 doi:10.20944/preprints202003.0021.v1

Virus population bottlenecks: what are they telling us?

Feng Qu[1,2], Limin Zheng[1,2], Shaoyan Zhang[1,2], Rong Sun[1,2], Jason Slot[1]

[1]Department of Plant Pathology, [2]Ohio Agricultural Research and Development Center, The Ohio State University

Contact: qu.28@osu.edu; 330-263-3835

 doi:10.20944/preprints202003.0021.v1

Summary

Stringent, stochastic viral population bottlenecks have been observed in the infections of many viruses, but exactly how and why they occur is unclear. A critical review of recent literature prompts us to propose a new hypothesis, designated Isolate, Amplify, and Select (IAS), that satisfactorily explains the bottlenecks in single-stranded, positive sense (+) RNA virus infections. This new hypothesis postulates that, unlike those in free-living organisms, the viral population bottlenecks are imposed by viruses themselves, inside the infected cells, through virus-encoded bottleneck-enforcing proteins (BNEPs) that function in a concentration-dependent manner. Most BNEPs are directly translated from the (+) RNA genomes of invading viruses, so that if numerous virions of the same virus invade a cell simultaneously, the bottleneck-ready concentration of BNEPs would be reached sufficiently early to arrest nearly all internalized viral genomes. As a result, in these cells very few (as few as one) viral genomes escape from the bottlenecks stochastically to initiate viral reproduction. Repetition of this process in successively infected cells ensures the progeny genomes in each cell descend from the same parental genome(s), hence isolating different mutant genomes in separate cells. This isolation precludes mutant viral genomes encoding defective replication proteins from exploiting the complementing proteins synthesized by sister genomes, leading to the prompt elimination of such mutants. Conversely, the IAS model ensures replication proteins with beneficial mutations exclusively amplify the viral genomes that harbor the very mutations. Reiteration of this process in consecutively infected cells enriches such beneficial mutations in the virus pool. In conclusion, the IAS hypothesis provides a compelling evolutionary model for population bottlenecks of (+) RNA viruses.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 March 2020 doi:10.20944/preprints202003.0021.v1

Introduction

For free-living organisms we define population bottlenecks as dramatic, but non-selective, decreases in the number of reproducing individuals in a population, often due to abrupt environmental changes, diseases, migration, geographic or sympatric isolation, resulting in stark loss of genetic diversity in subsequent generations. Population bottlenecks can lead to the fixation of certain sub-optimal mutations that would have been eliminated through natural selection had the population size been larger. Such mutations are well documented in certain human populations (e.g. ref[1]). However, classic population studies seem to have largely ignored an entire category of stringent population bottlenecks that are far more prevalent, namely the bottlenecks encountered by germline cell populations. Nearly all multicellular organisms that reproduce via single cells, asexually or sexually, produce huge numbers of germline cells, with very few of them contributing to the genetic diversities of future generations. This raises an intriguing prospect that some forms of population bottlenecks could be the consequence of natural selection.

Many viruses are known to experience severe and stochastic population bottlenecks at various stages of their reproduction, spread, and transmission (see later for examples). The intuitive thinking would be that such bottlenecks are detrimental to virus viability and adaptability because, similar to free-living organisms, they could lead to the fixation of deleterious mutations in subsequent generations of viruses. This problem would be exacerbated by the high error rate of the replication process of most viruses, eventually leading to the demise of the virus populations (the error catastrophe). If this is true, we could speculate that virus population bottlenecks might benefit the hosts. We might further speculate that successful viruses must have evolved ways to subvert population bottlenecks. An improved understanding of the bottleneck mechanism might therefore allow us to regulate their occurrence in order to control virus diseases.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 March 2020 doi:10.20944/preprints202003.0021.v1

However, exactly how virus populations bottleneck is not well understood. Here we propose a new, counterintuitive hypothesis. We summarize recently published studies suggesting that viral population bottlenecks often manifest a virus-encoded function that limits the number of genomes becoming replicated in each of the infected cells. We argue that such restriction is essential to virus persistence because it enables the purging of lethal mutations in protein-coding genes that frequently arise from the error-prone replication of most viruses, especially those with single-stranded, positive sense (+) RNA genomes. Our hypothesis further provides a mechanistic basis for superinfection exclusion between highly similar viruses, and the frequently observed *cis*-preference of many viral non-structural proteins. If validated, this model is expected to inspire novel virus control strategies.

Viral population bottlenecks are widespread in virus-infected hosts and tissues

Many viruses have been reported to experience stringent population bottlenecks in infected hosts [2,3]. An early estimate for tobacco mosaic virus (TMV), a (+) RNA virus, concluded that its population in each of systemically infected tobacco leaves was initiated by no more than 20 TMV genomes [2]. Separate studies examining wheat plants infected with wheat streak mosaic virus (WSMV), another (+) RNA virus, revealed even narrower population bottlenecks, showing that WSMV populations in each of the systemically infected wheat tillers originated from no more than five virions [4,5]. Another investigation of tobacco plants infected with cucumber mosaic virus (CMV), a (+) RNA virus with three genome segments, indicated that it likewise experienced severe population bottlenecks [6]. More recently, Zhang and colleagues [7] used a nine-variant population of turnip crinkle virus (TCV) to unveil dramatic, stochastic population bottlenecks in infected Arabidopsis plants that were independent of RNA silencing-mediated antiviral defense. While these earlier studies were inevitably low throughput, collectively they suggested that severe population bottlenecks accompanied systemic infections of multiple (+) RNA plant viruses. Notably, by examining small (1 mm^2) sections of the systemically infected

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 March 2020 doi:10.20944/preprints202003.0021.v1

wheat leaves, Hall and colleagues [4] found that different WSMV variants segregated into distinct "islands", thus providing the earliest hint that plant level population bottlenecks may have been caused by viral variants mutually excluding each other in the infected cells.

Severe, stochastic population bottlenecks were also reported in infections of animal viruses. For instance, Venezuelan equine encephalitis arbovirus (VEEV), a (+) RNA animal virus, was found to encounter repeated population bottlenecks as the virus moves through different tissues of the mosquito vector [8]. Separately, Kuss and colleagues [9,10] identified multiple population bottlenecks during poliovirus infections of mice, especially at the stage of central nervous system invasion. More recently, the advent of high throughput sequencing technology has permitted much more refined description of population bottlenecks in diverse virus infections. Taking advantage of the enormous sequencing depth, Grubaugh and colleagues [11] discovered that during the infection of the mosquito vectors, the Western Nile arbovirus (WHV) populations were repeatedly bottlenecked as it spread from one tissue type to another (e.g. midgut, hemolymph, salivary gland), even though the bottlenecked populations always expanded within the tissues they moved into. Importantly, such bottlenecks were stochastic and failed to enrich any particular viral genotypes. It is worth noting that stochasticity appears to be a shared feature for most of the documented viral population bottlenecks.

Finally, although population bottlenecks were mostly documented in infections of (+) RNA viruses, they were also reported in infections of viruses with other forms of genomes. For example, an artificial population of Influenza A virus (IAV), consisting of 100 variants of the negative sense (-), segmented RNA genome, encountered strong population bottlenecks in infected individuals that were only partially driven by genetic pressures [12]. Indeed, another study with IAV samples collected from naturally infected patients concluded that variations at the single nucleotide levels "provided little evidence for positive selection", and, highly restrictive yet stochastic population bottlenecks limited the number of the infection-originating IAV genomes in a given patient to no more than two [13].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 March 2020 doi:10.20944/preprints202003.0021.v1

Viral population bottlenecks likely have originated inside the virus-infected cells

The examples discussed above mostly deal with viral population bottlenecks at the levels of infected host individuals, or specific tissues/organs. But similar bottlenecks have also been observed within infected cells. As mentioned earlier, Hall and colleagues observed that mixed infection with two different WSMV variants led to the segregation of the two variants in tiny adjacent cell clusters in systemically infected leaves [4]. Frost and colleagues reported that human immunodeficiency virus (HIV) in the spleen of infected patients likewise segregated into numerous subpopulations that are genetically distinct from each other [3]. It is important to note that the tissues being examined in both of these studies, namely wheat leaves systemically infected by WSMV and spleens of HIV patients, most likely encountered large numbers of viruses that had been amplified in other tissues of the same host individual. The fact that these tissues then segregated into small cell clusters populated by genetically distinct viral variants offered two very important insights: (i) each of the individual cell clusters was founded by very few virions, possibly just one, thus hinting at a highly stringent cellular level population bottleneck; (ii) that the viral variants stayed constrained within the cell clusters suggests a mechanism that enables infected cells of one cluster to exclude virions produced from the neighboring clusters. Are these two events, namely population bottlenecking and mutual exclusion, controlled by the same mechanism?

Direct evidence for intracellular population bottlenecking was first provided by Miyashita and colleagues [14]. These authors constructed a population of TMV with more than 100,000 variants by inserting randomized sequences composed of 10 randomized nucleotides into TMV genomes. This TMV population was then brought into individual wall-less plant cells known as protoplasts to launch viral replication. 15 infected protoplasts were then processed individually and subjected to deep sequencing to reveal the identity of TMV variants in each of the cells. These experiments led to at least two

important observations. (i) No more than seven variants replicated in any given protoplast, even though the number of TMV genomes entering each protoplast was estimated to be more than 5,000. (ii) The identities of the TMV genomes isolated from all these cells were different from each other. Together these data revealed highly stringent, yet stochastic population bottlenecks encountered by TMV variants inside the cells they invaded.

Do natural viral infections entail the entry of large numbers of viral genomes per host cell? Indeed this is exactly what virologists have discovered in recent years. They found that many, if not most, viruses invade most cells in the form of so-called "collective infectious units" (CIUs) [15,16]. CIU can take many forms. HIV, for example, spreads between infected and uninfected T cells through intercellular connections known as virological synapses that transmit massive amounts of virions between two cells [17–19]. By contrast, poliovirus and coxsackievirus particles are non-lytically released from infected cells in the form of lipid membrane vesicles containing dozens of virions [20,21]. Such vesicles then fuse with uninfected cells to deliver all the virions "en bloc" [21]. Even more dramatically, respiratory syncytial virus (RSV) induces the fusion of infected cell membrane with that of neighboring cells to form giant multinucleated cells [22]. Still other viruses connect infected and uninfected cells through tubular or filamentous intercellular extensions [23]. It is now well accepted that such multi-genome transmission is the primary, and more efficient mode of viral intercellular spread inside infected individuals, and possibly also between host individuals [24,25]. Importantly, plant viruses spread cell-to-cell with high efficiency, through plasmodesmata channels modified by virus-encoded movement proteins. Such modifications are thought to shuffle large numbers of virions or viral genomes between adjacent cells.

Contrasting with the mass entry of viral genomes, only a small fraction of the entered genomes were found to embark on active replication. Such intracellular population bottlenecks have been observed not only in detached protoplast cells [14], but also in the cells of infected hosts. Several independent studies demonstrated this using virus variants labelled with different fluorescent proteins.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 March 2020 doi:10.20944/preprints202003.0021.v1

For example, Miyashita and Kishino [26] used yellow and cyan fluorescent proteins (YFP and CFP) to differentially label soil-borne wheat mosaic virus (SBWMV) in order to follow the fate of the two viral variants, focusing only on the cells co-infected by both variants at the very beginning. They found that the virus-borne fluorescence markers began to segregate into cell clusters with a single FP after seven to nine cell-to-cell movement events. They further estimated that the progeny viruses in these cells must have descended from no more than six founding genomes. Importantly, such severe cellular level bottlenecks are not unique to plant viruses, or RNA viruses. Kobiler and colleagues [27] engineered the PRV263 strain of pseudorabies virus, a herpesvirus with a large, double-stranded DNA genome, using a special tool called Brainbow cassette. The progeny genomes of the engineered PRV263 were each expected to encode one of the three FPs – CFP, GFP, and RFP. Using the engineered PRV263 to infect culture cells, the authors observed the segregation of cell clusters containing single or double FPs almost immediately after the viral spread into neighboring cells. These results led them to conclude that infections in new cells must have been founded with fewer than seven genomes.

Intracellular population bottlenecks are also evident in systemically infected plant leaves (SLs). Using two engineered plum pox virus (PPV) variants encoding GFP and RFP (DsRed), respectively, Dietrich and Maiss [28] observed that in SLs, the two PPV variants formed separate, single-colored cell clusters that are adjacent to each other, but co-infection by both occurred only in the one-cell-width borderlines between green and red cell clusters. The very fact that the each of the cell clusters contained a single variant betrays highly stringent population bottlenecks that permitted just one variant to infect a group of connected cells in an SL. Strikingly similar observations have since been made with other viruses, including WSMV, apple latent spherical virus (ALSV), tobacco etch virus (TEV), and a negative-strand RNA virus known as sonchus yellow net virus (SYNV) [29–33], underlining the highly conserved nature of intracellular bottlenecking encountered by virus populations.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 March 2020 doi:10.20944/preprints202003.0021.v1

How do viral populations become bottlenecked inside infected cells?

Answering this question requires an experimental approach capable of generating large intracellular virus populations. While a TMV population comprising more than 100,000 (4^{10}) variants was constructed by Miyashita and colleagues [14], that population was used to infect protoplast cells detached from plants. By contrast, it is not trivial to verify the internalization of multiple viral variants in cells of intact plants or animals. Fortunately, with (+) RNA viruses we can transfect host cells with transcribable viral cDNA. Viral cDNA of (+) RNA viruses, when equipped with promoters recognizable by RNA polymerases of the host cells (e.g. DNA-dependent RNA polymerase II, Pol II), directs the transcription of numerous viral genomic RNA copies that initiate potent viral infections. Indeed, plant virologists routinely use this approach to launch infections of (+) RNA viruses. It is commonly used in combination with the model plant *Nicotiana benthamiana*, thanks to the exceptional amenability of this plant to modified transfer DNA (T-DNA) exported from cells of *Agrobacterium tumefaciens*, a plant-infecting bacterium. This process is hence dubbed as "agro-infiltration". An additional advantage of agro-infiltration is that multiple *A. tumefaciens* strains, each carrying different modified T-DNA constructs, can be mixed to deliver multiple viral cDNA into the same *N. benthamiana* cells with up to 100% co-introduction efficiency.

Zhang and colleagues used agro-infiltration to generate large intracellular populations of TCV [34]. They placed the duplicated 35S promoter (2X35S) of cauliflower mosaic virus (CaMV), which has the potential to drive the transcription of up to 10,000 RNA copies [35], upstream of modified TCV cDNAs that encode GFP and mCherry, respectively, to obtain two constructs capable of launching TCV infections. Both TCV variants could replicate only in the cells entered by the cognate constructs, as they did not encode the capsid protein needed for TCV cell-to-cell movement in *N. benthamiana*.

The agro-infiltration-delivered TCV intracellular population, made up of thousands of copies of each of the two variants, encountered extremely narrow bottlenecks that permitted the replication of just one copy of one variant in nearly all treated cells, as deduced from the fact that less than 0.1% of the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 March 2020 doi:10.20944/preprints202003.0021.v1

cells expressed both FPs. Strikingly, TCV replication was not needed for the bottlenecking to occur – cells treated with one variant commenced viral replication in a much delayed, and asynchronous fashion, so that 3%, 10%, 23%, and 33% of the cells produced the replication-dependent red fluorescence at 48, 72, 96, and 120 hours post infiltration, respectively. By contrast, a co-delivered, non-viral construct expressed GFP in over 90% of the cells as early as 36 hours post infiltration. These findings indicated that a strongly repressive, bottlenecked state was in place almost immediately after agro-infiltration, blocking the overwhelming majority of the TCV genomes from initiating replication. As a result, very few TCV genomes managed to escape in a delayed, sporadic fashion, so that when multiple cells were examined, the escape in different cells was asynchronous.

Further investigations identified one single TCV-encoded protein as necessary and sufficient to establish the strongly repressive state. Incidentally, this protein, p28, is directly translated from the TCV genomic RNA. Thus, the strong transcription powered by the 2X35S promoter would have provided a sufficient amount of p28 mRNA in the cells, permitting fast accumulation of p28 protein, and swift establishment of the repressive state. In summary, viral population bottlenecks in TCV-infected cells can be satisfactorily explained by the rapid accumulation of p28 proteins, made possible, ironically, by the availability of high numbers of TCV genomic RNA copies in the cell.

Finally, we must note that p28 is also the auxiliary replication protein encoded by TCV. Hence, the single p28 protein plays two completely opposite roles in TCV infections, namely facilitating, *and* repressing, the replication of TCV genomes. We now have evidence to suggest that the key regulative switch may be the intracellular concentration of p28, with lower concentration favoring replication, whereas higher concentration favors repression [36]. These points will become important in the next section, when we attempt to answer the question of “why”.

Why do viruses bottleneck their own populations inside the cells they enter?

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 March 2020 doi:10.20944/preprints202003.0021.v1

By now we hope that you are convinced that viruses frequently deliver large quantities of genomes into individual host cells, where only a small fraction of the entered genomes end up launching successful infections. In fact new evidence suggests that some viruses with multiple genome segments may not even get to replicate all their segments in every infected cell [13,37]. It is also well known that established infections exclude secondary invasions by the same virus [7,38–40]. Finally, such exclusion/repression has been shown to be actively enforced by virus-encoded proteins in several cases [30,31,34]. Why would viruses actively block most copies of their own genomes from replicating in the same cell? The insights gleaned from the TCV-based system prompt us to propose a new hypothesis that answers this question. We name this hypothesis Isolate, Amplify, and Select (IAS). For now we limit the applicability of the IAS model to (+) RNA viruses, although it may prove to be more widely applicable as new data emerge.

First and foremost, the IAS model addresses the need for separating viral genomes with varying replication potentials. The major replication enzyme encoded by (+) RNA viruses, known as RNA-dependent RNA polymerase (RdRP), introduces mutations into newly synthesized genomes at a rate of roughly 10^{-4} per nucleotide. This mutation rate translates into at least one error per genome for most (+) viruses, meaning nearly all viral genomes received by an uninfected cell contain errors. These errors are randomly distributed through the entire genome, and their impacts on genome integrity range from lethal to highly beneficial. Mutations within *cis*-acting RNA elements yield phenotypes that are immediately acted on by natural selection. For example, a mutation causing disruption of an RNA motif required for replication leads to the immediate exclusion of the mutated genome from the replicable genome pool. By contrast, errors harming replication proteins, or other viral proteins for that matter, would have to be purged by a different mechanism. For simplicity, hereafter we limit the discussion to errors in the RdRP gene.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 March 2020 doi:10.20944/preprints202003.0021.v1

The central assumption of the IAS hypothesis is that nearly all of the internalized viral genomes will template the translation of proteins directly translatable from the genomic RNA, even though a majority of these RNA molecules do not have the chance to replicate. This is not really controversial because all kinds of non-replicating RNAs are known to undergo protein translation once inside a living cell. Moreover, the genomes of (+) RNA viruses need the replication proteins to be translated first in order to initiate replication. The only new proposition here is that aside from replication proteins, another class of proteins, designated bottleneck-enforcing proteins (BNEPs), are also translated from these genomic RNA molecules. In some viruses a single protein could act as both a replication protein and a BNEP (e.g. TCV p28). Another assumption, based on TCV p28 results, is that establishment of an operational bottleneck requires the viral BNEP to reach a certain concentration threshold [34,36].

With these assumptions in mind, let us first imagine a viral population that does not bottleneck due to the absence of a functional BNEP. As depicted in Fig. 1, this would mean virtually all genomes inside a cell could replicate themselves. Indeed even those containing fatal errors in RdRP can still replicate – by hitchhiking functional RdRP produced by sister genomes in the same cell. These fatal errors are thus retained in the pool of progeny genomes (Fig. 1, cell 1, column 2). To aggravate this problem, genomes that contain no errors in RdRP replicate to incur new errors in the progeny. Bear in mind that the number of new errors is not insignificant. Considering that (i), almost every newly replicated viral genome will contain one new error (see earlier), and (ii), a typical (+) RNA virus devotes at least one third of its genome to encode RdRP; we can expect that more than 30% of the newly synthesized genomes contain errors in RdRP. Since it is generally accepted that most mutations are deleterious [41,42], it would be rather generous to assume that one fifth of the errors in RdRP cause RdRP to malfunction. At this rate, approximately 6% of the newly synthesized genomes would be predicted to encode a defective RdRP. Therefore, even if there is just one cycle of replication in each cell, the repetition of replication in consecutively infected cells would still lead to fast increase of defective viral genomes harboring fatal

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 March 2020 doi:10.20944/preprints202003.0021.v1

RdRP errors (Fig. 1, cells 2 and 3), and simultaneous decrease of those still encoding intact RdRP. It is easy to see why such a population cannot survive for long even if the effect of host defense is not considered (Fig. 1).

Now imagine the intracellular bottlenecking is actively enforced by a BNEP, leaving as few as one viral genome in each cell to initiate replication in a stochastic manner (Fig. 2). Under this scenario, in most cells the genome(s) undergoing replication would probably contain no deleterious error in the RdRP, allowing replication to reiterate in successive cells (Fig. 2, A and B). In these cells the defective genomes would be purged by the bottleneck alone (those with a red star in Fig. 2, A and B). On the other hand, if a genome encoding a defective RdRP happens to pass through the bottleneck, it could still be amplified to high numbers in this cell because it has access to functional RdRP produced by sister genomes present in the same cell (Fig. 2, C1). However, the defective genome will be isolated in the next batch of cells because its progeny will be the exclusive residents of these new cells. It is in this next generation of cells that the defective genomes are eliminated by natural selection, as they can no longer hitchhike via functional RdRP produced by other genomes (Fig. 2, C2).

But there is an even more beautiful extrapolation of this model – it allows beneficial mutations to flourish! Extending from the logic of the previous paragraph, viral genomes with beneficial mutations are blocked from replication most of the time as well (Fig. 2, B and C). However, should one of such genomes escape from the bottleneck in one of the cells, its progeny will likewise become the exclusive residents of the next batch of cells becoming infected. Here the natural selection reveals its utmost elegance. Assuming this beneficial mutation enables the mutated RdRP to synthesize more genomic RNA copies than the original RdRP, in these cells this benefit is exclusively enjoyed by the very genomic RNA harboring this mutation! Therefore, the IAS model also ensures the proliferation of beneficial mutations by ensuring the genomes that carry these mutations become more numerous (Fig. 1, D1 and D2). In

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 March 2020 doi:10.20944/preprints202003.0021.v1

short, the intracellular bottlenecks of viral populations create the environment in which both purifying and positive selections reach their full potential.

In summary, the IAS hypothesis stipulates that (+) RNA viruses encode BNEPs that act in a concentration-dependent manner to establish intracellular population bottlenecks in order to limit the number of reproducing viral genomes in each cell. Counterintuitively, this model predicts that the larger the number of viral genomes (of the same virus) is in an infected cell, the more stringent is the bottleneck. Reiteration of such population bottlenecks in successive virus-infected cells isolates individual viral variants with mutations that either thwart or bolster the relative competitiveness of the virus. The isolated variant genomes can then be directly selected based on their ability to decrease or increase their own copy numbers. Such a mechanism synchronizes the unit of viral selection in most of the infected cells, and purges cheaters in the population in almost real-time. In conclusion, this model proposes a natural selection mechanism that operates in each of the infected cells to constantly surveying viral genome integrity and competitiveness.

A few pressing questions

#1: Wouldn't the intracellular population bottlenecks also act on newly synthesized progeny genomes? According to the IAS model, newly synthesized progeny genomes of (+) RNA viruses would be blocked from a second round of replication in the same cells because they would reinforce the bottlenecks by supplying more BNEPs. Nevertheless, our model does allow for limited occurrence of second cycle replication in cells invaded by very few viral genomes (Fig. 3), such as the very first cell(s) of a newly infected host. In these cells the concentration of viral BNEP would initially be short of the threshold needed for bottleneck establishment, due to the low number of genomes templating BNEP translation. This would not only permit the replication of multiple internalized genomes, but also allow some of the progeny genomes to undergo a second round of replication. However, the IAS model

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 March 2020 doi:10.20944/preprints202003.0021.v1

postulates that even in these cells, the second cycle of replication would likely take place with only a minority of newly synthesized progeny genomes (Fig. 3).

#2: Might the intracellular viral population bottlenecks be the root cause of superinfection exclusion (SIE)? The IAS model indeed predicts SIE as one of the manifestations of the intracellular viral population bottlenecks. This is because the narrow bottlenecks in most of the infected cells allow only very few viral genomes to replicate, even if they enter the cells at the same time. The bottlenecks are further tightened by active genome replication that produces more progeny genomes that in turn template the translation of more BNEPs. It is then no surprise that the superinfecting genomes will be excluded from replication in these cells.

#3: Isn't it a huge waste if most of viral genomes entering a cell are denied the chance to replicate (reproduce)? The idea of enlisting all genomic RNAs in a cell to enforce a bottleneck that blocks most of them from reproduction is not untenable. Indeed the entry of multiple genomes not only accelerates the establishment of bottlenecks – it also enables the rapid construction of intracellular structures needed for replication, known as replication organelles [43,44]. Thus, collaboration among multiple internalized genomes powers the swift mobilization of both structures, ensuring efficient and faithful reproduction of the virus. It follows that genes encoding such collaborative traits, include genes encoding BNEPs, but also those facilitating multi-virion entry, must be favored by natural selection. By contrast, since nearly all internalized genomes (except for those containing eventful mutations) encode genes that are phenotypically identical, exactly which one of them gets the chance to replicate is inconsequential to the evolutionary persistence of the virus. Indeed similar bottlenecks are also evident in single-celled bacteria like *Agrobacterium tumefaciens* – an overwhelming majority of bacterial cells that participate in the induction of crown galls become trapped inside the galls, with no chance to ever pass their genes to future generations [45]. Finally, bottlenecks of virus populations appear to be

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 March 2020 doi:10.20944/preprints202003.0021.v1

analogous to the bottlenecks encountered by germline cell populations of multi-cellular organisms that restrict the number of gametes that mate to produce viable offspring.

#4: How do the intracellular population bottlenecks relate to the tissue/organ level viral population bottlenecks? This question warrants a separate, detailed discussion, but the short answer is probably the former causes the latter. Very briefly, most tissues or organs in a host individual probably become infected by viruses amplified elsewhere in the same host. For example, a young plant leaf (YL) mostly becomes infected by viruses that underwent replication in an older leaf (OL). Due to bottlenecking events in the OL, different variants likely colonized separate cell clusters. Because these virus-inhabiting cell clusters are variously distributed throughout the OL, the ones with easier access to vascular bundles may release their variants for systemic spread earlier than others, delivering high numbers of the same variants to cells of SL ahead of other variants. The early-arriving variants would then establish intracellular bottlenecks in SL cells, and in turn exert SIE against the late-arriving variants.

#5: Does the intracellular viral population bottlenecking provide a plausible explanation for the *cis*-preference of some viral non-structural proteins (NSPs)? The answer would be yes if the IAS model holds. *Cis*-preference refers to the observation that some virus-encoded NSPs, especially those involved in genome replication, appear to preferentially or exclusively benefit the very genomic RNA from which they are translated [46–49]. This phenomenon was frequently observed when researchers tried to use one viral genome that encodes a functional NSP to complement a mutant genome encoding a defective or null form of the same NSP. Such complementation often led to preferential, or sometimes exclusive, accumulation of the former genome. According to our IAS model, one possibility is that the NSP may also serve as a BNEP, thus bottlenecking both genomes, especially when they entered the cell in high numbers. The former could more frequently evade the bottleneck due to its closer vicinity to low concentration of the NSP that is sufficient for initiating replication but insufficient for establishing the bottleneck.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 March 2020 doi:10.20944/preprints202003.0021.v1

#6: Why should we care about virus population bottlenecks? If high intracellular concentration of BNEP is indeed the culprit that causes viral populations to bottleneck, and such bottlenecks are essential for keeping viral mutation rate low, targeting BNEP for destruction or inactivation constitutes a previously unrecognized antiviral treatment strategy. Alternatively, the functionality of BNEP could be defeated by purposely introducing mutations into BNEP coding sequence. The resulting mutant viruses would be predicted to undergo transient infections in host cells before being overburdened by replication errors. Such viral mutants would be ideal live attenuated vaccines. Therefore, it is vitally important for us to have a thorough understanding of viral population bottlenecks in order to gain an upper hand in the constant battle against virus infections in plants, animals, and humans.

More questions, testable predictions, and future directions

There are obviously many more questions that we cannot answer at this point. For example, exactly how many viral genomes would be allowed to replicate in each cell? Or is the stochastic bottlenecking even capable of very precise regulation? Or is such precision even needed? How do the bottlenecks regulate the replication of multi-partite viruses made up of multiple genome segments? However, the IAS model does lead to predictions that may be easily testable, especially with the advent of the single-cell, high-throughput sequencing technology. For example, can we test the predication of single-cycle replication by examining the viral progenies in individual cells? Testing this and other predictions of the IAS model promises to open new opportunities to improve our understanding of viral evolution, and contribute to novel antiviral therapeutics.

Acknowledgement

We are thankful to current and former members of Qu lab for stimulating discussions. Research in our labs are supported by grants from the National Science Foundation, The US Department of

Agriculture Hatch funds, as well as grants from the United Soybean Board, the North Central Soybean Research Program, the Ohio Soybean Council, and the SEEDS awards of the Ohio Agricultural Research and Development Center.

References

1. Tester, D. J. *et al.* Identification of a Novel Homozygous Multi-Exon Duplication in RYR2 Among Children With Exertion-Related Unexplained Sudden Deaths in the Amish Community. *JAMA Cardiol.* (2020) doi:10.1001/jamacardio.2019.5400.
2. Sacristán, S., Malpica, J. M., Fraile, A. & García-Arenal, F. Estimation of Population Bottlenecks during Systemic Movement of Tobacco Mosaic Virus in Tobacco Plants. *J. Virol.* **77**, 9906 (2003).
3. Frost, S. D. W., Dumaurier, M.-J., Wain-Hobson, S. & Brown, A. J. L. Genetic drift and within-host metapopulation dynamics of HIV-1 infection. *Proc. Natl. Acad. Sci.* **98**, 6975 (2001).
4. Hall, J. S., French, R., Hein, G. L., Morris, T. J. & Stenger, D. C. Three Distinct Mechanisms Facilitate Genetic Isolation of Sympatric Wheat Streak Mosaic Virus Lineages. *Virology* **282**, 230–236 (2001).
5. French, R. & Stenger, D. C. Evolution of wheat streak mosaic virus: Dynamics of population growth within plants may explain limited variation. *Annu. Rev. Phytopathol.* **41**, 199–214 (2003).
6. Li, H. & Roossinck, M. J. Genetic Bottlenecks Reduce Population Variation in an Experimental RNA Virus Population. *J. Virol.* **78**, 10582–10587 (2004).
7. Zhang, X.-F. *et al.* Random Plant Viral Variants Attain Temporal Advantages During Systemic Infections and in Turn Resist other Variants of the Same Virus. *Sci. Rep.* **5**, (2015).
8. Forrester, N. L., Guerbois, M., Seymour, R. L., Spratt, H. & Weaver, S. C. Vector-Borne Transmission Imposes a Severe Bottleneck on an RNA Virus Population. *PLOS Pathog.* **8**, e1002897 (2012).

9. Pfeiffer, J. K. Innate host barriers to viral trafficking and population diversity: lessons learned from poliovirus. *Adv. Virus Res.* **77**, 85–118 (2010).
10. Kuss, S. K., Etheredge, C. A. & Pfeiffer, J. K. Multiple Host Barriers Restrict Poliovirus Trafficking in Mice. *PLOS Pathog.* **4**, e1000082 (2008).
11. Grubaugh, N. D. *et al.* Genetic Drift during Systemic Arbovirus Infection of Mosquito Vectors Leads to Decreased Relative Fitness during Host Switching. *Cell Host Microbe* **19**, 481–492 (2016).
12. Varble, A. *et al.* Influenza A Virus Transmission Bottlenecks Are Defined by Infection Route and Recipient Host. *Cell Host Microbe* **16**, 691–700 (2014).
13. McCrone, J. T. *et al.* Stochastic processes constrain the within and between host evolution of influenza virus. *eLife* **7**, e35962 (2018).
14. Miyashita, S., Ishibashi, K., Kishino, H. & Ishikawa, M. Viruses Roll the Dice: The Stochastic Behavior of Viral Genome Molecules Accelerates Viral Adaptation at the Cell and Tissue Levels. *PLOS Biol.* **13**, e1002094 (2015).
15. Sanjuán, R. Collective Infectious Units in Viruses. *Trends Microbiol.* **25**, 402–412 (2017).
16. Leeks, A., Sanjuán, R. & West, S. A. The evolution of collective infectious units in viruses. *Virus Res.* **265**, 94–101 (2019).
17. Chen, P., Hübner, W., Spinelli, M. A. & Chen, B. K. Predominant Mode of Human Immunodeficiency Virus Transfer between T Cells Is Mediated by Sustained Env-Dependent Neutralization-Resistant Virological Synapses. *J. Virol.* **81**, 12582 (2007).
18. Hübner, W. *et al.* Quantitative 3D Video Microscopy of HIV Transfer Across T Cell Virological Synapses. *Science* **323**, 1743 (2009).
19. Del Portillo, A. *et al.* Multiploid Inheritance of HIV-1 during Cell-to-Cell Infection. *J. Virol.* **85**, 7169 (2011).

20. Robinson, S. M. *et al.* Coxsackievirus B Exits the Host Cell in Shed Microvesicles Displaying Autophagosomal Markers. *PLOS Pathog.* **10**, e1004045 (2014).
21. Chen, Y.-H. *et al.* Phosphatidylserine Vesicles Enable Efficient En Bloc Transmission of Enteroviruses. *Cell* **160**, 619–630 (2015).
22. Cifuentes-Muñoz, N., Dutch, R. E. & Cattaneo, R. Direct cell-to-cell transmission of respiratory viruses: The fast lanes. *PLOS Pathog.* **14**, e1007015 (2018).
23. Martinez, M. G. & Kielian, M. Intercellular Extensions Are Induced by the Alphavirus Structural Proteins and Mediate Virus Transmission. *PLOS Pathog.* **12**, e1006061 (2016).
24. Feng, Z. *et al.* A pathogenic picornavirus acquires an envelope by hijacking cellular membranes. *Nature* **496**, 367–371 (2013).
25. Santiana, M. *et al.* Vesicle-Cloaked Virus Clusters Are Optimal Units for Inter-organismal Viral Transmission. *Cell Host Microbe* **24**, 208-220.e8 (2018).
26. Miyashita, S. & Kishino, H. Estimation of the size of genetic bottlenecks in cell-to-cell movement of soil-borne wheat mosaic virus and the possible role of the bottlenecks in speeding up selection of variations in trans-acting genes or elements. *J. Virol.* **84**, 1828–1837 (2010).
27. Kobiler, O., Lipman, Y., Therkelsen, K., Daubechies, I. & Enquist, L. W. Herpesviruses carrying a Brainbow cassette reveal replication and expression of limited numbers of incoming genomes. *Nat. Commun.* **1**, 146 (2010).
28. Dietrich, C. & Maiss, E. Fluorescent labelling reveals spatial separation of potyvirus populations in mixed infected Nicotiana benthamiana plants. *J. Gen. Virol.* **84**, 2871–2876 (2003).
29. Takahashi, T. *et al.* Analysis of the Spatial Distribution of Identical and Two Distinct Virus Populations Differently Labeled with Cyan and Yellow Fluorescent Proteins in Coinfected Plants. *Phytopathology* **97**, 1200–1206 (2007).

30. Tatineni, S. & French, R. The Coat Protein and NIa Protease of Two Potyviridae Family Members Independently Confer Superinfection Exclusion. *J. Virol.* **90**, 10886–10905 (2016).
31. Zhou, X., Sun, K., Zhou, X., Jackson, A. O. & Li, Z. The matrix protein of a plant rhabdovirus mediates superinfection exclusion by inhibiting viral transcription. *J. Virol.* JVI.00680-19 (2019) doi:10.1128/JVI.00680-19.
32. Zwart, M. P., Daròs, J.-A. & Elena, S. F. One Is Enough: In Vivo Effective Population Size Is Dose-Dependent for a Plant RNA Virus. *PLOS Pathog.* **7**, e1002122 (2011).
33. Tromas, N., Zwart, M. P., Lafforgue, G. & Elena, S. F. Within-Host Spatiotemporal Dynamics of Plant Virus Infection at the Cellular Level. *PLOS Genet.* **10**, e1004186 (2014).
34. Zhang, X.-F. *et al.* A self-perpetuating repressive state of a viral replication protein blocks superinfection by the same virus. *PLOS Pathog.* **13**, e1006253 (2017).
35. R. Hull, S. N. C., P. Dale. Genetically modified plants and the 35S promoter: assessing the risks and enhancing the debate. *Microb. Ecol. Health Dis.* **12**, 1–5 (2000).
36. Guo, Q. *et al.* Superinfection Exclusion by p28 of Turnip Crinkle Virus Is Separable from Its Replication Function. *Mol. Plant-Microbe Interactions®* MPMI-09-19-0258-R (2019) doi:10.1094/MPMI-09-19-0258-R.
37. Sicard, A. *et al.* A multicellular way of life for a multipartite virus. *eLife* **8**, e43599 (2019).
38. Folimonova, S. Y. Superinfection Exclusion Is an Active Virus-Controlled Function That Requires a Specific Viral Protein. *J. Virol.* **86**, 5554 (2012).
39. Zou, G. *et al.* Exclusion of West Nile Virus Superinfection through RNA Replication. *J. Virol.* **83**, 11765 (2009).
40. Webster, B., Ott, M. & Greene, W. C. Evasion of Superinfection Exclusion and Elimination of Primary Viral RNA by an Adapted Strain of Hepatitis C Virus. *J. Virol.* **87**, 13354–13369 (2013).

41. Duffy, S. Why are RNA virus mutation rates so damn high? *PLoS Biol.* **16**, e3000003–e3000003 (2018).
42. Elena, S. F. & Sanjuán, R. Adaptive value of high mutation rates of RNA viruses: separating causes from consequences. *J. Virol.* **79**, 11555–11558 (2005).
43. Ertel, K. J. *et al.* Cryo-electron tomography reveals novel features of a viral RNA replication compartment. *Elife* **6**, (2017).
44. Xu, K. & Nagy, P. D. Enrichment of Phosphatidylethanolamine in Viral Replication Compartments via Co-opting the Endosomal Rab5 Small GTPase by a Positive-Strand RNA Virus. *PLOS Biol.* **14**, e2000128 (2016).
45. Dawkins, R., 1941-. *The selfish gene*. (New edition. Oxford ; New York : Oxford University Press, 1989., 1989).
46. Novak, J. E. & Kirkegaard, K. Coupling between genome translation and replication in an RNA virus. *Genes Dev.* **8**, 1726–1737 (1994).
47. White, K. A., Skuzeski, J. M., Li, W., Wei, N. & Morris, T. J. Immunodetection, Expression Strategy and Complementation of Turnip Crinkle Virus p28 and p88 Replication Components. *Virology* **211**, 525–534 (1995).
48. Okamoto, K. *et al.* cis-Preferential requirement of a −1 frameshift product p88 for the replication of Red clover necrotic mosaic virus RNA1. *Virology* **375**, 205–212 (2008).
49. Kawamura-Nagaya, K., Ishibashi, K., Huang, Y.-P., Miyashita, S. & Ishikawa, M. Replication protein of tobacco mosaic virus cotranslationally binds the 5' untranslated region of genomic RNA to enable viral replication. *Proc. Natl. Acad. Sci.* **111**, E1620–E1628 (2014).

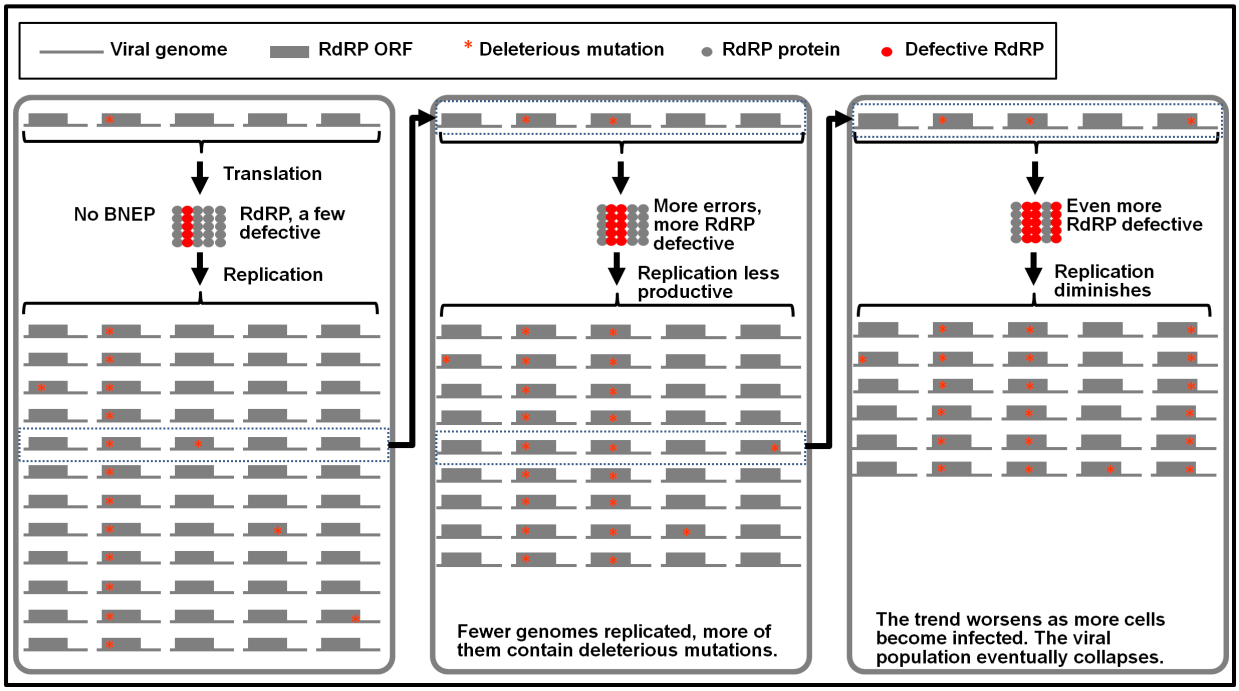
Figure legends

 doi:10.20944/preprints202003.0021.v1

501 **Figure 1**. Failure to establish population bottlenecks leads to fast accumulation of deleterious
502 mutations as viral infection progresses in consecutive cells. The drawings are schematic depiction of
503 three representative cells. Meanings of the symbols used are given on the top of the cells. Note that
504 new red stars at different positions of RdRP ORF denote new mutations incurred during replication.
505 Abbreviations: RdRP, RNA-dependent RNA polymerase; ORF, open reading frame; BNEP, bottleneck-
506 enforcing protein. For simplicity, only five founding genomes were drawn to represent the dozens, if not
507 hundreds, of viral genomes received by most of the susceptible cells in a typical host individual.
508 Furthermore, our discussions also omitted detrimental, yet non-lethal mutations.
509 **Figure 2**. Anticipated fates for different mutations in the RdRP gene incurred during viral replication
510 under the IAS model. The meanings of all symbols are given in the top left box. Note that replication in
511 all cells gives rise new mutations that are depicted as either red (deleterious) or yellow (beneficial)
512 mutations at different positions of the RdRP ORF.
513 **Figure 3**. Limited 2nd cycle replication in cells entered by a very small number (<10) of viral genomes.
514

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 March 2020 doi:10.20944/preprints202003.0021.v1

Qu et al 2020 Figure 1

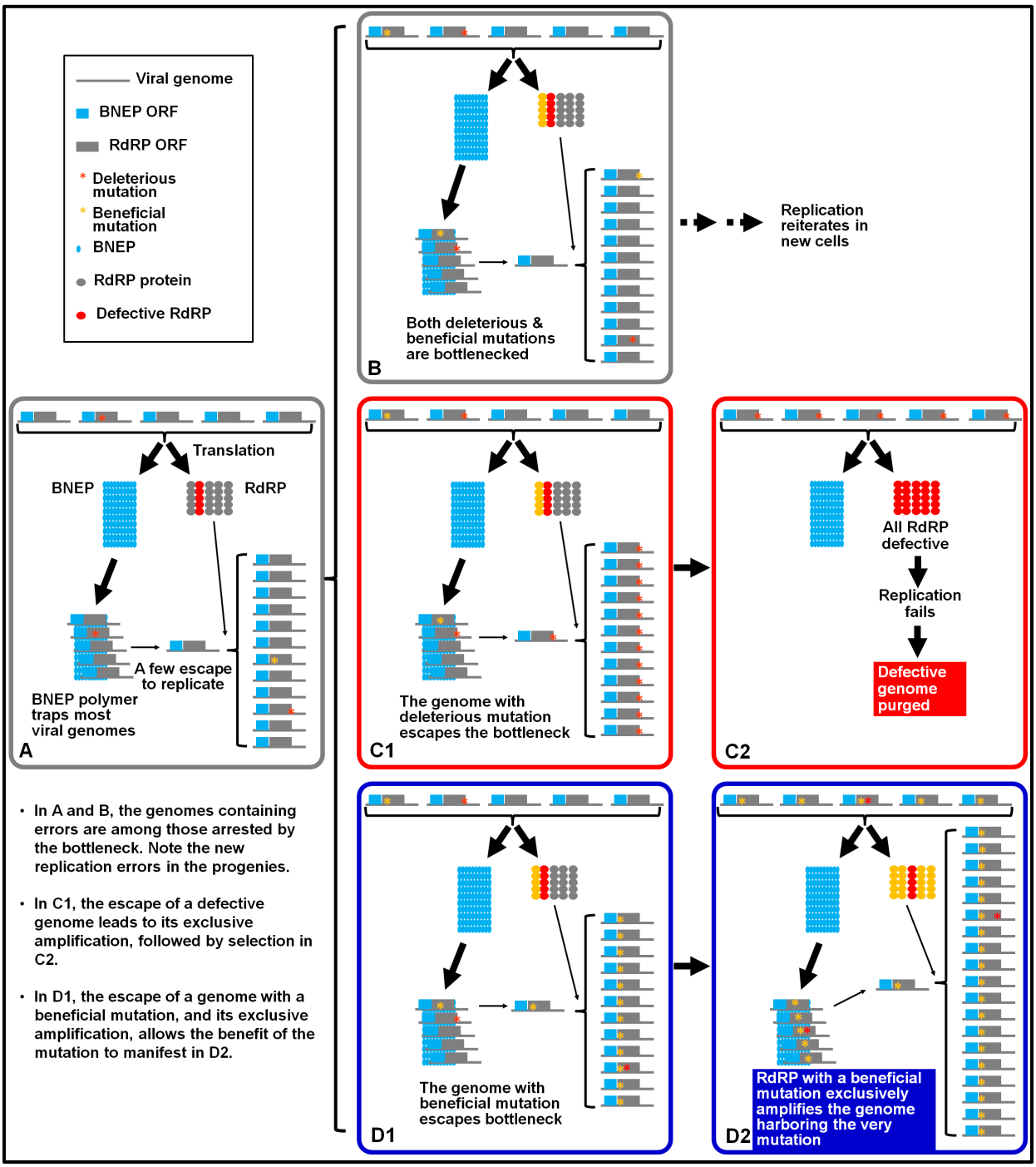


515

516

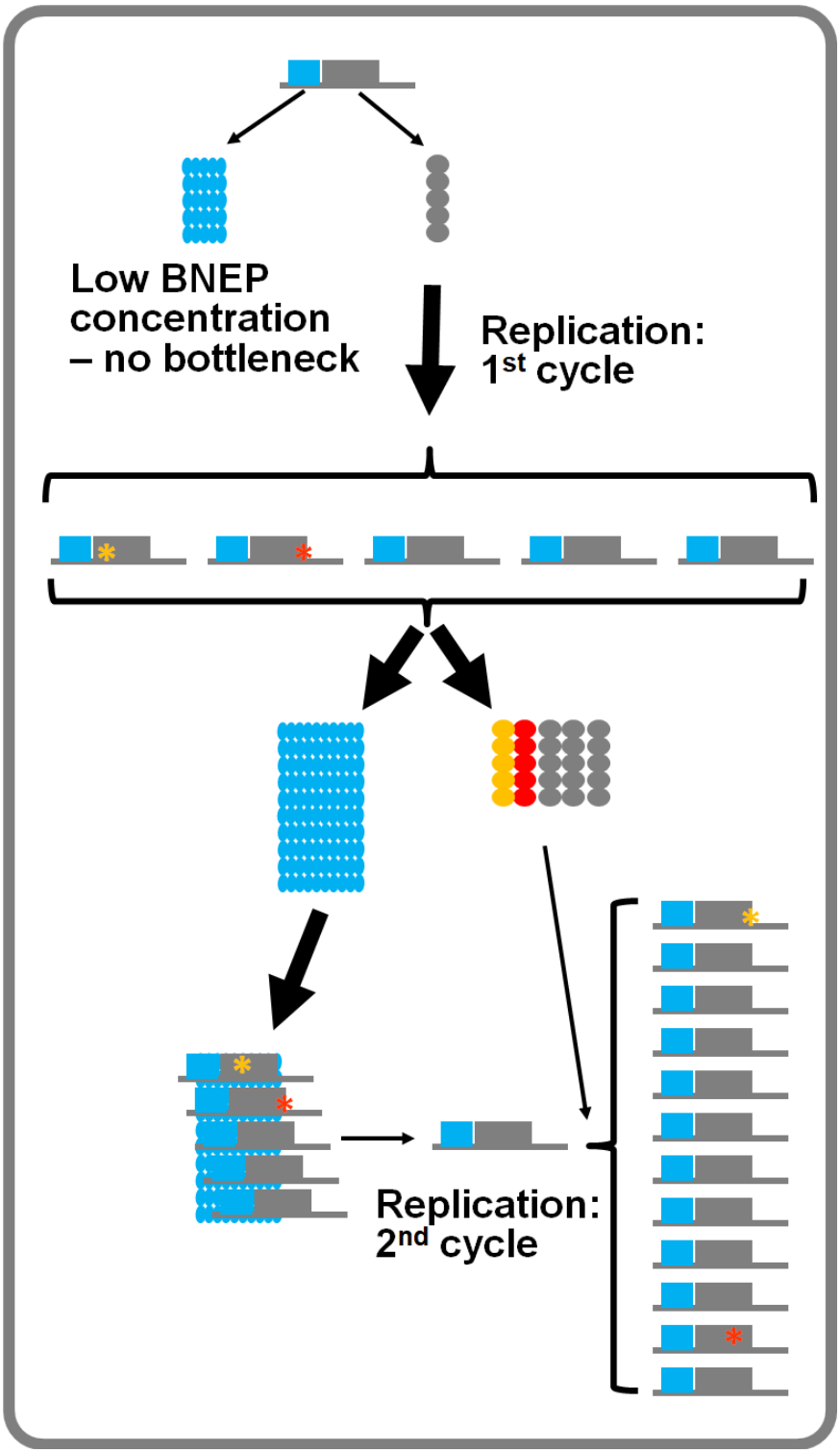
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 March 2020 doi:10.20944/preprints202003.0021.v1

Qu et al 2020 Figure 2



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 March 2020 doi:10.20944/preprints202003.0021.v1

Qu et al 2020 Figure 3



519