Micronesia, 2007. *Emerg. Infect. Dis.* 14, 1232–1239 (2008).

86. Musso, D. & Gubler, D. J. Zika virus. *Clin. Microbiol. Rev.* 29, 487–524 (2016).

87. Liu, Y. et al. Evolutionary enhancement of Zika virus infectivity in *Aedes aegypti* mosquitoes. *Nature* 545, 482–485 (2017).

88. Azar, S. R. & Weaver, S. C. High vector competence in *Aedes aegypti* for a *Zikavirus* collected during a Zika virus outbreak in Brazil. *PLoS Negl. Trop. Dis.* 10, e0005055 (2016).

89. Azar, S. R. & Weaver, S. C. Vector competence: what have we learned? What gaps remain? A report from the Global Virus Network. *Antivir. Res.* 144, 225–246 (2017).

90. Cassadou, S. et al. Emergence of chikungunya fever on the French side of Saint Martin island, October 2017. *PLoS Negl. Trop. Dis.* 11, e0005158 (2017).

91. Kumar, S., Stecher, G., Li, M., Knazy, C. & Tamura, K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549 (2018).

92. Smithburn, K. C. Neutralizing antibodies against arthropod-borne viruses in the sera of long-time residents of Malaya and Borneo. *Am. J. Trop. Med. Hyg.* 107, 157–163 (1954).

93. atlange, M. T. et al. Zika in the Americas, year 2: what have we learned? What gaps remain? A report from the Global Virus Network. *Antivir. Res.* 144, 225–246 (2017).

94. Bennett, S. Patterns, drivers, and challenges of vector-borne disease emergence. *Nature Rev. Microbiol.* 17, 145–150 (2018).

Acknowledgements
Research in the authors’ laboratories is supported by National Institutes of Health (NIH) grants AI120942 and AI214552 (to S.C.) and AI145918 (to N.V.).

Author contributions
The authors contributed equally to all aspects of the article

Competing interests
The authors declare no competing interests.

Peer review information
Nature Reviews Microbiology thanks D. Brandt, G. Ebel and L. Baumann for their contribution to the peer review of this work.

Publisher’s note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© Springer Nature Limited 2021