

Population bottlenecks in multicomponent viruses: First forays into the uncharted territory of genome-formula variation

Serafin Gutierrez^{1,2} & Mark P. Zwart³

¹ CIRAD, UMR ASTRE, F-34398 Montpellier, France ² ASTRE, CIRAD, INRA, Univ Montpellier, Montpellier, France ³ Department of Microbial Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, The Netherlands. Corresponding author: E-mail: serafin.gutierrez@cirad.fr

Current Opinion in Virology **33**:184–190 (2018), edited by Stéphane Blanc and Yannis Michalakis; <https://doi.org/10.1016/j.coviro.2018.09.001>

Abstract: Multicomponent viral systems face specific challenges when enduring population bottlenecks. These systems can lose coding information due to the lack of co-encapsidation of all the genetic information, at least in a proportion of the capsids in a population. Moreover, bottlenecks can also impact one of the main potential advantages of multicomponent systems: the regulation of gene expression through changes in gene copy frequencies at the population level. How these systems cope with population bottlenecks is far from being clear. Here, two non-exclusive scenarios are described. In the first scenario, population bottlenecks during host infection allow for the isolation of within-host populations with different gene frequencies, leaving the door opened for the selection of populations with adaptive gene frequencies. The second scenario postulates that viruses could influence bottleneck size, at least at certain steps of their life cycle, to limit random changes in gene frequencies. Examples of viral mechanism impacting bottleneck size at cell infection are available and, intriguingly, they can lead to either increases or reductions in bottleneck size. This situation opens the way for putative tradeoffs on both gene frequencies and bottleneck sizes that could differ among multicomponent systems.

Variation in gene copy number in virus populations

Viruses display an extremely diverse array of mechanisms for the regulation of gene expression. Nevertheless, copy number variation (CNV) is too seldom considered as a phenomenon taking place in viruses [1]. Copy number variants (CNVs) occur when a genomic region is repeated through recombination. The generated genotypes thus differ in the number of copies of the genes within this region compared to the wild type. Differences in copy number of a given gene can alter its expression and thus the phenotype. In general, a higher number of gene copies will lead to higher messenger and protein levels [2] – although theory indicates this may not always be the case [3]. Through its impact on gene expression, CNV plays an important role in the adaptation of organisms to changing environments [2,4–8], and their role in bacterial adaptation has been particularly well documented [8]. In addition, CNV can affect the mutation supply and hereby the chance that a beneficial mutation will be sampled, with a higher copy number leading to a larger mutation supply [6,8,9]. Finally, CNV can affect robustness and evolvability, as it can introduce genetic redundancy and hereby allow one or more gene copies to explore sequence space more effectively [2,7].

Pioneering work on poxviruses, a family of large DNA viruses with a monopartite genome, has shown that CNV occurs in viruses and can play an important role in adaptation. Experimental evolution of a poxvirus has uncovered an evolutionary pattern called a genomic accordion, previously observed in other organisms [8]: rapid selection for CNVs that amplify gene copies, followed by the occurrence of beneficial point mutations at these loci and the subsequent loss of multiple gene copies [10–12]. Nevertheless, many viruses have small and highly-streamlined genomes [13], and gene duplications are rare [14]. Moreover, when duplications are generated in the laboratory, they tend to have strong deleterious effects on viral fitness and are quickly purged from evolving populations [15]. Therefore, CNV probably plays a role in the adaptation of only those viruses with large DNA genomes (e.g. poxviruses).

However, recent work has shown that viruses with small genomes could also display a hitherto unknown and fascinating version of CNV. This new version operates, beyond the genotype level, at the within-host population level. Sicard and colleagues showed in a groundbreaking study [16] that populations of a nanovirus contain differences in copy number among the different viral genes. Nanoviruses possess a multipartite genome consisting of several segments. Contrary to segmented viruses (e.g. Influenza A virus), each segment of a multipartite virus is encapsidated alone in a virus particle [1,17,18]. These authors found that the differences in copy number among different nanovirus segments are not random: they are host-specific and reproducible no matter their value at inoculation [16,19]. The relative copy

numbers of the viral genes were called the genome formula, and the genome formula specific to a given host or environment was named the setpoint genome formula. In addition, they showed that within-host populations show higher accumulation when segment frequencies are closer to the setpoint genome formula, thus pointing to a correlation between the genome formula and virus fitness. Later work has shown similar results in other multipartite viruses, with now examples from both DNA and RNA viruses [16,20,21]. These variations in gene frequencies among virus populations could lead to gene expression regulation, and thus to phenotype optimization, similarly to CNV [1,16,20]. Since the term CNV is usually reserved to describe structural variation within a genome, here we will call this new type of population-level variation genome-formula variation (GFV).

The suitability of multipartite genomes for GFV control has been postulated as one of the few hypotheses specifically explaining the evolution of this genome organization. Whether GFV takes place in the two other viral genome architectures, the segmented and monopartite genomes, can be a matter of debate. Currently, there is a single example of a behavior reminiscent of GFV for each of these architectures. First, segmented viruses have multiple genome segments, each bearing different genes, which are packaged into the same virus particle. It has been shown that the relative frequencies of the segments in populations of Influenza A virus (IAV) can differ in a host-specific manner and that such differences can be adaptive [22–24]. Contrary to multipartite and segmented viruses, the genome of monopartite viruses is composed of a single molecule bearing all the genes and *stricto sensu* GFV is not possible. However, wild populations of a baculovirus, an insect virus with a monopartite genome, contain a proportion of non-infectious genotypes with deletions in the same genomic region. These populations thus contain differences in frequency among virus genes and these differences have been shown to increase viral fitness through an increase in infectivity of a mixed-genotype population [25,26]. These two examples thus show populations with uneven gene frequencies that are linked to viral fitness, just as expected in a GFV context. Given the paucity of examples from segmented and monopartite viruses, more studies are required to determine the potential conservation of GFV in genome architectures beyond the multipartite one.

An uneven genome formula may provide a benefit through the regulation of gene expression. However, it may also come to a cost: all other things equal, the more uneven a genome formula, the more likely the genes with low frequencies are lost during the population bottlenecks that virus populations usually encounter along their cycle.

Population bottlenecks and genome-formula drift

As obligate intra-cellular parasites, one of the main hurdles viruses must constantly surmount is transmission between the intracellular environments that support their replication. There are barriers to virus transmission at the between-cell, between-organ and between-host levels [27–30]. Even if there is a large number of virus particles present, typically only a small number of founders will effectively contribute to new infections because transmission is difficult, resulting in a population bottleneck. There are different approaches to demonstrate population bottlenecks and to estimate the number of infection founders [30–37]. Many different laboratory studies have shown that population bottlenecks are common at different steps of the viral life cycle and that they occur for a wide range of viruses [27,28]. Recent studies have highlighted that population bottlenecks also occur in vertical transmission [38], and that population bottlenecks are common in natural virus populations [39]. Although severe bottlenecks—in the order of units—are common in the current literature, it is important to note that bottleneck estimates are available for a limited number of viruses [27,28,40]. As more and more estimates become available, less severe bottlenecks have also been observed at all steps of viral cycles in different viruses [30,35,41–43]. Virus particle concentration, host susceptibility and the route of transmission probably all play important roles in determining bottleneck size [30,43,44], so although bottlenecks can be narrow, they need not always be in practice.

One important consequence of population bottlenecks is genetic drift: stochastic changes in allele frequencies (Figure 1). Genetic drift can lead to losses in genetic diversity, which in turn can have a wide range of effects on evolution [28]. The main negative effects of population bottlenecks are the fixation of deleterious mutations [45–47] and the loss of *de novo* beneficial variants [48]. By contrast, positive effects of bottlenecks include permitting selection for traits that act *in trans* [36], including the removal of defective genotypes [49]. These effects occur because narrow bottlenecks reduce possibilities for coinfections between genotypes, and hence selection can then act directly on the traits associated with a single genotype. A further advantage is more efficient adaptation on rugged fitness landscapes [50]. For multipartite viruses, population bottlenecks cannot only result in genetic drift but also in random changes in the genome formula. These changes occur because only a small number of virus particles contributes to infection. In the extreme, the frequency of one or more genes is zero and there is no infection. But even if all genes pass through a bottleneck and infection proceeds, the frequency at which these genes are present in the founding population can change from the setpoint genome formula [16] (Figure 1b). In the absence of any existing term for these changes in population composition, we propose to call such random changes in gene frequencies ‘genome-formula drift’. Genome-formula drift is therefore the random effects of

population bottlenecks on gene frequencies, analogous to genetic drift being the effect of population bottlenecks on allele frequencies. As such, we see genome-formula drift as a stochastic event and a non-directional evolutionary force. We expect that genetic drift and genome-formula drift will occur simultaneously (Figure 1c).

But what do we know about bottlenecks and multicomponent viruses? There are a limited number of bottleneck estimates for viruses showing a GFV-like behavior, even including non-multipartite examples. In most cases, bottleneck sizes are below 10 units no matter whether bottlenecks take place at cell infection [36,51,52], host colonization [53] or between-host transmission [53,54]. Only recent estimates of population sizes during between-host transmission of IAV show more relaxed bottlenecks, in the order of hundreds [43]. Note also that for the studies concerning multipartite viruses [36,53,54], bottleneck sizes are estimates of the number of founders for a single segment, meaning that the total number of particles, all genomic segments considered, will be higher. Overall, these few estimates suggest that the genome formula probably varies during bottlenecks, although more quantitative data are dearly needed to improve our understanding on the impact of bottlenecks on GFV.

Regulation of GFV along the viral cycle

If bottlenecks at different steps of the viral cycle lead to genome-formula drift and hereby potentially move a population at equilibrium away from the setpoint genome formula, how can populations maintain the setpoint genome formula? One possibility is that GFV requires bottlenecks to be generated in the first place [16,55] (Figure 1b). On the one hand, although GFV can move a population away from the setpoint genome formula, it also provides the variation on which selection acts to evolve to the setpoint. Under this hypothesis, bottlenecks during host infection lead to different populations within a host, a situation commonly observed in viruses (e.g. Ref. [56]). These populations will presumably differ in gene frequencies due to random processes behind GFV generation. Thus, selection will operate on the resulting diversity in genome formula leading to the increase in frequency of within-host populations with the most beneficial genome formula, that is, the setpoint genome formula. These beneficial frequencies will be driven in part by the regulation of gene expression and selection to maintain infectivity, in a frequency-dependent manner [16,20]. This scenario would also imply that GFV presents a limitation compared to other mechanisms regulating gene expression. Periodic bottlenecks would lead to genome formulae diverging from the setpoint, leading to transient periods of suboptimal regulation of gene expression. This situation does not take place in other regulatory mechanisms, for example promoter-based regulation of gene transcription.

Another possible scenario is that viruses actively regulate population bottlenecks to maintain the setpoint formula, at least at certain steps of their cycle. Larger population sizes at infection can be achieved through collective transmission [57]. This phenomenon implies that virus spread is mediated by structures that contain several viral particles, thus ensuring larger population sizes at cell infection. There are more and more examples of viruses using different mechanisms for collective transmission [57,58]. However, most viruses display mechanisms that have exactly the opposite effect: they reduce the size of colonizing populations. This limitation in population size is due to superinfection exclusion, a phenomenon in which a preexisting viral infection in a cell prevents a secondary infection. Again, the mechanisms underlying superinfection exclusion are very diverse and have been observed in an extremely large array of viruses (e.g. see Refs. [59–65]), including examples of viruses with a GFV-like behavior [66,67]. By tuning collective transmission and superinfection exclusion, multicomponent viruses could affect bottlenecks, either increasing or decreasing colonizing populations in different cell types or at different steps of the virus cycle. Further studies will surely deepen our understanding of the putative role of collective transmission and superinfection exclusion in GFV.

GFV and virus evolution

In bacteria and large DNA viruses, CNV has been shown to play a role in adaptation on intermediate time scales, whereas point mutations with larger benefits typically only occur and are fixed later, due to low mutation rates and the paucity of beneficial mutations [8,10,11]. It is not known yet whether similar patterns also take place in the evolution of virus showing GFV. Similarly to CNV, GFV can facilitate phenotypic plasticity but point mutations could bring about similar changes in a manner that is less costly. If gene expression needs to be upregulated, it seems more efficient to have—for example—a point mutation that increases promoter activity x -fold than to have an x -fold increase in the frequency of the gene in the population. This effect occurs because of possible losses of infectivity as genes with low frequencies can be lost during bottlenecks [68] (Figure 2). GFV adaptation seems a rapid but coarse regulatory mechanism, whereas mutation-driven adaptation would be less costly. Experimental evolution could help to define whether, in a new but constant environment, a multicomponent virus adapts only by GFV or, alternatively, in a two-step process involving initially GFV and later point mutations optimizing another mechanism regulating gene expression.

GFV has been suggested to influence evolution in different manners. The uneven genome formulae in GFV should lead to different evolution rates among genes through

different mechanisms [68]. First, genes at higher frequency would have a greater mutation supply, which can in turn drive more rapid evolution. Second, opportunities for intra-specific genetic exchange should also increase with gene frequency. Finally, the strength of genetic drift for each gene can differ depending on the gene frequency [53]. Excitingly, recent work on nanoviruses has shown that this relationship can be rather complex: estimates of the number of founders for different genome segments clearly depended on the gene frequency, whereas dose-dependent effects were seen for only one gene [53]. Such varying tempo and mode of evolution for the different genes of a multicomponent virus could lead to a wealth of interesting interactions, one of which might be loose associations between segments on longer evolutionary time scales [69].

Acknowledgements

SG acknowledges support from grant CuliOme (ANHIWA ERA-net) from the European Commission. MZ acknowledges a VIDI grant from the Netherlands Organisation for Scientific Research (NWO).

References

1. Sicard A, Michalakis Y, Gutiérrez S, Blanc S: The strange lifestyle of multipartite viruses. *PLoS Pathog* 2016, **12**:1-19.
2. Kondrashov FA: Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc R Soc B Biol Sci* 2012, **279**:5048-5057.
3. Mileyko Y, Joh RI, Weitz JS: Small-scale copy number variation and large-scale changes in gene expression. *Proc Natl Acad Sci U S A* 2008, **105**:16659-16664.
4. DeLuna A, Vetsigian K, Shoresh N, Hegreness M, Colón-González M, Chao S, Kishony R: Exposing the fitness contribution of duplicated genes. *Nat Genet* 2008, **40**:676-681.
5. Stevenson BS, Schmidt TM: Life history implications of rRNA gene copy number in *Escherichia coli*. *Appl Environ Microbiol* 2004, **70**:6670-6677.
6. San Millan A, Escudero JA, Gifford DR, Mazel D, MacLean RC: Multicopy plasmids potentiate the evolution of antibiotic resistance in bacteria. *Nat Ecol Evol* 2017, **1**:0010.
7. Ohno S: Evolution by Gene Duplication. Springer; 1970.
8. Sandegren L, Andersson DI: Bacterial gene amplification: implications for the evolution of antibiotic resistance. *Nat Rev Microbiol* 2009, **7**:578-588.
9. Rodriguez-Beltran J, Hernandez-Beltran JCR, DelaFuente J, Escudero JA, Fuentes-Hernandez A, Craig MacLean R, Pena-Miller R, San Millan A: Multicopy plasmids allow bacteria to escape from fitness trade-offs during evolutionary innovation. *Nat Ecol Evol* 2018, **2**:873-881.
10. Elde NC, Child SJ, Eickbush MT, Kitzman JO, Rogers KS, Shendure J, Geballe AP, Malik HS: Poxviruses deploy genomic accordions to adapt rapidly against host antiviral defenses. *Cell* 2012, **150**:831-841.
11. Cone KR, Kronenberg ZN, Yandell M, Elde NC: Emergence of a viral RNA polymerase variant during gene copy

- number amplification promotes rapid evolution of vaccinia virus. *J Virol* 2017, **91**:e01428-16.
12. Bayer A, Brennan G, Geballe A: Adaptation by copy number variation in monopartite viruses. *Curr Opin Virol* 2018, **33**:42-46.
 13. Lynch M: Streamlining and simplification of microbial genome architecture. *Annu Rev Microbiol* 2006, **60**:327-349.
 14. Simon-Loriere E, Holmes EC: Gene duplication is infrequent in the recent evolutionary history of RNA viruses. *Mol Biol Evol* 2013, **30**:1263-1269.
 15. Willemsen A, Zwart MP, Higuera P, Sardanyés J, Elena SF: Predicting the stability of homologous gene duplications in a plant RNA virus. *Genome Biol Evol* 2016, **8**:3065-3082.
 16. Sicard A, Yvon M, Timchenko T, Gronenborn B, Michalakis Y, Gutiérrez S, Blanc S: Gene copy number is differentially regulated in a multipartite virus. *Nat Commun* 2013, **4**:1-8.
 17. Fulton RW: The effect of dilution on Necrotic ringspot virus infectivity and the enhancement of infectivity by noninfective virus. *Virology* 1962, **18**:477-485.
 18. Ladner JT, Wiley MR, Beitzel B, Kramer LD, Tesh RB, Palacios G, Auguste AJ, Dupuis li AP, Lindquist ME, Sibley SD et al.: A multicomponent animal virus isolated from mosquitoes. *Cell Host Microbe* 2016, **20**:1-11.
 19. Sicard A, Zeddoum J-L, Yvon M, Michalakis Y, Gutiérrez S, Blanc S: Circulative nonpropagative aphid transmission of nanoviruses: an oversimplified view. *J Virol* 2015, **89**: 9719-9726.
 20. Wu B, Zwart MP, Sánchez-Navarro JA, Elena SF: Within-host evolution of segments ratio for the tripartite genome of alfalfa mosaic virus. *Sci Rep* 2017, **7**:5004.
 21. Hu Z, Zhang X, Liu W, Zhou Q, Zhang Q, Li G, Yao Q: Genome segments accumulate with different frequencies in *Bombyx mori* bidensovirus. *J Basic Microbiol* 2016, **56**:1338-1343.
 22. Brooke CB, Ince WL, Wei J, Bennink JR, Yewdell JW: Influenza A virus nucleoprotein selectively decreases neuraminidase gene-segment packaging while enhancing viral fitness and transmissibility. *Proc Natl Acad Sci* 2014, **111**:16854-16859.
 23. Brooke CB, Ince WL, Wrarmert J, Ahmed R, Wilson PC, Bennink JR, Yewdell JW: Most influenza A virions fail to express at least one essential viral protein. *J Virol* 2013, **87**:3155-3162.
 24. Diefenbacher M, Sun J, Brooke C: The parts are greater than the whole: the role of semi-infectious particles in influenza A virus biology. *Curr Opin Virol* 2018, **33**:42-46.
 25. Simón O, Williams T, Caballero P, López-Ferber M: Dynamics of deletion genotypes in an experimental insect virus population. *Proc Biol Sci* 2006, **273**:783-790.
 26. Clavijo G, Williams T, Simon O, Munoz D, Cerutti M, Lopez-Ferber M, Caballero P: Mixtures of complete and *pif1*- and *pif2*- deficient genotypes are required for increased potency of an insect nucleopolyhedrovirus. *J Virol* 2009, **83**:5127-5136.
 27. Gutiérrez S, Michalakis Y, Blanc S: Virus population bottlenecks during within-host progression and host-to-host transmission. *Curr Opin Virol* 2012, **2**:546-555.
 28. Zwart MP, Elena SF: Matters of size: genetic bottlenecks in virus infection and their potential impact on evolution. *Annu Rev Virol* 2015, **2**:161-179.
 29. Tromas N, Zwart MP, Lafforgue G, Elena SF: Within-host spatiotemporal dynamics of plant virus infection at the cellular level. *PLoS Genet* 2014, **10**:e1004186..
 30. Gutiérrez S, Pirolles E, Yvon M, Baecker V, Michalakis Y, Blanc S: The multiplicity of cellular infection changes depending on the route of cell infection in a plant virus. *J Virol* 2015, **89**: 9665-9675.
 31. Sacristán S, Malpica JM, Fraile A, García-Arenal F: Estimation of population bottlenecks during systemic movement of tobacco mosaic virus in tobacco. *Plants* 2003, **77**:9906-9911.
 32. Smith IR, Crook NE: In vivo isolation of baculovirus genotypes. *Virology* 1988, **166**:240-244.

33. González-Jara P, Fraile A, Canto T, García-Arenal F: The multiplicity of infection of a plant virus varies during colonization of its eukaryotic host. *J Virol* 2009, **83**:7487-7494.
34. Zwart MP, Daròs J-A, Elena SF: One is enough: in vivo effective population size is dose-dependent for a plant RNA virus. *PLoS Pathog* 2011, **7**:e1002122.
35. Monsion B, Froissart R, Michalakis Y, Blanc S: Large bottleneck size in cauliflower mosaic virus populations during host plant colonization. *PLoS Pathog* 2008, **4**:e1000174.
36. 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* 2010, **84**:1828-1837.
37. Lauring AS, Andino R: Exploring the fitness landscape of an RNA virus by using a universal barcode microarray. *J Virol* 2011, **85**:3780-3791.
38. Fabre F, Moury B, Johansen EI, Simon V, Jacquemond M, Senoussi R: Narrow bottlenecks affect Pea seedborne mosaic virus populations during vertical seed transmission but not during leaf colonization. *PLoS Pathog* 2014, **10**:e1003833.
39. Kennedy DA, Dwyer G: Effects of multiple sources of genetic drift on pathogen variation within hosts. *PLoS Biol* 2018, **16**:1-17.
40. Blanc S, Gutierrez S: The specifics of vector transmission of arboviruses of vertebrates and plants. *Curr Opin Virol* 2015, **15**:27-33.
41. Gutiérrez S, Yvon M, Thébaud G, Monsion B, Michalakis Y, Blanc S: Dynamics of the multiplicity of cellular infection in a plant virus. *PLoS Pathog* 2010, **6**:e1001113.
42. Gutiérrez S, Thébaud G, Smith DR, Kenney JL, Weaver SC: Demographics of natural oral infection of mosquitoes by Venezuelan equine encephalitis virus. *J Virol* 2015, **89**:4020-4022.
43. Sobel Leonard A, Weissman DB, Greenbaum B, Ghedin E, Koelle K: Transmission bottleneck size estimation from pathogen deep-sequencing data, with an application to human influenza A virus. *J Virol* 2017, **91**:e00171-17.
44. Zwart MP, Daròs J-A, Elena SF: One is enough: in vivo effective population size is dose-dependent for a plant RNA virus. *PLoS Pathog* 2011, **7**: e1002122.
45. de la Iglesia F, Elena SF: Fitness declines in tobacco etch virus upon serial bottleneck transfers. *J Virol* 2007, **81**:4941-4947.
46. Chao L: Fitness of RNA virus decreased by Muller's ratchet. *Nature* 1990, **348**:454-455.
47. Muller HJ: The relation of recombination to mutational advance. *Mutat Res* **1964**, **1**:2-9.
48. Willemsen A, Carrasco JL, Elena SF, Zwart MP: Going, going, gone: predicting the fate of genomic insertions in plant RNA viruses. *Heredity (Edinb)* 2018 <http://dx.doi.org/10.1038/s41437-018-0086-x>.
49. Zwart MP, Erro E, van Oers MM, de Visser JAGM, Vlak JM: Low multiplicity of infection in vivo results in purifying selection against baculovirus deletion mutants. *J Gen Virol* **89**:1220-1224.
50. Salverda MLM, Koomen J, Koopmanschap B, Zwart MP, De Visser JAGM: Adaptive benefits from small mutation supplies in an antibiotic resistance enzyme. *Proc Natl Acad Sci U S A* 2017 **114**:12773-12778.
51. Bull JC, Godfray HC, O'Reilly DR: Persistence of an occlusion-negative recombinant nucleopolyhedrovirus in *Trichoplusia ni* indicates high multiplicity of cellular infection. *Appl Environ Microbiol* 2001, **67**:5204-5209.
52. Bull JC, Godfray HCJ, O'Reilly DR: A few-polyhedra mutant and wild-type nucleopolyhedrovirus remain as a stable polymorphism during serial coinfection in *Trichoplusia ni*. *Appl Environ Microbiol* 2003, **69**:2052-2057.
53. R. Gallet, F. Fabre, G. Thébaud, M.T. Sofonea, A. Sicard, S. Blanc, Y. Michalakis, Small bottleneck size in a highly multipartite virus during a complete infectious cycle. *J Virol* doi: 10.1128/JVI.00139-18.
54. Betancourt M, Fereres A, Fraile A, García-Arenal F: Estimation of the effective number of founders that initiate an infection after aphid transmission of a multipartite plant virus. *J Virol* 2008, **82**:12416-12421.
55. 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* 2015, **13**:1-27.

56. Frost SDW, Dumaurier M-J, Wain-Hobson S, Brown AJL: Genetic drift and within-host metapopulation dynamics of HIV-1 infection. *Proc Natl Acad Sci* 2001, **98**:6975-6980.
57. Sanjuán R: Collective properties of viral infectivity. *Curr Opin Virol* 2018, **33**:1-6.
58. Sanjuán R: Collective infectious units of viruses. *Trends Microbiol* 2017, **25**:402-412.
59. Nethe M, Berkhout B, van der Kuyl AC: Retroviral superinfection resistance. *Retrovirology* 2005, **2**:52.
60. Schaller T, Appel N, Koutsoudakis G, Kallis S, Lohmann V, Pietschmann T, Bartenschlager R: Analysis of hepatitis C virus superinfection exclusion by using novel fluorochrome gene-tagged viral genomes. *J Virol* 2007, **81**:4591-4603.
61. Tscherne DM, Evans MJ, von Hahn T, Jones CT, Stamataki Z, McKeating JA, Lindenbach BD, Rice CM: Superinfection exclusion in cells infected with hepatitis C virus. *J Virol* 2007, **81**:3693-3703.
62. Zou G, Zhang B, Lim P-Y, Yuan Z, Bernard KA, P-Y Shi: Exclusion of west nile virus superinfection through RNA replication. *J Virol* 2009, **83**:11765-11776.
63. Simon K, Cardamone JJ, Whitaker-Dowling P, Youngner J, Widnell C: Cellular mechanisms in the superinfection exclusion of vesicular stomatitis virus. *Virology* 1990, **177**:375-379.
64. Kobiler O, Lipman Y, Therkelsen K, Daubechies I, Enquist LW: Herpesviruses carrying a Brainbow cassette reveal replication and expression of limited numbers of incoming genomes. *Nat Commun* 2010, **1**:146.
65. Laliberte JP, Moss B: A novel mode of poxvirus superinfection exclusion that prevents fusion of the lipid bilayers of viral and cellular membranes. *J Virol* 2014, **88**:9751-9768.
66. Huang I-C, Li W, Sui J, Marasco W, Choe H, Farzan M: Influenza A virus neuraminidase limits viral superinfection. *J Virol* 2008, **82**:4834-4843.
67. Beperet I, Irons SL, Simon O, King LA, Williams T, Possee RD, Lopez-Ferber M, Caballero P: Superinfection exclusion in Alphabaculovirus infections is concomitant with actin reorganization. *J Virol* 2014, **88**:3548-3556.
68. De Bruyn A, Villemot J, Lefeuvre P, Villar E, Hoareau M, Harimalala M, Abdoul-Karime AL, Abdou-Chakour C, Reynaud B, Harkins GW et al.: East African cassava mosaic-like viruses from Africa to Indian ocean islands: molecular diversity, evolutionary history and geographical dissemination of a bipartite begomovirus. *BMC Evol Biol* 2012, **12**:228.
69. Lucía-Sanz A, Manrubia S: Multipartite viruses: adaptive trick or evolutionary treat? *NPJ Syst Biol Appl* 2017, **3**:34.
70. Iranzo J, Manrubia SC: Evolutionary dynamics of genome segmentation in multipartite viruses. *Proc Biol Sci* 2012, **279**:3812-3819.
71. Sánchez-Navarro JA, Zwart MP, Elena SF: Effects of the number of genome segments on primary and systemic infections with a multipartite plant RNA virus. *J Virol* 2013 **87**:10805–10815.

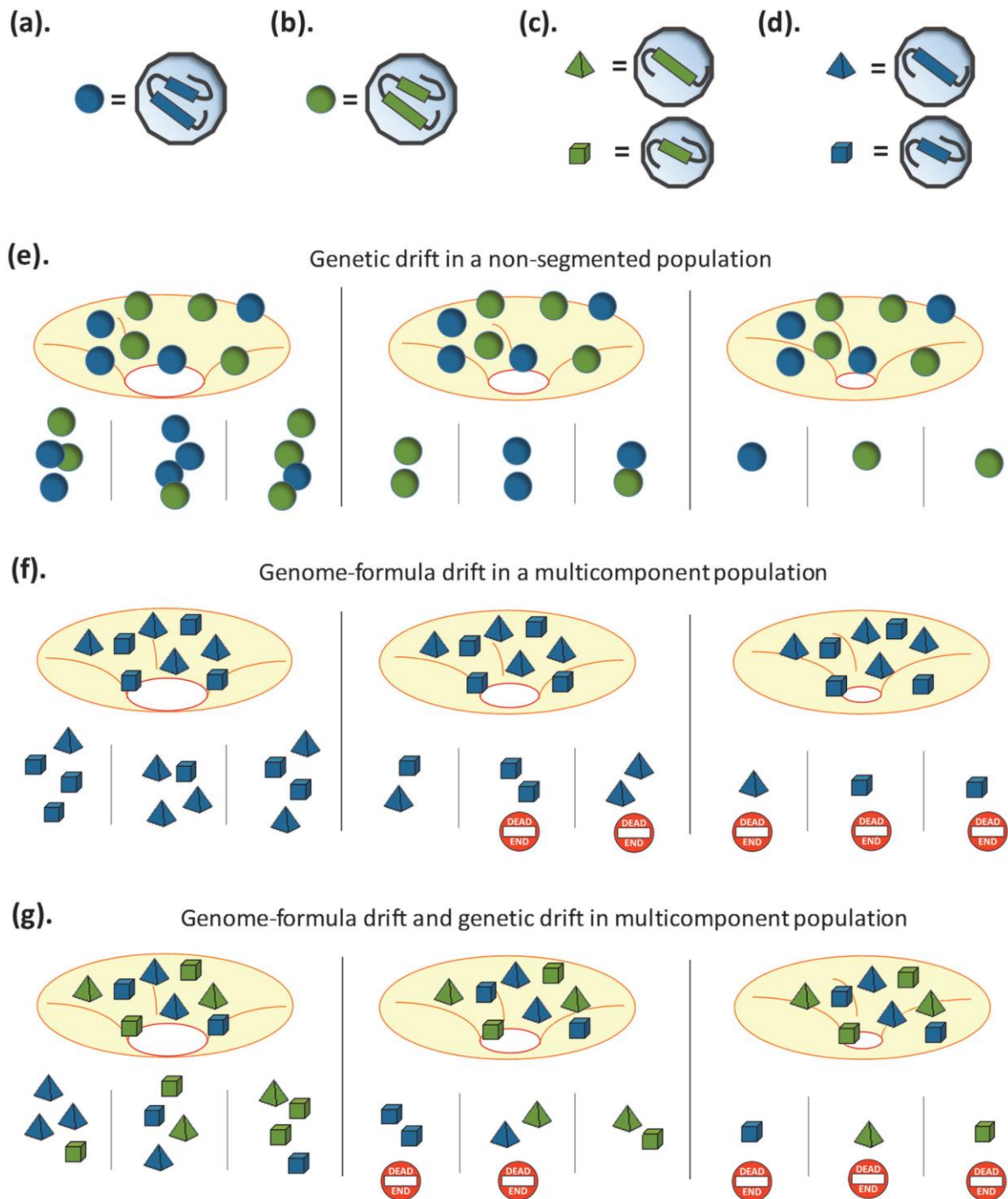


Figure 1: To illustrate the concept of gene-frequency drift, we consider a virus with two genes, a long and a short one, which both have green and blue alleles. Panels a–d are a legend for Panels e–g. Panels a and b show virus particles of a monopartite virus with the blue and green alleles, whereas c and d show virus particles of a bipartite virus. Panels e–g illustrate what happens with virus populations as they pass through population bottlenecks, that get narrower from left ($n = 4$) to middle ($n = 2$) to right ($n = 1$). The three populations below the bottleneck

illustrate three populations that have passed through the bottleneck. Panel e illustrates a monopartite virus, where each virus particle can cause infection and in the extreme case of $n = 1$ (right) one allele always becomes fixed. Panel f illustrates what happens when a clonal bipartite virus passes through a population bottleneck. If both types of virus particles do not pass through the bottleneck there will be no infection, as indicated by the traffic signs and reflecting the cost of multipartition. As there is no standing genetic variation, allele frequency will always be 1 in the infecting populations. However, the frequency of the genome segments can change. We call these random changes genome-formula drift. In Panel g, we illustrate what happens if a bipartite virus with two alleles for each genome segment passes through a bottleneck. Both segment frequencies and allele frequencies can change simultaneously, illustrating the co-occurrence of genetic and genome-formula drift.

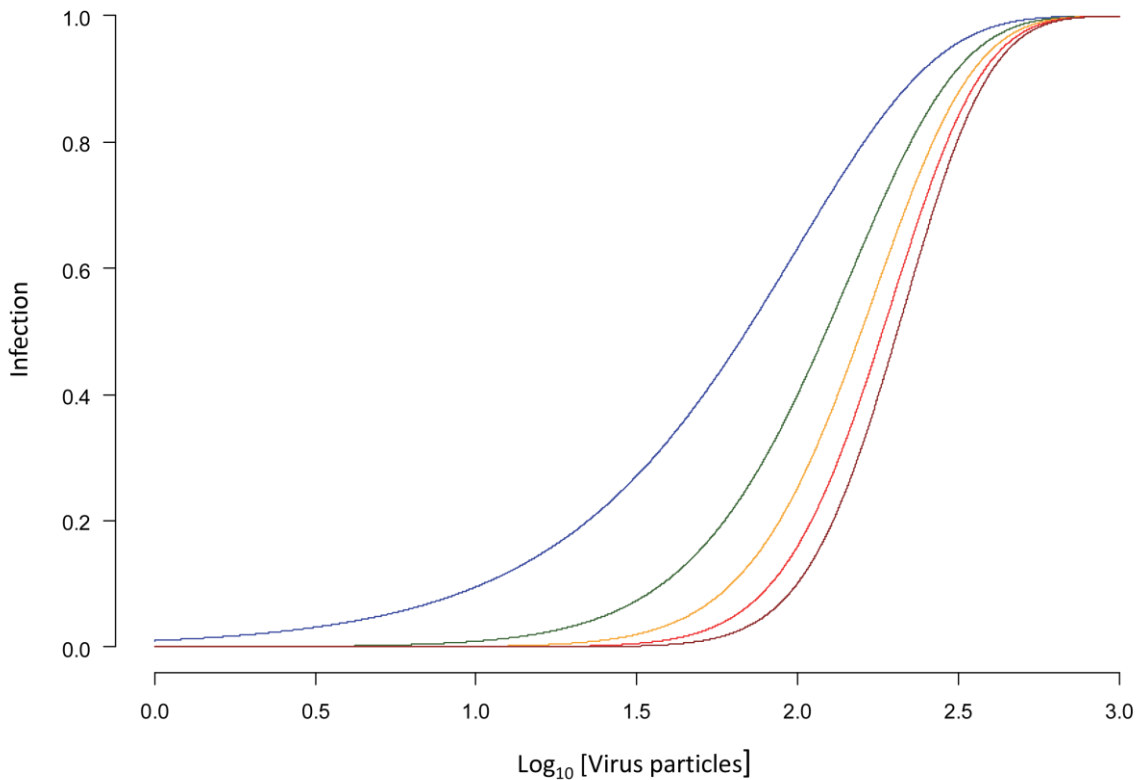


Figure 2: The cost of multipartition is illustrated. A multipartite organization comes at a high cost [70], because at low doses not all virus genome segments will be transmitted all the time leading to a lower infection rate. Consider the numerical example where all virus particles v have a transmission probability $\rho = 0.01$ and there are k genome segments. If virus particles are at equal frequency and transmitted independently, the rate of infection will be $(1 - e^{-\rho v})^k$ [71], and we can see that as the number of genome segments is increased from 1 (blue), to 2 (green), 3 (yellow), 4 (red) and 5 (dark red) the rate of infection decreases in the figure above. It should be noted that this cost of segmentation for infection will increase if segments differ in their frequencies within a population, all other things being equal. Moreover, this cost is strongly dose-dependent, eventually disappearing at high doses when all hosts are infected. All though this relationship holds up nicely in the case of mechanical transmission [71], it could be that in some cases or under some conditions virus particles clump together and hereby lower these costs.