Social Evolution in Viruses

Asher Leeks

St. John's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary 2021

For my Mum, Margaret Leeks

I declare that this thesis was composed by myself and that the work contained herein is my own except where explicitly stated in the text. This work has not been submitted for any degree or professional qualification except as specified.

Asher Leeks

Hilary term 2021

Acknowledgements

I would like to thank Stu, my supervisor, whose wisdom, patience, generosity, and inspiration, have shaped me so much as a scientist. I will be forever grateful.

In the same vein, I would like to thank my first supervisor at Oxford, Ashleigh, who gave me the opportunity to come to Oxford in the first place. I can't thank you enough for setting me on the path of science, and for supporting me when it mattered most.

I decided to work on viruses following a talk by Rafael Sanjuán. Since then, the ideas in this thesis have been shaped by conversations with a number of people, who have been very generous with their time: Samuel Díaz-Muñoz; Pilar Domingo-Calap; Santiago Elena; Sylvain Gandon; Melanie Ghoul; Alan Grafen; Ashleigh Griffin; Ryosuke Iritani; Katrina Lythgoe; Oliver Pybus; Rafael Sanjuán; Ernesto Segredo-Otero. I would also like to thank the organisers and attendees of TEEB and the Social Evolution Journal Club, and especially the regular members of Sums Club: Guy Cooper; Miguel dos Santos; Sam Levin; Matishalin Patel; Tom Scott; Gijsbert Werner.

Thanks also to many lab-mates past and present, for making it a genuine pleasure to come to the office each day: Alper; Anna; Becca; Berti; Chucky; Daniel; David; Gijsbert; Guy; James; JB; Josh; Louis; Mark; Mati; Max; Mel; Miguel; Ming; Ornela; Philip; Sam; Shana; Steffi; Tom. Thanks especially to Mati, for putting up with my endless technical questions, and Miguel, for showing me how to program.

Many thanks to Alan Grafen and Katia Koelle for a rigorous and enjoyable viva.

I would like to thank Ryosuke Iritani, Tomoko Iwanami, and Tetsuo Hatsuda, for making me feel so welcome at RIKEN in Japan, and for going out of their way to help me when the pandemic made things difficult.

Finally, I would like to thank my parents, Margaret and Derek, for a lifetime of support and encouragement. And, of course, Ailsa, for many years of love.

Abstract

Viruses are the most diverse and abundant lifeforms on Earth. In this thesis, I argue that they are also social organisms, and that it can be useful to study viruses within the framework of social evolution theory. Specifically, I: (i) define how cooperating and cheating can occur in viruses, review the diversity of viral cheats, and suggest why studying them can be useful for both virology and evolutionary biology; (ii) model how beneficial interactions between different viral variants can increase the genetic diversity of viral infections; (iii) model how beneficial interactions between viruses can promote the evolution of group dispersal in viruses; (iv) model distinct reasons for why vertically transmitted symbionts, viral or otherwise, are more cooperative than horizontally transmitted ones; (v) investigate the abundance of defective interfering 'cheat' genomes in natural infections of Influenza and SARS-CoV-2. Overall, I suggest that the nascent field of sociovirology has much to offer virologists and evolutionary biologists alike.

Publications

The following chapters have resulted in publications:

- Chapter Three
  - Leeks, A., Segredo-Otero, E.A., Sanjuán, R. West, S.A. 2018. Beneficial coinfection can promote within-host viral diversity. Virus Evol 4.
- Chapter Four
  - Leeks, A., Sanjuán, R. West, S.A. 2019. The evolution of collective infectious units in viruses. Virus Research 265: 94–101.
- Chapter Five
  - Leeks, A., dos Santos, M. West, S.A. 2019. Transmission, relatedness, and the evolution of cooperative symbionts. J Evol Biol jeb.13505.

In addition, Chapter Two has been accepted for publication at *Nature Communications*, and Chapter Six is being written up for publication. While producing this thesis, I also cowrote the following published commentary, which is included in Appendix A:

- Leeks, A. West, S.A. 2019. Altruism in a virus. Nat Microbiol 4: 910–911.

Contents

An inordinate fondness for ~~beetles~~ viruses

— JBS Haldane on the mind of the Creator,
re-imagined for the 21st century

1 Introduction

1.1 Why viruses?

To a close approximation, every living thing is a virus. Viruses infect humans, they infect our crops and livestock, they infect bacteria, fungi, and other microbes, and their fossilised remains litter our genomes. No environment on Earth is safe from viruses, from hydrothermal vents to hospital wards, from our own genomes to computer simulations of artificial life (Katzourakis & Gifford, 2010; Zaman et al., 2014). Their effects on human life are similarly universal, perhaps best exemplified by pandemics such as Covid-19 which, at the time of writing, has taken millions of lives and brought the world to a standstill.

However, despite their importance to us, human understanding of viruses is in its infancy. We have described a tiny fraction of the viral species that exist, and discoveries continue to overturn long-held assumptions about the fundamental biology of viruses (Fig. 1.1) (La Scola et al., 2008; Shi et al., 2016; Zhang et al., 2018; Sicard et al., 2019). Unlike other pathogens, viruses are perhaps a greater threat to us now than at any other point in human history. Many of the defining features of modern society, such air travel, globalisation, or the mass monoculture of crops, make us intrinsically vulnerable to viral outbreaks. At the same time, we still lack broad-spectrum antiviral therapeutics or even the ability to predict

when and where new outbreaks will occur. Viruses therefore share the unfortunate duality of being both the most important organisms for us to understand, and the organisms about which we understand the least.

In this thesis I present work on one of the least understood aspects of the least understood organisms; the social lives of viruses. When they are thought of as living organisms at all, viruses are usually assumed to be solitary. Textbooks depict single virions entering cells, releasing a single genome that then produces and uses its own gene products (Flint et al., 2015). However, this vision is incomplete. Viruses do not exist in isolation, but in a world of other viruses. Interactions between viruses are pervasive, and have profound effects on viral evolution, population dynamics, and pathogenicity (Díaz-Muñoz et al., 2017; Leeks et al., 2020). I suggest that we can study these interactions using social evolution theory, providing a new way to think about interactions between viruses.

1.2 Are viruses social?

Social interactions occur when a trait carried by one individual influences the fitness of another individual (West et al., 2007b). In this thesis, I consider each distinct copy of a viral genome to be an individual, and therefore consider interactions between different viral genomes (Díaz-Muñoz et al., 2017). Larger groupings, such as the virus-infected cell, or the collection of viral genomes infecting a host, generally contain multiple distinct entities that have separate routes to transmission, hence they are distinct coreplicons, each with their own maximand, that can be in evolutionary conflict with one another (Gardner & Grafen, 2009). Here, I will review viral social traits only briefly, since Chapter Two presents a more thorough overview.

Much of the time, viruses outnumber their host cells, creating a situation where multiple viral genomes infect each cell (Flint et al., 2015). In other cases, viruses infect new cells inside collective infectious units, some of which can deliver hundreds of viral genomes to the same cell (Sanjuán, 2017). Even when such coinfection doesn't occur, and just a single viral genome initially infects a cell, that genome is

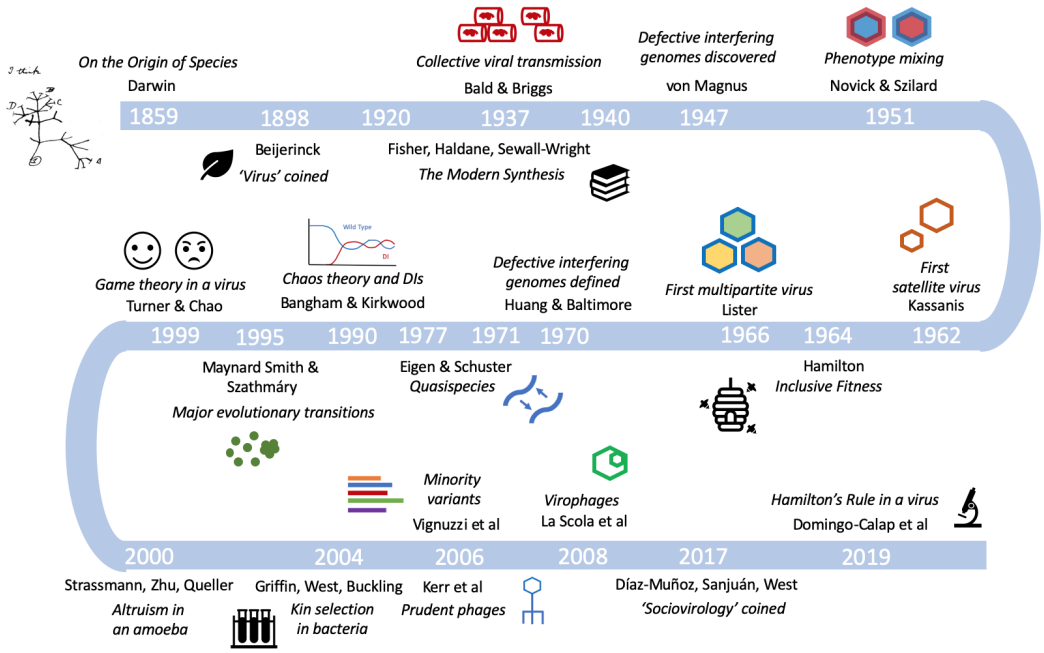


Figure 1.1: The path to sociovirology. A far-from-exhaustive, illustrative timeline of some of the advances within virology and evolutionary biology that have made it possible to apply social evolution theory in viruses.

then replicated, quickly creating a social environment in which dozens or hundreds of viral genomes can interact (Andreu-Moreno et al., 2020).

In each of these cases, some of the gene products produced by one viral genome can be used by other viral genomes, creating a common pool of resources. Typically, this common pool includes essential viral gene products, such as the replicase enzyme, that replicates the viral genome, and capsid proteins, that construct the viral capsid for infecting new cells. In virology, such shared gene products are known as 'trans-acting'; in evolution they are known as 'public goods', analogous to secreted molecules in bacteria, stashed acorns in woodpeckers, or shared farmland in humans (Davies et al., 2012; Flint et al., 2015). Public goods are vulnerable to exploitation by 'cheats', that avoid the cost of producing them, while still benefitting from the common pool of public goods produced by others (Ghoul et al., 2013; Leeks et al., 2020). Viruses are no exception, providing some of the best examples of cheats in the natural world; this is explored in more detail in Chapter Two.

Social interactions can also extend beyond the cell. One example is when animal viruses suppress host cell immune responses, keeping neighbouring cells susceptible to infection (Domingo-Calap et al., 2019). Another example is found in multipartite viruses, where viral gene products produced in one cell are exported into neighbouring cells (Sicard et al., 2019). Even virulence itself can be a social trait, since viral variants that exploit their host more rapidly can exhaust the supply of susceptible cells, reducing the fitness of other variants growing in the same host (Wild et al., 2009).

To some extent, the study of these virus-virus interactions is a novel field, driven by recent technological advances (Fig. 1.1). New techniques in microscopy have allowed us to see individual virions and study individual infected cells, while new genome sequencing technology allows us to track the evolutionary dynamics of different variants within a viral population. However, in other respects, this field is as old as virology itself, with many examples of viral social interactions coming from the very earliest days of virology (Fig. 1.1). Group dispersal in viruses has been known about for nearly a century (Bald & Briggs, 1937); the first virus dependent on another to complete its lifecycle was discovered more than fifty years ago (Kassanis, 1962); and viral 'cheats', that exploit gene products encoded by other viruses, have been observed since the very first days of viruses being grown in cell culture (von Magnus, 1947). It seems that viral social interactions are so common that they appear even when we're not looking for them.

But despite the importance and prevalence of virus-virus interactions, there is relatively little unifying theory that explains when different kinds of interaction evolve, or what their consequences might be. There are clear similarities between social interactions throughout the viral world, perhaps best exemplified by the fact that cheats with entirely independent origins exploit the same viral gene products across very different types of virus. In contrast with this, the study of most types of viral social interaction are confined to taxon-specific subdisciplines within virology. General theoretical principles that could cut across discipline-specific divides do exist, in the form of social evolution theory, but these ideas have rarely been applied

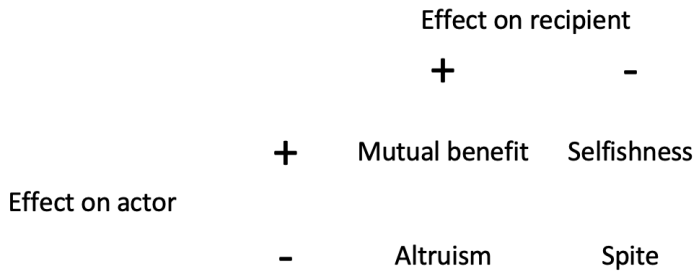


Figure 1.2: 'Hamilton's Square'. Social traits can be divided into four types, depending on the fitness consequences on the actor of the recipient. The actor refers to the individual bearing the trait, while recipient refers to an individual not bearing the trait who nevertheless suffers fitness consequences as a result of the actor's trait. Fitness consequences are defined in terms of lifetime number of offspring (West et al., 2007b).

in viruses (Bourke, 2011; Davies et al., 2012). I suggest that bridging this divide, by expanding social evolution theory to encompass the viral world, could provide the unifying theoretical framework needed.

1.3 Why social evolution theory?

Social evolution theory is a body of work that can explain why different kinds of social interaction evolve. By focussing on the fitness consequences of social interactions, we can dissect social traits into four categories, illustrated in a two-by-two square inspired by the work of Hamilton (West et al., 2007b) (Fig. 1.2).

Half of Hamilton's square concerns antagonistic traits, that have a negative impact on a recipient: selfish traits are those that are beneficial for the 'actor', who bears the trait, but have negative fitness consequences for the 'recipient' of the trait; while spiteful traits have negative fitness consequences for both the actor and recipient. This thesis contains a number of examples of selfish traits, but spiteful traits are much rarer throughout nature, evolving only under particular circumstances that have not yet been found in viruses (Patel et al., 2020).

The other half of Hamilton's square deals with traits that have a beneficial impact on the recipient: mutually beneficial traits, where both the actor and recipient benefit; and altruistic traits, where the actor pays a cost and the recipient

benefits. Traits that fall within these two squares, and that have evolved because of their fitness consequences for the recipient, are termed cooperative (West et al., 2007b). Cooperative traits have traditionally posed a problem for evolutionary theory, because they seem to break the basic principle of Darwinism: why help another when you could help yourself?

The answer to this problem depends on whether cooperation occurs via mutual benefit or altruism (West et al., 2007a). When cooperation involves mutual benefit, it can be explained by mechanisms that allow the costs of cooperation to be recouped by the actor. For example, if cooperating increases the likelihood of receiving reciprocal cooperation in the future, or if deviating from cooperation results in being sanctioned by the other party. When a trait is altruistic, the long-term fitness consequences for the actor are negative, and such feedback mechanisms cannot explain cooperation. In this case, kin selection provides the key: altruistic traits are favoured when they preferentially benefit individuals who are genetically related to the actor (Hamilton, 1964).

A defining feature of this classification is that it is based on the evolutionary consequences of traits or behaviours, rather than mechanistic details. Consequently, answers to problems that are highlighted in one group of organisms can also apply in other groups of organisms, providing a coherent body of theory to explain the evolution of social traits wherever they occur (Maynard Smith & Szathmary, 1995; Bourke, 2011; Davies et al., 2012). Empirically, this approach has been extraordinarily successful, with the same factors promoting cooperation across the natural world, from bacteria, through bees, birds, and beetles, all the way to human behaviour (West et al., 2021). In some areas, such as sex allocation, empirical data fits theoretical predictions to a degree of precision that is rarely seen in biology (West, 2009).

Perhaps the best example of the explanatory power of social evolution theory, however, is the field of sociomicrobiology (West et al., 2006, 2021). Twenty years ago, it would have been highly unusual to talk about bacteria and other microorganisms as social organisms, capable of ‘complex’ traits such as cooperation. But now it

is so commonplace that microbes are used in textbooks as canonical examples of social traits, such as cheats in the bacterium *Pseudomonas aeruginosa*, or altruistic self-sacrifice in the slime mould *Dictyostelium discoideum* (Fig. 1.1) (Strassmann et al., 2000; Griffin et al., 2004; Davies et al., 2012). The practice of applying principles originally developed to explain animal behaviours, to organisms as distantly related to animals as bacteria, has allowed us to understand the evolution of these traits in an intuitive way, that stimulates useful empirical work, and that hasn't required developing new theoretical frameworks. On a pragmatic level, this field has also stimulated new areas of applied research that form a key branch of evolutionary medicine, such as by explaining the clinical course of chronic bacterial infections, offering new 'evolution-proof' targets for antibiotics, or opening up the possibility of disrupting social dynamics for therapeutic purposes (Brown et al., 2009; Andersen et al., 2015; Dieltjens et al., 2020).

Applying social evolution theory so broadly has also strengthened the body of theory itself. A key test of any scientific theory is its ability to predict novel empirical patterns (Bacon, 1620). Consequently, the widespread empirical successes of social evolution theory, particularly in organisms that are so different from those that the theory was originally developed for, improve our confidence in the underlying tenets of the theory (West et al., 2021). At the same time, various unique features of microbes have prompted new theoretical developments, such as the importance of mutation in cooperation, virulence as a social trait, or population structures that allow traits such as spite to evolve (Gardner et al., 2004; Wild et al., 2009; dos Santos et al., 2018).

1.4 Sociovirology

I suggest that the features that make social evolution theory so empirically successful elsewhere in nature make it particularly suitable for studying viruses. More so than any other group of organisms, viruses have a unique biology that is mechanistically distinct from the rest of the living world – unlike other lifeforms, viruses don't even have their own cells. Viruses' unique biology is mirrored in their diversity,

perhaps best exemplified by the fact that we cannot construct a single phylogeny that contains all viruses, dealing a hatchet blow to Darwin's vision of a single tree of life (Holmes, 2009; Wolf et al., 2018). In the face of such diversity, there is a need for theoretical ideas that are based on fitness consequences of traits, even when the mechanistic details vary considerably. Social evolution theory does this, and can explain social traits in viruses, no matter where they are found or what they look like.

The unique and often surprising biology of viruses also offers excellent opportunities to test and expand social evolution theory itself. Some examples of viral sociality appear quite alien, with no clear parallels elsewhere in the living world. One such example would be multipartite viruses, where a single viral genome is split into multiple segments, each packaged inside its own capsid (Lucía-Sanz & Manrubia, 2017). Other types of viral sociality are familiar, but can be pushed to the extremes in comparison with other organisms. For example, when a cuckoo parasitizes a nest, it can eject a bird's whole brood, exploiting the host's parental care to produce a single cuckoo chick (Davies, 2010). But when a defective interfering genome exploits a cooperative virus, it can produce thousands of defective interfering genomes for each cooperative genome (Shirogane et al., 2021). Viral cheats can therefore gain fitness advantages that are orders of magnitude higher than those seen elsewhere in nature, while at the same time emerging spontaneously via mutation. This means that no cooperative virus, no matter how clonal the infection, is safe from the spectre of cheating. The combination of the unique and the extreme creates interesting evolutionary puzzles for the theoretician.

But perhaps the best reason to use social evolution theory to study viruses is because social processes underpin the most fundamental aspects of viral biology. Take public goods games as an example: bacteria play public goods games when scavenging for scarce resources such as iron, and humans play public goods games when farming common land, or when put in a room together by a behavioural economist (Davies et al., 2012). But viruses play public goods games every time they infect a cell (Díaz-Muñoz et al., 2017). The most basic viral traits, that are shared by all lytic viruses – replicating their genome and packaging it to infect

a new cell – depend on public goods. Viruses are not just social organisms; they may be the most social organisms.

1.5 Breakdown of thesis chapters

This thesis contains some theoretical and empirical work relevant to studying social traits found in viruses. I briefly summarise the remaining chapters below, and list my contribution to each chapter:

- Chapter Two
  - Chapter two is a perspective on cheating in viruses. This chapter provides a general introduction to cooperation and cheating in viruses, including how to test experimentally whether a viral trait is cooperative. The rest of the manuscript then provides an overview of the diversity of viral cheats, explores what is known about different types of viral cheat, and discusses what virologists and evolutionary biologists can each gain from studying cheating in viruses.
  - Melanie Ghoul and I conceived the project and I wrote the first draft of the manuscript. Stuart West, Melanie Ghoul, and I all contributed to the preparation of the manuscript in its final form (Leeks et al., 2020).
- Chapter Three
  - This chapter presents some theoretical work on the consequences of a beneficial viral social interaction, namely the mutually beneficial sharing of different versions of a gene product. Many viral infections show high levels of diversity, with many different variants of the same virus coexisting inside a host. This diversity presents an evolutionary problem: why don't the faster-growing variants out-compete the slower-growing variants? We showed that diversity can be maintained if host cells infected by multiple different variants are more productive than cells infected by just one variant, a phenomenon that could be relatively

common in RNA viruses. This mechanism is directly analogous to the concept of heterozygote advantage in population genetics.

  - I conceived the project, performed the Mathematical modelling, and wrote the first draft of the manuscript. Ernesto Segredo-Otero and Rafael Sanjuán wrote the spatial simulation and corresponding section of the manuscript. Stuart West and I both contributed to the preparation of the manuscript in its final form (Leeks et al., 2018).

- Chapter Four

  - This chapter focuses on how beneficial virus-virus interactions could drive the evolution of group dispersal in viruses. Many viruses disperse as groups, inside 'collective infectious units'. Although these structures appear across many different viruses, there are few ideas for why they are favoured over individual transmission. We modelled some evolutionary hypotheses that could drive the evolution of collective infectious units, finding that beneficial interactions are key, and that the way in which these interactions manifest can influence the size of the collective infectious units that evolve. We found that these advantages can be enhanced by efficiency benefits to packaging multiple viral genomes inside the same structure, but that the possibility of defective interfering 'cheat' viruses can destroy the benefits of dispersing in a group.
  - I conceived the project, performed the Mathematical modelling, and wrote the first draft of the manuscript. Stuart West, Rafael Sanjuán and I all contributed to the preparation of the manuscript in its final form (Leeks et al., 2019b).

- Chapter Five

  - While many viruses are obligate parasites, some act as mutualists, providing benefits to their host. Some authors estimate that the vast majority of plant viruses (upwards of 99%) are mutualists, while phages have

long been used as model organisms to study the mutualism-parasitism continuum. In both of these cases, viruses have been shown to become more mutualistic when transmitting vertically, from parents to offspring. However, two distinct explanations have been proposed for this: firstly, vertical transmission aligns the fitness interests of host and symbiont; secondly, vertical transmission reduces mixing of symbiont lineages, increasing the genetic relatedness between symbionts sharing a host. This chapter presents a model in which both mechanisms can operate and finds that the second mechanism tends to overshadow the first.

  - Stuart West and I conceived the project, I performed the Mathematical modelling, and Miguel dos Santos performed the computational simulation. All authors contributed to the preparation of the manuscript in its final form (Leeks et al., 2019a).

- Chapter Six

  - Defective interfering genomes are viral cheats that can emerge spontaneously via mutation and can have drastic consequences for viral infections. They are extremely common in viral tissue culture infections, but it is unclear whether they are also common in natural viral infections. This chapter presents bioinformatic analyses of viral sequencing data in which we detected defective interfering genomes in Influenza and SARS-CoV-2 infections.

  - For the work on Influenza, I conceived of the project and performed the bioinformatic analyses. For the work on SARS-CoV-2, I conceived of the project and performed most of the bioinformatic analyses presented. Mariá José Olmo Uceda wrote a Python script that removed subgenomic mRNAs. Manish Choudhary and Jonathan Li obtained and provided access to the longitudinal dataset. Mariá José Olmo Uceda, Julia Hilung, Santiago Elena, Manish Choudhary, Jonathan Li, Stuart West, and I

contributed to the interpretation of the bioinformatic analyses. I wrote the manuscript with feedback from Stuart West.

- Chapter Seven
  - This chapter discusses some of the implications of the preceding chapters and introduces some directions for future work.
  - I wrote the discussion with feedback from Stuart West.
- Appendix A
  - This appendix is an invited commentary on an empirical study by Domingo-Calap et al, that was the first to test Hamilton's Rule in a virus.
  - Stuart West and I wrote the manuscript (Leeks & West, 2019).

1.6 References

- Andersen, S.B., Marvig, R.L., Molin, S., Krogh Johansen, H. & Griffin, A.S. 2015. Long-term social dynamics drive loss of function in pathogenic bacteria. Proc. Natl. Acad. Sci. 112: 10756–10761.
- Andreu-Moreno, I., Bou, J.-V. & Sanjuán, R. 2020. Cooperative nature of viral replication. Sci. Adv. 6: eabd4942. American Association for the Advancement of Science.
- Bacon, F. 1620. Novum Organum. ed. by Joseph Devey. New York: P.F. Collier, 1902.
- Bald, J.G. & Briggs, G.E. 1937. Aggregation of Virus Particles. Nature 140: 111.
- Bourke, A.F.G. 2011. Principles of Social Evolution. Oxford University Press, Oxford, New York.

- Brown, S.P., West, S.A., Diggle, S.P. & Griffin, A.S. 2009. Social evolution in micro-organisms and a Trojan horse approach to medical intervention strategies. Philos. Trans. R. Soc. B Biol. Sci. 364: 3157–3168.
- Davies, N.B. 2010. Cuckoos, Cowbirds and Other Cheats. T & AD Poyser.
- Davies, N.B., Krebs, J.R. & West, S.A. 2012. An Introduction to Behavioural Ecology, 4th ed. John Wiley & Sons.
- Díaz-Muñoz, S.L., Sanjuán, R. & West, S.A. 2017. Sociovirology: Conflict, Cooperation, and Communication among Viruses. Cell Host Microbe 22: 437–441.
- Dieltjens, L., Appermans, K., Lissens, M., Lories, B., Kim, W., Van der Eycken, E.V., et al. 2020. Inhibiting bacterial cooperation is an evolutionarily robust anti-biofilm strategy. Nat. Commun. 11: 107.
- Domingo-Calap, P., Segredo-Otero, E., Durán-Moreno, M. & Sanjuán, R. 2019. Social evolution of innate immunity evasion in a virus. Nat. Microbiol. 1.
- dos Santos, M., Ghoul, M. & West, S.A. 2018. Pleiotropy, cooperation, and the social evolution of genetic architecture. PLOS Biol. 16: e2006671.
- Flint, J., Racaniello, V.R., Rall, G.F. & Skalka, A.M. 2015. Principles of Virology, 4th ed. American Society of Microbiology.
- Gardner, A. & Grafen, A. 2009. Capturing the superorganism: a formal theory of group adaptation. J. Evol. Biol. 22: 659–671.
- Gardner, A., West, S.A. & Buckling, A. 2004. Bacteriocins, spite and virulence. Proc. R. Soc. Lond. B Biol. Sci. 271: 1529–1535.
- Ghoul, M., Griffin, A.S. & West, S.A. 2013. Toward an evolutionary definition of cheating. Evolution 68: 318–331.

- Griffin, A.S., West, S.A. & Buckling, A. 2004. Cooperation and competition in pathogenic bacteria. Nature 430: 1024–1027.
- Hamilton, W.D. 1964. The genetical evolution of social behaviour. I. J. Theor. Biol. 7: 1–16.
- Holmes, E.C. 2009. The Evolution and Emergence of RNA Viruses. Oxford University Press, Oxford, New York.
- Kassanis, B. 1962. Properties and behaviour of a virus depending for its multiplication on another. J. Gen. Microbiol., doi: 10.1099/00221287-27-3-477.
- Katzourakis, A. & Gifford, R.J. 2010. Endogenous Viral Elements in Animal Genomes. PLOS Genet. 6: e1001191. Public Library of Science.
- La Scola, B., Desnues, C., Pagnier, I., Robert, C., Barrassi, L., Fournous, G., et al. 2008. The virophage as a unique parasite of the giant mimivirus. Nature 455: 100–104.
- Leeks, A., dos Santos, M. & West, S.A. 2019a. Transmission, relatedness, and the evolution of cooperative symbionts. J. Evol. Biol. jeb.13505.
- Leeks, A., Sanjuán, R. & West, S.A. 2019b. The evolution of collective infectious units in viruses. Virus Res. 265: 94–101.
- Leeks, A., Segredo-Otero, E.A., Sanjuán, R. & West, S.A. 2018. Beneficial coinfection can promote within-host viral diversity. Virus Evol. 4.
- Leeks, A. & West, S.A. 2019. Altruism in a virus. Nat. Microbiol. 4: 910–911.
- Leeks, A., West, S.A. & Ghoul, M. 2020. Cheating in the Viral World. , doi: 10.20944/preprints201906.0106.v2. Preprints.
- Lucía-Sanz, A. & Manrubia, S. 2017. Multipartite viruses: adaptive trick or evolutionary treat? Npj Syst. Biol. Appl. 3: 34.

- Maynard Smith, J. & Szathmary, E. 1995. The Major Transitions in Evolution. Oxford University Press, New York.
- Patel, M., West, S.A. & Biernaskie, J.M. 2020. Kin discrimination, negative relatedness, and how to distinguish between selfishness and spite. Evol. Lett. 4: 65–72.
- Sanjuán, R. 2017. Collective Infectious Units in Viruses. Trends Microbiol. 25: 402–412.
- Shi, M., Lin, X.-D., Tian, J.-H., Chen, L.-J., Chen, X., Li, C.-X., et al. 2016. Redefining the invertebrate RNA virosphere. Nature, doi: 10.1038/nature20167.
- Shirogane, Y., Rousseau, E., Voznica, J., Xiao, Y., Su, W., Catching, A., et al. 2021. Experimental and mathematical insights on the interactions between poliovirus and a defective interfering genome. bioRxiv. doi: 10.1101/2021.01.11.426198
- Sicard, A., Pirolles, E., Gallet, R., Vernerey, M.-S., Yvon, M., Urbino, C., et al. 2019. A multicellular way of life for a multipartite virus. eLife 8: e43599.
- Strassmann, J.E., Zhu, Y. & Queller, D.C. 2000. Altruism and social cheating in the social amoeba Dictyostelium discoideum. Nature 408: 965–967.
- von Magnus, P. 1947. Studies on interference in experimental influenza. Almqvist & Wiksell.
- West, S.A. 2009. Sex Allocation. Princeton University Press.
- West, S.A., Cooper, G.A., Ghoul, M.B. & Griffin, A.S. 2021. Ten recent insights for our understanding of cooperation. Nat. Ecol. Evol. 1–12. Nature Publishing Group.
- West, S.A., Griffin, A.S. & Gardner, A. 2007a. Evolutionary Explanations for Cooperation. Curr. Biol. 17: R661–R672.

- West, S.A., Griffin, A.S. & Gardner, A. 2007b. Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. J. Evol. Biol. 20: 415–432.
- West, S.A., Griffin, A.S., Gardner, A. & Diggle, S.P. 2006. Social evolution theory for microorganisms. Nat. Rev. Microbiol. 4: 597–607.
- Wild, G., Gardner, A. & West, S.A. 2009. Adaptation and the evolution of parasite virulence in a connected world. Nature 459: 983–986.
- Wolf, Y.I., Kazlauskas, D., Iranzo, J., Lucía-Sanz, A., Kuhn, J.H., Krupovic, M., et al. 2018. Origins and Evolution of the Global RNA Virome. mBio 9.
- Zaman, L., Meyer, J.R., Devangam, S., Bryson, D.M., Lenski, R.E. & Ofria, C. 2014. Coevolution Drives the Emergence of Complex Traits and Promotes Evolvability. PLOS Biol. 12: e1002023. Public Library of Science.
- Zhang, Y.-Z., Shi, M. & Holmes, E.C. 2018. Using Metagenomics to Characterize an Expanding Virosphere. Cell 172: 1168–1172.

2

The Evolution of Cheating in Viruses

The Evolution of Cheating in Viruses

Authors: Asher Leeks*, Stuart A. West, Melanie Ghoul

*Corresponding author

Affiliations

A.L., S.A.W., M.G.: Department of Zoology, University of Oxford, Zoology Research and Administration Building, 11a Mansfield Road, Oxford, OX1 3SZ

Keywords:

Cheating; cooperation; social evolution; virus evolution; defective interfering genome; satellite virus

Abstract

The success of many viruses depends upon cooperative interactions between viral genomes. However, whenever cooperation occurs, there is the potential for 'cheats' to exploit that cooperation. We suggest that: (1) the biology of viruses makes viral cooperation particularly susceptible to cheating; (2) cheats are common across a wide range of viruses, including viral entities that are already well studied, such as defective interfering genomes, and satellite viruses. Consequently, the evolutionary theory of cheating could help us understand and manipulate viral dynamics, while viruses also offer new opportunities to study the evolution of cheating.

Introduction

Cooperation can be observed at all scales of biology. Bacteria secrete molecules that scavenge resources for the local group of cells, worker ants forage for food to rear the offspring produced by their queen, and subordinate meerkats babysit the offspring of the dominant individuals in their group[1–3]. We term these kinds of traits cooperative because their evolution is influenced by the benefit they provide to other individuals, not just the individual that performs the trait[4].

Whenever cooperation occurs, there is the potential for exploitation by 'cheats', that avoid the cost of cooperating, but still gain the benefits of others cooperating[5,6]. For example, cuckoo chicks that deceive parents of a different species of bird to feed them, or bacterial mutants that have lost the ability to produce and secrete a molecule that benefits the local group, but can still benefit from the molecule secreted by neighbouring cells[7–9]. However, the extent to which cheating occurs in nature has proved contentious, and empirical examples of

cheats are relatively rare[6,10]. The prevalence of cheating matters, because it determines the extent to which different individuals are in conflict over cooperation, and whether successful cooperation needs mechanisms to counter cheating[5,6,11,12].

In this perspective, we suggest that, in contrast to elsewhere in the living world, cheats are both common and relatively easy to detect in viruses. Several examples of viral cheats are already well studied within virology, including defective interfering genomes and satellite viruses. We synthesise the relevant evolutionary and virology literatures, showing how similar issues have been examined in these two fields, but from very different perspectives.

The widespread prevalence of cheating in viruses poses novel evolutionary problems, which also have direct implications for our ability to manage viral infections. From an evolutionary perspective, why is cheating so common in viruses? Are there different types of cheats, that need different types of explanation? And can cheating in viruses help us better understand the evolution of cheating more generally? From a virology perspective, cheating can have a substantial impact on the epidemiology and outcome of viral infections. Can evolutionary theory tell us how viral cheats will evolve and when they will spread? Or how viruses will evolve in response to cheating? And can cheats be exploited or manipulated to help control viral infections?

What is Cooperation?

Before discussing viruses, it is useful to define exactly what we mean by both cooperation and cheating. When an individual is cooperating, it is performing a behaviour or trait that evolved because it benefits another individual[4]. Cooperation poses an evolutionary problem

because, all else being equal, it should reduce the relative fitness of the individual performing the cooperation, and hence be selected against (Box 1).

There are two broad solutions to this problem of cooperation[1]. Firstly, cooperation can be favoured if it also provides a benefit to the individual performing the cooperation. For example, in humans, cooperation is often reciprocated, so that when A helps B this will lead to B helping A at some later date. Consequently, in the long-term, cooperation can provide a direct benefit to the cooperator, because they also get help.

The second solution is that cooperation can be favoured if it provides a benefit to other individuals that carry the cooperative gene. This process is termed kin selection, because the easiest and most common way for individuals to carry the same genes is through common descent. By helping a close relative reproduce, an individual is still passing on its genes to the next generation, just indirectly[13]. Such indirect benefits (kin selection) have been shown to explain many forms of cooperation, from the production of shared molecules in bacteria, to the evolution of sterile workers in the social insects[14].

What is a Cheat?

Cheats are individuals that exploit cooperators, by avoiding paying the cost of cooperation, while still benefiting from the cooperation of others[5,6] (Box 2).

The simplest possible form of cheating is to just not cooperate. Bacteria have provided numerous examples of individuals that cheat by 'not cooperating'. For example, when iron is limited, bacteria produce and release siderophores, which scavenge iron and make it available to the bacteria. Siderophores provide a benefit to the local group of cells,

not just the cell that produced them, and so they represent a form of cooperation that is termed a 'public good' (Box 1)[8,15,16]. Cells that do not produce siderophores are still able to take up iron via siderophores produced by other cells, and so represent a form of cheat. Cheats that exploit siderophores, and other 'not-producing' cheats that exploit similar bacterial public goods, have been observed in both laboratory and natural populations of bacteria (Fig. 1)[7,17,18].

Cheating can also take more active or devious forms, such as in cuckoos and other avian brood parasites (Fig. 1). Here, individuals of one species trick parents of a different species into neglecting their own chicks and instead feeding the brood parasite's chicks[19]. This trickery can range from cuckoo parents laying eggs that closely mimic those of the host species, to cuckoo chicks actively ejecting the offspring of the host parent.

However, while these are examples of common and well-studied cheats, the extent to which cheating is prevalent in nature has proved contentious[6,10] (Box 2).

How do Viruses Cooperate?

When examining cooperation, we need to think about interactions between different 'individuals'. In viruses, we consider a physically distinct copy of a viral genome to be an individual, because it is the largest unit which we can consider to be evolving as a single agent. Larger groupings such as a virion, or a cloud of viral genome sequences, can contain multiple distinct genetic entities that can be in evolutionary conflict[20–22].

The simplest and most common form of cooperation in viruses occurs when multiple viral genomes infect the same host cell and share gene products (Box 1). For example, when one genome produces a replicase enzyme, this will commonly replicate all the genomes in the host cell, and not just the genome that produced it (Fig. 2). Consequently, the replicase enzyme can be a form of 'public good' which provides both a direct benefit to the genome that produced it, and a shared benefit to any other genomes in the cell (Box 1).

Other viral gene products can also act as public goods, provided they are shared within a cell ('trans-complementable') (Box 1; Fig. 2). For example, capsid proteins build the viral capsid (or virion) that transports viral progeny to new cells, and can contain genomes other than those that produced the capsids. Replicase enzymes and capsid proteins together represent two potentially very common forms of cooperation in viruses. More generally, cooperation is important for the evolution of any shared viral gene product that has evolved at least partially because of a group benefit that it provides[23].

Viral cooperation can also extend beyond the cell, to include cases where benefits are shared between viral genomes infecting different cells (Fig. 2). For example, animal viruses block the release of interferon from host cells, to suppress the host immune response (Fig. 2)[24]. This suppression is costly to the viral genome encoding the gene for suppression, slowing its replication within a cell, but it provides a public benefit by keeping the local population of host cells susceptible to infection by neighbouring viruses[25]. A number of analogous examples appear to exist elsewhere in viruses, such as when phages encode 'anti-CRISPR' proteins that partially overcome bacterial host defences[26,27].

How to Test for Cooperation and Cheating

We have claimed that viral traits such as producing replicase enzymes can be a form of cooperation, which could be cheated by genomes that do not perform these traits (Box 1). This requires that these traits have evolved at least partially due to the benefits they provide to other viral genomes. How can this be tested experimentally?

The first step is to carry out fitness assays on strains that do and do not perform the putative cooperative trait, on their own and in a mixed culture[5]. For example, we might have a strain that did produce the replicase enzyme (putative cooperator), and a strain that did not produce replicase (putative cheat). When the benefits of cooperation go to other cooperators, a cooperative trait is selected for; but when those benefits go to non-cooperative individuals (cheats), cooperation is selected against. Therefore, if these strains really represent a cooperator and a cheat, we would observe three results:

(i) when grown separately, the cheat would not be able to exploit cooperation, and so would grow slower than the cooperator (Fig. 3);

(ii) when grown together, in a mixture, the cheat would be able to exploit and outcompete the cooperator (Fig. 3);

(iii) the exploitation by the cheat would reduce the fitness of the cooperator, compared to when the cooperator is alone (interference) (Fig. 3).

In viruses, these three results have been clearly demonstrated in experiments on DI PV1, a cheat of poliovirus that uses capsid proteins produced by cooperative wild-type poliovirus genomes (Box 3). Similar experiments in a wide range of different viruses have determined that replicase, capsid proteins, and other shared gene products, are cooperative public goods, that are commonly exploited by viral cheats such as defective interfering

genomes and satellite viruses[28–30]. 'Accidental' experiments have reinforced these findings, by showing that cheats such as defective interfering genomes rapidly spread when viruses are cultured in conditions that favour high coinfection and low relatedness.

Virologists have begun using these kinds of experiment to examine a range of other potentially cooperative traits, from the suppression of host immune systems, to the rate at which phages lyse their hosts[25,31,32]. Despite this, the cooperative nature of many viral traits remains speculative, with formal experiments required[20].

More broadly, over the last 20 years, these experimental methods have revolutionised our understanding of cooperation in bacteria and other microorganisms[14]. Are we at the start of a similar revolution with viruses?

Where are Viral Cheats Found?

Cheats are common throughout the viral world. We can divide and classify cheats based on the different kinds of cooperation that they exploit (Fig. 4), or how they are created (Fig. 5).

Intracellular Cooperation

Our first division is between cheats that exploit cooperation between viruses within the same cell (intracellular cooperation) or different cells (extracellular cooperation).

Defective Interfering Genomes

Defective interfering genomes, such as DI PV1, are literally defined by the features that make them cheats (Box 2)[29]. They are defective, because they grow less well, if at all, on their own

194 (result i of the three results that define a cheat), and they are interfering genomes because
195 they exploit and interfere with the growth of the 'normal' cooperative strain (results ii & iii)
196 (Box 3)[29]. Defective interfering genomes are the most prominent kind of viral cheat, having
197 been studied for decades and observed in tissue culture for almost all animal and plant
198 viruses[28,29,33–35]. They can be found in non-segmented, segmented, and multipartite viruses,
199 viruses infecting every major type of host (animal, plant, microbial), and in viruses from
200 across the Baltimore classification (Fig. 4).
201
202 Defective interfering genomes emerge spontaneously during infections, by mutations
203 that delete genes for intracellular viral public goods, such as the replicase enzyme, capsid
204 proteins, or proteins that manipulate host cell machinery for the benefit of the infecting
205 viruses[35]. They are closely analogous to public goods cheats that have been commonly
206 observed in bacteria (Box 1)[7,17,18]. Because they arise through large mutations, defective
207 interfering genomes are much shorter than the wild-type virus, which is a key reason why
208 they gain such a large replication advantage in coinfection[36–42]. In short RNA viruses,
209 cooperative genes can comprise a large proportion of the genome, and so defective
210 interfering genomes achieve this shorter length by deleting the cooperative gene[43]. In larger
211 DNA viruses, defective interfering genomes often delete a large number of genes, keeping
212 only small number of genes, suggesting that most of the wild-type genome consists of
213 cooperative genes[44].
214
215 ## Free-living satellites
216 Satellite viruses are small viral entities that encode only some, and sometimes none, of the
217 genes required for successful infection, instead relying on exploiting gene products encoded
218 by complete 'helper' viruses[30,45]. Satellites are varied and can have a range of different

effects on their helper viruses, with some satellites being cheats, but others appearing to be mutualists, and some that even encode novel genes not found in the helper virus[30,45–47]. Cheat satellites share many functional similarities with defective interfering genomes, exploiting similar gene products such as replicase and capsid proteins, and gaining similar advantages through having a shorter length. However, unlike defective interfering genomes, we do not know how most satellites originate, since they tend to share little to no sequence homology with their helper viruses.

Satellite viruses are also much longer-lived than defective interfering genomes, transmitting between hosts and persisting over long evolutionary timescales. Consequently, they are more analogous to cheats such as cuckoos, that have evolved over long timescales to exploit their hosts. In some cases, satellites are themselves exploited by other satellites, or by defective interfering genomes[48]. Satellites are very common in plant viruses, although they can also be found in phages and in animal viruses, including those that infect humans[49,50].

Integrated Satellites

Some satellite viruses employ a 'sit-and-wait' strategy, in which they integrate into host genomes and become dormant, replicating only when a cooperative virus infects their host cell[50,51]. Examples include: adeno-associated viruses, which are present in up to 90% of human genomes; Staphylococcal pathogenicity islands, which are widespread throughout the genomes of gram-positive bacteria, and show evidence of long-term convergent evolution towards cheating; and phage-inducible chromosomal island-like elements (PLEs) in the genomes of *Vibrio cholerae* bacteria, where the wild-type phage has evolved resistance by encoding CRISPR-Cas to combat the satellite, resulting in cooperator-cheat arms race dynamics that influence when cholera outbreaks occur[47,50,52,53]. This 'sit-and-wait' strategy

appears to be novel form of cheating, not known in other organisms, and which represents an elegant solution to the problem of a low likelihood of co-transmission with a cooperative virus.

Virophages

One of the most exciting recent discoveries in virology has been that of giant viruses and their associated satellites, the virophages[54,55]. Giant viruses have very large genomes that can be larger than some bacterial genomes, and construct capsids big enough to be seen with a light microscope[51]. Not long after giant viruses were discovered, virophages were found, which parasitize giant viruses, and appear to be highly abundant[55–57]. They share many similarities with satellite viruses, including a reliance on giant viruses to replicate, origins that are phylogenetically distinct from giant viruses, effects ranging from negative to neutral on giant virus replication, and integration into the genomes of Eukaryotic hosts[58]. Based on what is currently known about virophages, it seems that most known virophages are cheats, because they exploit the cooperative replication machinery of giant viruses, and potentially also capsids[55]. However, our knowledge of the molecular mechanisms by which virophages exploit giant viruses is limited, and they may also engage in other forms of parasitism, for instance if they co-opted giant virus resources for novel functions.

Partial Cheats

Defective interfering genomes and satellites represent extreme forms of cheating, where the cheat has completely lost the ability to replicate on its own, and is entirely dependent on the cooperator. Less extreme forms of cheating are also possible in viruses, where the cheat does better when it exploits the cooperator, but can still survive and replicate on its own. For example, PhiH2 is a cheat of the phage Phi6, out-competing Phi6 in mixed infections and

losing in single infections. However, PhiH2 is only a partial cheat, because it still retains some ability to replicate itself in the absence of Phi6[59]. This is analogous to many kinds of facultative cheating elsewhere in nature, such as bacteria that only partially downregulate the production of a public good[19,60]. Partial cheating may be rarer in viruses than in other organisms, perhaps because viruses' smaller genomes lend themselves more easily to simpler 'all-or-nothing' mutations that completely knock out a function, resulting in complete cheating[43].

Inter-cellular

Cooperation and cheating also occur between viral genomes infecting different cells.

Host Immune Defences

In animals, cells which are infected with a virus often produce interferons, a group of signalling molecules that spread to nearby cells and trigger antiviral defences[24]. Many viruses block the release of interferon – this blocking involves producing a costly molecule, but it also provides a benefit to viruses in nearby cells, by keeping the local population of host cells susceptible to infection[25]. Consequently, interferon-blocking represents a form of costly cooperation between viruses in different host cells[61].

In Vesicular Stomatitis Virus (VSV), D51 is a cheat mutant that exploits cooperative interferon blocking. D51 avoids the cost of blocking interferon, consequently replicating more quickly in infected cells, and spreads at the expense of wild-type VSV when both are grown together[25,61]. However, when D51 is grown on its own, no interferon-blocking proteins are produced, and so it quickly becomes extinct because local host cells activate their antiviral defences. Viral mutants that are less effective at suppressing interferon are common

294 in natural infections of many different viruses, including important human pathogens such as
295 SARS-CoV-2 and Influenza A, potentially suggesting that this is a widespread form of viral
296 cheating[62,63].
297
298 ## Virulence
299 In parasites such as viruses, slow growth is another potentially cooperative trait, because it
300 avoids hosts being overexploited, allowing for more transmission opportunities in the long
301 run[64]. Faster growing parasite strains can be seen as cheats, because they can outcompete
302 slower growing strains in the short term, but also exhaust the local supply of hosts[65,66].
303 Examples of virulence cheats in viruses include the fast-growing 'rapacious' phages, which
304 burst their bacterial hosts especially quickly, and have been described in a number of
305 different phage species[32,66,67].
306
307 It is possible that some satellite viruses are a form of virulence cheat. Many satellites
308 are considered to be mutualistic, because they appear to have a positive impact on the
309 replication of the wild-type virus, resulting in higher viral loads and more severe symptoms
310 for the host[47]. However, an alternative explanation could be that they favour a higher level of
311 virulence than the wild-type, because a satellite virus will always exist in a mixed infection
312 with a helper virus, whereas helper viruses can infect plants on their own. Consequently, the
313 satellite virus could be under greater selective pressure than the helper virus towards faster
314 short-term replication within a host. Within-host replication is only one aspect of viral fitness,
315 and so although these satellites appear beneficial for helper virus replication in the short run,
316 they may be costly for long-term fitness.
317

Where else could viral cheats be found?

We expect that the cooperative viral traits we have described are only a fraction of those that exist (Fig. 4)[68–71]. As we explore more of the viral universe, we expect to find new kinds of cooperative traits, that may be exploited by new kinds of viral cheat. Some types of trait that could be cooperative but where cheats have not yet been found include: the production of anti-CRISPR proteins by phages of *Pseudomonas* bacteria[31]; the production of Arbitrium quorum-sensing molecules in phages of *Bacillus* bacteria[72]; and the production of depolymerase enzymes, that many phages produce to break down bacterial cell walls[73].

More broadly, a comparative approach to viral cheats is currently held back by technical limitations and taxonomic bias. For example, we have relatively few examples of 'point mutation' cheats compared to defective interfering genomes, but is this because they are rarer, or just because they are harder to detect? Satellite viruses are found commonly in plants and phages, but less commonly in animal viruses; while defective interfering genomes are found commonly in animal and plant viruses, but less commonly in phages. Are these real patterns, or just statistical artefacts stemming from the fact that we have only studied a small and biased subset of viruses in depth? Technical advances that are allowing unbiased metagenomic sequencing across a broad range of viruses, coupled with sequencing technology that can reveal within-host viral variation, could help to solve these limitations.

An Evolutionary Approach to Cheating

Cheats and Parasites

Cheating can be thought of as a special form of parasitism (Box 2). Some of the viral entities we discuss have also been modelled as parasites in the past, and both perspectives can

342 be helpful[38,39]. However, what is special about cheating is that cheats exploit social traits, and
343 benefit from them in the same way that the trait has evolved to be used. For example, when a
344 defective interfering genome is replicated by a replicase enzyme that was encoded by a
345 cooperative virus. In contrast, parasitism can evolve many other forms of exploitation, such
346 as just eating a host. In microorganisms such as viruses, cheating may evolve more easily
347 than other forms of parasitism, because the cheat does not need to evolve a new way of using
348 a gene product, and so can evolve from just a single loss-of-function mutation.
349
350 This distinction between cheating and other forms of parasitism means that cheating
351 can influence selection on cooperative traits. In the extreme, when cheating is common,
352 cooperation can be disfavoured, even for traits that are relatively essential. In contrast, being
353 eaten by a parasite doesn't select against cooperation. In other cases, the possibility for
354 cheating can shape the evolution of cooperative traits, such as when they evolve to be
355 directed more specifically towards cooperators, so that they are less likely to benefit cheats.
356 More broadly, this means that the social environment influences whether cooperation is
357 favoured, and the details of the cooperative traits that evolve.
358
359 ## The Cheating Framework
360 Our approach highlights the evolutionary similarities between the different types of viral
361 cheat, despite the differences that exist in the biological details of how they exploit
362 cooperation. Consequently, the cheating perspective provides a single framework for
363 studying these entities, that have historically been studied within their respective disciplines
364 or subdisciplines[28,30,47,74]. It also draws links with the broader field of cheating within
365 evolutionary biology, highlighting conceptual analogies, as well as theoretical and empirical
366 methods, that can be applied across disciplines[5,6].

For viruses specifically, one further advantage of using an evolutionary definition of cheating is that we can classify new viral entities and make predictions even when we do not yet understand all of the biological details. For example, using simple experiments such as in Fig. 3, Turner & Chao determined that PhiH2 is a partial cheat of Phi6, without fully uncovering the mechanisms by which it gains an advantage[59,75]. Such experiments allow us to place newly discovered viral entities within an existing framework, and to draw common links between otherwise disparate parts of the rapidly expanding virosphere. This offers one way to help bring 'order to the viral universe'[76].

When are Cheats Favoured?

Successful cheats and relatedness

The success of a cheat depends upon its ability to interact with and exploit cooperators. This will largely be determined by the extent to which different genomes are able to interact and share resources, which is summarised by the genetic relatedness between interacting genomes (Box 4). When relatedness is high, kin selection favours cooperation[77]; when relatedness is low, cheats can invade, leading to either coexistence of cooperators and cheats, or the loss of cooperation.

Relatedness has a clear and consistent effect on the evolution of cooperation and cheating across all levels of biology[14]. We suggest it can also be useful for understanding when viral cheats will spread. When a single viral genome infects a cell, relatedness is high, and cooperation favoured, since all of the viral genomes are likely to be highly genetically similar (effectively clonal, r=1) (Box 4). In contrast, when coinfection allows genetically

391 distinct viral genomes to infect the same cells, then relatedness will be low, and cheating
392 potentially favoured. We can predict which of these scenarios will occur, and therefore
393 whether or not cheating is favoured, based on biological properties of viruses, and physical
394 properties of the environments they exist in.
395
396 ## Viral Biology and Relatedness
397 Biological properties of viruses can influence relatedness indirectly. For example, smaller
398 virions may disperse further, decreasing spatial structure and leading to lower relatedness
399 conditions[78]. Higher mutation rates could also lower relatedness by generating high amounts
400 of genetic diversity, even in cells or hosts initially infected by just a single viral genome[64].
401
402 In some cases, viruses have traits that directly influence relatedness. Commonly, the
403 first viral genome to infect a cell excludes further viral genomes from infecting that cell
404 (superinfection exclusion)[79–81]. Superinfection exclusion maintains a high relatedness,
405 allowing cooperators to avoid interacting with cheats, and may have evolved for that
406 reason[79].
407
408 In other cases, viruses have traits that could directly lower relatedness, by allowing
409 multiple viral genomes to infect the same cell[82,83]. These mechanisms include 'collective
410 infectious units', such as when one virion contains multiple viral genomes, or when multiple
411 virions aggregate after leaving host cells[82]. Direct cell-cell transmission also occurs in many
412 plant and animal viruses, potentially allowing hundreds to thousands of viral genomes to
413 infect the same cell simultaneously[84–87]. When these collective infection mechanisms allow
414 coinfection between viral genomes that come from different cells, they can lower relatedness
415 and select for cheats[83,88].

Viral Environments and Relatedness

Within-Host Relatedness

What is relatedness in natural viral populations? This will depend on both the degree to which coinfection occurs, and the degree to which coinfection allows distinct viral genomes to infect the same cells (Box 4). Empirically, quantitative estimates suggest that coinfection occurs frequently in many natural viral populations. In Guinea pigs infected with influenza A[89], and turnip plants infected by cauliflower mosaic virus[90], 5-15 and 2-13 viral genomes infected each host cell respectively. In marine *Gammaproteobacteria*, half of the infected bacterial cells contained multiple actively replicating phage species[91].

The extent to which these instances of coinfection will lead to low relatedness will depend on spatial structuring. When viruses grow in an environment with strong spatial structuring, coinfecting viruses are more likely to have come from the same original cell and be genetically identical, so relatedness can be high even when coinfection is common. Many natural environments appear to have strong spatial structuring, with viruses growing in specific tissues, or in patches within each tissue[78,92,93].

However, despite evidence for spatial structuring in some host environments, coinfection appears to occur relatively frequently between distinct viral genomes in natural viral populations, suggesting that relatedness can often be low. For example: defective viruses are maintained in natural infections; viruses such as influenza and HIV show relatively high rates of reassortment and recombination between different genomes; and viruses modified to be entirely dependent on coinfection can grow robustly in animal hosts[94–99].

It seems likely that relatedness varies both between different host types, and between different tissue types within the same host. This could be one explanation for why defective interfering genomes accumulate at different rates in different tissues of the same host[100]. More precise estimates of relatedness within hosts could be inferred from high-coverage sequencing of viral populations, especially sequencing of multiple infected tissues within the same host.

Relatedness Between Hosts

Although it seems that relatedness is often low within hosts, this is only one part of the viral environment, since viruses also spread between hosts. When viruses spread from one host to another, usually only a small number of distinct viral genomes are transmitted to the new host - a population 'bottleneck'[101]. When bottlenecks are narrow, consisting of only a few viral genomes, relatedness will be high and so it is less likely that cheats will coinfect a new host alongside cooperators (Box 4). In contrast, wider bottlenecks will lower relatedness at the start of a new host infection, potentially making it more likely that cheats will spread between hosts.

Empirically, bottlenecks are often particularly narrow in animal viruses, such as HIV, SARS-CoV-2, and Hepatitis C, where infections tend to be initiated by just a single viral genome[101–105]. In plant viruses, bottlenecks are often wider. For example, in Cauliflower Mosaic Virus, infections can be initiated by 1-13 genomes[90,104]. Bottlenecks can also vary within the same virus, depending on the transmission route. For example, in influenza A, bottlenecks are larger when transmission is via direct contact, than when it is via aerosol[106]. Some of the widest between-host bottlenecks are found when viruses transmit between hosts

using collective infectious units, sometimes allowing hundreds of viral genomes to infect a new host[82,107]. These patterns could explain why cheats that spread between hosts, such as satellite viruses, are more common in plant viruses. Similarly, defective interfering genomes commonly transmit between hosts in baculoviruses, which transmit between their insect hosts within large collective infectious units, but defective interfering genomes appear to transmit between hosts less commonly in other animal viruses[108].

Relatedness in Man-Made Viral Environments

The role of relatedness has been demonstrated unintentionally numerous times, when growing viruses. Artificial environments, such as tissue cultures or bioreactors, often involve large numbers of viruses compared to the number of host cells, and can be well-mixed, with weak spatial structuring. This means that viruses often achieve very high rates of coinfection in artificial environments, with potentially hundreds to thousands of viruses infecting each host cell, and where those viruses are likely to have come from different host cells. These are low relatedness conditions that should favour cheats spreading.

Consistent with this, viral cheats such as defective interfering genomes are extremely common in viral tissue culture infections, and have long been an issue in industrial processes that depend on culturing viruses, such as the production of vaccines, biopesticides, or vectors for gene therapy[109–111]. Many techniques that industrial producers use to increase yields, such as periodic bottlenecking, are effective because they increase relatedness, making it harder for these cheats to spread[112]. The close parallels between these methods and the formal tests for cooperation and cheating highlight the importance of cooperation and cheating in viral population dynamics (Fig. 3).

Why don't cheats take over?

Given the potential benefits of cheating, what stops viral cheats from spreading to fixation after they have arisen and started to spread? When relatedness is low, do cheats inevitably win, or can cooperators and cheats coexist?

In cases where relatedness is low enough for cheats to initially spread, they can be prevented from spreading to fixation by frequency dependence. A common feature of cheating is that the relative fitness of cheats decreases as they become more common – termed negative frequency dependence[77]. Because cheats spread by exploiting cooperators, they experience the greatest fitness advantages when rare, when most other individuals they interact with are cooperators. In contrast, as cheats become more common, they interact with other cheats more frequently than with cooperators, and so their fitness advantage decreases. Consequently, cheating can be self-limiting, and even cheats that have substantial fitness advantages when rare, may end up coexisting with cooperators rather than driving cooperators extinct.

Another possibility is that cooperators can adapt to the presence of cheats, in a way that limits their spread[113–115]. In vesicular stomatitis virus (VSV), wild-type cooperators can evolve a form of resistance to cheats, by changing the recognition sequence for the replicase enzyme, so that it still replicates the wild-type cooperator, but no longer replicates the defective interfering cheat genome[116]. Alternatively, viruses could evolve to manipulate population structure in ways that increase relatedness, preventing cheats from spreading (Box 4). For example, viruses could decrease the number of viral genomes that collectively transmit to new cells (smaller collective infectious units)[83,88], or exclude additional viral genomes from infecting the same host cell (superinfection exclusion)[79–81].

Why Should Evolutionary Biologists Care?

Viruses versus Other Lifeforms

A comparison of cheating in viruses versus other organisms raises the question of whether cheating in viruses is the same as cheating elsewhere in the living world (Fig. 1)? We argue that while it is clearly analogous, viral biology leads to important differences. These include:

(1) The high mutation rate and simple genome of viruses means that mutations to cheating can happen relatively easily[117]. For example, defective interfering genomes regularly emerge *de novo* in viral infections[118]. This high mutation rate allows cheats to frequently arise and spread, even when they would not be maintained long-term.

(2) Viruses can benefit from cheating in a unique way: not only do they benefit through avoiding cooperation, but also through losing or otherwise modifying those now-redundant cooperative genes[41,119,120]. Cheat genomes can therefore be replicated and encapsidated much faster than complete, cooperative viruses, giving an additional advantage with no clear analogue elsewhere in the natural world.

(3) Consequently, the short-term advantages of cheating in viruses can be exceptionally high. Viral cheats can achieve a 1,000-fold or higher replicative advantage over cooperators, which is orders of magnitude higher than the fitness advantages seen in cuckoos, non-producing bacteria, or other cheats[8,19,121].

(4) However, this fitness advantage of cheats is often transient at a local scale. For example, many viral cheats emerge easily, and spread rapidly, within a host, but then show poor or even non-existent transmission to new hosts[88].

Taken together, these features mean that cheating can be both common and transient in many viruses. Viral cheats are therefore special in the extent to which they are often characterised by 'boom and bust' dynamics. More so than in other organisms, viruses could be selected to evolve mechanisms to avoid generating cheats, and/or reduce exploitation by cheats.

Not all viral cheats are transient. We can place viral cheats on a continuum between 'short-sighted' and 'long-sighted' cheats[122]. Defective interfering genomes are short-sighted cheats, that arise and spread transiently, mostly within but not between hosts, with boom and bust dynamics. Satellite viruses are long-sighted cheats that spread both within and between hosts, allowing persistence over long evolutionary timescales. Long-sighted cheats are more similar to forms of cheating observed in animals, such as cuckoos, whereas short-sighted cheats may be closer to public goods cheats in bacteria[5,19,113].

Viruses Make Model Cheats

The unique features of viral cheating make viruses excellent model organisms for studying cheating. Cheats may be both more common and easier to find in viruses than in other organisms. The relatively small genomes and short generation times of viruses mean that it is often easy to link genotype with phenotype, allowing us to identify cheats relatively easily, and to follow evolutionary dynamics over time[123]. The large amounts of clinical and environmental genomic data allow the ecological and coevolutionary dynamics of cheating to be studied in nature[12,96]. These studies can then be complemented with manipulative laboratory experiments that are more feasible in viruses than in other organisms[123].

Novel Evolutionary Problems

Cheating in viruses raises novel evolutionary problems. In the laboratory, viruses can be genetically modified so that it is harder for cheats to arise through mutation[124]. Have viral genomes naturally evolved similar strategies to limit the emergence of cheats, for instance by linking cooperative genes with essential private functions, that cannot be cheated[125]? There are a number of instances where it looks like this might have happened: in measles virus, the C protein inhibits the formation of defective interfering genomes[126]; in polioviruses, defective genomes that lack sections of the replicase gene are unable to be incorporated into virions, and so 'replicase-cheats' do not evolve (although 'capsid cheats' do)[127]; in Flock House Virus, successful cheats contain two large deletions in the genome, because cheats with just one large deletion lose essential functions in the middle of the genome that cannot be complemented by coinfecting with another genome[128].

Why Should Virologists Care?

Adaptation in Viruses

An understanding of cheat-cooperator dynamics can inform how we think about viral populations. Viral cheats appear to be very common, but also have strongly negative consequences for viral infections, and for other viral variants. This challenges the idea that viruses should be defined at the group or 'quasispecies' level, because it suggests that the potential for conflict is likely to prevent adaptations that are solely for the benefit of the group of viruses[129,130]. Therefore, we should not necessarily expect viral populations to evolve as coherent groups, nor to be adapted towards any collective goal.

New Predictions: Molecular and Genomic

Social evolution theory makes broad predictions about the genomic and molecular biology of viral cheats, but in many cases more specific predictions will depend on linking more detailed viral biology with the evolutionary biology of cheating.

For instance, we can predict that intracellular viral cheats are likely to exploit trans-acting gene products such as replicase enzymes and capsid proteins, since these can be public goods (Box 1). Molecular details could then allow more precise predictions to be developed, such as why viral cheats sometimes exploit capsid proteins, sometimes replicase, and sometimes both.

We can also predict which viruses are most likely to be affected by cheats. Viruses that encode many trans-acting (social) gene products should be more likely to be exploited by viral cheats, as should viruses that experience higher rates of coinfection, such as those with mechanisms for collective infection. Pleiotropy will make it harder for viral cheats to emerge when it links trans-acting (social) with cis-acting (non-social) functions, such as in the replicase gene of poliovirus[127]. However, pleiotropy may make it easier for viral cheats to emerge when it links multiple social gene products, as occurs in the phage M2, since this allows a single mutation to exploit multiple forms of cooperation[120].

We might expect viral cheats to be more common in short RNA viruses, where mutations rates are high and a single cooperative gene can comprise a large fraction of the viral genome[117]. On the other hand, cheating could be easier in large DNA viruses, which are less likely to have pleiotropy linking social with non-social traits, and may contain a larger number of cooperative genes, allowing cheats to lose many genes and become much shorter

than cooperators[44]. Formal tests of these kinds of evolutionary prediction will become possible when we have more unbiased sampling of cheats across different types of virus, and more thoroughly annotated viral genomes.

We would also expect to see long-term genomic consequences of cheating. A number of resistance mechanisms are possible in response to viral cheating, such as pleiotropy that links social (trans-acting) to non-social (cis-acting) genes, modifying trans-acting genes to be more cis-acting, or decreasing the size of collective infectious units[116,124]. Different types of virus may vary in the extent to which resistance is possible. For instance, it may be less feasible for viruses with segmented genomes to evolve replicase enzymes that are more cis-acting, because this would require simultaneous mutations on several genome segments at once. Resistance mechanisms may also complicate evolutionary predictions, because they are most likely to evolve in viruses that experience lots of cheating, but they are then likely to decrease the prevalence of cheating in those viruses, potentially creating a 'chicken-and-egg' problem.

Explaining clinical outcomes

Understanding cheating could help explain why the same virus can lead to different clinical outcomes in different patients. For example, if cheats spread during an infection, then this could lead to a lower viral load and lower virulence. Consistent with this, infections with both influenza and ebola viruses that contain a larger number of defective interfering genomes are less likely to lead to severe clinical outcomes, such as admittance to the intensive care unit[131,132]. However, one potential complication here is that higher viral loads can allow cheats to better exploit cooperators, in which case cheats might be more likely to be found in infections with higher viral loads, and more severe clinical outcomes.

638

639 Social dynamics could help to explain why variants that arise during individual
640 clinical infections often fail to spread between hosts to become dominant on an
641 epidemiological scale[133,134]. One possibility is that some of these variants could be short-
642 sighted cheats, that can spread within a host, but that are unlikely to co-transmit with
643 cooperative viruses between hosts, due to small between-host bottlenecks[101,104].

644

645 ## New Treatments

646 Cheating can be exploited as a mechanism to disrupt viral infections, and social evolution
647 could inform how we approach this. Therapeutic interfering particles (TIPs) are synthetic
648 viruses designed to exploit wild-type virus cooperation, and to suppress viral infections by
649 acting as a cheat, mimicking defective interfering genomes[135,136]. They are analogous to other
650 types of 'cheat therapy' being developed against bacteria[137,138]. Animal challenge studies
651 suggest that therapeutic interfering particles can be highly effective, both as prophylactic and
652 as a treatment post-infection, against viruses including Lassa virus, Chikungunya virus,
653 Influenza A virus, and SARS-CoV-2[100,139–143].

654

655 One advantage of therapeutic interfering particles is that they exploit predictable
656 features that are common to all viruses, and so it could be relatively quick and
657 straightforward to develop therapeutic interfering particles against a novel virus. Another
658 advantage is that some therapeutic interfering particles are effective even against viruses that
659 are relatively distantly related to the virus that they came from – a feature they share with
660 many satellite viruses[136]. For example, defective interfering genomes of Chikungunya virus
661 can be effective against other alphaviruses, such as Sindbis virus and O'nyong-nyong

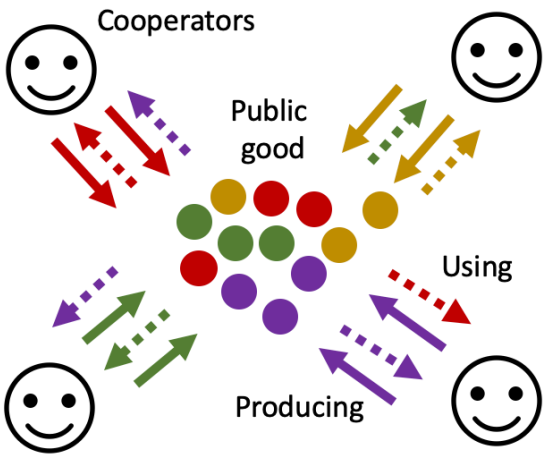
662 virus[100,143]. Consequently, it could be relatively quick to deploy therapeutic interfering
663 particles against a novel virus, either by designing new ones or adapting existing ones.
664
665 Social evolution theory could help to design effective therapeutic interfering particles
666 (TIPs), because it offers tools for answering analogous questions to those being posed in
667 therapeutic interfering particle research. For example, compare "when do TIPs suppress wild-
668 type viruses?", "can TIPs be maintained in the population?", and "can TIPs revert to being
669 fully infectious viruses?", with "when do cheats win?", "when does frequency dependent
670 selection maintain cheats and cooperators at equilibrium?", and "can cooperation be
671 regained?".
672
673 A social evolution perspective could also help us determine how to use therapeutic
674 interfering particles effectively, by focusing on the evolutionary dynamics of natural viral
675 cheats. Before using a synthetic cheat to control a viral infection, we would first want to
676 know what kinds of cheat affect the virus naturally, and how the virus responds to them. Are
677 some viruses more susceptible than others to being exploited by cheats? Which viral cheats
678 are able to spread between hosts, and why? In what ways do viruses evolve resistance to
679 cheats, and can cheats coevolve in response? These are all pressing questions about the
680 natural history of viral cheats, which also have clear applications in informing how
681 therapeutic interfering particles could be used safely and effectively.
682

683 ## Where next?

684 The study of viral cheats offers amazing opportunities to both evolutionary biology and
685 virology. However, the biggest obstacle in this field is that most of the existing empirical
686 work has been done in the laboratory, either in tissue culture or in model hosts. To move

forward, we need to understand the role that cheats play in the epidemiological and evolutionary dynamics of viruses in their natural environment, which includes humans, crops, and livestock. Fortunately, exactly the right kind of data is now being collected, as next-generation sequencing technology is increasingly being used to monitor viral outbreaks and to chart the enormous unexplored diversity of viruses[69–71,144–146]. The next steps will involve harnessing this rapidly advancing technology in order to test and expand evolutionary theory about viral cheats.

Box 1: Cooperation as a Public Goods Game



(Fig legend begin)

Public goods provide benefits to all of the individuals in the group, not just the individual that produced the public good. The benefits of producing a public good are shared with increasing numbers of individuals as the size of the group increases. In microbes, well-studied examples of public goods include elastase, siderophores, beta-lactamase, and others. We suggest that many shared viral gene products, such as replicase and capsid proteins, also act as public goods (Fig. 2).

(Fig legend end)

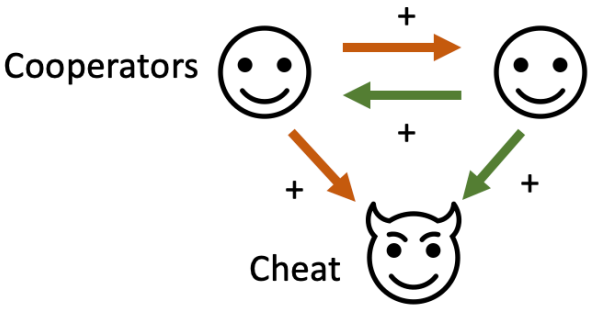
Many forms of cooperation in microorganisms such as bacteria and viruses are analogous to what economists and evolutionary biologists call a public goods game. In the simplest public goods game, there are *N* unrelated group members who can each contribute some resources to a group project. Those resources are then multiplied by a factor *M*, and divided out amongst each member of the group, such that each individual gains *M/N* per unit of resources contributed.

This game illustrates the problem of cooperation. Cooperation by producing public goods is favoured at the group level – if all individuals cooperate, everyone does better. However, if each individual gains back less than one unit for each unit that they invested (when M/N<1), then each individual does better if they invest nothing (not cooperate). Selfish interests are increasingly likely to outweigh the benefits of cooperation as N increases, since higher N means that the benefits of cooperation have to be shared with increasing numbers of other individuals. Even though each individual gets a return on their own investment, they can still be selected to invest nothing.

Viruses play public goods games whenever there are multiple viral genomes in a cell (N > 1). In that case, the production of shareable gene products, such as replicase enzymes or capsid proteins, becomes a public goods game that is open to cheating (*N*>1). Multiple viral genomes will generally be present inside an infected cell, since even if a cell is initially infected by just a single viral genome, that genome will be replicated, quickly resulting in large numbers of viral genomes[147]. Public goods games in viruses are closely analogous to those in bacteria and other microorganisms, which produce shared gene products such as iron-scavenging siderophore molecules.

Public goods are so common in viruses that virologists already have a term for them. 'Trans-complementable' or 'trans-acting' describes gene products that are shared between different viral genomes in the same cell, as opposed to 'cis-acting' gene products that only affect the genome that encoded them[23]. Common examples that all viruses need include replicase enzymes, that replicate the viral genome, and capsid proteins, that construct the capsid that transports viral genomes to new cells. From an evolutionary perspective, trans-complementable viral gene products are public goods, and almost all viruses depend on some kind of public good to complete their lifecycle[23].

Box 2: An Evolutionary Definition of Cheating



(Fig legend begin)

Cheats exploit cooperation.

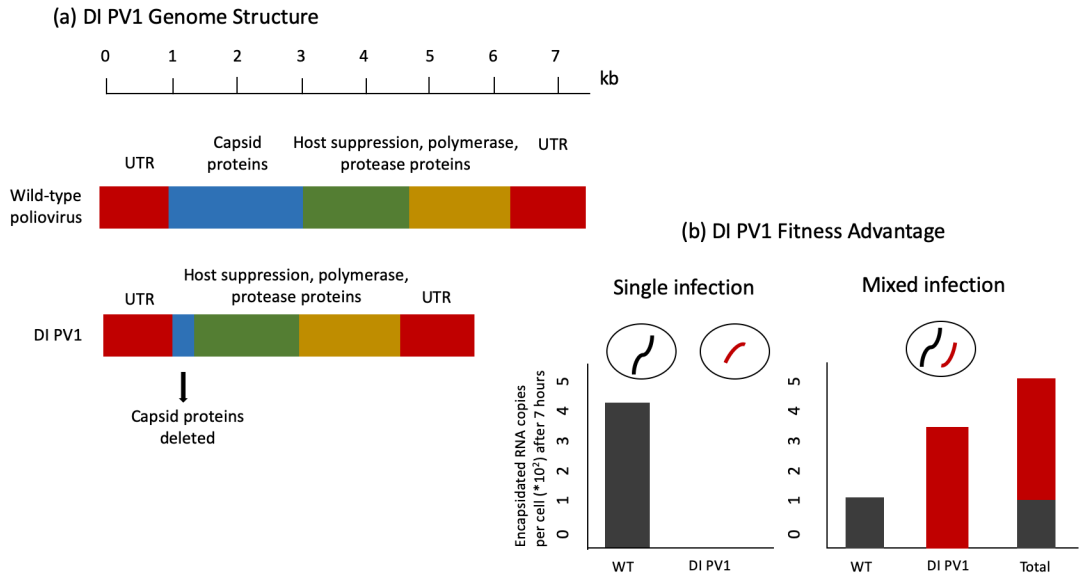
(Fig legend end)

Cheats exploit cooperators. More formally, we follow the evolutionary definition of cheating as: (i) a trait that is beneficial to a cheat and costly to a cooperator in terms of inclusive

748 fitness; (ii) when these benefits and costs arise from the actor directing a cooperative
749 behaviour toward the cheat, rather than the intended recipient[5].
750
751 Our definition of cheating is relatively broad, focusing on when cooperation can be
752 exploited. An alternative, narrower definition is also possible, in which a cheat also has to
753 have evolved from a cooperative lineage[6]. When applied to viruses, this narrower definition
754 would still classify a range of cases as cheating, especially defective interfering genomes, and
755 point-mutation mutants such as D51 and PhiH2. However, the narrower definition would
756 exclude satellite viruses, since these have unclear origins and may not have evolved from the
757 cooperator they exploit, and therapeutic interfering particles, since these are manufactured
758 rather than naturally evolved. It could also discount cases where a defective interfering
759 genome that evolved from one viral variant or species exploits a different type of virus. The
760 broader definition that we use counts all of these as cheats, but then classifies them into
761 different types of cheat (Fig. 4; Fig. 5)[5].
762
763 The use of different definitions depends on the questions being asked. The narrower
764 definition places the emphasis on asking when cheating can evolve *de novo* within a system
765 that is originally purely cooperative. In contrast, the broader definition focuses on when
766 cooperation can be exploited and potentially broken down more widely. In the case of
767 viruses, the broader definition emphasises functional similarities, in terms of fitness
768 consequences, between entities that may otherwise appear quite different, such as defective
769 interfering genomes and satellite viruses, or between defective interfering genomes and
770 therapeutic interfering genomes (Fig. 4; Fig. 5). In addition, the broader definition provides a
771 framework for classifying different types of cheat, such as short versus long-sighted cheats
772 (Fig. 4; Fig. 5).

By our definition, cheating is context-dependent, meaning that an individual that produces less of something can be a relative cheat, compared to an individual that produces more[60]. Consequently, a spectrum of cheating is possible, between 'full cheats' that completely lose the ability to cooperate, and 'partial cheats' that keep some or most ability to cooperate.

Box 3: A Typical Viral Cheat



(Fig Legend Start)

DI PV1 is a model viral cheat.

DI PV1, a defective interfering genome, is a cheat of poliovirus. (a) DI PV1 lacks the section of genome that encodes capsid proteins, resulting in a substantially shorter genome than the cooperative wild-type. (b) Consequently, DI PV1 gains more than a 1,000-fold replication advantage over the wild-type cooperator when both coinfect the same cell. Data for (b) taken from Shirogane et al 2021[121].

(Fig Legend End)

The defective interfering genome 'DI PV1' is cheat of poliovirus. DI PV1 contains a large deletion that removes the entire capsid protein region[121,148]. When grown on its own, DI PV1 therefore produces no viral capsids, and so is unable to spread between host cells. However, when wild-type poliovirus and DI PV1 are grown together, copies of DI PV1 can be incorporated into viral capsids produced by the wild-type cooperator. In coinfected cells, the shorter length of DI PV1 means that it is replicated substantially faster than the wild-type, and it is also able to enter virions more effectively than the wild-type. Consequently, DI PV1 is able to achieve more than 1,000 times as many genomes inside viral capsids as the wild-type cooperator, which is a huge fitness advantage[121].

DI PV1 is a well-known viral entity that has been studied for decades, that also provides a clear fit to the evolutionary definition of a cheat[5,121,127,148–152]. It avoids encoding a cooperative trait (producing capsid proteins), but it is able to exploit the cooperation of other genomes (by using capsid proteins they encoded). There are direct parallels between the experiments that virologists used to investigate DI PV1, with the experiments that evolutionary biologists typically conduct to examine cheating in bacteria[8].

Box 4: Genetic Relatedness

Genetic relatedness is a statistical concept, describing the degree of genetic similarity between social partners, over and above genetic similarity to the average individual in the population[153]. In the simplest case, if N genetically distinct viral genomes infect a cell, then average relatedness will be $r=1/N$. This comes from the average of individuals being related by $r=1$ to their clonemates, and by $r=0$ to individuals in the other N-1 lineages.

814

815 Relatedness changes the benefit of investing into public goods (Box 2). When
816 relatedness is high, interacting individuals (social partners) are more likely to share genes,
817 and so the benefits of cooperation are likely to be returned to other individuals who also carry
818 the gene for cooperation. Consequently, cooperation provides an indirect (kin selected)
819 benefit. At the same time, high relatedness means that cheats are likely to interact with other
820 cheats, meaning they will be unable to exploit cooperation, and so will not spread.
821

822 In contrast, when relatedness is low, interacting individuals are less likely to share
823 genes, and so cooperation provides a smaller indirect (kin selected) benefit. At the same time,
824 low relatedness means that cheats are more likely to interact with cooperators, meaning they
825 will be able to exploit them, and hence spread.
826

827 Empirically, relatedness has been shown to have a clear and consistent influence on
828 the evolution of cooperation, at all levels of biology, from humans, through birds, bees, and
829 insects, all the way down to bacteria, viruses, and simple RNA replicators[14].
830
831
832
833
834
835
836

References

1. West, S. A., Griffin, A. S. & Gardner, A. Evolutionary Explanations for Cooperation. *Curr. Biol.* **17**, R661–R672 (2007).
2. Bourke, A. F. G. *Principles of Social Evolution*. (Oxford University Press, 2011).
3. Davies, N. B., Krebs, J. R. & West, S. A. *An Introduction to Behavioural Ecology*. (John Wiley & Sons, 2012).
4. West, S. A., Griffin, A. S. & Gardner, A. Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. *J. Evol. Biol.* **20**, 415–432 (2007).
5. Ghoul, M., Griffin, A. S. & West, S. A. Toward an evolutionary definition of cheating. *Evolution* **68**, 318–331 (2013).
6. Jones, E. I. *et al.* Cheaters must prosper: reconciling theoretical and empirical perspectives on cheating in mutualism. *Ecol. Lett.* **18**, 1270–1284 (2015).
7. Andersen, S. B., Marvig, R. L., Molin, S., Krogh Johansen, H. & Griffin, A. S. Long-term social dynamics drive loss of function in pathogenic bacteria. *Proc. Natl. Acad. Sci.* **112**, 10756–10761 (2015).
8. Griffin, A. S., West, S. A. & Buckling, A. Cooperation and competition in pathogenic bacteria. *Nature* **430**, 1024–1027 (2004).
9. Flower, T. Fork-tailed drongos use deceptive mimicked alarm calls to steal food. *Proc. R. Soc. B Biol. Sci.* **278**, 1548–1555 (2011).
10. Frederickson, M. E. Mutualisms Are Not on the Verge of Breakdown. *Trends Ecol. Evol.* (2017) doi:10.1016/j.tree.2017.07.001.
11. Jandér, K. C. & Herre, E. A. Host sanctions and pollinator cheating in the fig tree–fig wasp mutualism. *Proc. R. Soc. B Biol. Sci.* **277**, 1481–1488 (2010).
12. Ostrowski, E. A. *et al.* Genomic Signatures of Cooperation and Conflict in the Social Amoeba. *Curr. Biol.* **25**, 1661–1665 (2015).

13. Hamilton, W. D. The genetical evolution of social behaviour. I. *J. Theor. Biol.* **7**, 1–16 (1964).
14. West, S. A., Cooper, G. A., Ghoul, M. B. & Griffin, A. S. Ten recent insights for our understanding of cooperation. *Nat. Ecol. Evol.* 1–12 (2021) doi:10.1038/s41559-020-01384-x.
15. West, S. A., Griffin, A. S., Gardner, A. & Diggle, S. P. Social evolution theory for microorganisms. *Nat. Rev. Microbiol.* **4**, 597–607 (2006).
16. Kramer, J., Özkaya, Ö. & Kümmerli, R. Bacterial siderophores in community and host interactions. *Nat. Rev. Microbiol.* 1–12 (2019) doi:10.1038/s41579-019-0284-4.
17. Gano-Cohen, K. A. *et al.* Recurrent mutualism breakdown events in a legume rhizobia metapopulation. *Proc. R. Soc. B Biol. Sci.* **287**, 20192549 (2020).
18. Cordero, O. X., Ventouras, L.-A., DeLong, E. F. & Polz, M. F. Public good dynamics drive evolution of iron acquisition strategies in natural bacterioplankton populations. *Proc. Natl. Acad. Sci.* **109**, 20059–20064 (2012).
19. Davies, N. B. *Cuckoos, Cowbirds and Other Cheats*. (T & AD Poyser, 2010).
20. Díaz-Muñoz, S. L., Sanjuán, R. & West, S. A. Sociovirology: Conflict, Cooperation, and Communication among Viruses. *Cell Host Microbe* **22**, 437–441 (2017).
21. West, S. A., Fisher, R. M., Gardner, A. & Kiers, E. T. Major evolutionary transitions in individuality. *Proc. Natl. Acad. Sci.* **112**, 10112–10119 (2015).
22. Queller, D. C. & Strassmann, J. E. Beyond society: the evolution of organismality. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **364**, 3143–3155 (2009).
23. Flint, J., Racaniello, V. R., Rall, G. F. & Skalka, A. M. *Principles of Virology*. (American Society of Microbiology, 2015). doi:10.1128/9781555819521.
24. Murphy, K. *Janeway's Immunobiology*. (Garland Science, 2011).

25. Domingo-Calap, P., Segredo-Otero, E., Durán-Moreno, M. & Sanjuán, R. Social evolution of innate immunity evasion in a virus. *Nat. Microbiol.* 1 (2019) doi:10.1038/s41564-019-0379-8.
26. Landsberger, M. *et al.* Anti-CRISPR Phages Cooperate to Overcome CRISPR-Cas Immunity. *Cell* **174**, 908-916.e12 (2018).
27. Borges, A. L. *et al.* Bacteriophage Cooperation Suppresses CRISPR-Cas3 and Cas9 Immunity. *Cell* **174**, 917-925.e10 (2018).
28. Vignuzzi, M. & López, C. B. Defective viral genomes are key drivers of the virus–host interaction. *Nat. Microbiol.* 1 (2019) doi:10.1038/s41564-019-0465-y.
29. Huang, A. S. & Baltimore, D. Defective Viral Particles and Viral Disease Processes. *Nature* **226**, 325–327 (1970).
30. Simon, A. E., Roossinck, M. J. & Havelda, Z. Plant virus satellite and defective interfering RNAs: new paradigms for a new century. *Annu. Rev. Phytopathol.* **42**, 415–437 (2004).
31. Chevallereau, A. *et al.* Exploitation of the Cooperative Behaviors of Anti-CRISPR Phages. *Cell Host Microbe* **0**, (2019).
32. Kerr, B., Neuhauser, C., Bohannan, B. J. M. & Dean, A. M. Local migration promotes competitive restraint in a host–pathogen 'tragedy of the commons'. *Nature* **442**, 75 (2006).
33. von Magnus, P. *Studies on interference in experimental influenza*. vol. 1 (Almqvist & Wiksell, 1947).
34. Rezelj, V. V., Levi, L. I. & Vignuzzi, M. The defective component of viral populations. *Curr. Opin. Virol.* **33**, 74–80 (2018).
35. Pathak, K. B. & Nagy, P. D. Defective Interfering RNAs: Foes of Viruses and Friends of Virologists. *Viruses* **1**, 895–919 (2009).

911 36. Nee, S. & Maynard Smith, J. The evolutionary biology of molecular parasites.
912 *Parasitology* **100**, S5–S18 (1990).
913 37. Szathmáry, E. Co-operation and defection: playing the field in virus dynamics. *J. Theor.*
914 *Biol.* **165**, 341–356 (1993).
915 38. Kirkwood, T. B. & Bangham, C. R. Cycles, chaos, and evolution in virus cultures: a
916 model of defective interfering particles. *Proc. Natl. Acad. Sci.* **91**, 8685–8689 (1994).
917 39. Frank, S. A. Within-host spatial dynamics of viruses and defective interfering particles.
918 *J. Theor. Biol.* **206**, 279–290 (2000).
919 40. Brown, S. P. Collective action in an RNA virus. *J. Evol. Biol.* **14**, 821–828 (2001).
920 41. Chao, L. & Elena, S. F. Nonlinear trade-offs allow the cooperation game to evolve from
921 Prisoner's Dilemma to Snowdrift. *Proc R Soc B* **284**, 20170228 (2017).
922 42. Rüdiger, D., Kupke, S. Y., Laske, T., Zmora, P. & Reichl, U. Multiscale modeling of
923 influenza A virus replication in cell cultures predicts infection dynamics for highly
924 different infection conditions. *PLOS Comput. Biol.* **15**, e1006819 (2019).
925 43. Holmes, E. C. *The Evolution and Emergence of RNA Viruses*. (Oxford University Press,
926 2009).
927 44. Schröder, C. H., Fürst, B., Weise, K. & Gray, C. P. A Study of Interfering Herpes
928 Simplex Virus DNA Preparations Containing Defective Genomes of Either Class I or II
929 and the Identification of Minimal Requirements for Interference. *J. Gen. Virol.* **65**, 493–
930 506 (1984).
931 45. Vogt, P. K. & Jackson, A. O. *Satellites and defective viral RNAs*. (Springer, 1999).
932 46. Roossinck, M. J., Sleat, D. & Palukaitis, P. Satellite RNAs of plant viruses: structures
933 and biological effects. *Microbiol. Rev.* **56**, 265–279 (1992).
934 47. Gnanasekaran, P. & Chakraborty, S. Biology of viral satellites and their role in
935 pathogenesis. *Curr. Opin. Virol.* **33**, 96–105 (2018).

936 48. Qiu, W. & Scholthof, K.-B. G. Defective Interfering RNAs of a Satellite Virus. *J. Virol.*
937 **75**, 5429–5432 (2001).
938 49. Chang, W.-S. *et al.* Novel hepatitis D-like agents in vertebrates and invertebrates.
939 *bioRxiv* 539924 (2019) doi:10.1101/539924.
940 50. Penadés, J. R. & Christie, G. E. The Phage-Inducible Chromosomal Islands: A Family
941 of Highly Evolved Molecular Parasites. *Annu. Rev. Virol.* **2**, 181–201 (2015).
942 51. Mougari, S., Sahmi-Bounsiar, D., Levasseur, A., Colson, P. & La Scola, B. Virophages
943 of Giant Viruses: An Update at Eleven. *Viruses* **11**, 733 (2019).
944 52. Parks, W. P., Casazza, A. M., Alcott, J. & Melnick, J. L. Adeno-associated satellite
945 virus interference with the replication of its helper adenovirus. *J. Exp. Med.* **127**, 91–
946 108 (1968).
947 53. McKitterick, A. C. & Seed, K. D. Anti-phage islands force their target phage to directly
948 mediate island excision and spread. *Nat. Commun.* **9**, 2348 (2018).
949 54. La Scola, B. *et al.* A Giant Virus in Amoebae. *Science* **299**, 2033–2033 (2003).
950 55. La Scola, B. *et al.* The virophage as a unique parasite of the giant mimivirus. *Nature*
951 **455**, 100–104 (2008).
952 56. Zhou, J. *et al.* Diversity of Virophages in Metagenomic Data Sets. *J. Virol.* **87**, 4225–
953 4236 (2013).
954 57. Paez-Espino, D. *et al.* Diversity, evolution, and classification of virophages uncovered
955 through global metagenomics. *Microbiome* **7**, 157 (2019).
956 58. Duponchel, S. & Fischer, M. G. Viva lavidaviruses! Five features of virophages that
957 parasitize giant DNA viruses. *PLOS Pathog.* **15**, e1007592 (2019).
958 59. Turner, P. E. & Chao, L. Prisoner's dilemma in an RNA virus. *Nature* **398**, 441–443
959 (1999).

60. Ghoul, M., West, S. A., Diggle, S. P. & Griffin, A. S. An experimental test of whether cheating is context dependent. *J. Evol. Biol.* **27**, 551–556 (2014).
61. Leeks, A. & West, S. A. Altruism in a virus. *Nat. Microbiol.* **4**, 910–911 (2019).
62. Yuen, C.-K. *et al.* SARS-CoV-2 nsp13, nsp14, nsp15 and orf6 function as potent interferon antagonists. *Emerg. Microbes Infect.* **0**, 1–29 (2020).
63. Russell, A. B., Elshina, E., Kowalsky, J. R., Velthuis, A. J. W. te & Bloom, J. D. Single-cell virus sequencing of influenza infections that trigger innate immunity. *J. Virol.* JVI.00500-19 (2019) doi:10.1128/JVI.00500-19.
64. Frank, S. A. Models of Parasite Virulence. *Q. Rev. Biol.* **71**, 37–78 (1996).
65. Boots, M. & Sasaki, A. 'Small worlds' and the evolution of virulence: infection occurs locally and at a distance. *Proc. R. Soc. Lond. B Biol. Sci.* **266**, 1933–1938 (1999).
66. Wild, G., Gardner, A. & West, S. A. Adaptation and the evolution of parasite virulence in a connected world. *Nature* **459**, 983–986 (2009).
67. Berngruber, T. W., Lion, S. & Gandon, S. Spatial Structure, Transmission Modes and the Evolution of Viral Exploitation Strategies. *PLOS Pathog.* **11**, e1004810 (2015).
68. Wolf, Y. I. *et al.* Origins and Evolution of the Global RNA Virome. *mBio* **9**, (2018).
69. Wolf, Y. I. *et al.* Doubling of the known set of RNA viruses by metagenomic analysis of an aquatic virome. *Nat. Microbiol.* **5**, 1262–1270 (2020).
70. Zhang, Y.-Z., Shi, M. & Holmes, E. C. Using Metagenomics to Characterize an Expanding Virosphere. *Cell* **172**, 1168–1172 (2018).
71. Shi, M. *et al.* Redefining the invertebrate RNA virosphere. *Nature* (2016) doi:10.1038/nature20167.
72. Erez, Z. *et al.* Communication between viruses guides lysis–lysogeny decisions. *Nature* **541**, 488–493 (2017).

984 73. Schmerer, M., Molineux, I. J. & Bull, J. J. Synergy as a rationale for phage therapy
985 using phage cocktails. *PeerJ* **2**, e590 (2014).
986 74. Christie, G. E. & Dokland, T. Pirates of the Caudovirales. *Virology* **434**, 210–221
987 (2012).
988 75. Dennehy, J. J. & Turner, P. E. Reduced fecundity is the cost of cheating in RNA virus 6.
989 *Proc. R. Soc. B Biol. Sci.* **271**, 2275–2282 (2004).
990 76. Geoghegan, J. L. & Holmes, E. C. Evolutionary Virology at 40. *Genetics* **210**, 1151–
991 1162 (2018).
992 77. Ross-Gillespie, A., Gardner, A., West, S. A. & Griffin, A. S. Frequency Dependence
993 and Cooperation: Theory and a Test with Bacteria. *Am. Nat.* **170**, 331–342 (2007).
994 78. Gallagher, M. E., Brooke, C. B., Ke, R. & Koelle, K. Causes and Consequences of
995 Spatial Within-Host Viral Spread. *Viruses* **10**, 627 (2018).
996 79. Bondy-Denomy, J. *et al.* Prophages mediate defense against phage infection through
997 diverse mechanisms. *ISME J.* **10**, 2854–2866 (2016).
998 80. Folimonova, S. Y. Superinfection Exclusion Is an Active Virus-Controlled Function
999 That Requires a Specific Viral Protein. *J. Virol.* **86**, 5554–5561 (2012).
1000 81. Doceul, V., Hollinshead, M., van der Linden, L. & Smith, G. L. Repulsion of
1001 superinfecting virions: a mechanism for rapid virus spread. *Science* **327**, 873–876
1002 (2010).
1003 82. Sanjuán, R. Collective Infectious Units in Viruses. *Trends Microbiol.* **25**, 402–412
1004 (2017).
1005 83. Leeks, A., Sanjuán, R. & West, S. A. The evolution of collective infectious units in
1006 viruses. *Virus Res.* **265**, 94–101 (2019).
1007 84. Hull, R. *Comparative Plant Virology*. (Elsevier Academic Press, 2009).

85. Shirogane, Y., Watanabe, S. & Yanagi, Y. Cooperation between different RNA virus genomes produces a new phenotype. *Nat. Commun.* **3**, 1235 (2012).
86. Graw, F. & Perelson, A. S. Modeling Viral Spread. *Annu. Rev. Virol.* **3**, 555–572 (2016).
87. Law, K. M. *et al.* In Vivo HIV-1 Cell-to-Cell Transmission Promotes Multicopy Micro-compartmentalized Infection. *Cell Rep.* **15**, 2771–2783 (2016).
88. Andreu-Moreno, I. & Sanjuán, R. Collective Viral Spread Mediated by Virion Aggregates Promotes the Evolution of Defective Interfering Particles. *mBio* **11**, (2020).
89. Brooke, C. B., Ince, W. L., Wei, J., Bennink, J. R. & Yewdell, J. W. Influenza A virus nucleoprotein selectively decreases neuraminidase gene-segment packaging while enhancing viral fitness and transmissibility. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 16854–16859 (2014).
90. Gutiérrez, S. *et al.* Dynamics of the Multiplicity of Cellular Infection in a Plant Virus. *PLOS Pathog.* **6**, e1001113 (2010).
91. Díaz-Muñoz, S. L. Viral coinfection is shaped by host ecology and virus–virus interactions across diverse microbial taxa and environments. *Virus Evol.* **3**, (2017).
92. Ball, J. K., Holmes, E. C., Whitwell, H. & Desselberger, U. Genomic variation of human immunodeficiency virus type 1 (HIV-1): molecular analyses of HIV-1 in sequential blood samples and various organs obtained at autopsy. *J. Gen. Virol.* **75 (Pt 4)**, 67–79 (1994).
93. Graw, F. *et al.* Inferring Viral Dynamics in Chronically HCV Infected Patients from the Spatial Distribution of Infected Hepatocytes. *PLOS Comput. Biol.* **10**, e1003934 (2014).
94. Saira, K. *et al.* Sequence Analysis of In Vivo Defective Interfering-Like RNA of Influenza A H1N1 Pandemic Virus. *J. Virol.* **87**, 8064–8074 (2013).

95. Li, D. *et al.* Defective Interfering Viral Particles in Acute Dengue Infections. *PLOS ONE* **6**, e19447 (2011).
96. Gelbart, M. *et al.* Accurate in vivo population sequencing uncovers drivers of within-host genetic diversity in viruses. *bioRxiv* 349498 (2019) doi:10.1101/349498.
97. Shriner, D., Rodrigo, A. G., Nickle, D. C. & Mullins, J. I. Pervasive genomic recombination of HIV-1 in vivo. *Genetics* **167**, 1573–1583 (2004).
98. Lowen, A. C. It's in the mix: Reassortment of segmented viral genomes. *PLOS Pathog.* **14**, e1007200 (2018).
99. Jacobs, N. T. *et al.* Incomplete influenza A virus genomes occur frequently but are readily complemented during localized viral spread. *Nat. Commun.* **10**, 1–17 (2019).
100. Levi, L. I. *et al.* Defective viral genomes from chikungunya virus are broad-spectrum antivirals and prevent virus dissemination in mosquitoes. *PLOS Pathog.* **17**, e1009110 (2021).
101. McCrone, J. T. & Lauring, A. S. Genetic bottlenecks in intraspecies virus transmission. *Curr. Opin. Virol.* **28**, 20–25 (2018).
102. Bull, R. A. *et al.* Sequential Bottlenecks Drive Viral Evolution in Early Acute Hepatitis C Virus Infection. *PLOS Pathog.* **7**, e1002243 (2011).
103. Martin, M. A. & Koelle, K. Reanalysis of deep-sequencing data from Austria points towards a small SARS- COV-2 transmission bottleneck on the order of one to three virions. 7 (2021).
104. Zwart, M. P. & Elena, S. F. Matters of Size: Genetic Bottlenecks in Virus Infection and Their Potential Impact on Evolution. *Annu. Rev. Virol.* **2**, 161–179 (2015).
105. Keele, B. F. *et al.* Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 7552–7557 (2008).

106. Varble, A. *et al.* Influenza A Virus Transmission Bottlenecks Are Defined by Infection Route and Recipient Host. *Cell Host Microbe* **16**, 691–700 (2014).
107. Sexton, N. R. *et al.* Genome Number and Size Polymorphism in Zika Virus Infectious Units. *J. Virol.* **95**, (2021).
108. López-Ferber, M., Simón, O., Williams, T. & Caballero, P. Defective or effective? Mutualistic interactions between virus genotypes. *Proc. R. Soc. Lond. B Biol. Sci.* **270**, 2249–2255 (2003).
109. Tramper, J. & Vlak, J. M. Some Engineering and Economic Aspects of Continuous Cultivation of Insect Cells for the Production of Baculoviruses. *Ann. N. Y. Acad. Sci.* **469**, 279–288 (1986).
110. Frensing, T., Pflugmacher, A., Bachmann, M., Peschel, B. & Reichl, U. Impact of defective interfering particles on virus replication and antiviral host response in cell culture-based influenza vaccine production. *Appl. Microbiol. Biotechnol.* **98**, 8999–9008 (2014).
111. Tapia, F. *et al.* Continuous influenza virus production in a tubular bioreactor system provides stable titers and avoids the “von Magnus effect”. *PLOS ONE* **14**, e0224317 (2019).
112. Frensing, T. Defective interfering viruses and their impact on vaccines and viral vectors. *Biotechnol. J.* **10**, 681–689 (2015).
113. Andersen, S. B. *et al.* Privatisation rescues function following loss of cooperation. *eLife* **7**, e38594 (2018).
114. Butaitė, E., Baumgartner, M., Wyder, S. & Kümmerli, R. Siderophore cheating and cheating resistance shape competition for iron in soil and freshwater Pseudomonas communities. *Nat. Commun.* **8**, 1–12 (2017).

115. Ostrowski, E. A. Enforcing Cooperation in the Social Amoebae. *Curr. Biol.* **29**, R474–R484 (2019).

116. DePolo, N. J., Giachetti, C. & Holland, J. J. Continuing coevolution of virus and defective interfering particles and of viral genome sequences during undiluted passages: virus mutants exhibiting nearly complete resistance to formerly dominant defective interfering particles. *J. Virol.* **61**, 454–464 (1987).

117. Sanjuán, R., Nebot, M. R., Chirico, N., Mansky, L. M. & Belshaw, R. Viral Mutation Rates. *J. Virol.* **84**, 9733–9748 (2010).

118. Martin, M. A., Kaul, D., Tan, G. S., Woods, C. W. & Koelle, K. The Dynamics of Influenza A H3N2 Defective Viral Genomes from a Human Challenge Study. *bioRxiv* (2019) doi:10.1101/814673.

119. Spiegelman, S., Haruna, I., Holland, I. B., Beaudreau, G. & Mills, D. The synthesis of a self-propagating and infectious nucleic acid with a purified enzyme. *Proc. Natl. Acad. Sci. U. S. A.* **54**, 919–927 (1965).

120. Meir, M. *et al.* Competition between social cheater viruses is driven by mechanistically different cheating strategies. *Sci. Adv.* **6**, eabb7990 (2020).

121. Shirogane, Y. *et al.* Experimental and mathematical insights on the interactions between poliovirus and a defective interfering genome. *bioRxiv* (2021) doi:10.1101/2021.01.11.426198.

122. Lythgoe, K. A., Gardner, A., Pybus, O. G. & Grove, J. Short-Sighted Virus Evolution and a Germline Hypothesis for Chronic Viral Infections. *Trends Microbiol.* **25**, 336–348 (2017).

123. Elena, S. F. & Sanjuán, R. Virus Evolution: Insights from an Experimental Approach. *Annu. Rev. Ecol. Evol. Syst.* **38**, 27–52 (2007).

124. Giri, L., Feiss, M. G., Bonning, B. C. & Murhammer, D. W. Production of baculovirus defective interfering particles during serial passage is delayed by removing transposon target sites in fp25k. *J. Gen. Virol.* **93**, 389–399 (2012).
125. dos Santos, M., Ghoul, M. & West, S. A. Pleiotropy, cooperation, and the social evolution of genetic architecture. *PLOS Biol.* **16**, e2006671 (2018).
126. Pfaller, C. K. *et al.* Measles Virus Defective Interfering RNAs Are Generated Frequently and Early in the Absence of C Protein and Can Be Destabilized by Adenosine Deaminase Acting on RNA-1-Like Hypermutations. *J. Virol.* **89**, 7735–7747 (2015).
127. Novak, J. E. & Kirkegaard, K. Coupling between genome translation and replication in an RNA virus. *Genes Dev.* **8**, 1726–1737 (1994).
128. Jaworski, E. & Routh, A. Parallel ClickSeq and Nanopore sequencing elucidates the rapid evolution of defective-interfering RNAs in Flock House virus. *PLOS Pathog.* **13**, e1006365 (2017).
129. Gardner, A. & Grafen, A. Capturing the superorganism: a formal theory of group adaptation. *J. Evol. Biol.* **22**, 659–671 (2009).
130. Andino, R. & Domingo, E. Viral quasispecies. *Virology* **479–480**, 46–51 (2015).
131. Vasilijevic, J. *et al.* Reduced accumulation of defective viral genomes contributes to severe outcome in influenza virus infected patients. *PLOS Pathog.* **13**, e1006650 (2017).
132. Dong, X. *et al.* Variation around the dominant viral genome sequence contributes to viral load and outcome in patients with Ebola virus disease. *Genome Biol.* **21**, 238 (2020).
133. Valesano, A. L. *et al.* Temporal dynamics of SARS-CoV-2 mutation accumulation within and across infected hosts. *bioRxiv* (2021) doi:10.1101/2021.01.19.427330.

134. Lauring, A. S. Within-Host Viral Diversity: A Window into Viral Evolution. *Annu. Rev. Virol.* **7**, (2020).

135. Metzger, V. T., Lloyd-Smith, J. O. & Weinberger, L. S. Autonomous Targeting of Infectious Superspreaders Using Engineered Transmissible Therapies. *PLOS Comput. Biol.* **7**, e1002015 (2011).

136. Dimmock, N. J. & Easton, A. J. Defective Interfering Influenza Virus RNAs: Time To Reevaluate Their Clinical Potential as Broad-Spectrum Antivirals? *J. Virol.* **88**, 5217–5227 (2014).

137. Gu, S. *et al.* Competition for iron drives phytopathogen control by natural rhizosphere microbiomes. *Nat. Microbiol.* 1–9 (2020) doi:10.1038/s41564-020-0719-8.

138. Brown, S. P., West, S. A., Diggle, S. P. & Griffin, A. S. Social evolution in micro-organisms and a Trojan horse approach to medical intervention strategies. *Philos. Trans. R. Soc. B Biol. Sci.* **364**, 3157–3168 (2009).

139. Zhao, H. *et al.* Dual-functional peptide with defective interfering genes effectively protects mice against avian and seasonal influenza. *Nat. Commun.* **9**, 2358 (2018).

140. Noble, S. & Dimmock, N. J. Defective interfering type A equine influenza virus (H3N8) protects mice from morbidity and mortality caused by homologous and heterologous subtypes of influenza A virus. *J. Gen. Virol.* **75 (Pt 12)**, 3485–3491 (1994).

141. Weinberger, L. S. & Notton, T. J. Methods and Compositions for Generating a Deletion Library and for Identifying a Defective Interfering Particle (DIP). United States Patent Application 20210024921. (2021).

142. Johnson, D. M., Cubitt, B., Pfeffer, T. L., de la Torre, J. C. & Lukashevich, I. S. Lassa Virus Vaccine Candidate ML29 Generates Truncated Viral RNAs Which Contribute to Interfering Activity and Attenuation. *Viruses* **13**, 214 (2021).

143. Rand, U. *et al.* Antiviral activity of influenza A virus defective interfering particles against SARS-CoV-2 replication in vitro through stimulation of innate immunity. *bioRxiv* (2021) doi:10.1101/2021.02.19.431972.
144. Zhang, Y. *et al.* Influenza Research Database: An integrated bioinformatics resource for influenza virus research. *Nucleic Acids Res.* **45**, D466–D474 (2017).
145. Benson, D. A. *et al.* GenBank. *Nucleic Acids Res.* **46**, D41–D47 (2018).
146. Grubaugh, N. D. *et al.* Tracking virus outbreaks in the twenty-first century. *Nat. Microbiol.* **4**, 10–19 (2019).
147. Andreu-Moreno, I., Bou, J.-V. & Sanjuán, R. Cooperative nature of viral replication. *Sci. Adv.* **6**, eabd4942 (2020).
148. Cole, C. N. & Baltimore, D. Defective interfering particles of poliovirus: II. Nature of the defect. *J. Mol. Biol.* **76**, 325–343 (1973).
149. Cole, C. N. & Baltimore, D. Defective interfering particles of poliovirus: III. Interference and enrichment. *J. Mol. Biol.* **76**, 345–361 (1973).
150. Cole, C. N., Smoler, D., Wimmer, E. & Baltimore, D. Defective Interfering Particles of Poliovirus I. Isolation and Physical Properties. *J. Virol.* **7**, 478–485 (1971).
151. Song, Y., Paul, A. V. & Wimmer, E. Evolution of Poliovirus Defective Interfering Particles Expressing Gaussia Luciferase. *J. Virol.* **86**, 1999–2010 (2012).
152. Cole, C. N. & Baltimore, D. Defective Interfering Particles of Poliovirus IV. Mechanisms of Enrichment. *J. Virol.* **12**, 1414–1426 (1973).
153. Grafen, A. A geometric view of relatedness. *Oxf. Surv. Evol. Biol.* **2**, 28–89 (1985).
154. Olsen, P. H. *English: Reed Warbler feeding a Common Cuckoo chick in a nest. Brood parasitism. Wikimedia Commons.*
155. Baltes, A., Akpinar, F., Inankur, B. & Yin, J. Inhibition of infection spread by co-transmitted defective interfering particles. *PLOS ONE* **12**, e0184029 (2017).

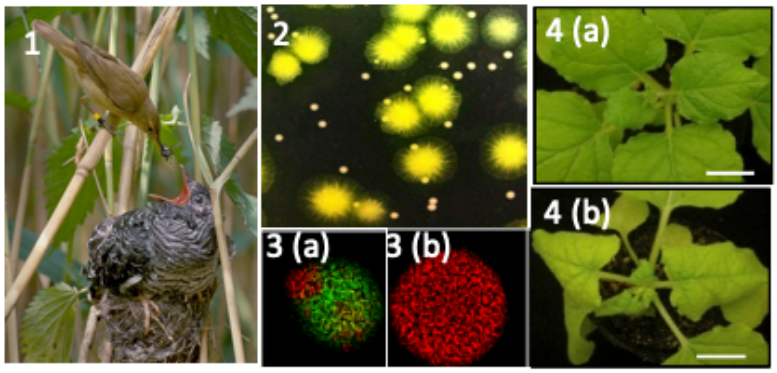
156. He, L. *et al.* A conserved RNA structure is essential for a satellite RNA-mediated inhibition of helper virus accumulation. *Nucleic Acids Res.* (2019) doi:10.1093/nar/gkz564.
157. Huang, A. S. & Wagner, R. R. Defective T particles of vesicular stomatitis virus: II. Biologic role in homologous interference. *Virology* **30**, 173–181 (1966).
158. Kawai, A. & Matsumoto, S. Interfering and noninterfering defective particles generated by a rabies small plaque variant virus. *Virology* **76**, 60–71 (1977).
159. Potter, J. N., Faulkner, P. & MacKinnon, E. A. Strain selection during serial passage of Trichoplusia in nuclear polyhedrosis virus. *J. Virol.* **18**, 1040–1050 (1976).
160. Ram, G. *et al.* Staphylococcal pathogenicity island interference with helper phage reproduction is a paradigm of molecular parasitism. *Proc. Natl. Acad. Sci.* **109**, 16300–16305 (2012).

Acknowledgements

We thank: thank Rafael Sanjuán, Ashleigh Griffin, Helen Leggett, John Bruce, and four anonymous reviewers for comments and discussion; the Clarendon Fund, St. John's College, Oxford, and the BBSRC (grant number BB/M011224/1) (A.L.) & ERC (S.A.W. & M.G.) for funding.

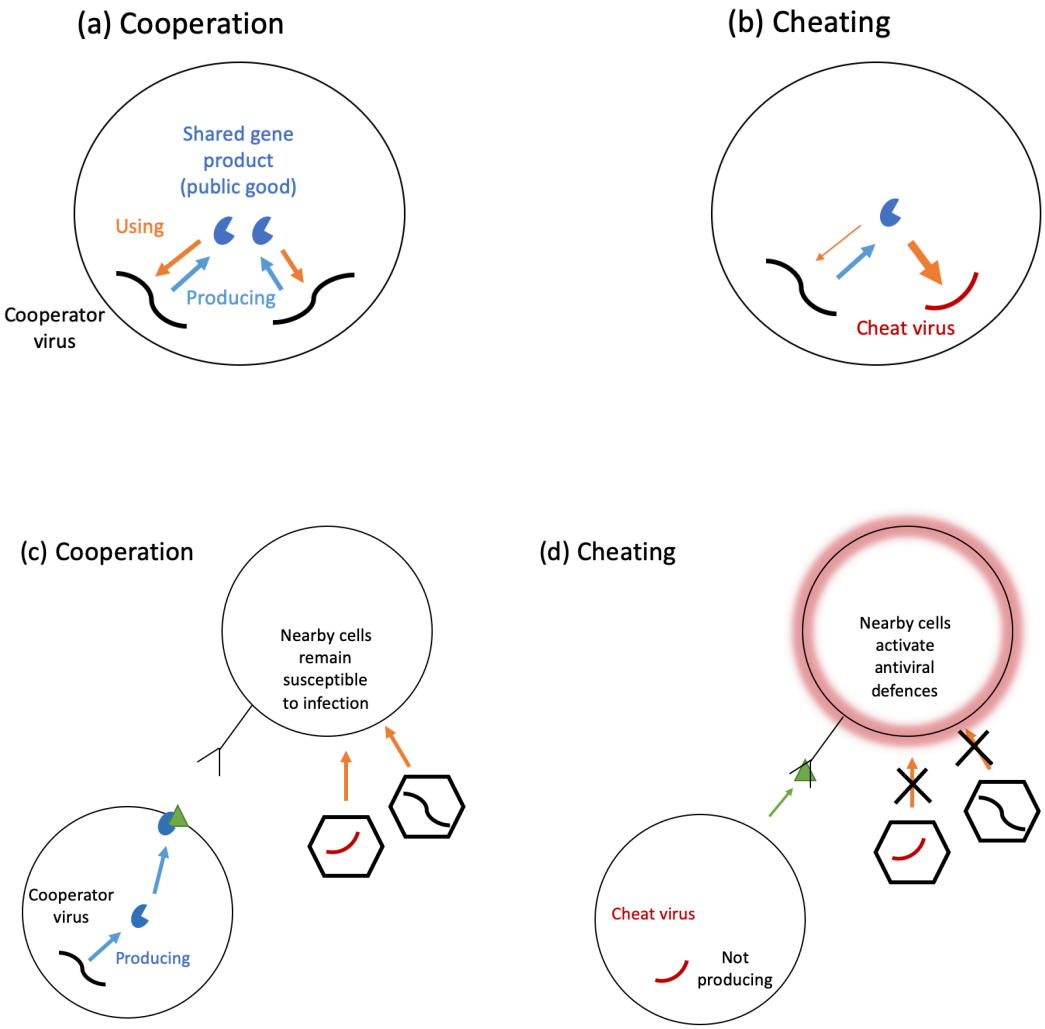
Figures and Legends

Figure 1: cheating throughout the natural world



Cheating occurs throughout the natural world, including in viruses: (1) the common cuckoo (*Cuculus canorus*) lays eggs in other birds' nests, here tricking a reed warbler (*Acrocephalus scirpaceus*) into taking care of a much larger cuckoo chick[19,154]; (2) Cells of the bacterial pathogen *Pseudomonas aeruginosa* which do not produce iron-scavenging molecules (labelled in green) are able to exploit those produced by others (labelled in white), and consequently grow much larger colonies; (3) in Vesicular Stomatitis Virus (VSV), when a defective interfering genome (labelled in green) is grown in a mixed infection with wild-type VSV (labelled in red), the defective interfering genome exploits replicase proteins encoded by the wild-type cooperator, resulting in a colony (a) that is dominated by the defective interfering genome, and grows less effectively than a colony consisting just of the cooperative wild-type (b)[155]; (4) in cucumber mosaic virus (CMV) infections, a satellite (satCMV) exploits gene products encoded by the wild-type, substantially reducing the overall viral load and leading to less severe infections in plants infected by both satellite and wild type (a) compared to plants infected by just the wild type (b)[156].

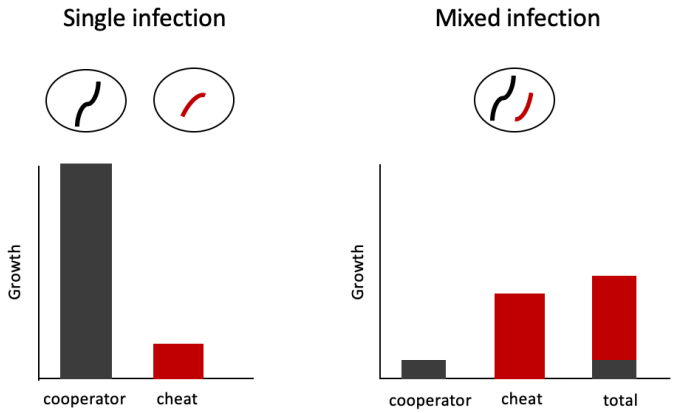
Figure 2: Viruses can cooperate within and between cells.



Viruses can cooperate when infecting the same cell, and also when infecting different cells. (a) In coinfection, shared viral gene products, such as replicase enzymes or capsid proteins, have the potential to benefit other viral genomes, and hence act as cooperative 'public goods' (Box 1). (b) When multiple viral genomes infect a host cell, there is also the potential for cheating, where some individuals benefit from the public good without producing it (cheats). (c) Cooperation can also occur between viral genomes that infect different host cells. Here, cooperative viruses produce a gene product that prevents host cells releasing interferon,

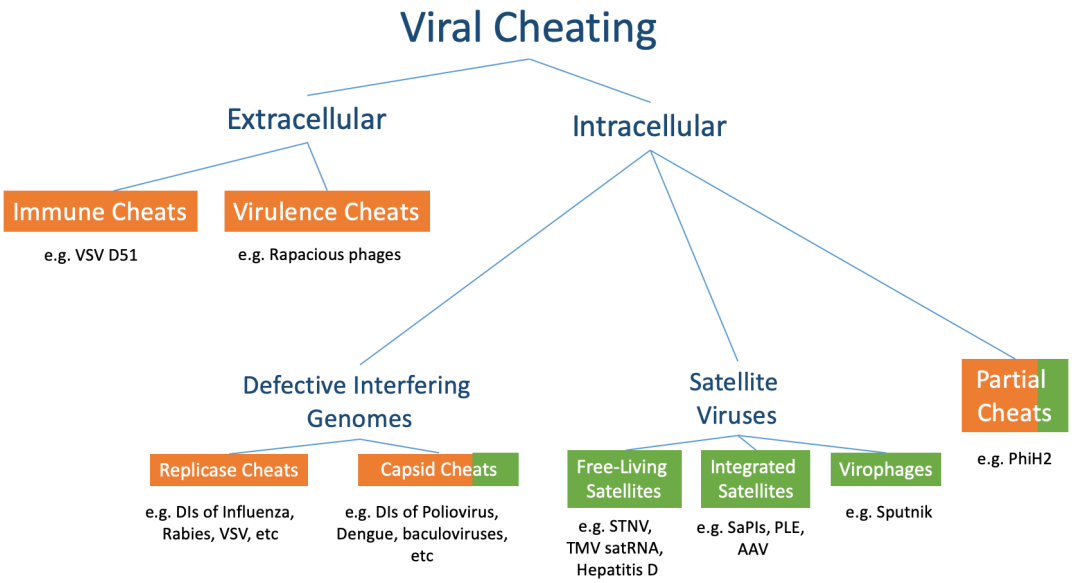
keeping the local population of host cells susceptible to infection, and hence providing a benefit to other viral genomes infecting different cells[5]. (d) Cheat viruses do not block interferon, and hence replicate faster than cooperators. This cheating is costly for the viral population as a whole, because interferon is released from the infected cell, binding to nearby cells, which become resistant to infection by other viral genomes.

Figure 3: how to test for a cheat



For two individuals to count as a cheat and cooperator respectively, three conditions must be met: (1) the cooperator must have a higher fitness than the cheat when each are alone; (2) the cheat must have a higher fitness than the cooperator when both are mixed; (3) the cooperator must have a lower fitness in a mixture than it did when it was alone.

Figure 4: A Classification of Viral Cheats



A variety of different kinds of viral entity are cheats. We first divide viral cheats depending on whether the cooperation that is being exploited is intracellular or extracellular, then according to their origins, and finally the specific gene products exploited, if known. We also denote the evolutionary timescales at which cheats exist, dividing between: short-sighted cheats (in orange), which arise and spread de novo within hosts, without spreading between hosts; and long-sighted cheats (in green), which spread both within and between hosts, and persist over longer evolutionary timescales. We offer a few illustrative, but not exhaustive, examples of each type of cheat[25,32,33,45,47,51–53,59,95,150,157–160].

Figure 5: Viral Cheat Origins.

	Defective interfering genome	Point mutation cheat	Satellite virus cheat
Cooperator genome			
Cheat genome			
How is the cheat created?	Large deletion(s)	A point mutation or small deletion leading to loss or change of function	Independent origin, with a completely different genome sequence
Where found?	Almost all animal and plant viruses, many phages	Common in phages, sometimes in animal viruses, difficult to say how common	Commonly in plant viruses, phages, and giant viruses, but only rarely in animal viruses
Short vs long-sighted	Usually short, some evidence for long-term transmission	Usually short, potentially long	Long-sighted

We show three different types of cheat, that arise in different ways: large deletion; point mutation; and independent origin.

3

Beneficial coinfection can promote within-host viral diversity

VIRUS EVOLUTION

Virus Evolution, 2018, 4(2): vey028

doi: 10.1093/ve/vey028
Research article

Beneficial coinfection can promote within-host viral diversity

Asher Leeks,[1,*,†], Ernesto A. Segredo-Otero,[2] Rafael Sanjuán,[2,‡] and Stuart A. West,[1]

[1]Department of Zoology, University of Oxford, Oxford, UK and [2]Institute for Integrative Systems Biology (I2SysBio), Universitat de València, València, Spain

*Corresponding author: E-mail: asherleeks@gmail.com

†http://orcid.org/0000-0002-1175-7610

‡http://orcid.org/0000-0002-1844-545X

Abstract

In many viral infections, a large number of different genetic variants can coexist within a host, leading to more virulent infections that are better able to evolve antiviral resistance and adapt to new hosts. But how is this diversity maintained? Why do faster-growing variants not outcompete slower-growing variants, and erode this diversity? One hypothesis is if there are mutually beneficial interactions between variants, with host cells infected by multiple different viral genomes producing more, or more effective, virions. We modelled this hypothesis with both mathematical models and simulations, and found that moderate levels of beneficial coinfection can maintain high levels of coexistence, even when coinfection is relatively rare, and when there are significant fitness differences between competing variants. Rare variants are more likely to be coinfecting with a different variant, and hence beneficial coinfection increases the relative fitness of rare variants through negative frequency dependence, and maintains diversity. We further find that coexisting variants sometimes reach unequal frequencies, depending on the extent to which different variants benefit from coinfection, and the ratio of variants which leads to the most productive infected cells. These factors could help drive the evolution of defective interfering particles, and help to explain why the different segments of multipartite viruses persist at different equilibrium frequencies.

Key words: phenotype mixing; diversity; evolution; multipartite; frequency dependence; coinfection

1. Introduction

Viruses can form exceptionally diverse populations inside hosts, with thousands of distinct genetic variants infecting a single host (Lauring and Andino 2010). Infections with a high viral diversity can be more virulent, for example, by infecting more tissue types or through reaching higher viral titres (Vignuzzi et al. 2006; Coffey et al. 2011; Shirogane, Watanabe, and Yanagi 2012; Cao et al. 2014; Bordería et al. 2015; Skums, Bunimovich, and Khudyakov 2015; Xue et al. 2016). Infection diversity also influences virus evolution, as more diverse populations may develop antiviral resistance more rapidly, may be more likely to adapt to new hosts, and can recombine, leading to the emergence of novel pathogens (Bonhoeffer et al. 1997; Bordería, Stapleford, and Vignuzzi, 2011; Ke et al. 2015; Pérez-Losada et al. 2015). The existence of high within-host diversity presents an evolutionary problem, because different variants of the same virus compete to infect a limited population of host cells (Gause 1934; Clarke et al. 1994; Moya et al. 2000). Consequently, why do faster-replicating variants not out-compete slower-replicating variants, leading to a loss of variant diversity?

Several mechanisms have been suggested to promote variant coexistence. If different variants specialise on infecting different cell types, then this could reduce competition, allowing variants to coexist (Elena, Miralles, and Moya 1997; Yuste, Moya,

and López-Galíndez 2002; Wilke, Reissig, and Novella 2004; Arbiza, Mirazo, and Fort 2010). Alternatively, if the mutation rate is high enough, a diverse set of variants could be maintained through mutation-selection balance (Wilke 2005; Domingo, Sheldon, and Perales 2012; Andino and Domingo 2015). The relative importance of these hypotheses depends upon the extent to which different variants do infect different tissues, and whether the mutation rate is high enough, respectively. Another possibility to explain variant coexistence is if cells infected by multiple viral variants produce more virions, or more effective virions, than singly infected cells (Roossinck, Sleat, and Palukaitis 1992; Qiu and Scholthof 2001; López-Ferber et al. 2003; Shirogane, Watanabe, and Yanagi 2012; Andino and Domingo 2015; Xue et al. 2016; Díaz-Muñoz, Sanjuán, and West 2017; Hisano et al. 2018). This 'beneficial coinfection' hypothesis could work via viruses sharing gene products when infecting the same cell, resulting in phenotype mixing (Novella, Reissig, and Wilke 2004; Závada 1976). For example, if multiple mutations are beneficial, but result in different changes to the same gene, then they may not be compatible in the same genome. Therefore, variants with different beneficial mutations could mix in a synergistic way.

However, the viability of the beneficial coinfection hypothesis is not clear. Beneficial coinfection might just slow down the extinction of less fit variants by 'masking' fitness differences, rather than allow the long-term coexistence of different variants (Godfray, O'Reilly, and Briggs 1997; Wilke and Novella 2003; Froissart et al. 2004; Wilke, Reissig, and Novella 2004; Gao and Feldman 2009; Loverdo and Lloyd-Smith 2013). Alternatively, this hypothesis might require unrealistically high rates of coinfection, or unrealistically large benefits of coinfection, in order to allow variants to coexist. Could population bottlenecks, a common feature of virus life cycles, reduce the extent to which different variants can interact beneficially (Zwart and Elena 2015; McCrone and Lauring 2018)? Finally, if different variants benefitted differently from coinfection, then the variant which benefitted the most could be favoured disproportionately, reducing coexistence.

We investigated the theoretical plausibility of the beneficial coinfection hypothesis. Our specific aims were to: (1) test how frequent and how beneficial coinfection needs to be for a slower-replicating variant to coexist at equilibrium with a faster-replicating variant; (2) test how bottlenecks in the virus population affect coexistence; (3) investigate the effect of asymmetries in how variants benefit from and contribute to beneficial coinfection. We use an equilibrium modelling approach based on population genetics which we attempt to parameterise using real data. We then follow this up with more realistic simulations of virus growth in cell culture.

2. Equilibrium model

2.1 Model overview

We have deliberately kept our model as simple as possible, to capture the possible role of beneficial coinfection in a manner that does not depend upon the biological details of certain viruses. For example, we do not model a specific mechanism for coinfection benefit, since this would require making assumptions based on a particular system. Instead, we choose parameters which could result from many different specific mechanisms for coinfection benefit.

We assume that two variants of a virus, A and B, infect the same kinds of cells inside a host. We assume that the rate of spread of each variant within the host depends on the number of virions each variant produces in a given amount of time, as well as the chance that these virions successfully infect further host cells (Klasse 2015). Fitness differences between the variants could therefore stem from mutations which: increase the total number of virions released from an infected host cell; increase the speed of the infection cycle; or which increase the effectiveness of the virions produced. In order to avoid making specific assumptions about the nature of fitness differences between the variants, we capture these factors in a single parameter, 'productivity'. The productivity of a focal host cell is the relative number of further host cells which are successfully infected by virions produced in the focal host cell in a given amount of time; therefore productivity is analogous to the basic reproductive ratio (R_0) at a cellular level (Bonhoeffer et al. 1997; Nowak et al. 1997). We can express the rate of spread of each variant within the host in terms of its share of the productivity of the cells it infects. This method is formally analogous to treating the two variants as different alleles at a locus in a population genetics model, where phenotypes are infected host cells, alleles are the different viral variants, genotypes are the combinations of viral variants infecting each host cell, the ploidy is specified by the likelihoods of different multiplicities of infection, and fitness is productivity of infected host cells (Chao 1991; Wilke 2005; Otto and Day 2007; Elena et al. 2011).

2.2 Lifecycle

We are interested in the maintenance of diversity within a host, and so we model the evolution of an infection inside a single host. We examine the situation when a host is infected by two variants, and ask when this will lead to coexistence, or to one variant outcompeting the other. The two variants could arise through both initially coinfecting the host, or by mutation.

We assume that virions infect host cells according to a Poisson process where the Poisson parameter λ represents the ratio of virions to host cells. We assume that host cells and virions are well mixed, and that each virion contains only one viral genome, such that the multiplicity of infection (MOI) is equivalent to the ratio of virions to host cells (λ). The relative likelihood of a cell being infected by k virions is therefore given by the function $P(k)$, defined as $\left(e^{-\lambda}\ \frac{\lambda^k}{k!}\right)/\left(\sum_{n=1}^{m} e^{-\lambda}\ \frac{\lambda^n}{n!}\right)$, where λ is the ratio of virions to host cells, k is the number of virions infecting each host cell, and m is defined as $\lambda + 3\sqrt{\lambda}$ and is chosen to ensure that we consider >99 per cent of possible infection states. The numerator of $P(k)$ gives the likelihood of a cell infected by k virions where k is a Poisson-distributed discrete random variable around λ, and the denominator is a normalisation factor to ensure that we consider the relative likelihoods of each infection state. For full details, see the Supplementary data. $P(k)$ is illustrated in Fig. 1a for different MOI values.

We further assume that the likelihood of different variant combinations infecting each cell is described by a Binomial process. To achieve this we use a function $B(x_t, k, i) = \binom{k}{i} {x_t}^i (1 - x_t)^{k-i}$, where x_t is the relative frequency of variant A at time t (and so $1 - x_t$ is the relative frequency of variant B), k is the number of virions infecting the host cell, and i is the number of virions infecting the host cell which are variant A. For example, when two virions infect a host cell ($k = 2$), the possible infection outcomes are AA ($i = 2$), AB ($i = 1$), and BB ($i = 0$). The relative likelihoods of these are given by $pr(AA) = B(x_t,\ 2,\ 2) = {x_t}^2$; $pr(AB) = B(x_t,\ 2,\ 1) = 2\ x_t\ (1 - x_t)$; $pr(BB) = B(x_t,\ 2,\ 0) = (1 - x_t)^2$.

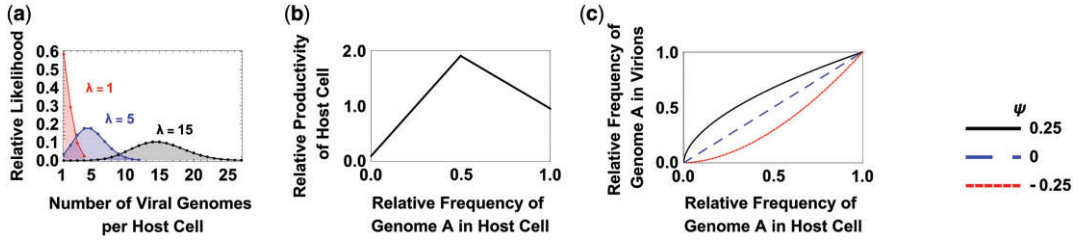


Figure 1. Equilibrium model assumptions. (a) We assume that infection of host cells is well described by a Poisson distribution ($P(\lambda)$), where the Poisson parameter λ is given by the ratio of virions to susceptible host cells (MOI). We truncate our Poisson distribution at 1, to focus on infected cells, and at three standard deviations above the mean, to avoid a potentially infinite number of infection states. (b) We assume mixed infections are more productive. The productivity of an infected host cell (y-axis) shows a peaked relationship with relative proportion of variant A virions in the initial infection mixture (p_A; x-axis) according to the function $\Phi(p_A)$. (c) We assume that the proportion of A genomes in the virions produced by that cell (y-axis) can be either linearly or non-linearly related to the proportion of A genomes that initially infected the host cell (p_A; x-axis), according to the function $\Pi(p_A)$.

We assume that the productivity of an infected host cell i (W_i) depends on the viral genes which are expressed. Therefore, the productivity depends on the viral genomes that initially infect the cell. For example, if a host cell is infected only by genomes of variant A, we assume that its productivity is W_A. For consistency, we assume that variant A has a higher productivity when in pure infection than variant B, and consequently spreads more quickly inside the host, all else being equal ($W_A > W_B$). As we are focusing on the effects of coinfection of different variants, we also assume that cells infected by multiple genomes of the same variant have the same productivity as cells infected by just one genome of that variant (W_A or W_B) (Timm and Yin 2012).

When a host cell is infected by a mixture of genomes of both variants, then genes from both variants A and B can be expressed (Novella, Reissig, and Wilke 2004; Závada 1976). Therefore, the total productivity of the cell could be different from W_A or W_B. We capture beneficial coinfection by allowing these mixed infections to have a higher maximum productivity than the most productive pure infections ($W_M > W_A$). Therefore, coinfection benefit could arise from different kinds of biological interactions: host cells infected by both variants may produce more virions, may produce virions that are more effective at infecting new cells, or may have a faster infection cycle.

To determine the productivity of cells infected by different combinations of variants A and B, we use a discontinuous function where we independently specify the productivities of cells infected only by variant A, only by variant B, or by a mixture of both variants. This is given by

$$\varphi(P_A) = \begin{cases} W_B + p_A \dfrac{W_M - W_B}{\tau_A} & 0 \le p_A < \tau_A \\ W_M & \tau_A \le p_A \le \tau_B \\ \dfrac{\tau_B W_A - W_M}{\tau_B - 1} + p_A \dfrac{W_M - W_A}{\tau_B - 1} & \tau_B < p_A \le 1 \end{cases}$$

Here, p_A is the relative proportion of variant A infecting a given host cell (and is given by i/k), W_A and W_B give the productivities of a cell infected entirely by variant A or entirely by variant B, respectively, W_M gives the maximum possible productivity of a cell infected by both variants, and τ_A and τ_B determine the threshold proportions of variant A and variant B, respectively, that are required for the most productive coinfections. This function results in productivity increasing linearly from W_B at p_A=0 to W_M at $p_A = \tau_A$, and then decreases linearly from W_M at $p_A = \tau_B$ to W_A at p_A=1 (Fig. 1b). Throughout the rest of the paper, we use 'coinfection benefit' to refer to W_M, the relative productivity of mixed infections relative to the most productive single infection. Initially, we assume that the highest coinfection benefit (W_M) occurs when both variants infect the host cell in equal proportion ($\tau_A = \tau_B = 0.5$; Fig. 1b). However, we later relax this assumption.

We assume that, in mixed infections, the virions produced can contain the genome of either variant. Initially we assume that variants A and B are replicated and encapsidated at the same rate inside a mixed infection, and so the ratio of virions leaving the cell containing each variant's genome is the same as the initial ratio of virions of each variant that infected the cell. However, we later relax this assumption and allow the two variants to benefit differently from the virions produced by cells in mixed infection. To do this we model the output ratio of A:B with the function $\Pi(P_A) = p_A^{(W_A/W_B)^{\psi}}$, where p_A, W_A, and W_B are as defined above, and ψ is a parameter that determines the shape of the relationship between input and output proportions of virions. When ψ is positive, it indicates that the variant which is more productive in a pure infection gains a greater share of the virions in a mixed infection; when ψ is negative it indicates that the variant which is more productive in pure infection gains a smaller share of the virions in a mixed infection (Fig. 1c). This allows us to capture a range of biological scenarios, including defective interfering particles (DIPs), which have negligible productivity in pure infection (W_{DIP} =0), but gain a disproportionate share of the productivity of a mixed infection ($\psi < 0$).

2.3 Dynamics

In order to determine the dynamics of variants A and B, we write an expression for the rate of change in the relative abundance of variant A within the host (x) over time:

$$x_{t+1} = \frac{\sum_{k=1}^{m}\left(P(k)\left(\sum_{i=1}^{k} B(x_t,k,i)\left(\Phi\left(\frac{i}{k}\right)\Pi\left(\frac{i}{k}\right)\right)\right)\right)}{\sum_{k=1}^{m}\left(P(k)\left(\sum_{i=0}^{k} B(x_t,k,i)\Phi\left(\frac{i}{k}\right)\right)\right)} \quad (1)$$

Here, $P(k)$ gives the relative likelihood of different numbers of virions infecting each host cell (Fig. 1a), $B(x_t, k, i)$ gives the relative likelihood that i A-virions infect a host cell that is infected by k total virions, $\Phi(i/k)$; gives the relative productivity of a host cell infected by i/k variant A virions (Fig. 1b), and $\Pi(i/k)$ gives the proportion of variant A virions produced by a host cell infected by i/k variant A virions (Fig. 1c).

The numerator of Equation (1) captures all of the potential ways in which variant A can be produced in each timestep,

weighted by the relative likelihood of each of these ways. The denominator captures all of the ways in which either variant can be produced per timestep. Equation (1) therefore gives the relative frequency of variant A in the next generation as a function of the relative frequency of variant A in the current generation.

2.4 Equilibrium model results

We want to determine when variants A and B coexist stably. Therefore, we solve $x_{t+1} = x_t$ and find stable values of x_t which are between 0 and 1. When a stable solution is found in this region it indicates that both genotypes are maintained within the host at an equilibrium frequency. Through a thorough search of the parameter space that we explore, we find that when an equilibrium frequency exists between 0 and 1, it will be reached from any initial frequency of the two variants. Therefore, the findings that we present here do not rely on assumptions about the initial frequencies of the two variants. Our general method was to find numerical solutions to Equation (1) for different sets of parameter values, as plotted in Figs 2–5, since finding a general analytical solution was not possible. However, our analytical solutions for a version of Equation (1) that uses a simpler function to determine coinfection are consistent with our numerical findings (Supplementary Fig. S1).

2.4.1. *Beneficial coinfection promotes coexistence*

We find that coexistence is favoured by high coinfection benefit ($W_M > W_A$) and high MOI (λ) (Fig. 2; Supplementary Fig. S2). The requirements for an appreciable MOI and coinfection benefit depended upon each other: when coinfection benefit was large, and mixed infections were an order of magnitude more productive than pure infections ($W_M \gg W_A$), coexistence could be maintained at relatively low MOI ($\lambda = 0.5$–1; Fig. 2); when coinfection benefit was smaller, and mixed infections were only slightly more productive than the most productive pure infections ($W_M = W_A$), coexistence required relatively high MOI ($\lambda > 2$; Fig. 2). Additionally, coexistence was relatively unaffected by the productivities of pure infections of variant A and variant B (W_A and W_B; Fig. 2). This meant that coexistence could occur even when pure infections of variant A were orders of magnitude more productive than pure infections of variant B ($W_A \gg W_B$; Fig. 2b).

Negative frequency dependence arises because when a variant is rare, it usually experiences a mixed infection together with genomes from the more common variant. In contrast, when a variant is common, it mostly experiences pure infections, with multiple genomes of its own variant. Since we assumed that mixed infections are more productive than pure infections, the rarer variant therefore experiences a higher average productivity, and subsequently has a higher mean fitness, than the more common variant (Fig. 3). This mechanism requires coinfection to be sufficiently common; when coinfection is rare and most host cells are only infected by a single virion, then the variant which has a higher productivity in pure infection has a higher average productivity and can drive the less productive variant extinct (Fig. 3). In our model, fitness refers to the expected number of progeny belonging to an individual sequence of each variant. To obtain this, we calculated

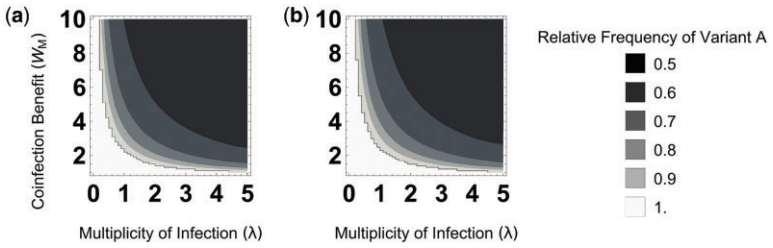


Figure 2. Coexistence. The x-axis is the multiplicity of infection (MOI; λ), which represents the ratio of virions to host cells. The y-axis is the productivity of mixed infections (W_M), relative to the productivity of a cell infected only with genome A (W_A). (a) Variant A spreads 10 times more quickly than variant B in pure infections (W_A=1, W_B=0.1). Coexistence is favoured by high multiplicity of infection (λ) and productivity in mixed infection (W_M). (b) Variant A spreads 1,000 times more quickly than variant B in pure infections (W_A=1, W_B=0.001). Even though variant A is three orders of magnitude more productive than variant B in pure infection, provided coinfections are frequent and beneficial enough, variants A and B coexist at approximately equal frequencies.

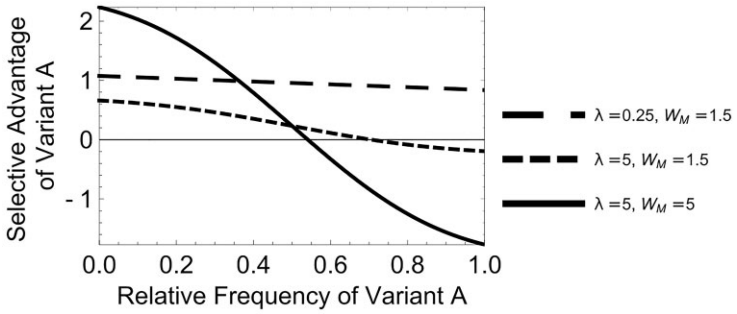


Figure 3. Negative frequency dependence. The selective advantage of variant A is plotted against the relative frequency of variant A in the population. As variant A becomes more common, its relative fitness decreases (negative frequency dependence). When the selective advantage of variant A is >0 it will increase in frequency, and when it is <0, it will decrease in frequency.

the total production of each variant, and divided by the abundance of the variant; for full details see the Supplementary data.

We checked that it is this ratio of mixed to pure infections, and not the mean number of viral genomes per cell, that determines coexistence. To do this, we used a truncated Geometric function to determine the likelihood of different infection states (Godfray, O'Reilly, and Briggs 1997). In this function, coinfection likelihood and number of genomes per cell can be varied independently, and we were able to obtain analytical solutions using this function. We found that the maximum number of genomes per cell makes a very small difference to coexistence, whereas the likelihood of coinfection makes a very big difference (Supplementary Fig. S1). In reality, many different factors can influence the likelihood that multiple viral genomes infect each host cell, including superinfection exclusion and collective infection (Doceul et al. 2010; Folimonova 2012; Bergua et al. 2014; Díaz-Muñoz, Sanjuán, and West 2017; Sanjuán 2017, 2018; Erickson et al. 2018). Our results suggest that beneficial coinfection depends on the relative likelihood that multiple different viral genomes infect a host cell, regardless of the route by which this occurs.

2.5. Equilibrium model extensions

We next consider two extensions to our equilibrium model which might change the predicted level of coexistence. First, we consider bottlenecking, and then we consider asymmetries in how the two variants contribute to, and benefit from, coinfection.

2.5.1 *Bottlenecking disfavours coexistence*

So far, we have assumed that the ratio of viral particles to host particles (MOI) remains constant throughout an infection. In reality, the MOI changes over the course of an infection, and viral populations can go through strong bottlenecks (Wilke, Reissig, and Novella 2004; Gutiérrez et al. 2010, 2015; Zwart and Elena 2015; McCrone and Lauring 2018). These changes in MOI could influence the likelihood of multiple infections, and consequently change the conditions when coexistence is favoured. Since there are many ways in which MOI could vary in reality, we do not simulate specific cases. Instead, we examine how the frequency-dependent process outlined in this paper operates at different MOIs.

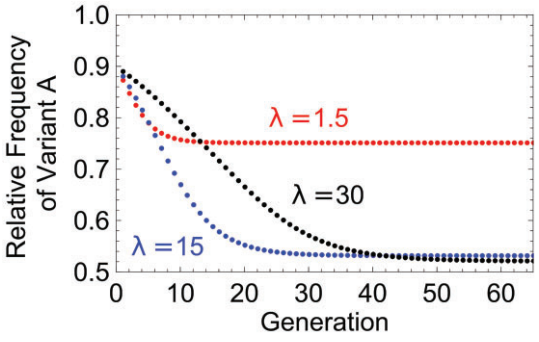


Figure 4. Time to reach equilibrium. The relative frequency of variant *A* is plotted against time, for different multiplicities of infection (MOI; λ). At low MOI (red), the system quickly reaches an equilibrium state. At higher MOI (black and blue), the system reaches an equilibrium that is closer to an even distribution of the two variants. At the highest MOI (black) it takes longer to reach this equilibrium. Therefore, while the highest MOI gives the most even equilibrium ratio of *A*:*B*, if the system is observed before it has reached equilibrium (e.g. generation 30), higher MOI may result in a more uneven ratio of *A*:*B*.

We found that as MOI increases, the equilibrium frequency of *A*: *B* becomes more even, but the time it takes to reach equilibrium increases (Fig. 4). This occurs because at high MOIs, most cells are infected by at least one of each genotype. Therefore, both variants benefit from most infections, and so it takes longer for variants to change in frequency. Regular bottlenecking events could therefore have a bigger influence on the overall frequency of *A*:*B* at very high MOI than at low MOI, because the system takes longer to return to equilibrium after a perturbation. However, it is worth noting that we only saw appreciable differences in the time to equilibrium at very high MOI ($\lambda > 15$), and so in reality this may only matter in cases where such MOIs are typically very high, such as in tissue culture or in some plant viruses (Wilke, Reissig, and Novella 2004; Gutiérrez et al. 2010, 2015).

Bottlenecking may have additional effects that we do not consider in this analysis. For example, if a bottleneck results in a temporary reduction in viral population size, then by chance one variant could be lost from the viral population. In this case, coexistence would only be observed when the lost variant has been regained, for example, through mutation. We do not consider this stochastic effect of bottlenecking since it requires making specific assumptions about bottleneck sizes and the rates of spontaneous generation of the different variants.

2.5.2 *Unequal coexistence*

So far, we have found that the variants tended to coexist in approximately equal proportions at high MOI and high coinfection benefit. This may reflect our assumptions that host cells produce the most virions when infected with an equal mixture of the two types, and that both variants receive a fair share of the productivity of mixed infections. In the next two sections, we relax these assumptions.

2.5.3 *Productivity thresholds*

We examined the consequences of allowing the variants to contribute differently to coinfection benefit, by varying the ratio of *A*:*B* at which cells are most productive (W_M). We did this in three different ways (Fig. 5b–d).

First, we assumed that productivity 'plateaus' such that only a small proportion of either genome is required for the highest coinfection benefit (Fig. 5b). We found that this leads to a slightly higher level of coexistence being maintained at both high and low MOI (Fig. 5b). This is because a higher proportion of mixed infections have the maximal productivity, so mixed infections exert a slightly stronger frequency dependent effect. When coinfection was very common (high MOI), we found that the equilibrium ratio of variant *A* to variant *B* approached 0.5.

Second, we assumed that a small amount of the more productive variant (*A*), but a large amount of the less productive variant (*B*), is required for the highest coinfection benefit (Fig. 5c). We found that when coinfection was relatively rare (low MOI), coexistence was disfavoured. This was because the most productive mixed infections occurred when lots of *B*-virions, and few *A*-virions, infected host cells. This outcome is unlikely when coinfections are rare, as variant *A* is always more frequent than variant *B*. Therefore, mixed infections were on average less productive than when the optimal threshold was more even (Fig. 5a). This lower average productivity of mixed infections leads to a lower equilibrium frequency of the variant which is weaker in pure infection (*B*).

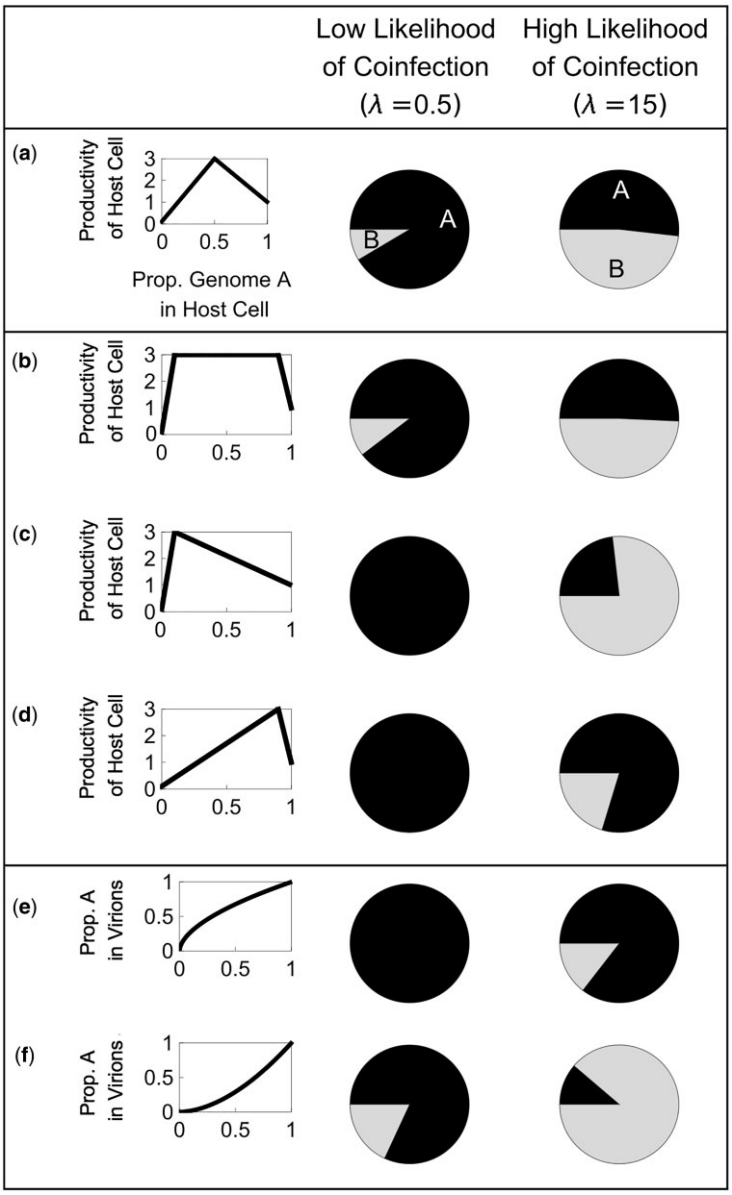


Figure 5. Within-cell processes. Shaded pie charts illustrate the relative frequency of each variant at equilibrium. (a) When both variants contribute equally to coinfection benefit, variant *A* dominates at low MOI, while both variants coexist at high MOI. (b) A similar pattern is seen when a small amount of either a variant maximises coinfection benefit. (c) When a small amount of the more productive variant (A) maximises coinfection benefit, variant *A* dominates at low MOI while variant *B* dominates at high MOI. (d) When a small amount of the less productive variant (B) maximises coinfection benefit, variant *A* dominates at both low and high MOI. (e) When the more productive variant (A) gains a greater share of the virions produced in mixed infection, variant *A* dominates at both low and high MOI. (f) When the less productive variant (B) gains a greater share of the virions produced in mixed infection, variant *A* dominates at low MOI but variant *B* dominates at high MOI. Overall, asymmetries in how each variant contributes to and benefits from coinfection benefit tend to disfavour coexistence.

However, we found a different result when coinfections were very common: at high MOI, the equilibrium frequency of variant *A* decreased below 0.5, favouring the variant that was less productive in pure infection (*B*). This occurred because when coinfections dominate, the overall frequency of the variants in the population depends on the virions released by mixed infections. Mixed infections which have a lower frequency of variant *A* release more virions because they are closer to the optimal A-genome frequency of 0.1, and they release more variant *B* virions than variant *A* virions. Therefore, when coinfections dominated, the overall frequency of the two genomes approached the ratio which maximises productivity in a mixed infection. In this case, this ratio resulted in more of variant *B* than variant *A*.

Third, we assumed that a small amount of the less productive variant (*B*), but a large amount of the more productive variant (*A*), is required for maximal productivity (Fig. 5d). In this scenario, coexistence was again disfavoured when coinfection was relatively rare (low MOI). This was because the most productive mixed infections were most likely to occur when *B* was rare (0.1 relative frequency). Therefore, variant *B* was unable to increase in frequency much above this value, because as variant *B* became more common, mixed infections became on average less productive, and so variant *B* was selected against.

When coinfections were very common (high MOI), coexistence was once again disfavoured: variant *A* became more common than variant *B* (Fig. 5d). This occurred for the same reason as before; when coinfection is very common, the equilibrium frequency of the two variants approaches the ratio that leads to the highest productivity of mixed infections. However, this process is not able to drive variant *B* extinct, since cells with a small fraction of variant *B* produce more virions than cells with no variant *B*.

Overall, these findings suggest that the equilibrium frequency of the two variants can be influenced by the ratio of the two variants that leads to maximum mixed productivity (W_M). When this ratio is asymmetric, implying that a higher proportion of one variant is required than the other, coexistence is disfavoured at both high and low MOI. Furthermore, the variant required in the smaller proportion for maximum productivity will persist at a lower relative frequency at high MOI.

2.5.4 *Within-cell competition*

We investigated the consequences of allowing one variant to gain a disproportionate share of the virions produced in a mixed infection. This could be the case if one variant's genome replicates faster within a cell, for example if it is shorter, or if one variant's genome is incorporated into virions at a faster rate. We consider two cases (Fig. 5e and f). First, we examined when the variant that is more productive in pure infection (*A*) produces a greater share of the virions in mixed infections (Fig. 5e). This could be the case if variant *A* replicates more efficiently than variant *B*, and so produces more genome copies than *B* in both pure and mixed infections. We find that in this case, the variant that does better in both pure and mixed infections (*A*) is more likely to outcompete the other variant (*B*), and so coexistence is disfavoured over the whole parameter space (Fig. 5e).

We then considered the opposite scenario, where the variant that is less productive in pure-infection (*B*) gains a greater share of the productivity of mixed infections (Fig. 5f). This could be the case if variant *B* lacks a key gene and consequently has a shorter genome, for example, if it is a DIP. In this case variant *B* could produce fewer viable virions when in pure infection, but might replicate more rapidly than genome A when in mixed infection. In this scenario, coexistence is favoured at low MOI, since mixed infections provide a stronger frequency dependent force to counteract variant *A*'s pure-infection advantage. However, when coinfection becomes common, variant *B* is able to out-compete variant *A*, reducing diversity in the opposite direction (Fig. 5f). At very high MOI, variant *B* can even drive variant *A* extinct, which agrees with previous theoretical predictions of DIP dynamics (Szathmáry 1993; Kirkwood and Bangham 1994; Frank 2000; Nee 2000; Chao and Elena 2017).

Overall, these findings suggest that when one variant benefits more from coinfection than the other, coexistence is generally disfavoured. When both variants do coexist, then the variant which gains more from mixed infection is likely to reach a higher relative frequency, even if that variant has a lower productivity in pure infection.

3. Parameterising the equilibrium model

We examined whether our equilibrium model led to coexistence when parameterised with real data. A caveat here is that we have not developed a model for a specific species, and there are important biological features, such as spatial structure, that we have left out of the equilibrium model. Also, we need to infer the parameters indirectly. Consequently, our aim here is to test the extent to which a specific case can be accounted for with just the simple processes included in our equilibrium model, and with data that we hope to be the right order of magnitude.

We obtained data from the literature on the H3N2 strain of human influenza A virus. In one study, two variants, that differed at a single amino acid residue in neuraminidase, D-151 and G-151, coexisted at approximately equal frequencies across multiple serial passages in tissue culture (Xue et al. 2016). We are interested in whether this pattern of coexistence can be explained through the negative frequency dependence described by our model. To do this, we need to estimate MOI and coinfection benefit.

The initial MOI is determined by the researchers and fixed at 0.2 at the start of each serial passage. As the infection progresses, the MOI will increase, since the number of viral particles increases while the number of susceptible host cells decreases. This change in MOI was not recorded, and is difficult to infer from the parameters that were recorded. However, previous theoretical work has used an MOI of 10 to reflect the higher MOI values reached over the course of a tissue culture infection (Wilke, Reissig, and Novella 2004). To allow for a conservative test of our model, we will consider MOI in the range 0.2–10 for the first viral growth phase recorded, between 8 and 16 h post-infection. If we assume higher MOI values in the initial growth period, this decreases our estimate of how productive mixed infections are. Therefore, considering an MOI as high as 10 leads to a conservative estimate for coinfection benefit, which decreases the likelihood that our model will predict coexistence.

The magnitude of coinfection benefit depends on the relative productivities (analogous to cellular R_0) of host cells infected by either or both variants. Although we cannot estimate all of the parameters which contribute to cellular R_0, we can infer differences in the fastest viral growth rate, *r*, observed in pure and mixed populations. We account for the fact that in the mixed population, some host cells will be infected by just one variant, by using our upper- and lower-bound estimates for MOI 8–16 h post-infection (0.2–10) and the Poisson function to determine the proportion of host cells in the mixed population treatment that were infected by both variants (full details are in the Supplementary data). We therefore obtain the following estimates for the relative productivities of cells infected by only *D*, only *G*, or both *D* and *G*: $W_D = 1$; $W_G = 0.007$; $W_M = 3.2$ (if MOI=10) or 59 (if MOI=0.2).

With these parameter values, both our upper- and lower-bound estimates for coinfection benefit (58.1 and 3.2, respectively) predict appreciable coexistence between D-151 and G-151 provided MOI is above 2 (Supplementary Fig. S3). Xue et al. (2016) found that coexistence was found between the variants in serial passages starting with an MOI of 0.2. Our upper-bound estimate for coinfection benefit predicts that the stable equilibrium at MOI=0.2 contains both variants whereas our lower-bound estimate predicts that variant D should out-compete

variant G when the MOI is 0.2. However, if we take into account the fact that MOI is likely to increase over the course of the experiment, then both of our estimates for coinfection benefit predict stable coexistence between the two variants, provided that the MOI increases above 0.7 before variant G is lost from the population. Consequently, our model shows that, even with relatively rough calculations of the relevant parameters, to parameterise a relatively simple model, beneficial coinfection can explain the coexistence of multiple genetic variants. The data suggest that D-151 and G-151 coexist at roughly equal frequencies across multiple passages (Xue et al. 2016). Our model predicts that D will be slightly more common, even at high MOI, but there are many parameters, that we have not been able to estimate, which will influence the ratio of the two variants, and our model was not designed to match this specific system (Fig. 5). Our aim here was test our model qualitatively (can it explain coexistence?), not quantitatively (what fraction will be variant D-151?).

4. The spatial simulation

So far, we have taken a relatively simple approach that has allowed us to investigate the role of coinfection likelihood and coinfection benefit in the maintenance of viral diversity. However, this approach is limited in two key ways. Firstly, we have assumed that the relative frequency of each variant and the ratio of virions to susceptible host cells (MOI) are independent. However, if coinfection benefit is above one, then the viral population will increase in size as the ratio of the two variants becomes more even. Consequently, if the MOI in the equilibrium model is taken as the starting MOI, then our model is likely to underestimate the degree of coexistence arising from different levels of coinfection benefit. Secondly, our model does not include spatial structure, which could play an important role in influencing the likelihood of coinfections involving both variants. On the one hand, spatial structure could increase coexistence, since the different viral variants would infect different cells, and so they might not be in direct competition. On the other hand, spatial structure could decrease coexistence, by reducing the likelihood that host cells are infected by both variants, and so reducing the importance of coinfection synergy. To test whether our key predictions still hold when these factors are taken into account, we applied our model in a spatial simulation of viral growth in a two-dimensional grid of cells.

4.1. Simulation description

We have a diffusion-reaction model which we parameterised using values typical for a fast-replicating lytic animal virus. We considered a population of cells in a two-dimensional grid, each of which could be susceptible, infected but not yet producing viruses (eclipse phase), producing viruses, or dead (Fig. 6). We can consider one, two or several virus variants, with characteristics that we can control. Cells can be infected by one or several variants of the virus, and we can use different models to calculate the number of virions of each variant produced by multiply infected cells. We modelled infection as a second-order, Poisson stochastic process that depended on binding to virions. All other cellular state transitions were first-order random processes occurring at a fixed mean time. Infection spread was governed by a diffusion-reaction process in the two-dimensional grid of cells.

In order to generate a chronic infection model, we established values of cell supply rate (r_B) and virus outflow rate (δ_V) that resulted in a stable equilibrium concentration of viruses and cells. In order to calculate a stable MOI, we adapted the infectivity of the virions (k_V). The generation time is the mean eclipse time plus the mean virus production time ($\tau_{EP}+\tau_{PD}$; here 12 h). The infection probability is: $P=1-\exp(-\text{MOI}) = 1-\exp(-\lambda\Delta t)$. So we have that $\text{MOI}=\lambda\Delta t$ and we take $\Delta t=12$ h. In the model, $\lambda=k_V V N_S$, where N_s can only be 0 or 1 (which means that, in a given grid subunit, a cell is either susceptible or unsusceptible/dead) and V is the local virion concentration. This way, we control the MOI using the infectivity parameter, k_V, but, unlike in the equilibrium model, the MOI is also affected by the virus equilibrium concentration (V). The dynamics of V are described by a reaction–diffusion process of the form $\partial V/\partial t=r_v N_p-\delta_V V + D\Delta V$, where N_P are producer cells (0 or 1, as above), r_v is the virus production rate of infected cells, δ_V is the virus degradation/outflow rate, D is the diffusion coefficient, as defined by the Stokes–Einstein equation, and ΔV is the virus concentration gradient (we ignored loss of viruses due to adsorption). We perform dynamic simulations to find the equilibrium coexistence between variants with different fitness values, so we can investigate coexistence when both infectivity

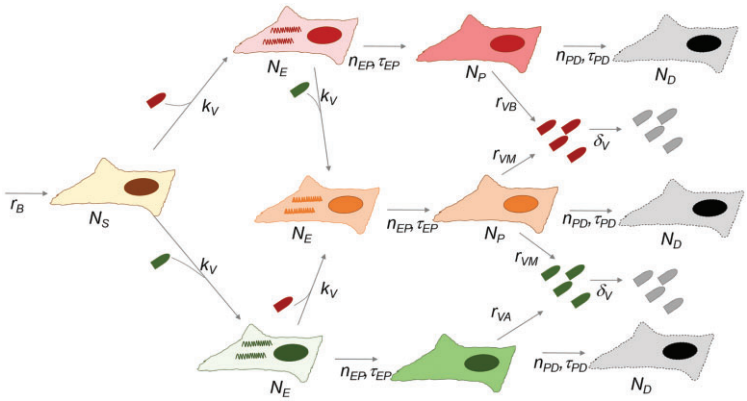


Figure 6. Scheme of the simulation. A two-dimensional grid contained N_S susceptible, N_E eclipse phase, N_P virus-producing, and N_D dead cells. Infection ($N_S \rightarrow N_E$) was a second order, Poisson stochastic processes occurring with probability $P=1-\exp(-\lambda\Delta t)$ for each cell and simulation time unit Δt (0.1 min). For infection, $\lambda=k_V V N_S$, where k_V is the infection rate (infectivity), V is the local virus concentration, and N_S is 1 or 0.

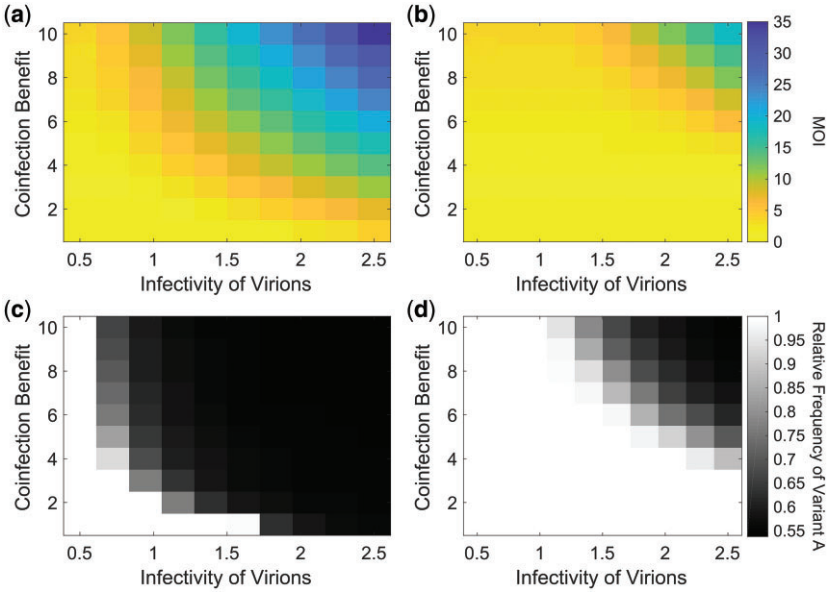


Figure 7. Coexistence in the simulations. The heatmaps represent equilibrium values, which were typically reached after 20 generations. In (b) and (d), superinfection exclusion occurs 3 h after a cell is infected. (a) The final MOI (ratio of viral particles: susceptible host cells) depends on both the infectivity of viral particles and the coinfection benefit. The infectivity is the likelihood that a viral particle will successfully infect a host cell upon contact. (b) Superinfection exclusion reduces the MOI since it makes multiple infection less likely. (c) Coexistence between the two variants is most likely when the coinfection benefit is large and viral particles are highly infectious. (d) Superinfection exclusion reduces the parameter space under which coexistence is found. All parameters were as shown in Supplementary Table S1 except those varied in the graph; r_v was 10 times lower for variant B.

and viral population growth can influence MOI. We allow variant A and variant B to have different fitness values by scaling r_V, the rate at which virions are produced in cells, by W_A or W_B. We allow for coinfection benefit by allowing cells infected by both variants to have the highest r_V values. Parameter values used are shown in Supplementary Table S1 and correspond to a typical fast-replicating lytic animal virus.

4.2. Simulation results

We found that, as predicted, the MOI increased as the infection progressed, and that the final MOI reached depended on both the coinfection benefit and the infectivity of viral particles (Fig. 7a). The degree of coexistence between the two variants was therefore determined by both the coinfection benefit and infectivity (Fig. 7c). We also investigated superinfection exclusion, using a superinfection exclusion time of 3 h post-infection. We found that superinfection exclusion resulted in a lower MOI, which reduced the parameter space under which coexistence occurred, although we still found coexistence at high coinfection benefit and when viral particles were highly infectious (Fig. 7b and d). Our simulation generally reached an equilibrium quickly, within twenty viral generations, which indicates that at least in a two-dimensional infection process, spatial structure may only temporarily reduce the likelihood of multiple infection. We further confirmed that both MOI and coinfection benefit contributed to coexistence in the simulation (Supplementary Fig. S4).

Overall, the simulation results agree with the predictions from the equilibrium model that coinfection benefit and the likelihood of multiple infection can contribute to coexistence and highlight that different factors can contribute to the likelihood of multiple infection.

5. Discussion

We investigated theoretically how mutually beneficial interactions between viral variants influence coexistence. We found that coexistence could occur when mixed infections were frequent relative to pure infections and when they were more productive than pure infections (Figs 2, 3, and 7; Supplementary Figs S1 and S2). This effect did not depend on the initial frequencies of the two variants and it was able to counteract even very significant fitness differences between variants when in pure infections (Fig. 2b). Furthermore, we found that when coinfections were very common, coexistence between variants was determined by two factors: the ratio of the two variants that maximised productivity in mixed infections (Fig. 5b–d); and the relative benefit each variant gains in mixed infections (Fig. 5e and f). We parameterised our model using data from the H3N2 strain of human influenza A virus, and found that it could explain coexistence (Supplementary Fig. S3) (Xue et al. 2016). We also developed a more realistic spatial simulation, and found that in this, mutually beneficial interactions also led to coexistence (Fig. 7).

The extent to which our model predicts coexistence depends upon two main factors. First, coexistence requires coinfection of the same cell by the different variants. Empirical estimates have found that MOI, and hence the possibility for coinfection, is higher in plant viruses and in tissue culture, which could explain why empirical examples of beneficial coinfection mostly come from observational studies on plant viruses, or from cell culture experiments on animal viruses (Wilke, Reissig, and Novella 2004; Gutiérrez et al. 2010; Shirogane, Watanabe, and Yanagi 2012; Bordería et al. 2015; Xue et al. 2016, 2018). Second, we require that mixed coinfections are more productive than pure infections. Both large and small coinfection benefits have

been observed, and these frequently arise when virions from cells in mixed infection are more effective because they contain proteins encoded by multiple viral genomes ('phenotype mixing') (Závada 1976; Roossinck, Sleat, and Palukaitis 1992; López-Ferber et al. 2003; Hull 2009; Shirogane, Watanabe, and Yanagi 2012; Bordería et al. 2015; Xue et al. 2016). One common cause of coinfection benefit could be if beneficial mutations exhibit antagonistic epistasis when in the same genome, but not in different genomes—this appears to occur more commonly in RNA viruses than other organisms (Holmes 2003; Sanjuán and Elena 2006).

Our model highlights how the evolutionary consequences of coinfection depend on the details of how gene products are shared in coinfection. In our model, gene product mixing is synergistic, allowing for beneficial coinfection in which cells infected by both variants are more productive than cells infected by either variant alone. Previous models have either assumed full complementation, in which cells infected by both variants have the same productivity as cells infected by just one variant, or intermediate complementation, in which cells infected by both variants have the mean productivity of cells infected by either variant alone (Chao 1991; Szathmáry 1993; Kirkwood and Bangham 1994; Godfray, O'Reilly, and Briggs 1997; Frank 2000; Bull, Godfray, and O'Reilly 2001; Wilke and Novella 2003; Novella, Reissig, and Wilke 2004; Wilke, Reissig, and Novella 2004; Gao and Feldman 2009). Consequently, these previous models require another mechanism to explain stable coexistence, such as the less fit variant being able to exploit the other (Szathmáry 1992; Kirkwood and Bangham 1994). We have shown that synergistic mixing of gene products, which we have called beneficial coinfection, is enough on its own to allow for the stable coexistence of different variants (Fig. 2).

We found that when coinfection was common, variants that gained a greater share of coinfection could reach very high frequencies, even if they had very low productivity in pure infections. This potential advantage was greatest when coinfection benefit was highest, and so beneficial coinfection could favour mutants that trade off productivity in pure infection for a greater share of the virions produced in coinfection. This could lead to greater selection for DIPs, which lack key genome sections, and so are unable to replicate in pure infections, but can gain a disproportionate advantage in mixed infections (von Magnus 1954; Huang and Baltimore 1970; Nee and Maynard Smith 1990; Szathmáry 1992, 1993; Pathak and Nagy 2009; Nee 2016). Therefore, it is possible that beneficial coinfection could promote coexistence only transiently, with mutually beneficial variants eventually being replaced by DIPs, or that cycles would occur with wild-types evolving resistance to DIPs (DePolo, Giachetti, and Holland, 1987).

Our model can also be applied to help explain the evolutionary stability of multipartite viruses. Multipartite viruses have a genome which is split into multiple segments, and each segment is packaged into a separate virion (van Kammen 1972; Fulton 1980; Lucía-Sanz and Manrubia 2017). Empirical studies have found that these different genome segments of multipartite viruses can exist at different equilibrium frequencies within a host, despite the fact that every segment is required for successful infection of host cells (Sicard et al. 2013; Hu et al. 2016; Wu et al. 2017). In our model, multipartite viruses are captured by the case where neither variant can replicate on its own ($W_A = W_B \sim 0$; $W_M > 0$), and so they represent an extreme example of beneficial coinfection. The potential advantages of being multipartite are captured by our coinfection benefit (W_M) parameter (Iranzo and Manrubia 2012). Our model suggests two mechanisms by which multipartite virus segments could coexist at unequal frequencies within a host (Chao 1991; Szathmáry 1992; Iranzo and Manrubia 2012). The first mechanism was also suggested by Szathmáry (1992) and depends on one segment obtaining a greater benefit from mixed infection than the other (Fig. 5e and f) (Szathmáry 1992). There is evidence that different segments of multipartite viruses could achieve this through different rates of replication or encapsidation (Herzog and Hirth 1978; Loesch-Fries and Hall 1980; Dore, Pinck, and Pinck 1989). The second mechanism is that the segments contribute asymmetrically to the productivity of infected cells, such that cells are most productive when infected by an uneven ratio of segments. This could occur if segments encode different gene products which are required in different amounts. In this case, cells infected by the optimal ratio of segments will produce the most virions, and will also produce virions in this optimal ratio, provided that one genome is not encapsidated substantially faster than the other. Therefore, the equilibrium frequency of segments in the system as a whole converges upon the frequency which maximises the productivity of infected cells (Fig. 5b–d).

To conclude, there are a number of ways that our equilibrium model could be expanded, to match the biology of specific virus-host systems. One possibility is that if mutation rates are high, then natural selection can act on 'clouds' of mutationally linked genotypes (quasispecies theory), rather than on individual genotypes (Wilke, Reissig, and Novella 2001; Wilke 2005; Lauring and Andino 2010; Domingo, Sheldon, and Perales 2012). Our model shows how coexistence can emerge without invoking high mutation rates, and so it may be applicable in a wide range of viruses. Different modes of viral spread can also influence the likelihood of multiple infections, and so a natural extension of our model could incorporate cell–cell spread, virion aggregation, and other modes of collective infection (Sanjuán 2017, 2018). Finally, our model has focused on evolution within hosts, whereas coexistence of viral variants at the epidemiological level is likely to depend on both evolution within hosts and transmission between hosts.

Supplementary data

Supplementary data are available at Virus Evolution online. Supplementary code is available at https://osf.io/akrmp/.

Acknowledgements

A.L. was supported by funding from the Clarendon Fund, St. John's College, Oxford, and the BBSRC (grant number BB/M011224/1). E.A.S-O. and R.S. were supported by ERC Consolidator Grant 724519 (ViS-a-ViS). For useful comments and discussion, we thank Santiago F. Elena, Matishalin Patel, the Oxford social evolution group, Susanna Manrubia, and two anonymous referees. Supplementary information and code is available at https://osf.io/akrmp/.

Conflict of interest: None declared.

References

Andino, R., and Domingo, E. (2015) 'Viral Quasispecies', *Virology*, 479–480: 46–51.

Arbiza, J., Mirazo, S., and Fort, H. (2010) 'Viral Quasispecies Profiles as the Result of the Interplay of Competition and Cooperation', *BMC Evolutionary Biology*, 10: 137.

Bergua, M. et al. (2014) 'A Viral Protein Mediates Superinfection Exclusion at the Whole-Organism Level but Is Not Required for Exclusion at the Cellular Level', *Journal of Virology*, 88: 11327–38.

Bonhoeffer, S. et al. (1997) 'Virus Dynamics and Drug Therapy', *Proceedings of the National Academy of Sciences of the United States of America*, 94: 6971–6.

Bordería, A. V. et al. (2015) 'Group Selection and Contribution of Minority Variants during Virus Adaptation Determines Virus Fitness and Phenotype', *PLOS Pathogens*, 11: e1004838.

——, Stapleford, K. A., and Vignuzzi, M. (2011) 'RNA Virus Population Diversity: Implications for Inter-Species Transmission', *Current Opinion in Virology*, 1: 643–8.

Bull, J. C., Godfray, H. C. J., and O'Reilly, D. R. (2001) 'Persistence of an Occlusion-Negative Recombinant Nucleopolyhedrovirus in Trichoplusia Ni Indicates High Multiplicity of Cellular Infection', *Applied and Environmental Microbiology*, 67: 5204–9.

Cao, L. et al. (2014) 'Coexistence of Hepatitis B Virus Quasispecies Enhances Viral Replication and the Ability to Induce Host Antibody and Cellular Immune Responses', *Journal of Virology*, 88: 8656–66.

Chao, L. (1991) 'Levels of Selection, Evolution of Sex in RNA Viruses, and the Origin of Life', *Journal of Theoretical Biology*, 153: 229–46.

—— & Elena, S. F. (2017) 'Nonlinear Trade-Offs Allow the Cooperation Game to Evolve from Prisoner's Dilemma to Snowdrift', *Proceedings of the Royal Society B: Biological Sciences*, 284: 20170228.

Clarke, D. K. et al. (1994) 'The Red Queen Reigns in the Kingdom of RNA Viruses', *Proceedings of the National Academy of Sciences of the United States of America*, 91: 4821–4.

Coffey, L. L. et al. (2011) 'Arbovirus High Fidelity Variant Loses Fitness in Mosquitoes and Mice', *Proceedings of the National Academy of Sciences*, 108: 16038–43.

DePolo, N. J., Giachetti, C., and Holland, J. J. (1987) 'Continuing Coevolution of Virus and Defective Interfering Particles and of Viral Genome Sequences during Undiluted Passages: Virus Mutants Exhibiting Nearly Complete Resistance to Formerly Dominant Defective Interfering Particles', *Journal of Virology*, 61: 454–64.

Díaz-Muñoz, S. L. (2017) 'Viral Coinfection Is Shaped by Host Ecology and Virus–Virus Interactions across Diverse Microbial Taxa and Environments', *Virus Evolution*, 3, DOI: 10.1093/ve/vex011.

——, Sanjuán, R., and West, S. A. (2017) 'Sociovirology: Conflict, Cooperation, and Communication among Viruses', *Cell Host & Microbe*, 22: 437–41.

Doceul, V. et al. (2010) 'Repulsion of Superinfecting Virions: A Mechanism for Rapid Virus Spread', *Science (New York, N.Y.)*, 327: 873–6.

Domingo, E., Sheldon, J., and Perales, C. (2012) 'Viral Quasispecies Evolution', *Microbiology and Molecular Biology Reviews*, 76: 159–216.

Dore, J. M., Pinck, M., and Pinck, L. (1989) 'Competitive Multiplication of RNA3 Species of Different Strains of Alfalfa Mosaic Virus', *The Journal of General Virology*, 70: 777–82.

Elena, S. F. et al. (2011) 'The Evolutionary Genetics of Emerging Plant RNA Viruses', *Molecular Plant-Microbe Interactions*, 24: 287–93.

——, Miralles, R., and Moya, A. (1997) 'Frequency-Dependent Selection in a Mammalian RNA Virus', *Evolution*, 51: 984–7.

Erickson, A. K. et al. (2018) 'Bacteria Facilitate Enteric Virus Co-Infection of Mammalian Cells and Promote Genetic Recombination', *Cell Host & Microbe*, 23: 77–88.e5.

Folimonova, S. Y. (2012) 'Superinfection Exclusion Is an Active Virus-Controlled Function That Requires a Specific Viral Protein', *Journal of Virology*, 86: 5554–61.

Frank, S. A. (2000) 'Within-Host Spatial Dynamics of Viruses and Defective Interfering Particles', *Journal of Theoretical Biology*, 206: 279–90.

Froissart, R. et al. (2004) 'Co-Infection Weakens Selection against Epistatic Mutations in RNA Viruses', *Genetics*, 168: 9–19.

Fulton, R. W. (1980) 'Biological Significance of Multicomponent Viruses', *Annual Review of Phytopathology*, 18: 131–46.

Gao, H., and Feldman, M. W. (2009) 'Complementation and Epistasis in Viral Coinfection Dynamics', *Genetics*, 182: 251–63.

Gause, G. F. (1934). *The Struggle for Existence*. Baltimore: The Williams & Wilkins company.

Godfray, H. C. J., O'Reilly, D. R., and Briggs, C. J. (1997) 'A Model of Nucleopolyhedrovirus (NPV) Population Genetics Applied to Co–Occlusion and the Spread of the Few Polyhedra (FP) Phenotype', *Proceedings of the Royal Society of London B: Biological Sciences*, 264: 315–22.

Gutiérrez, S. et al. (2015) 'The Multiplicity of Cellular Infection Changes Depending on the Route of Cell Infection in a Plant Virus', *Journal of Virology*, 89: 9665–75.

—— et al. (2010) 'Dynamics of the Multiplicity of Cellular Infection in a Plant Virus', *PLOS Pathogens*, 6: e1001113.

Herzog, M., and Hirth, L. (1978) 'In Vitro Encapsidation of the Four RNA Species of Brome Mosaic Virus', *Virology*, 86: 48–56.

Hisano, S. et al. (2018) 'A Neo-Virus Lifestyle Exhibited by a (+)ssRNA Virus Hosted in an Unrelated dsRNA Virus: Taxonomic and Evolutionary Considerations', *Virus Research*, 244: 75–83.

Holmes, E. C. (2003) 'Error Thresholds and the Constraints to RNA Virus Evolution', *Trends in Microbiology*, 11: 543–6.

Hu, Z. et al. (2016) 'Genome Segments Accumulate with Different Frequencies in Bombyx mori Bidensovirus', *Journal of Basic Microbiology*, 56: 1338–43.

Huang, A. S., and Baltimore, D. (1970) 'Defective Viral Particles and Viral Disease Processes', *Nature*, 226: 325–7.

Hull, R. (2009). *Comparative Plant Virology*, 2nd edn. Amsterdam; Boston: Elsevier Academic Press.

Iranzo, J., and Manrubia, S. C. (2012) 'Evolutionary Dynamics of Genome Segmentation in Multipartite Viruses', *Proceedings of the Royal Society of London B: Biological Sciences*, 279: 3812–9.

van Kammen, A. (1972) 'Plant Viruses with a Divided Genome', *Annual Review of Phytopathology*, 10: 125–50.

Ke, R. et al. (2015) 'Rational Design and Adaptive Management of Combination Therapies for Hepatitis C Virus Infection', *PLOS Computational Biology*, 11: e1004040.

Kirkwood, T. B., and Bangham, C. R. (1994) 'Cycles, Chaos, and Evolution in Virus Cultures: A Model of Defective Interfering Particles', *Proceedings of the National Academy of Sciences*, 91: 8685–9.

Klasse, P. J. (2015) 'Molecular Determinants of the Ratio of Inert to Infectious Virus Particles', *Progress in Molecular Biology and Translational Science*, 129: 285–326.

Lauring, A. S., and Andino, R. (2010) 'Quasispecies Theory and the Behavior of RNA Viruses', *PLOS Pathogens*, 6: e1001005.

Loesch-Fries, L. S., and Hall, T. C. (1980) 'Synthesis, Accumulation and Encapsidation of Individual Brome Mosaic Virus RNA Components in Barley Protoplasts', *Journal of General Virology*, 47: 323–32.

López-Ferber, M. et al. (2003) 'Defective or Effective? Mutualistic Interactions between Virus Genotypes', *Proceedings of the Royal Society of London B: Biological Sciences*, 270: 2249–55.

Loverdo, C., and Lloyd-Smith, J. O. (2013) 'Inter-Generational Phenotypic Mixing in Viral Evolution', *Evolution; International Journal of Organic Evolution*', *Evolution*, 67: 1815–22.

Lucía-Sanz, A., and Manrubia, S. (2017) 'Multipartite Viruses: Adaptive Trick or Evolutionary Treat?', *NPJ Systems Biology and Applications*, 3: 34.

von Magnus, P. (1954) 'Incomplete Forms of Influenza Virus', *Advances in Virus Research*, 2: 59–79.

McCrone, J. T., and Lauring, A. S. (2018) 'Genetic Bottlenecks in Intraspecies Virus Transmission', *Current Opinion in Virology, Emerging Viruses: intraspecies Transmission • Viral Immunology*, 28: 20–5.

Moya, A. et al. (2000) 'The Evolution of RNA Viruses: A Population Genetics View', *Proceedings of the National Academy of Sciences*, 97: 6967–73.

Nee, S. (2000) 'Mutualism, Parasitism and Competition in the Evolution of Coviruses', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 355: 1607–13.

—— (2016) 'The Evolutionary Ecology of Molecular Replicators', *Royal Society Open Science*, 3: 160235.

—— & Maynard Smith, J. (1990) 'The Evolutionary Biology of Molecular Parasites', *Parasitology*, 100: S5–18.

Novella, I. S., Reissig, D. D., and Wilke, C. O. (2004) 'Density-Dependent Selection in Vesicular Stomatitis Virus', *Journal of Virology*, 78: 5799–804.

Nowak, M. A. et al. (1997) 'Viral Dynamics of Primary Viremia and Antiretroviral Therapy in Simian Immunodeficiency Virus Infection', *Journal of Virology*, 71: 7518–25.

Otto, S. P., and Day, T. (2007). *A Biologist's Guide to Mathematical Modeling in Ecology and Evolution*. Princeton: Princeton University Press.

Pathak, K. B., and Nagy, P. D. (2009) 'Defective Interfering RNAs: Foes of Viruses and Friends of Virologists', *Viruses*, 1: 895–919.

Pérez-Losada, M. et al. (2015) 'Recombination in Viruses: Mechanisms, Methods of Study, and Evolutionary Consequences', *Infection, Genetics and Evolution*, 30: 296–307.

Qiu, W., and Scholthof, K.-B. G. (2001) 'Defective Interfering RNAs of a Satellite Virus', *Journal of Virology*, 75: 5429–32.

Roossinck, M. J., Sleat, D., and Palukaitis, P. (1992) 'Satellite RNAs of Plant Viruses: Structures and Biological Effects', *Microbiological Reviews*, 56: 265–79.

Sanjuán, R. (2017) 'Collective Infectious Units in Viruses', *Trends in Microbiology*, 25: 402–12.

—— (2018) 'Collective Properties of Viral Infectivity', *Current Opinion in Virology, Virus Vector Interactions • Special Section: Multicomponent Viral Systems*, 33: 1–6.

—— & Elena, S. F. (2006) 'Epistasis Correlates to Genomic Complexity', *Proceedings of the National Academy of Sciences*, 103: 14402–5.

Shirogane, Y., Watanabe, S., and Yanagi, Y. (2012) 'Cooperation between Different RNA Virus Genomes Produces a New Phenotype', *Nature Communications*, 3: 1235.

Sicard, A. et al. (2013) 'Gene Copy Number Is Differentially Regulated in a Multipartite Virus', *Nature Communications*, 4: 2248.

Skums, P., Bunimovich, L., and Khudyakov, Y. (2015) 'Antigenic Cooperation among Intrahost HCV Variants Organized into a Complex Network of Cross-Immunoreactivity', *Proceedings of the National Academy of Sciences*, 112: 6653–8.

Szathmáry, E. (1992) 'Natural Selection and Dynamical Coexistence of Defective and Complementing Virus Segments', *Journal of Theoretical Biology*, 157: 383–406.

—— (1993) 'Co-Operation and Defection: Playing the Field in Virus Dynamics', *Journal of Theoretical Biology*, 165: 341–56.

Timm, A., and Yin, J. (2012) 'Kinetics of Virus Production from Single Cells', *Virology*, 424: 11–7.

Vignuzzi, M. et al. (2006) 'Quasispecies Diversity Determines Pathogenesis through Cooperative Interactions in a Viral Population', *Nature*, 439: 344–8.

Wilke, C. O. (2005) 'Quasispecies Theory in the Context of Population Genetics', *BMC Evolutionary Biology*, 5: 44.

—— & Novella, I. S. (2003) 'Phenotypic Mixing and Hiding May Contribute to Memory in Viral Quasispecies', *BMC Microbiology*, 3: 11.

——, Reissig, D. D., and Novella, I. S. (2004) 'Replication at Periodically Changing Multiplicity of Infection Promotes Stable Coexistence of Competing Viral Populations', *Evolution; International Journal of Organic Evolution*, 58: 900–5.

—— et al. (2001) 'Evolution of Digital Organisms at High Mutation Rates Leads to Survival of the Flattest', *Nature*, 412: 331–3.

Wu, B. et al. (2017) 'Within-Host Evolution of Segments Ratio for the Tripartite Genome of Alfalfa Mosaic Virus', *Scientific Reports*, 7: 5004.

Xue, K. S. et al. (2018) 'Cooperating H3N2 Influenza Virus Variants Are Not Detectable in Primary Clinical Samples', *mSphere*, 3: e00552–17.

—— et al. (2016) 'Cooperation between Distinct Viral Variants Promotes Growth of H3N2 Influenza in Cell Culture', *eLife*, 5: e13974.

Yuste, E., Moya, A., and López-Galíndez, C. (2002) 'Frequency-Dependent Selection in Human Immunodeficiency Virus Type 1', *Journal of General Virology*, 83: 103–6.

Závada, J. (1976) 'Viral Pseudotypes and Phenotypic Mixing', *Archives of Virology*, 50: 1–15.

Zwart, M. P., and Elena, S. F. (2015) 'Matters of Size: Genetic Bottlenecks in Virus Infection and Their Potential Impact on Evolution', *Annual Review of Virology*, 2: 161–79.

4

The Evolution of Collective Infectious Units in Viruses

Virus Research 265 (2019) 94–101

Contents lists available at ScienceDirect

Virus Research

journal homepage: www.elsevier.com/locate/virusres

The evolution of collective infectious units in viruses

Asher Leeks[a,*], Rafael Sanjuán[b], Stuart A. West[a]

[a] University of Oxford, Department of Zoology, Zoology Research and Administration, Oxford, OX1 3SZ, United Kingdom
[b] Institute for Integrative Systems Biology (I2SysBio), Universitat de València, València, Spain

ARTICLE INFO

Keywords:
Collective infection
Collective infectious unit
Virus evolution
Multiplicity of infection
Defective interfering genome
Bloc transmission

ABSTRACT

Viruses frequently spread among cells or hosts in groups, with multiple viral genomes inside the same infectious unit. These collective infectious units can consist of multiple viral genomes inside the same virion, or multiple virions inside a larger structure such as a vesicle. Collective infectious units deliver multiple viral genomes to the same cell simultaneously, which can have important implications for viral pathogenesis, antiviral resistance, and social evolution. However, little is known about why some viruses transmit in collective infectious units, whereas others do not. We used a simple evolutionary approach to model the potential costs and benefits of transmitting in a collective infectious unit. We found that collective infectious units could be favoured if cells infected by multiple viral genomes were significantly more productive than cells infected by just one viral genome, and especially if there were also efficiency benefits to packaging multiple viral genomes inside the same infectious unit. We also found that if some viral sequences are defective, then collective infectious units could evolve to become very large, but that if these defective sequences interfered with wild-type virus replication, then collective infectious units were disfavoured.

1. Introduction

Viruses disperse from host cells in many different ways. Some viruses disperse in single virions which each contain one genome. Other viruses can disperse in groups, with multiple genomes in the same virion, or multiple virions inside a larger structure. These are called collective infectious units (CIUs), and are characterised by multiple viral genomes transmitting as part of the same infective structure (Sanjuán, 2017). The simplest collective infectious units are virions containing multiple genomes ("polyploid" virions), and in some cases these can contain a variable number of genome copies (Hosaka et al., 1966; Dahlberg and Simon, 1969; Luque et al., 2009; Rager et al., 2002). In other cases, collective infectious units can comprise larger structures containing multiple virions. These can form through free virions aggregating after dispersal, either through direct contact with one another or through collectively binding to a vector, such as a bacterial cell (Bald and Briggs, 1937; Cuevas et al., 2017; Erickson et al., 2018). Alternatively, multiple virions can collectively disperse from the same host cell, for example inside extra-cellular vesicles formed of sections of host cell membrane, or inside protein-coated occlusion bodies (Altan-Bonnet and Chen, 2015; Chen et al., 2015; Santiana et al., 2018; Slack and Arif, 2007). These various kinds of collective infectious units appear to have evolved independently many times, since they exist in many different viral families and take a range of structural forms.

Transmitting as part of a CIU can have important consequences for viral evolution. By allowing the same host cell to be infected by multiple viral genomes simultaneously, CIUs allow for interactions between viruses even when we would otherwise expect coinfection to be rare, such as when there are strong population bottlenecks or low ratios of infectious viral particles to susceptible host cells (McCrone and Lauring, 2018; Sanjuán, 2018). Interactions between viral sequences can have important consequences for viral pathogenesis, diversity, and the evolution of antiviral resistance (Bordería et al., 2015; Leeks et al., 2018; Tanner et al., 2014; Vignuzzi et al., 2006; Xue et al., 2016). Furthermore, CIUs allow for repeated interactions between viral sequences, and this sets the stage for viral social adaptations. This can include cooperation, where viruses evolve adaptations that benefit other viruses, but may more commonly facilitate conflict, as in the case of defective interfering (DI) genomes, which exploit the cellular machinery of coinfecting viruses (Chao and Elena, 2017; Díaz-Muñoz et al., 2017; Huang and Baltimore, 1970; Sanjuán, 2018; Turner and Chao, 1999).

A number of hypotheses have been proposed for why viruses might transmit in CIUs. One possibility is that cells infected by multiple viral genomes might lead to more productive infections than cells infected by

* Corresponding author.
E-mail address: asher.leeks@zoo.ox.ac.uk (A. Leeks).

https://doi.org/10.1016/j.virusres.2019.03.013
Received 10 December 2018; Received in revised form 15 March 2019; Accepted 16 March 2019
Available online 17 March 2019

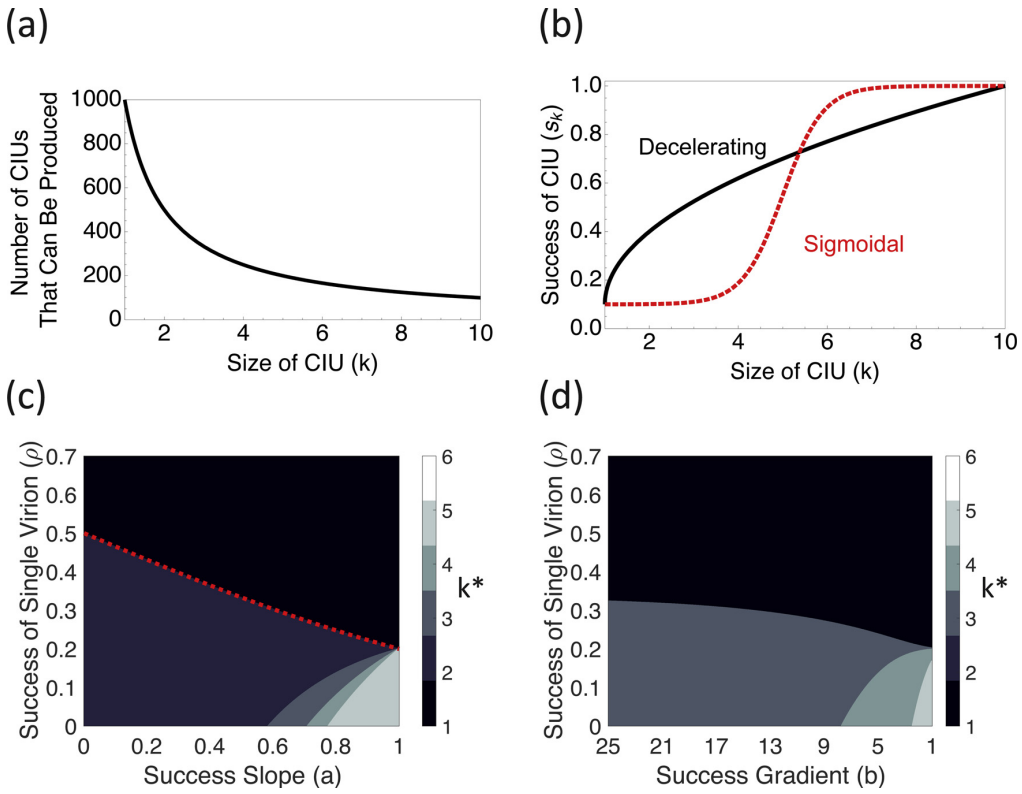


Fig. 1. Group infection benefits and CIU evolution. (a) plots the opportunity cost of larger CIUs. All else being equal, fewer CIUs can be produced if each CIU contains more genomes. In our model, we only use integer values of k. (b) plots the relationship between the success of a CIU and the number of genomes it contains. (c) and (d) plot the optimal size of a CIU (k^*) when CIU success has a diminishing (c) or threshold (d) relationship with CIU size. The red dashed line in (c) plots the analytical condition for when CIUs evolve. When infectious unit success has diminishing returns (c), larger CIUs ($k^* > 2$) only evolve when the success slope is relatively flat (a is high). In contrast, when there are threshold effects (d), only larger CIUs ($k^* > 2$) are found, but these are found over less of the parameter space.

just one viral genome (Andreu-Moreno and Sanjuán, 2018; Borges et al., 2018; Guo et al., 2017; Landsberger et al., 2018; Stiefel et al., 2012; Xue et al., 2016). In this case, CIUs might evolve if they are an effective way of delivering multiple viral genomes to the same cell (Sanjuán, 2017). A second mechanism could be if CIUs allow for more efficient use of limited resources, and therefore viruses could evolve larger burst sizes by packaging multiple genomes into the same infectious unit. A third mechanism could be if viruses have a high likelihood of producing defective genomes. In that case, CIUs could be favoured to ensure that at least one functional copy of each gene is delivered to a cell, or to increase the chance that one or more complete genomes arrive in a host cell (Andino and Domingo, 2015; Stiefel et al., 2012).

We model the theoretical plausibility of these three types of hypothesis: (i) if cells infected by multiple viral genomes are more productive (group infection benefits); (ii) if packaging multiple genomes into the same unit is more efficient (efficiency benefits); (iii) if there is a high likelihood that genomes are defective (insurance benefits). We ask whether each kind of hypothesis can plausibly favour the evolution of collective infectious units. For each case, we investigate what conditions are required for CIUs to be favoured as well as what sizes of CIU are favoured. Do we expect to see CIUs in all viruses, most viruses, or only under special conditions? Are some kinds of viruses more likely to evolve CIUs than others? And when CIUs do evolve, do we expect them to be small, containing just a few viral genomes, or large?

2. Model

Our goal is to examine the general theoretical plausibility of potential mechanisms, rather than to capture the specific details of a single species. We have therefore purposefully left out a number of potentially important details, such as complementation between defective mutants and beneficial interactions between different variants (Andino and Domingo, 2015; Leeks et al., 2018; Xue et al., 2016). We have chosen to model hypotheses which could apply to many viruses, and which could vary in predictable ways. Furthermore, we have focused on modelling the number of genomes that a generic infectious unit should contain, where an infectious unit is any structure that can deliver viral genomes to new host cells. Therefore, infectious units could reflect different biological structures, including virions, extracellular vesicles, or occlusion bodies. Our aim is to generate testable predictions across a range of different CIUs and consequently to encourage interplay between theory and data in the study of collective infectious units.

2.1. Model lifecycle

We imagine an acute, lytic virus spreading within a host. We assume that natural selection acts in order to maximise the rate at which it spreads. We therefore define a viral genotype's fitness as equivalent to the expected number of future infected cells from a given infected cell. We assume that superinfection is rare enough to be ignored and that the viral progeny which leave a cell are identical to the viral genotype which initially infected the cell. Consequently, we express viral fitness,

W, as:

$$W = \sum_{k=1}^{\infty} n_k s_k \quad (1)$$

Where k is the number of genomes inside each infectious unit, n_k is the number of infectious units of size k produced and s_k is the expected number of future cellular infections each of these virions will lead to, scaled between 0 and 1. Next, we simplify our fitness equation so that we can compare the fitness of viral variants that transmit in infectious units of different sizes k:

$$W_k = n_k s_k \quad (2)$$

We assume that the number of infectious units that can be produced per unit time (n_k) depends on both the number of viral genomes produced in the cell and the number of genomes that are packaged into each infectious unit. The total number of viral genomes produced by a virus may depend on the size of the infectious unit, because viruses with larger infectious units may use gene products more efficiently and so produce more genomes (see section 2.3: efficiency benefits). Consequently, we arrive at our general fitness equation:

$$W = \frac{n_g(k)}{k} s_k \quad (3)$$

Where $n_g(k)$ is the number of genomes produced by a virus which disperses in infectious units of size k.

Eq. 3 reveals that there is a trade-off between the number of infectious units that can be produced and the number of genomes inside each infectious unit (Fig. 1a). This trade-off is analogous to that between the number and size of offspring (clutch size) produced by animals: with all other factors equal, the larger the clutch size, the fewer clutches can be produced (Godfray et al., 1991; Lack, 1947). We will now consider three factors that could potentially favour CIUs.

2.2. Group infection benefits

We first consider the possibility that infections initiated with multiple viral genomes are more successful. We assume that the expected number of future infections is larger for infectious units containing more genomes, by making $s(k)$ an increasing function of k. This could capture different biological mechanisms, including: larger infectious units lasting longer in the environment and so surviving longer to infect a host cell; larger numbers of initial genomes leading to a faster rate of viral production throughout the course of the cellular infection; larger infectious units having a greater likelihood of initially establishing an infection, for example through overcoming cellular immune responses, or if stochastic events early in infection can cause infections to fail (Andreu-Moreno and Sanjuán, 2018; Stiefel et al., 2012). The benefit to infectious units with more genomes is analogous to when animals experience benefits through dispersing in groups, rather than alone (Davies et al., 2012; Hamilton, 1971).

We assume that there is a limit to the potential benefit of multiple viral genomes infecting the same cell, and consequently that beyond a certain number of genomes, defined as k_t, additional genomes no longer increase the productivity of an infected cell. Since we are interested in the relative fitness of different infectious unit sizes, we set the maximum potential benefit of larger CIUs, which is found at k_t, equal to 1 and we express the success of infectious units of different sizes relative to this maximum potential benefit (y-axis of Fig. 1b). We also assume that the number of viral genomes produced per infected cell is constant, and that there is consequently a linear trade-off between the number of infectious units that can be produced and the number of genomes in each infectious unit (Fig. 1a). We consider the cases where the relationship between the number of viral genomes (k) and the productivity of an infected cell (s_k) either shows diminishing returns, or a threshold effect (Fig. 1b; Appendix 1).

We are interested both in when CIUs evolve, and in the size of CIUs that evolve. We therefore search for the size of infectious unit (k) that maximises viral fitness as defined in Eq. 3. We denote this value k^*, and it represents a candidate evolutionarily stable strategy, meaning that it could not be outcompeted by a virus employing any other strategy (Maynard Smith and Price, 1973). When $k^* > 1$, CIUs are favoured over individual transmission. To find k^*, we evaluate our fitness equation (Eq. 3) numerically at a large number of different values of k to determine the value which results in the highest fitness (Fig. 1 c–d). In section 2 of the Appendix, we derive an analytical condition for when collective transmission can be favoured over individual transmission when the group benefit shows diminishing returns, which we overlay in Fig. 1c.

We found that CIUs were more likely to evolve when: (i) infections initiated by a single viral genome are relatively unsuccessful (low ρ) (Fig. 1c–d); (ii) a small number of initial infecting genomes can reach the maximal infection efficiency (low k_t); (iii) additional genomes have a greater influence on infection success when there are fewer genomes infecting a cell (a steeper success gradient; Fig. 1b–d); (iv) additional genomes result in a diminishing relationship with infection success (Fig. 1b–d).

We found that the conditions that favoured large CIUs were not the same as those that favoured CIUs *per* se. In particular, larger CIUs were favoured when: (i) infections initiated by a single viral genome are relatively unsuccessful (low ρ) (Fig. 1c–d); (ii) a large number of initial infecting genomes are required for a successful infection (high k_t); (iii) additional genomes have a constant influence on infection success (a shallower success gradient; Fig. 1b–d); (iv) additional genomes show a threshold effect, resulting in a sigmoidal relationship with infection success (Fig. 1b–d).

For many factors (ii-iv above), we found that conditions that allowed CIUs to evolve more easily also favoured the evolution of smaller CIUs (Table 1). This pattern occurred because viruses are able to produce more CIUs if those CIUs are smaller (Fig. 1a). Consequently, CIUs were more likely to evolve when smaller CIUs were more successful, since in these cases viruses could achieve both the advantages of collective benefit and the advantages of transmitting large numbers of infectious units. In contrast, when the advantages of collective benefit were only possible with large numbers of genomes, viruses were able to

Table 1
Summary of theoretical predictions.

Class of Benefit	Factor	Effect on likelihood of CIU	Effect on size of CIU
Group Infection Benefits	Diminishing returns (decelerating relationship)	More likely	Smaller
	Threshold effect (sigmoidal relationship)	Less likely	Larger
	Steeper relationship	More likely	Smaller
	Low success rate of individual virion	More likely	Larger
	High threshold number of genomes	Less likely	Larger
Efficiency Benefits	Larger CIU transmits genomes more efficiently than multiple smaller CIUs	More likely	Larger
	More efficient use of CIU allows viral genome to be replicated more	More likely	Larger
Insurance Benefits	Lots of defective sequences	No effect	Larger
	Lots of defective interfering sequences	Less likely	Smaller

transmit fewer of these collective units, and so CIUs were less likely to evolve.

2.3. Efficiency benefits

A second hypothesis for the evolution of CIUs is that they may allow a more efficient way of packaging genomes into infectious units. There are two ways that efficiency benefits could result in increased viral fitness. The first way is that there could be a limited number of structures available for collective transmission, and so packaging more genomes inside each structure could allow for more genomes to be transmitted via the collective route. This mechanism assumes either that more viral genomes are produced than can be transmitted (if CIUs are essential for transmission), or that there is an intrinsic benefit to transmitting in a CIU as opposed to transmitting as an individual virion (if CIUs are non-essential for transmission). However, there seemed no reason to assume that more viral genomes are produced than transmitted, and the second condition requires that there is already a benefit to transmitting collectively.

Therefore, we instead focus on a second hypothesis for efficiency benefits, that viruses with more efficient packaging can evolve to produce higher numbers of genomes. This hypothesis requires two assumptions. First, that larger infectious units are more efficient at packaging genomes than smaller infectious units. This occurs when a single CIU containing multiple genomes costs fewer resources than the equivalent number of infectious units containing one genome. A general way that this could occur is if the resources required to produce an infectious unit increase with surface area, and the number of genomes that it can carry depends on its volume, in which case the potential efficiency gains depend on the ratio of volume to surface area as the infectious unit increases in size. Therefore, potential efficiency gains will be greatest in infectious units which are more spherical, and which enlarge by lengthening in all dimensions simultaneously, rather than by lengthening just in one dimension.

The second requirement for this hypothesis is that the increased efficiency benefits allow for a greater number of viral genomes to be produced. One way in which this might occur is if the infectious unit is constructed from virus-derived gene products, such as structural proteins, as occurs with polyploid virions and baculovirus occlusion bodies. In this case, a more efficient use of viral proteins would result in fewer viral proteins being required to transmit the same number of viral genomes. Since viral genome copies and viral genome products are both produced by transcription of the viral genome, viruses which use these structural proteins more efficiently could evolve to produce more viral genomes (Chao and Elena, 2017).

We found that the greatest efficiency benefits occurred when infectious units were spherical and when more efficient infectious units allowed more genomes to be produced. In that case, efficiency benefits scaled with the cubic root of k (Fig. 2a; Appendix 3). These maximum efficiency benefits therefore increased more slowly than the cost of including additional genomes (Fig. 1a), and so efficiency benefits alone were not able to favour the evolution of collective infectious units.

However, we did find that efficiency benefits were able to favour CIUs, and lead to larger CIUs, if combined with group infection benefits (Fig. 2b; section 2.2). This suggests that the requirements for CIUs to be favoured by group infection benefits may be lower when there are greater potential efficiency gains from CIUs. We found that this result critically depended on the assumption that more efficient infectious units could result in more genomes being produced (Fig. 2b).

2.4. Defective and defective interfering genomes

The third hypothesis we investigate rests on the fact that viral replication is error prone, and so some proportion of viral progeny are defective, meaning that they lack functional copies of genes required for successful infection. A high error rate could favour the evolution of

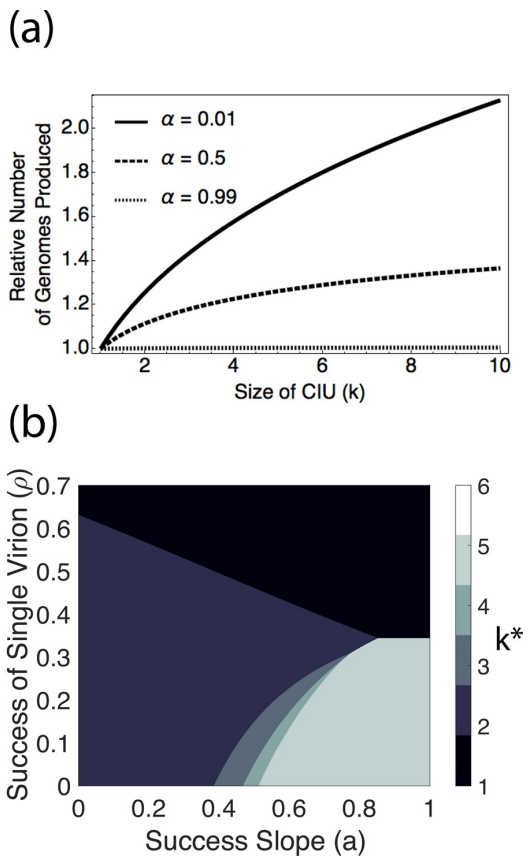


Fig. 2. The influence of efficiency benefits on CIU evolution. (a) plots the potential increase in genome availability which comes from transmitting in CIUs of larger sizes. The increased genome availability depends on α, which reflects the extent to which increased efficiency of genome packaging results in more viral genome copies being produced. (b) plots the optimal size of CIU (k^*) which is reached for a spherical CIU with $\alpha = 0$, reflecting the largest possible efficiency gains from larger CIUs. Compared to Fig. 1c, where there are no efficiency gains, CIUS evolve in a larger region of parameter space and are larger when they do evolve.

CIUs, since a larger infectious unit may have a greater likelihood of containing at least one functional genome. However, in most viruses, some fraction of defective genomes are also interfering, meaning that they reduce the accumulation of the wild-type, for example if they are preferentially replicated at the expense of the wild-type genome (Huang and Baltimore, 1970; Jaworski and Routh, 2017; Manzoni and López, 2018; Rezelj et al., 2018). Defective interfering genomes may disfavour the evolution of CIUs since larger infectious units may be more likely to contain an interfering genome. Here we incorporate both defective genomes and defective interfering genomes to see how these factors influence CIU evolution.

We investigate the possibility for defective genomes by assuming that a proportion μ of genomes produced are defective. For mathematical simplicity, we assume that these defective genomes are unable to be replicated in infected cells, and consequently that they don't contribute to the success of infectious units. Therefore, this model captures the idea that collective infection could make it more likely that at least one complete genome infects a host cell ('insurance benefits'), but this model does not allow for defective genes to be trans-complemented by functional copies of the same gene in a different genome ('trans-complementation') (Andino and Domingo, 2015; Stiefel et al., 2012). By ignoring trans-complementation, our model may over- or

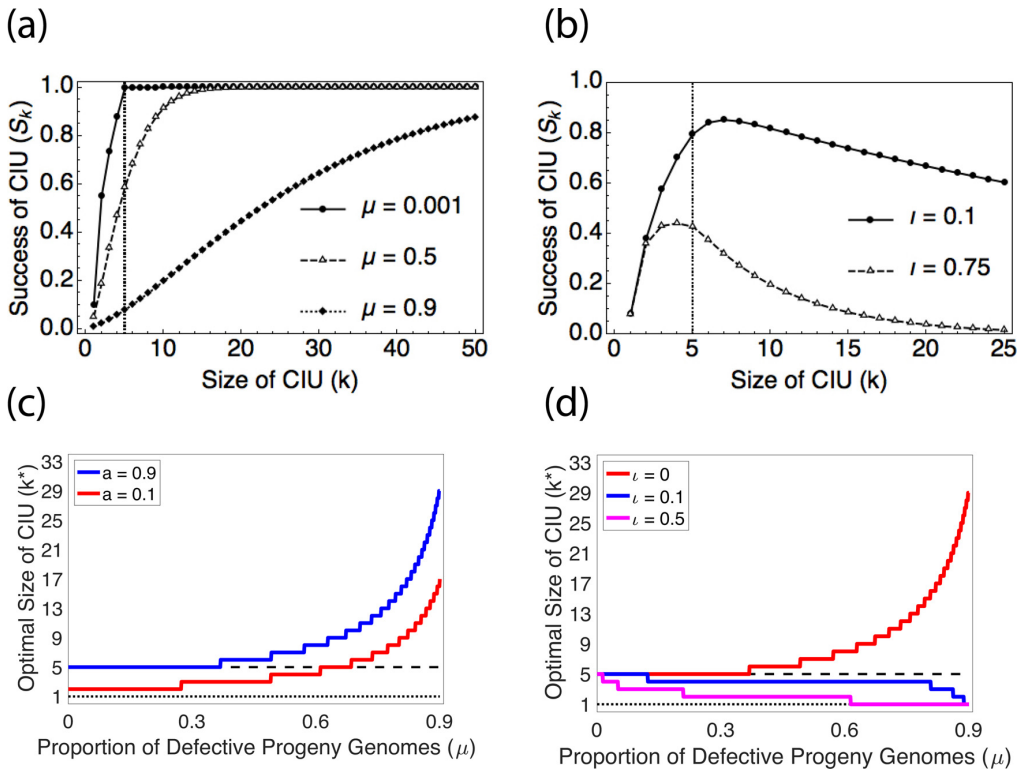


Fig. 3. Defective and interfering genomes and CIU evolution. (a) and (b) plot the relationship between the success of an infectious unit and its size when the proportion of genomes which are defective (μ) (a) or which are defective and interfering (ι) (b) varies. In (a), as μ increases, virions need to be larger to achieve the same success, because there is a larger chance that the genomes inside a virion are defective. However, when some defective genomes are interfering ($\iota > 0$) (b), there is a cost to larger CIUs, because larger CIUs have a greater chance of including an interfering genome. This cost reduces both the value of k at which success peaks and the success experienced by an infectious unit containing k genomes. In (b), 25% of viral progeny are defective ($\mu = 0.25$). (c) and (d) plot the optimal size of infectious unit (k^*) as the proportion of defective genomes (μ) increases. The dashed line plots k_t (the number of complete genomes that results in maximum infectious unit success) and the dotted line plots $k^* = 1$ (when CIUs are not favoured). In (c), a is the shape parameter for the diminishing returns success curve, with higher values indicating a more linear curve. As defective genomes become more prevalent, the optimal size of CIU increases and can reach values which are substantially higher than k_t. However, increases in μ by themselves cannot drive the evolution of CIUs from no CIUs. In (d), higher values of ι indicate that a higher proportion of defective genomes are interfering. As the proportion of interfering genomes (ι) increases, the optimal size of CIU decreases, and the likelihood that CIUs are favoured at all also decreases. Interfering genomes (ι) have a larger impact on CIU evolution when defective genomes are common (high μ).

underestimate the cost of defective genomes, since we do not allow defective genomes to contribute to group infection benefits, but we also do not allow defective genomes to build up over multiple generations. However, incorporating these additional complexities within our model would require a different model structure, since the model would need to track different classes of virus over multiple generations.

We found that defective genomes did not make CIUs more likely to evolve, but that they did influence the size of CIU that evolved (Fig. 3c; Appendix 4). When there was a very high likelihood of progeny genomes being defective (high μ), CIUs could be favoured to become very large, up to a second threshold, k_t', which is given by the value of k at which the likelihood of containing at least k_t complete genomes is approximately 1 (Fig. 3a).

Next, we investigated the consequences of interference by assuming that a fraction ι of defective genomes are also interfering. For mathematical simplicity, we assume that defective interfering genomes are completely interfering, such that a cell infected by at least one defective interfering genome produces only defective interfering genomes (Kirkwood and Bangham, 1994). This scenario represents an extreme case, but it allows us to capture the qualitative influence of defective interfering genomes while keeping our model tractable (Appendix 4) (Cole and Baltimore, 1973).

In contrast to our findings for defective genomes, we found that interfering genomes both: (i) made CIUs less likely to evolve, and (ii) decreased the size of the CIU which evolved when CIUs were favoured (Fig. 3d). This is because larger infectious units have a greater likelihood of incorporating a defective interfering genome, which then outcompetes the wild-type virus. In our model, the cost of defective interfering genomes depended on the product of the rate of defective mutant production (μ) and the chance that each defective genome is interfering (ι; Fig. 3c). Therefore, CIUs could only be favoured in viruses that had high rates of defective genome production if there was also a very low chance that these defective genomes were interfering (Fig. 3d).

3. Discussion

We tested the theoretical plausibility of three mechanisms that could favour the evolution of group dispersal in viruses inside collective infectious units (CIUs). Our models confirmed the hypothesis that if a greater number of complete viral genomes lead to more productive infections (group infection benefits), then CIUs could be favoured (Fig. 1). However, in contrast to predictions from verbal arguments, we found that: (1) the conditions which select for CIUs tend to favour smaller CIUs rather than larger ones (Table 1); (2) in the absence of group infection benefits, neither the production of defective viruses, nor

more efficient packaging of genomes, favour the evolution of CIUs (Figs. 2 & 3). Furthermore, if some fraction of progeny sequences are defective interfering genomes, then this disfavours the evolution of CIUs (Fig. 3). More generally, our results illustrate that by forcing assumptions to be made explicit, formal theoretical models can lead to different predictions than simple verbal arguments.

3.1. Predictions and data

Our 'group infection benefits' model suggested that CIUs should be favoured when cells infected with multiple copies of the same viral genome lead to more productive viral infections (Fig. 1). At least two experimental studies have directly investigated group infection benefits in different viruses, in vaccinia virus (VACV) and vesicular stomatitis virus (VSV) (Andreu-Moreno and Sanjuán, 2018; Stiefel et al., 2012). These studies suggest that at least two mechanisms can lead to group infection benefits: (i) if multiple genome copies are able to overwhelm cellular immunity responses; (ii) if stochastic events can prevent key viral gene products being expressed early in infection. One of these studies found a sigmoidal relationship between infectious unit size and infection success (threshold effects), and in both studies, the benefits of collective infection were large, increased relatively quickly, and saturated at relatively low numbers of genomes (k_t = approximately three genomes in Andreu-Moreno & Sanjuán; k_t = approximately eight genomes in Stiefel et al.). If these studies are representative, then group benefits to infection could provide a relatively general explanation for the evolution of collective infectious units (Fig. 1).

Our 'efficiency benefits' model predicts that polyploid virions may evolve more readily in isometric viruses than in rod-shaped viruses. This is because isometric viruses, which have approximately spherical virions, may transmit multiple genomes more efficiently than rod-shaped virions. Furthermore, since polyploid virions are derived from virus-encoded capsid proteins, this efficiency benefit could feasibly allow these viruses to evolve to produce more genome copies (Fig. 2). However, it is unclear whether this prediction is borne out by data, and there are a number of caveats that could complicate this prediction, including: smaller capsids may be more stable; smaller capsids may be required for direct cell-cell transmission; capsid size may have antigenic consequences; rod-shaped capsids may enlarge to incorporate extra genetic material more easily (Flint et al., 2015; Graw and Perelson, 2016; Hull, 2009; Ojosnegros et al., 2011).

To what extent can the models that we considered explain the pattern of CIUs in nature? While we found that CIUs could evolve to a range of different sizes under the models that we considered, in reality most CIUs are known to be large, containing many viral genomes. For example, baculovirus occlusion bodies are known to contain dozens of individual virions, while enterovirus vesicles are large enough to potentially contain hundreds of virions (Chen et al., 2015; Slack and Arif, 2007). We found that large CIUs such as these can evolve if: (i) group infection benefits increase slowly with the number of viral genomes (Fig. 1); (ii) group infection benefits require a high threshold number of genomes to accumulate (Fig. 1); (iii) viral progeny are frequently defective, but only rarely interfering (Fig. 3).

Our models have assumed that CIUs evolve due to the benefits of collective transmission. However, an alternative possibility is that collective transmission could be a by-product of selection for infectious units that are favoured for other reasons, such as increased infectivity or particle stability (Sanjuán, 2017; Santiana et al., 2018). In that case, collective infection would be a consequence, but not a cause, of the evolution of CIUs, and so different kinds of explanations would be required to explain when CIUs evolve in nature.

3.2. Further implications

It has been suggested that collective infectious units may evolve due to the benefits of trans-complementation between defective viral genomes (Andino and Domingo, 2015; Stiefel et al., 2012). While we did not model the possibility of trans-complementation, the potential for complementation to occur is greatest when defective mutation rates are high, since in that case a large fraction of the viral population could potentially benefit from trans-complementation. However, our model predicts that high rates of defective mutation may disfavour CIUs, by increasing the rate at which defective interfering (DI) genomes are produced (Fig. 3d). Consequently, our model predicts that the conditions which allow high levels of complementation to take place may also favour DIs, and so make it harder for CIUs to evolve.

Our model further suggests potential coevolutionary interactions between defective infectious (DI) genomes and CIUs. One possibility is that collective infectious units could favour the evolution of DIs by increasing the rate of cellular coinfection (Sanjuán, 2017). This could mean that in some cases, CIUs can evolve only temporarily, since they then favour the evolution of DIs and consequently create conditions under which CIUs are no longer favoured. An alternative possibility is that viruses which transmit collectively may have evolved mechanisms of resistance which prevent their exploitation by DIs, despite the higher level of coinfection that would otherwise favour DIs. One such mechanism of resistance could be if CIUs mainly transmit sister genomes, for example due to intracellular compartmentalisation. Alternatively, if CIUs were only used episodically, for example during transmission between hosts, then DIs may be unable to accumulate, since they would be selected against during within-host transmission, when CIUs were not used.

3.3. Optimality models in viruses

We have used an 'optimality' modelling approach that is common in behavioural ecology, but has been employed less often in microbiology (Davies et al., 2012; Bull and Wang, 2010). This has meant neglecting details that may be important in specific cases, in favour of focusing on the underlying selective forces that are likely to be important across a range of viruses (Grafen, 1991). Our aim is to examine broad trends of collective infection across multiple viral species, especially since each species may use a different specific mechanism to achieve collective infection. The next step requires empirical work, to: (1) verify experimentally whether the mechanisms that we consider are relevant in specific viral species (Andreu-Moreno and Sanjuán, 2018; Stiefel et al., 2012); and (2) examine whether our theory can explain the variation in when CIUs occur, across different viruses.

Supplementary data

MATLAB scripts for the numerical model and figures are available at https://osf.io/v3ru8/.

Acknowledgements

The authors would like to thank Ernesto Segredo-Otero, Esteban Domingo, Oliver Pybus, and Katrina Lythgoe for useful comments and discussion. A.L. was supported by funding from the Clarendon Fund, St. John's College, Oxford, and the BBSRC (grant number BB/M011224/1). R.S. was supported by ERC Consolidator Grant 724519 (ViS-a-ViS).

Appendix A

Appendix 1: Group infection benefits

In this section we define the two ways in which the success of a CIU can depend on the number of genomes inside it. We assume that there is a collective benefit to multiple infection, such that cells infected by more viral genomes lead to more productive viral infections. Since we assume that superinfection is rare, the cellular multiplicity of infection depends only on the number of genomes that initially infect a cell, i.e.

the size of the CIU (k). We assume that this collective benefit follows one of two possible relationships: diminishing returns, according to a decelerating function given by $s_D(k) = \rho + (1-\rho)(\frac{k-1}{k_t-1})^a$; or a threshold effect, according to a sigmoidal function given by $s_S(k) = \rho^2 + (1-\rho)(\rho + \frac{s_\sigma(k)-\rho}{1-\rho})$ where $s_\sigma(k) = (1-\rho)(\rho + \frac{1}{1+e^{b(\frac{k_i-k}{k_t-1})}})$. Here, ρ gives the relative success of a single virion, k is the number of genomes inside a CIU, k_t is the number of genomes inside a CIU at which success asymptotes, k_i is the value of k at which the inflection point of the sigmoidal curve is found, and a and b are the shape parameters for the two curves. These functions are chosen so that they always intercept the y-axis at ρ when $k = 1$ and at 1 when $k = k_t$. All functions are set to have a gradient of zero when $k > k_t$.

Appendix 2: When are CIUs Favoured?

In this section we calculate when collective infectious units are favoured when there is collective benefit that follows a law of diminishing returns. When the success of CIUs of size k is given by a decelerating function, the fitness of a viral genotype producing CIUs of size k is given by $\frac{n_g s_D(k)}{k}$ where $s_D(k) = \rho + (1-\rho)(\frac{k-1}{k_t-1})^a$ and n_g is a constant. We found that in this case there was always a single fitness peak with respect to size of CIU (k) and so CIUs evolved when the fitness of a CIU of size 2 is greater than the fitness of a CIU of size 1. We can find this by finding when w(2) > w(1), which evaluates to finding when $-\frac{1}{2}(\rho + (\rho-1)\left(\frac{1}{k_t-1}\right)^a) > 0$. This is plotted in Fig. 2c for a given value of k_t and is a decreasing function of a, ρ, and k_t, indicating that CIUs are more likely to be favoured when single virions are relatively less successful, when the benefits to additional genomes diminish rapidly, and when the benefits to collective infection can be achieved with a lower number of genomes (supplementary figure).

Appendix 3: Efficiency Benefits

In this section, we derive the potential efficiency benefits that can be achieved by larger CIUs, and we calculate how these can translate into additional viral genomes. First, we assume that a CIU is spherical, since this allows for the largest possible efficiency gain. In this case, the volume of the CIU is given by $\frac{4}{3}\pi r^3$ and its surface area is given by $4\pi r^2$. We assume that the number of genomes that a CIU can carry depends linearly on its volume, and that the number of constituent units that are required to build the CIU depends linearly on its surface area. We are therefore interested in the scaling relationship between the volume and surface area of the CIU; to increase the volume of the CIU by a factor of k, how much must its surface area increase?

We know that the volume of a CIU of size k is equal to k times the volume of a CIU of size 1. We can use this relationship to obtain an expression for the radius of a CIU of size k, and consequently calculate the corresponding surface area as follows. Firstly, we have that $\frac{4}{3}\pi {r_k}^3 = k\frac{4}{3}\pi {r_1}^3$, and so by rearranging, the radius of a CIU of volume k is $r_k = r_1\sqrt[3]{k}$. By substitution, the corresponding surface area of a CIU of volume k is $4\pi(r_1\sqrt[3]{k})^2$. We are interested in the number of genomes which can be transmitted per unit required for constructing the CIU ('structural units', such as capsid proteins if the CIU is a virion). This is given by the number of genomes transmitted per CIU multiplied by the number of CIUs which can be produced from a given amount of structural unit, which we can write as $k\frac{n_p}{c_p 4\pi(r_1\sqrt[3]{k})^2}$ where n_p is a constant giving the number of structural units available, and c_p is a constant giving the amount of surface area of CIU that can be produced by one structural unit. We now obtain the relative efficiency advantage by dividing the number of genomes transmitted with a constant amount of structural unit available for a genotype producing CIUs of size k by the same expression when $k = 1$, which gives us our scaling relationship $k\frac{n_p}{c_p 4\pi(r_1\sqrt[3]{k})^2} / \frac{n_p}{c_p 4\pi {r_1}^2} = \sqrt[3]{k}$.

To translate the efficiency benefit of a larger CIU into viral genome availability, we are interested in the relative number of genomes that a virus with CIUs of size k can produce. We first assume that the structural units required to build a CIU are virus-derived and depend on the viral genome being transcribed. We further assume that there is a linear trade-off between transcriptional events that produce viral gene products, resulting in structural units, and transcriptional events that replicate the viral genome, resulting in viral genome copies. Therefore, we assume that a reduced requirement for viral-derived structural units can result in an increased number of viral genome copies (Chao and Elena, 2017). Since we are interested in the relative viral genome production, we first assume that viruses are adapted to be efficient such that when they have CIUs of size 1 ($k = 1$), genomes and CIU structural units are produced in the ratio α:1-α, and that at this ratio, the right number of genomes are produced to fill every CIU constructed. We now assume that when $k > 1$, a larger number of genomes can be packaged per structural unit, according to the ratio $\sqrt[3]{k}$ derived above. With this increased efficiency of using structural proteins, the optimal ratio of genome:structural protein production now deviates from α:1-α to become $\frac{\alpha\sqrt[3]{k}}{\alpha\sqrt[3]{k}+(1-\alpha)}$: $\frac{1-\alpha}{\alpha\sqrt[3]{k}+(1-\alpha)}$ where we divide both sides of the ratio by the total amount of structural protein production and genome production, since we assume that this remains constant. Assuming that viruses with CIUs of size k quickly evolve to an optimal ratio of genome:structural protein production, we can express the relative number of genomes available to a virus with CIUs of size k by dividing the left hand side of this ratio for a general value of k by α, which yields $\frac{\sqrt[3]{k}}{1-\alpha+\alpha\sqrt[3]{k}}$. This relationship varies between 0 when $\alpha = 1$ and $\sqrt[3]{k}$ when $\alpha = 0$. This is because when α is low, relatively fewer genome copies are produced relative to viral gene products when CIUs are small, and so there is a larger potential gain in the number of genomes produced by viruses with larger CIUs. The relationship $\frac{\sqrt[3]{k}}{1-\alpha+\alpha\sqrt[3]{k}}$ is plotted in Fig. 2a.

Appendix 4: Defective and Defective Interfering Genomes

In this section, we calculate the impact of defective and defective interfering viral genomes on the optimal size of CIU that evolves. First, we calculate the impact of defective genomes. We assume that a fraction μ of progeny viral genomes are defective, and are not replicated inside an infected host cell and also do not contribute to the success of larger virions. We also assume that these defective genomes are just as likely as complete genomes to be incorporated into a CIU, and so the distribution of genomes inside CIUs is well described by a Binomial distribution, where the number of trials is the size of CIU and 'successes' represent incorporation of a complete genome, which happens with likelihood 1-μ. The success of a CIU is now given by the sum of the product of the likelihood of each possible combinations of complete and defective genomes (each state c where $c \in \{1...k\}$), and the success of each state c: $\sum_{c=1}^{k}(\frac{k}{c})(1-\mu)^c \mu^{k-c} s_D(c)$ where k is the total number of genomes (both complete and defective) inside a CIU, c is the number of complete genomes inside a CIU, μ is the likelihood that a progeny genome is defective, and SD(c) describes the success of a virion with c complete genomes. This success function is plotted in Fig. 3a and leads to a new value of k at which success asymptotes, k_t', which we can find by finding a value of k for which $\sum_{c=1}^{k_t-1}(\frac{k}{c})(1-\mu)^c \mu^{k-c} s_D(c) \approx 0$. k_t' can reach very large values when μ is high, indicating that CIUs may evolve to become very large if defective mutations are very common.

We next incorporate the possibility that a fraction ι of the defective genomes are also interfering. We assume that interference is total such that any infectious unit that contains at least one interfering genome produces no wild-type genomes upon infection of a new cell. We therefore weight the likelihood of each infection state c by the likelihood that none of the defective genomes are interfering, and so our new expression for the success of a CIU of size k becomes $\sum_{c=1}^{k}(\frac{k}{c})(1-\mu)^c \mu^{k-c}(1-\iota)^{k-c} s_D(c)$, which we plot in Fig. 3b.

References

Altan-Bonnet, N., Chen, Y.-H., 2015. Intercellular transmission of viral populations with vesicles. J. Virol. 89/24, 12242–12244. https://doi.org/10.1128/JVI.01452-15.

Andino, R., Domingo, E., 2015. 'Viral quasispecies'. Virology 479–480, 46–51. https://doi.org/10.1016/j.virol.2015.03.022.

Andreu-Moreno, I., Sanjuán, R., 2018. 'Collective infection of cells by viral aggregates promotes early viral proliferation and reveals a cellular-level allee effect'. Curr. Biol. https://doi.org/10.1016/j.cub.2018.08.028.

Bald, J.G., Briggs, G.E., 1937. Aggregation of virus particles. Nature 140/3533, 111. https://doi.org/10.1038/140111a0.

Bordería, A.V., Isakov, O., Moratorio, G., Henningsson, R., Agüera-González, S., Organtini, L., Gnädig, N.F., et al., 2015. Group selection and contribution of minority variants during virus adaptation determines virus fitness and phenotype. PLoS Pathog. 11/5, e1004838. https://doi.org/10.1371/journal.ppat.1004838.

Borges, A.L., Zhang, J.Y., Rollins, M.F., Osuna, B.A., Wiedenheft, B., Bondy-Denomy, J., 2018. Bacteriophage cooperation suppresses CRISPR-Cas3 and Cas9 immunity. Cell 174/4, 917–925. https://doi.org/10.1016/j.cell.2018.06.013. e10.

Bull, J.J., Wang, I.-N., 2010. Optimality models in the age of experimental evolution and genomics: experimental evolution and genomics. J. Evol. Biol. 23/9, 1820–1838. https://doi.org/10.1111/j.1420-9101.2010.02054.x.

Chao, L., Elena, S.F., 2017. Nonlinear trade-offs allow the cooperation game to evolve from Prisoner's Dilemma to Snowdrift. Proc. R. Soc. B 284/1854https://doi.org/10.1098/rspb.2017.0228. 20170228.

Chen, Y.-H., Du, W., Hagemeijer, M.C., Takvorian, P.M., Pau, C., Cali, A., Brantner, C.A., et al., 2015. Phosphatidylserine vesicles enable efficient en bloc transmission of enteroviruses. Cell 160/4, 619–630. https://doi.org/10.1016/j.cell.2015.01.032.

Cole, C.N., Baltimore, D., 1973. Defective interfering particles of poliovirus: III. Interference and enrichment. J. Mol. Biol. 76/3, 345–361. https://doi.org/10.1016/0022-2836(73)90509-3.

Cuevas, J.M., Durán-Moreno, M., Sanjuán, R., 2017. Multi-virion infectious units arise from free viral particles in an enveloped virus. Nat. Microbiol. 2/7, 17078. https://doi.org/10.1038/nmicrobiol.2017.78.

Dahlberg, J.E., Simon, E.H., 1969. Physical and genetic studies of Newcastle disease virus: evidence for multiploid particles. Virology 38/4, 666–678. https://doi.org/10.1016/0042-6822(69)90185-8.

Davies, N.B., Krebs, J.R., West, S.A., 2012. An Introduction to Behavioural Ecology. John Wiley & Sons.

Díaz-Muñoz, S.L., Sanjuán, R., West, S.A., 2017. Sociovirology: conflict, cooperation, and communication among viruses. Cell Host Microbe 22/4, 437–441. https://doi.org/10.1016/j.chom.2017.09.012.

Erickson, A.K., Jesudhasan, P.R., Mayer, M.J., Narbad, A., Winter, S.E., Pfeiffer, J.K., 2018. Bacteria Facilitate enteric virus co-infection of mammalian cells and promote genetic recombination. Cell Host Microbe 23/1, 77–88. https://doi.org/10.1016/j.chom.2017.11.007. e5.

Flint, J., Racaniello, V.R., Rall, G.F., Skalka, A.M., 2015. Principles of Virology, Fourth Edition, Bundle. American Society of Microbiologyhttps://doi.org/10.1128/9781555819521.

Godfray, H.C.J., Partridge, L., Harvey, P.H., 1991. Clutch size. Annu. Rev. Ecol. Syst. 22, 409–429.

Grafen, A., 1991. 'Modelling in Behavioural Ecology'. Behavioural Ecology, 3rd ed. Blackwell Scientific Publications, Oxford, pp. 5–31.

Graw, F., Perelson, A.S., 2016. Modeling viral spread. Annu. Rev. Virol. 3/1, 555–572. https://doi.org/10.1146/annurev-virology-110615-042249.

Guo, F., Li, S., Caglar, M.U., Mao, Z., Liu, W., Woodman, A., Arnold, J.J., et al., 2017. Single-Cell virology: on-chip investigation of viral infection dynamics. Cell Rep. 21/6, 1692–1704. https://doi.org/10.1016/j.celrep.2017.10.051.

Hamilton, W.D., 1971. Geometry for the selfish herd. J. Theor. Biol. 31/2, 295–311. https://doi.org/10.1016/0022-5193(71)90189-5.

Hosaka, Y., Kitano, H., Ikeguchi, S., 1966. Studies on the pleomorphism of HVJ virions. Virology 29/2, 205–221. https://doi.org/10.1016/0042-6822(66)90027-4.

Huang, A.S., Baltimore, D., 1970. Defective viral particles and viral disease processes. Nature 226/5243, 325–327. https://doi.org/10.1038/226325a0.

Hull, R., 2009. Comparative Plant Virology, 2nd ed. Elsevier Academic Press, Amsterdam; Boston.

Jaworski, E., Routh, A., 2017. Parallel ClickSeq and Nanopore sequencing elucidates the rapid evolution of defective-interfering RNAs in Flock House virus. PLoS Pathog. 13/5, e1006365. https://doi.org/10.1371/journal.ppat.1006365.

Kirkwood, T.B., Bangham, C.R., 1994. Cycles, chaos, and evolution in virus cultures: a model of defective interfering particles. Proc. Natl. Acad. Sci. 91/18, 8685–8689.

Lack, D., 1947. The significance of clutch-size. Ibis 89/2, 302–352. https://doi.org/10.1111/j.1474-919X.1947.tb04155.x.

Landsberger, M., Gandon, S., Meaden, S., Rollie, C., Chevallereau, A., Chabas, H., Buckling, A., et al., 2018. Anti-CRISPR phages cooperate to overcome CRISPR-Cas immunity. Cell 174/4, 908–916. https://doi.org/10.1016/j.cell.2018.05.058. e12.

Leeks, A., Segredo-Otero, E.A., Sanjuán, R., West, S.A., 2018. Beneficial coinfection can promote within-host viral diversity. Virus Evol. 4/2. https://doi.org/10.1093/ve/vey028.

Luque, D., Rivas, G., Alfonso, C., Carrascosa, J.L., Rodríguez, J.F., Castón, J.R., 2009. Infectious bursal disease virus is an icosahedral polyploid dsRNA virus. Proc. Natl. Acad. Sci. 106/7, 2148–2152. https://doi.org/10.1073/pnas.0808498106.

Manzoni, T.B., López, C.B., 2018. Defective (interfering) viral genomes re-explored: impact on antiviral immunity and virus persistence. Future Virol. https://doi.org/10.2217/fvl-2018-0021.

Maynard Smith, J., Price, G.R., 1973. The logic of animal conflict. Nature 246/5427, 15–18. https://doi.org/10.1038/246015a0.

McCrone, J.T., Lauring, A.S., 2018. Genetic bottlenecks in intraspecies virus transmission. Curr. Opin. Virol. 28, 20–25. https://doi.org/10.1016/j.coviro.2017.10.008.

Ojosnegros, S., García-Arriaza, J., Escarmís, C., Manrubia, S.C., Perales, C., Arias, A., Mateu, M.G., et al., 2011. Viral genome segmentation can result from a trade-off between genetic content and particle stability. PLoS Genet. 7/3, e1001344. https://doi.org/10.1371/journal.pgen.1001344.

Rager, M., Vongpunsawad, S., Duprex, W.P., Cattaneo, R., 2002. Polyploid measles virus with hexameric genome length. EMBO J. 21/10, 2364–2372. https://doi.org/10.1093/emboj/21.10.2364.

Rezelj, V.V., Levi, L.I., Vignuzzi, M., 2018. The defective component of viral populations. Curr. Opin. Virol. 33, 74–80. https://doi.org/10.1016/j.coviro.2018.07.014.

Sanjuán, R., 2017. Collective infectious units in viruses. Trends Microbiol. 25/5, 402–412. https://doi.org/10.1016/j.tim.2017.02.003.

Sanjuán, R., 2018. Collective properties of viral infectivity. Curr. Opin. Virol. 33, 1–6. https://doi.org/10.1016/j.coviro.2018.06.001.

Santiana, M., Ghosh, S., Ho, B.A., Rajasekaran, V., Du, W.-L., Mutsafi, Y., De Jésus-Diaz, D.A., et al., 2018. Vesicle-cloaked virus clusters are optimal units for inter-organismal viral transmission. Cell Host Microbe 24/2, 208–220. https://doi.org/10.1016/j.chom.2018.07.006. e8.

Slack, J., Arif, B.M., 2007. The baculoviruses occlusion-derived virus: virion structure and function. Adv. Virus Res. 69, 99–165. https://doi.org/10.1016/S0065-3527(06)69003-9. Elsevier.

Stiefel, P., Schmidt, F.I., Dörig, P., Behr, P., Zambelli, T., Vorholt, J.A., Mercer, J., 2012. Cooperative vaccinia infection demonstrated at the single-cell level using FluidFM. Nano Lett. 12/8, 4219–4227. https://doi.org/10.1021/nl3018109.

Tanner, E.J., Liu, H., Oberste, M.S., Pallansch, M., Collett, M.S., Kirkegaard, K., 2014. Dominant drug targets suppress the emergence of antiviral resistance. eLife 3, e03830. https://doi.org/10.7554/eLife.03830.

Turner, P.E., Chao, L., 1999. Prisoner's dilemma in an RNA virus. Nature 398/6726, 441–443. https://doi.org/10.1038/18913.

Vignuzzi, M., Stone, J.K., Arnold, J.J., Cameron, C.E., Andino, R., 2006. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. Nature 439/7074, 344–348. https://doi.org/10.1038/nature04388.

Xue, K.S., Hooper, K.A., Ollodart, A.R., Dingens, A.S., Bloom, J.D., 2016. Cooperation between distinct viral variants promotes growth of H3N2 influenza in cell culture. eLife 5, e13974. https://doi.org/10.7554/eLife.13974.

5

Transmission, relatedness, and the evolution of cooperative symbionts

Received: 21 January 2019 | Revised: 15 June 2019 | Accepted: 24 June 2019

DOI: 10.1111/jeb.13505

RESEARCH PAPER

Journal of Evolutionary Biology eseb WILEY

Transmission, relatedness, and the evolution of cooperative symbionts

Asher Leeks[1] | Miguel dos Santos[2] | Stuart A. West[1]

[1]Department of Zoology, University of Oxford, Oxford, UK

[2]Department of Social Psychology and Social Neuroscience, University of Bern, Bern, UK

Correspondence
Asher Leeks, Department of Zoology, Zoology Research and Administration Building, 11a Mansfield Road, OX1 3SZ Oxford, UK.
Email: asherleeks@gmail.com

Funding information
Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, Grant/Award Number: P2LAP3-158669; Biotechnology and Biological Sciences Research Council, Grant/Award Number: BB/M011224/1

Abstract

Cooperative interactions between species, termed mutualisms, play a key role in shaping natural ecosystems, economically important agricultural systems, and in influencing human health. Across different mutualisms, there is significant variation in the benefit that hosts receive from their symbionts. Empirical data suggest that transmission mode can help explain this variation: vertical transmission, where symbionts infect their host's offspring, leads to symbionts that provide greater benefits to their hosts than horizontal transmission, where symbionts leave their host and infect other hosts in the population. However, two different theoretical explanations have been given for this pattern: firstly, vertical transmission aligns the fitness interests of hosts and their symbionts; secondly, vertical transmission leads to increased relatedness between symbionts sharing a host, favouring cooperation between symbionts. We used a combination of analytical models and dynamic simulations to tease these factors apart, in order to compare their separate influences and see how they interact. We found that relatedness between symbionts sharing a host, rather than transmission mode per se, was the most important factor driving symbiont cooperation. Transmission mode mattered mainly because it determined relatedness. We also found evolutionary branching throughout much of our simulation, suggesting that a combination of transmission mode and multiplicity of infections could lead to the stable coexistence of different symbiont strategies.

KEYWORDS

cooperation, evolutionary branching, kin selection, mutualism, symbiosis

1 | INTRODUCTION

There is considerable variation in the benefit that hosts gain from their symbionts. In some cases, hosts are completely dependent upon their symbionts. For example, aphids cannot survive or reproduce without *Buchnera* symbionts, which provide essential amino acids (Buchner, 1965; Douglas, 1998). In other cases, symbionts appear to provide relatively minor benefits. For example, the removal of *Chlorella* symbionts from *Paramecium bursaria* leads to just a reduction in growth rate, and only under certain conditions (Karakashian, 1963; Lowe et al., 2016). Empirical studies have suggested that the way in which symbionts are transmitted between hosts plays an important role in explaining this variation (Bull et al., 1991; Bull & Molineux, 1992; Herre, 1995; Messenger et al., 1999; Sachs & Wilcox, 2006; Fisher et al., 2017). Specifically, that vertical transmission, where hosts transmit symbionts to their offspring, selects for more cooperative symbionts than horizontal transmission, where symbionts can leave their host and be transmitted to other individuals in the population.

Symbionts which are more cooperative could in turn provide greater benefits to their hosts, by investing more of their resources into functions which benefit their hosts or by refraining from overexploiting their hosts' resources (Frank, 1994, 1996).

Two different mechanisms have been given for why the mode of symbiont transmission matters (Frank, 1996). One mechanism is that if symbiont offspring are likely to be transmitted to host offspring, then symbionts benefit when the host has more offspring (Ewald, 1987; Yamamura, 1993, 1996; Ferdy & Godelle, 2005). In this "transmission" scenario, it is vertical transmission per se that selects for higher levels of symbiont cooperation, through aligning the fitness interests of hosts and symbionts—vertical transmission makes symbionts more dependent upon their hosts. The other mechanism is that the transmission route determines the genetic diversity or relatedness between the symbionts and that this determines selection for cooperation (Hamilton, 1964; Frank, 1994, 1996; Herre et al., 1999; West et al., 2002; Foster & Wenseleers, 2006). Greater horizontal transmission will lead to a lower relatedness between symbionts. As relatedness between symbionts goes down, this can favour symbionts who avoided the cost of helping their hosts, but could still benefit from the benefits provided to the hosts by other symbionts. In this "relatedness" scenario, transmission mode matters, but it does so through its influence on relatedness—vertical transmission reduces conflict between symbionts.

Both of these mechanisms, "transmission" and "relatedness", could operate, and both their relative importance and the extent to which one influences the other remain unclear. The empirical observation that vertically transmitted symbionts provide greater benefits to their hosts could be explained by either mechanism, or by both acting simultaneously. Theoretical studies tend to make simplifying assumptions that allow them to focus on just one of these mechanisms (Frank, 1996). For example, some of the studies that emphasize transmission mode assume that hosts can only be infected by one strain of symbiont at a time, ignoring the possibility for conflict between symbionts within a host (Yamamura, 1993, 1996). Similarly, models that examine the influence of variable relatedness do not usually explicitly model horizontal and vertical transmission (Frank, 1994, 2010). In nature, both mechanisms are likely to occur, and we have a poor understanding of the consequences. For example, would they have distinct and different influences, or would they interact; would one drive the other, or would one tend to dominate?

We use a three-pronged theoretical approach to investigate how these different mechanisms could interact, and their relative importance (Frank, 1996). We first build an analytical model of a specified symbiont life cycle in which we can tease apart the separate causal influences of relatedness and transmission mode. This allows us to test which mechanism plays the larger causal role in the evolution of cooperation. Then, by expressing relatedness in terms of symbiont transmission mode and bottlenecking between symbiont generations ("closing" the model), we allow transmission mode to influence relatedness. This allows us to partition the influence of transmission mode per se, and via its effect on relatedness (Cooper et al., 2018). Finally, we test the robustness of our conclusions with an individual-based simulation. This simulation allows us to relax several assumptions, including that mutations are of small size, and that the trait value for cooperation does not influence relatedness. Our simulation also allows us to investigate whether evolutionary branching can occur, as has been observed in the early stages of experimentally evolved mutualisms (Harcombe et al., 2018).

2 | MODELS AND RESULTS

2.1 | Assumptions and model life cycle

We assume a mutualism in which symbionts live inside hosts and potentially provide them with some benefit. We assume that the symbionts cannot survive long enough to reproduce outside the hosts, and so they are obligately dependent on the hosts. We assume that there is an infinite population of hosts with nonoverlapping generations and that there is no host population structure.

We assume that the cooperative symbiont trait x denotes the amount of resources contributed towards a service which benefits the host, but which does not directly benefit the symbiont. For example, this trait could be the production of a key nutrient that the host needs. We assume that hosts with more cooperative symbionts are more likely to survive to reproductive maturity and are more likely to produce more offspring after reaching reproductive maturity. Therefore, we assume that symbiont cooperation can benefit both host survival and host fecundity, according to the functions $s(x_g)$ and $f(x_g)$, respectively, where x_g refers to the mean investment into cooperation of all of the symbionts inside a focal host. We use mean, and not total, symbiont investment into cooperation, for the sake of simplicity, and to be consistent with previous work (Frank, 1994, 1996). We also assume that this trait is costly to the symbiont, by assuming that a focal symbiont's growth rate inside a host depends negatively on its investment into cooperation, according to the expression $\frac{1-x_i}{1-x_g}$, where x_i is a focal symbiont's investment into cooperation.

We assume that a symbiont can potentially transmit offspring to future generations via two routes, vertical or horizontal: vertical transmission occurs when a symbiont's offspring remain in their host and are passed on to the host's offspring; horizontal transmission is when a symbiont's offspring can infect the offspring of any host in the population. We assume that increased host survival increases the transmission opportunities for horizontally transmitting symbionts, and so we weight the horizontal component of symbiont fitness by a focal symbiont's host's relative survival, $\frac{s(x_g)}{s(\bar{x})}$, where $\bar{x}$ is the mean level of symbiont cooperation in the population as a whole. We assume that both host survival and host fecundity per unit time increase the transmission of vertically transmitting symbionts, and so we weight the vertical component of symbiont fitness by $\frac{s(x_g)}{s(\bar{x})}\frac{f(x_g)}{f(\bar{x})}$.

Finally, we use a parameter λ to capture the relative likelihood of horizontal (λ) compared to vertical (1- λ) transmission. λ could be influenced by a number of different biological factors, including if hosts are more likely to reject symbionts from one route than the other, or if one mode of transmission involves higher symbiont mortality. The fitness of a focal symbiont is then:

$$W=(1-\lambda)\frac{(1-x_i)}{(1-x_g)}\frac{s(x_g)}{s(\bar{x})}\frac{f(x_g)}{f(\bar{x})}+\lambda\frac{(1-x_i)}{(1-x_g)}\frac{s(x_g)}{s(\bar{x})}. \quad (1)$$

This fitness equation sets up a trade-off similar to other models of cooperative traits (Frank, 1994, 2010). Figures were produced using Wolfram Mathematica 11.3 (Harrower & Brewer, 2003; Wang, 2016).

2.2 | Equilibrium analysis

We are interested in the level of investment into cooperation (x^*) which, if adopted by all symbionts in the population, could not be beaten by any alternative value of x, which is termed an evolutionarily stable strategy (ESS). We used a neighbour-modulated fitness approach to obtain the inclusive fitness effect, Δ_{IF}, of small changes in the trait value for cooperation on the inclusive fitness of a focal individual, assuming the limit of weak selection (Taylor & Frank, 1996):

$$\Delta IF=\frac{\partial W}{\partial x_i}+R\frac{\partial W}{\partial x_g} \quad (2a)$$

We solved $\triangle IF = 0$ for x^*, evaluating all derivatives at $x_i=x_g=\bar{x}=x^*$ (Maynard Smith & Price, 1973). To allow for a wide range of relationships between symbiont cooperation and host survival or fecundity, we assume that $s(x_g)=x_g^s$ and $f(x_g)=x_g^f$, where $s > 0$ and $f > 0$, and so arrive at:

$$\Delta IF_{x_i=x_g=\bar{x}=x^*}=-\frac{1}{1-x^*}+R\left[\frac{s+f(1-\lambda)}{x^*}+\frac{1}{1-x^*}\right] \quad (2b)$$

where higher values of f or s indicate that host fecundity or survival respectively increases more quickly with symbiont cooperation, and R is the whole-group relatedness coefficient (Taylor & Frank, 1996; Pepper, 2000).

Equation 2b allows us to see the different effects of changes in cooperation (x^*) on the inclusive fitness of a focal individual. The first term in Equation 2b is the cost of cooperation (x^*), which reflects reduced symbiont competitiveness within a host. The second term in Equation 2b is the benefit of cooperation that goes to the other symbionts sharing the focal symbiont's host, weighted by the genetic relatedness between the focal symbiont and its neighbours (R). This benefit stems from the fact that more cooperative (higher x^*) groups of symbionts will have hosts that live longer (in a way that scales with s) and have more offspring (in a way that scales with f).

By taking the second derivative $\frac{\partial \Delta IF}{\partial x^{*2}}$, we find solutions which are local maxima, and hence candidate ESSs, over the relevant parameter space($0 \le R \le 1$, $0 \le \lambda \le 1$), which we denote x_0^* (Maynard Smith & Price, 1973; Taylor & Frank, 1996; Otto & Day, 2007; Lehmann & Rousset, 2014; Biernaskie & West, 2015):

$$x_0^*=\frac{R[f(1-\lambda)+s]}{R[f(1-\lambda)+s-1]+1}, \quad (3)$$

We found that both relatedness and transmission mode influenced the final level of cooperation in this model (Figure 1). Relatedness increases cooperation because it increases the extent to which the benefits of cooperation go to genetic relatives of the actor. This is reflected by an increased weighting of the second term in Equation 2b, resulting in a higher level of cooperation (x^*) when fitness is at equilibrium. Vertical transmission increases cooperation because higher levels of vertical transmission increase the extent to which host fecundity can benefit symbionts (Equation 1). This is reflected in Equation 2b by the fact that vertical transmission (lower λ) increases the $f(1-\lambda)$ component of the group symbiont benefit (second term of Equation 2b). These findings are consistent with previous work that looked just at transmission mode or just at relatedness (Yamamura, 1993; Frank, 1994).

2.3 | Transmission or relatedness: open model

At this stage, we are interested in asking two different questions of our model. The first question is whether relatedness or transmission mode plays the larger role in determining cooperation. To answer this question, we keep relatedness as an open parameter in our model, allowing us to examine the separate causal influences of relatedness (R) and transmission mode (λ). However, in reality, these factors are not independent, since transmission mode can determine relatedness (Taylor, 1992; Frank, 1996; Cooper *et al.*, 2018). We can capture this by "closing" the model and expressing relatedness in

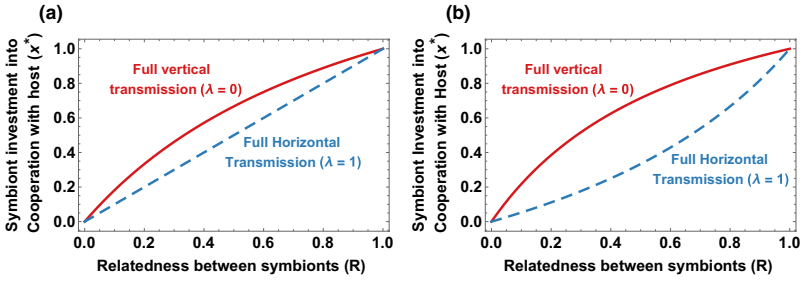


FIGURE 1 Both transmission mode and relatedness influenced the final level of cooperation that emerged (Equation 3). In (a), host survival and fecundity both increase in the same way with symbiont cooperation (s = f = 1). In (b), host fecundity increases more quickly with symbiont cooperation than host survival does ($s = 0.5$, $f = 2$)

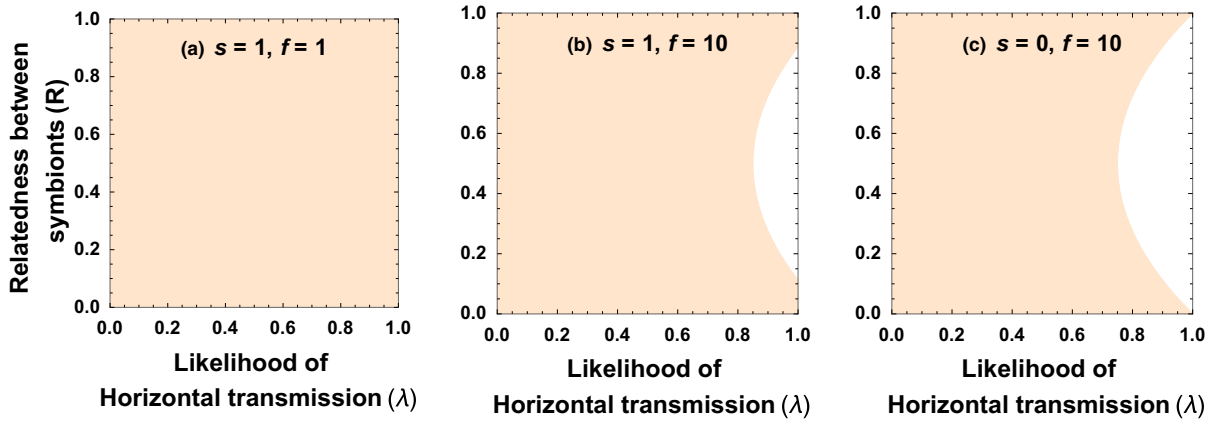


FIGURE 2 In the first analytical model (Equation 3), relatedness (R) usually had a larger influence on the final level of cooperation than transmission mode (λ) did. In the orange regions plotted, relatedness had a larger influence than transmission mode, whereas in the white regions, transmission mode had a larger influence than relatedness. Transmission mode only had a larger influence when transmission was mostly horizontal ($\lambda > 0.75$) and when host fecundity increased more rapidly with symbiont cooperation than host survival did ($f >> s$)

terms of demographic parameters. Closing the model allows us to ask our second question of why transmission mode influences cooperation: primarily through its direct influence on cooperation per se, or primarily through its influence on relatedness?

To start with, we keep relatedness as an open parameter. Both relatedness and transmission mode influence the equilibrium level of cooperation (Equation 3). For the parameters chosen in Figure 1, it appears that relatedness plays a larger role than transmission mode, in the sense that small changes to relatedness influence the equilibrium level of cooperation more than small changes in transmission mode do (Figure 1). To extend this comparison over all of the potential parameter space, we compared the marginal effect of changes in transmission mode (λ) or relatedness (R) on the equilibrium level of cooperation.

We calculated the marginal effects by taking the differential of the equilibrium level of cooperation with respect to either relatedness ($\frac{\partial x_0^*}{\partial R}$) or transmission mode ($\frac{\partial x_0^*}{\partial \lambda}$). The first of these differentials ($\frac{\partial x_0^*}{\partial R}$reflects the alignment of fitness interests between symbionts within a host—to what extent should more highly related groups of symbionts cooperate more? The second of these differentials ($\frac{\partial x_0^*}{\partial \lambda}$) reflects the alignment of fitness interests between a host and its symbionts—to what extent does increased vertical transmission favour a host's symbionts to cooperate more? By comparing the value of the two differentials, we can determine whether relatedness $\left(\left|\frac{\partial x_0^*}{\partial R}\right| > \left|\frac{\partial x_0^*}{\partial \lambda}\right|\right)$ or transmission mode $\left(\left|\frac{\partial x_0^*}{\partial R}\right| < \left|\frac{\partial x_0^*}{\partial \lambda}\right|\right)$ has a larger influence on the equilibrium level of cooperation.

In Appendix 1, we show that, for most of the possible parameter space, relatedness (R) plays a bigger role than transmission mode (λ) in determining the final level of cooperation (Figure 2). Specifically, transmission mode only plays a larger role if three conditions are all met: (a) horizontal transmission dominates ($\lambda > 0.75$); (b) host fecundity accelerates substantially faster with symbiont cooperation than host survival ($f > 4s$); and (c) relatedness is neither maximal nor minimal ($0 < R < 1$; Figure 2).

2.4 | Transmission and relatedness: closed model

Our next step is to "close" the model by expressing relatedness in terms of demographic parameters (Cooper *et al.*, 2018). We assume that hosts infected by symbionts horizontally are infected by k_h symbionts and that vertically infected hosts are infected by k_v symbionts. In Appendix 2, we show that whole-group relatedness can now be expressed as:

$$R = \frac{k_h(1-\lambda) + \lambda k_v}{k_h\left[1 + (k_v - 1)\lambda\right]}, \tag{4}$$

where k_h and k_v give the horizontal and vertical bottleneck sizes, respectively, and λ gives the fraction of host offspring that are infected horizontally.

Relatedness depends on the extent to which transmission is vertical or horizontal. Under full horizontal transmission ($\lambda = 1$), Equation 4 simplifies to $\frac{1}{k_h}$, whereas under full vertical transmission, Equation 4 simplifies to 1 (full relatedness). This occurs because horizontal transmission "resets" relatedness by enforcing complete mixing of unrelated symbionts, whereas vertical transmission allows relatedness to increase each generation, since symbionts interact only within a local group.

Next, we further simplify Equation 4 by assuming that horizontally and vertically transmitting symbionts experience the same bottleneck size ($k_h = k_v = k$) to arrive at:

$$R = \frac{1}{1 + \lambda(k-1)}. \tag{5}$$

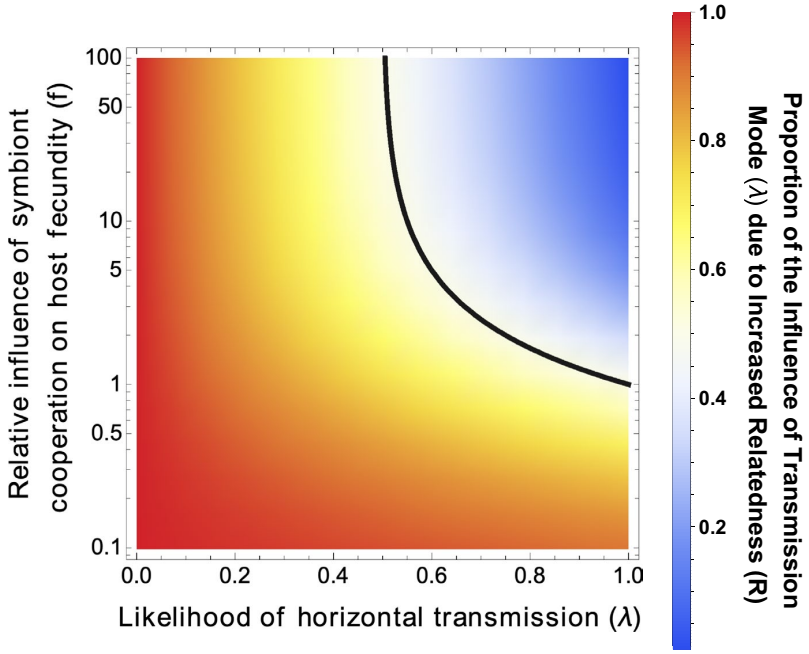


FIGURE 3 In the second analytical model (Equation 3), transmission mode influenced cooperation primarily through its influence on relatedness. Transmission mode always influenced cooperation more via *R* when transmission was mostly vertical (λ < 0.5) or when host survival increased with symbiont cooperation more quickly than host fecundity did (f < s). The dark line plots the point at which transmission mode influences cooperation equally through both routes. For this plot, s = 1

By substituting our expression for relatedness (Equation 5) into our expression for the equilibrium level of cooperation (x_0^*; Equation 3), we arrive at a new expression for the equilibrium level of cooperation, which we denote x_c^*:

$$x_c^* = \frac{f(1-\lambda)+s}{f(1-\lambda)+s+(k-1)\lambda}. \quad (6)$$

We then compared the extent to which transmission mode influences cooperation via its direct influence and via its influence on relatedness. To do this, we first calculated, as before, the marginal effect of changes in transmission mode on the equilibrium level of cooperation for the model with relatedness left open ($\frac{\partial x_0^*}{\partial \lambda}$). Then, we calculated the total effect of changes in transmission mode on the equilibrium level of cooperation, by taking the differential of the expression for equilibrium cooperation after the model has been closed ($\frac{\partial x_c^*}{\partial \lambda}$). These two partial derivatives represent, respectively, the influence of transmission mode via its direct influence and the total influence of transmission mode via both influences. We isolate the effect of transmission mode via its influence on *R* by subtracting the first partial derivative ($\frac{\partial x_0^*}{\partial \lambda}$) from the second ($\frac{\partial x_c^*}{\partial \lambda}$). By comparing these derivatives, we can then test whether transmission mode matters mostly because it aligns the interests of symbionts sharing a host (by increasing relatedness) or mostly by aligning the interests of symbionts and hosts. In Appendix 3, we show that transmission mode always had a larger influence via its influence on relatedness than via its direct influence $\left(\frac{\frac{\partial x_c^*}{\partial \lambda}-\frac{\partial x_0^*}{\partial \lambda}}{\frac{\partial x_0^*}{\partial \lambda}} > 1\right)$, unless: (i) symbiont cooperation increases host fecundity faster than it increases host survival (f > s); (ii) and transmission is mostly horizontal (λ > 0.5; Figure 3).

Our closed model highlights how focusing just on transmission mode could lead to misleading predictions about the level of cooperation. Equation 5 shows that if transmission is mostly vertical (low λ), then relatedness will always be high, because the λ(k-1) term will be small. However, if transmission is mostly horizontal (high λ), then relatedness can either be high or low, depending on the degree of bottlenecking (the value of k) (Equation 5). Consequently, if transmission is mostly horizontal, then focusing just on transmission mode erroneously predicts that a low level of cooperation will evolve, when in fact high levels of cooperation can sometimes evolve (Equation 6).

2.5 | Simulation

We next wrote an individual-based simulation in order to check whether our predicted equilibria were evolutionarily stable. Our simulation closely followed our analytical model life cycle (section 2.1), except that we specified the number of hosts and the frequency and size of mutations (Appendix 4). In the simulation, our transmission mode parameter λ is the likelihood that each new host receives symbionts horizontally (from the adult host population at large). Correspondingly, 1-λ gives the chance that each host receives symbionts vertically (from its parent).

Our simulation led to two different outcomes. In some simulation runs, the symbiont population remained at our predicted equilibrium level of cooperation, forming a monomorphic population. In these runs, the simulation results closely agreed with the analytical models (Figure 4).

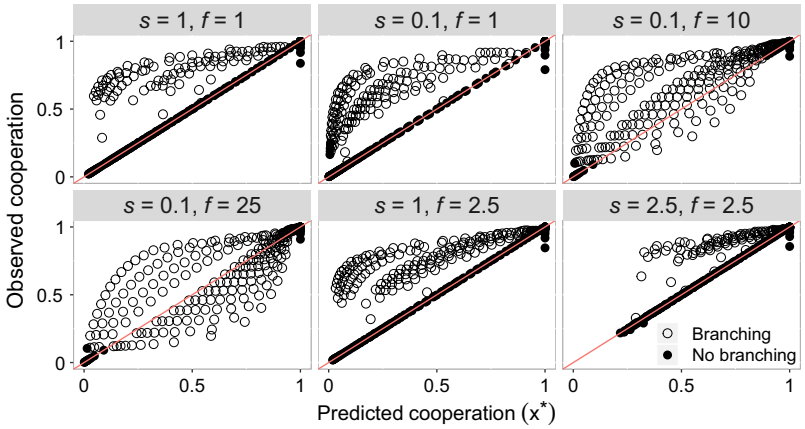


FIGURE 4 The mean level of cooperation in the symbiont population predicted by our analytical solution (red line) and observed in our simulation (circles). When evolutionary branching did not occur (closed circles), the simulation results closely match the analytical predictions. When evolutionary branching occurred (open circles), the simulation results diverged from the model predictions, and this generally led to a higher level of cooperation than predicted. Predictions are obtained using Equation 6

In other simulation runs, the symbiont population diverged to form a stable polymorphism between strains that cooperated to different degrees (evolutionary branching). In runs when branching occurred, the final mean level of cooperation was usually higher, but occasionally lower, than our predicted equilibrium (Figure 5). In these runs, the final level of cooperation correlated with both transmission mode and relatedness, and it was not possible to disentangle the causal influence of each.

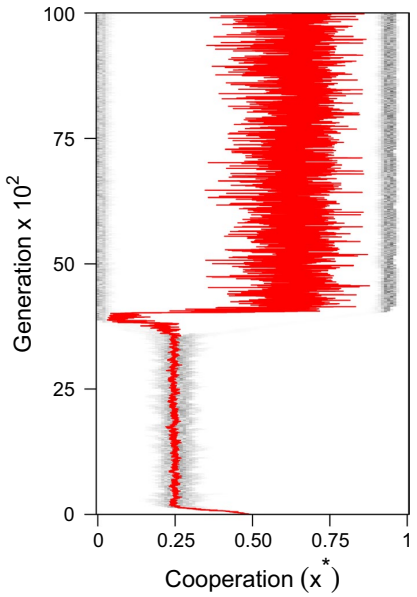


FIGURE 5 Evolutionary branching. An example simulation where the level of symbiont cooperation branched. In this plot, darker points reflect higher densities of symbionts at each cooperation strategy, and the red line gives the mean level of cooperation in the symbiont population as a whole. The level of cooperation initially approaches a monomorphic value of cooperation predicted by our analytical model at x^*~0.25, but then, between generations 2,500 and 5,000, it branches to a bimodal equilibrium, with some symbionts cooperating significantly more than others. In this simulation run, following the branching event, the mean level of symbiont cooperation (red line) is significantly higher than it was before. For this simulation run, $s(x) = x$ and $f(x) = x$

In the simulation runs when branching occurred, the symbiont population first reached the monomorphic equilibrium predicted by our analytical models, but then diverged to form a stable polymorphism between strains that cooperated to different degrees (Figure 5). In most runs, this resulted in a population of "super-defectors" that invested the minimum in cooperation. Additionally, in some runs, there were further branching events, leading to more than two populations of symbionts coexisting. When branching occurred, the resulting level of relatedness differed substantially from the relatedness that we predicted based on the demographic parameters (Equation 4; Figure 6). This indicates that in the simulations, unlike in the analytical models, the trait value for cooperation could influence relatedness. We suggest that this may be occur because less cooperative strains are more likely to be in mixed infections than more cooperative strains, since hosts infected only by cooperative strains are more likely to survive than those infected only by noncooperative strains. Consequently, positive feedback could drive more cooperative strains to cooperate more, and less cooperative strains to cooperate less. This feedback cannot occur in the analytical model, because we assume that the symbiont population is at a single equilibrium; however, it can occur in the simulation, where symbionts with very different values for cooperation can interact (Figure 5).

3 | DISCUSSION

We found, in both our analytical and simulation models, that the relatedness between symbionts in a host was a major determinant of the level at which symbionts cooperate with their hosts (Figures 1, 2 and 4). In contrast, while transmission mode was correlated with the level of symbiont cooperation, this was mainly through its influence on relatedness (Figure 3). Consequently, transmission mode can be a less useful predictor of the level of cooperation, because it is just one of a number of factors that determine relatedness—other factors include the degree of bottlenecking that occurs when symbionts infect new hosts.

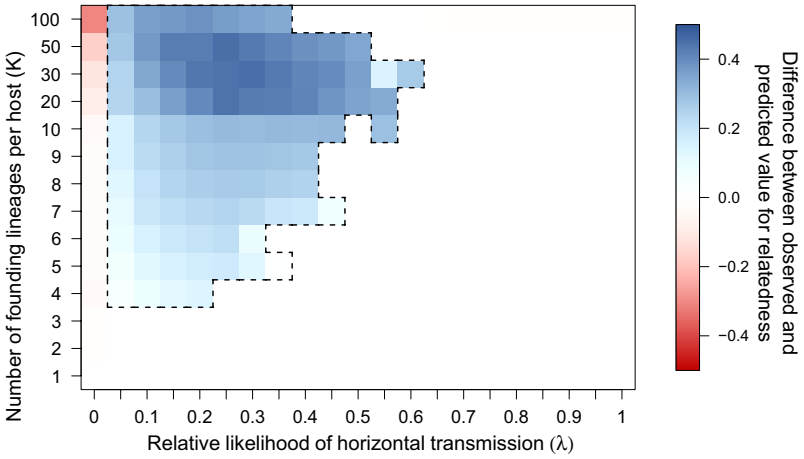


FIGURE 6 The distribution of evolutionary branching across the simulation. The region of parameter space in which evolutionary branching occurred is outlined by the black dotted line. When evolutionary branching occurred, the observed values of relatedness were significantly higher than the predicted values of relatedness. The red region reflects the fact that when there is no horizontal transmission, and bottlenecking is weak (high *k*), mutation–selection balance can keep noncooperative symbionts in the population transiently (Frank, 1994). For this figure, $s(x) = x$ and $f(x) = x$

Both experimental and across species comparative studies have suggested vertical transmission leads to symbionts that provide greater benefits to hosts (Sachs & Wilcox, 2006; Fisher *et al.*, 2017). Analogous patterns have been found in many parasitic systems, where vertical transmission commonly leads to reduced virulence in both experimental and comparative studies (Bull *et al.*, 1991; Herre, 1993; Messenger *et al.*, 1999; Stewart et al., 2005; Lambrechts & Scott, 2009). Our results suggest that the influence of transmission mode is primarily because of its influence on the relatedness between symbionts sharing a host (Figures 2 and 3). Although we have not modelled every possible scenario, and different life-history assumptions could lead to different results, we deliberately kept our model simple in order to focus on mechanisms which are likely to be of widespread importance, such as within-host competition for resources. Consequently, we expect our conclusions to be widely applicable (Herre, 1993; Frank, 1996; West & Buckling, 2003; Alizon et al., 2013; Speare *et al.*, 2018).

We found that evolutionary branching occurred across much of the parameter space in our simulations, leading to stable coexistence between two strains, which cooperate to different degrees (Figure 5). Evolutionary branching has been observed in game theory models in which there are saturating benefit and cost functions near the equilibrium, or where cooperation is linked with another trait under evolution, such as dispersal, as well as in models of parasite virulence (Nowak & May, 1994; Doebeli et al., 2004; El Mouden & Gardner, 2008; Wakano & Lehmann, 2014; Mullon, Keller, & Lehmann, 2016, 2018). Evolutionary branching has also been observed in the early stages of experimentally evolved mutualisms, resulting in variation in the extent to which members of one species cooperate with the other (Harcombe *et al.*, 2018). However, it is unclear whether this variation is likely to be sustained over evolutionary time periods, leading to variance in symbiont partner quality, or whether this variance will be eroded. This is because variation in the level of cooperation could select for hosts to preferentially reward cooperators and/or sanction noncooperators, as has been observed in a number of mutualisms (Noë & Hammerstein, 1995; Johnstone & Bshary, 2002; West et al., 2002; West *et al.*, 2002; Kiers et al., 2003; Jandér & Herre, 2010; Kiers *et al.*, 2011; Wyatt et al., 2014). The consequence of such rewarding and sanctions would select against less cooperative symbionts, reducing the variance in the level of cooperation (West *et al.*, 2002), which could reduce the likelihood that we observe coexistence in nature. Other explanations for the coexistence of symbionts which cooperate to different degrees include different symbiont genotypes adapted to different hosts (Bever et al., 2009; Gubry-Rangin et al., 2010; Gordon *et al.*, 2016).

To conclude, our results also emphasize the role of transmission route and relatedness in major evolutionary transitions. We predict that when symbionts are clonal ($R = 1$), they should cooperate at the highest level possible with their hosts ($x^*=1$). In this case, there is no conflict between symbionts, and the interests of the hosts and symbionts can be perfectly aligned with regard to how much the symbionts should cooperate (Bordenstein & Theis, 2015; Moran & Sloan, 2015). An alignment of interests between hosts and symbionts is one of the factors required for a major evolutionary transition to a higher level organism/individual (Maynard Smith & Szathmary, 1995; Gardner & Grafen, 2009; Bourke, 2011; West et al., 2015). Examples of such major transitions include the evolution of the eukaryotic cell, plastid endosymbiosis, and some obligate endosymbionts in insects (West *et al.*, 2015). Our results suggest that vertical transmission, combined with population bottlenecks, leading to clonal populations of symbionts within hosts, could play a key role in driving major transitions involving hosts and their symbionts. Furthermore, this is analogous to how clonality or monogamy can align interests and hence drive major transitions between members of the same species (Boomsma, 2007, 2009; Fisher et al., 2013; West *et al.*, 2015).

ORCID

Asher Leeks https://orcid.org/0000-0002-1175-7610
Miguel dos Santos https://orcid.org/0000-0002-2198-1560

REFERENCES

Alizon, S., de Roode, J. C., & Michalakis, Y. (2013). Multiple infections and the evolution of virulence. *Ecol. Lett.*, *16*, 556–567. https://doi.org/10.1111/ele.12076

Bever, J. D., Richardson, S. C., Lawrence, B. M., Holmes, J., & Watson, M. (2009). Preferential allocation to beneficial symbiont with spatial structure maintains mycorrhizal mutualism. *Ecol. Lett.*, *12*, 13–21. https://doi.org/10.1111/j.1461-0248.2008.01254.x

Biernaskie, J. M., & West, S. A. (2015). Cooperation, clumping and the evolution of multicellularity. *Proceedings of the Royal Society B: Biological Sciences*, *282*, 20151075.

Boomsma, J. J. (2007). Kin selection versus sexual selection: Why the ends do not meet. *Curr. Biol.*, *17*, R673–683. https://doi.org/10.1016/j.cub.2007.06.033

Boomsma, J. (2009). Lifetime monogamy and the evolution of eusociality. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*, 3191–3207. https://doi.org/10.1098/rstb.2009.0101

Bordenstein, S. R., & Theis, K. R. (2015). Host Biology in Light of the Microbiome: Ten Principles of Holobionts and Hologenomes. *PLoS Biol.*, *13*, e1002226. https://doi.org/10.1371/journal.pbio.1002226

Bourke, A. F. G. (2011). *Principles of social evolution*. Oxford, New York: Oxford University Press.

Buchner, P. (1965). *Endosymbiosis of animals with plant microorganisms*, Revised ed. New York, NY: Interscience Publishers / John Wiley.

Bull, J. J., & Molineux, I. J. (1992). Molecular Genetics of Adaptation in an Experimental Model of Cooperation. *Evolution*, *46*, 882–895. https://doi.org/10.1111/j.1558-5646.1992.tb00606.x

Bull, J. J., Molineux, I. J., & Rice, W. R. (1991). Selection of benevolence in a host-parasite system. *Evolution*, *45*(4), 875–882. https://doi.org/10.1111/j.1558-5646.1991.tb04356.x

Cooper, G. A., Levin, S. R., Wild, G., & West, S. A. (2018). Modeling relatedness and demography in social evolution. *Evolution Letters*, *2*, 260–271. https://doi.org/10.1002/evl3.69

Doebeli, M., Hauert, C., & Killingback, T. (2004). The evolutionary origin of cooperators and defectors. *Science*, *306*, 859–862. https://doi.org/10.1126/science.1101456

Douglas, A. E. (1998). Nutritional Interactions in Insect-Microbial Symbioses: Aphids and Their Symbiotic Bacteria Buchnera. *Annual Review Entomology*, *43*, 17–37.

Ewald, P. W. (1987). Transmission modes and evolution of the parasitism-mutualism continuuma. *Ann. N. Y. Acad. Sci.*, *503*, 295–306. https://doi.org/10.1111/j.1749-6632.1987.tb40616.x

Ferdy, J., & Godelle, B. (2005). Diversification of transmission modes and the evolution of mutualism. *Am. Nat.*, *166*, 613–627. https://doi.org/10.1086/491799

Fisher, R. M., Cornwallis, C. K., & West, S. A. (2013). Group formation, relatedness, and the evolution of multicellularity. *Curr. Biol.*, *23*, 1120–1125.

Fisher, R. M., Henry, L. M., Cornwallis, C. K., Kiers, E. T., & West, S. A. (2017). The evolution of host-symbiont dependence. *Nat. Commun.*, *8*, https://doi.org/10.1038/ncomms15973

Foster, K. R., & Wenseleers, T. (2006). A general model for the evolution of mutualisms. *J. Evol. Biol.*, *19*, 1283–1293. https://doi.org/10.1111/j.1420-9101.2005.01073.x

Frank, S. A. (1994). Kin selection and virulence in the evolution of protocells and parasites. *Proceedings of the Royal Society B: Biological Sciences*, *258*, 153–161.

Frank, S. A. (1996). Models of parasite virulence. *Q. Rev. Biol.*, *71*, 37–78. https://doi.org/10.1086/419267

Frank, S. A. (2010). A general model of the public goods dilemma. *J. Evol. Biol.*, *23*, 1245–1250. https://doi.org/10.1111/j.1420-9101.2010.01986.x

Gardner, A., & Grafen, A. (2009). Capturing the superorganism: A formal theory of group adaptation. *J. Evol. Biol.*, *22*, 659–671. https://doi.org/10.1111/j.1420-9101.2008.01681.x

Gordon, B. R., Klinger, C. R., Weese, D. J., Lau, J. A., Burke, P. V., Dentinger, B. T. M., & et al (2016). Decoupled genomic elements and the evolution of partner quality in nitrogen-fixing rhizobia. *Ecol. Evol.*, *6*, 1317–1327. https://doi.org/10.1002/ece3.1953

Gubry-Rangin, C., Garcia, M., & Béna, G. (2010). Partner choice in Medicago Truncatula-Sinorhizobium symbiosis. *Proceedings of the Royal Society B: Biological Sciences*, *277*, 1947–1951.

Hamilton, W. D. (1964). The genetical evolution of social behaviour. I. *J. Theor. Biol.*, *7*, 1–16. https://doi.org/10.1016/0022-5193(64)90038-4

Harcombe, W. R., Chacón, J. M., Adamowicz, E. M., Chubiz, L. M., & Marx, C. J. (2018). Evolution of bidirectional costly mutualism from byproduct consumption. *Proceedings of the National Academy of Sciences USA*, *115*(47), 12000–12004. https://doi.org/10.1073/pnas.1810949115

Harrower, M., & Brewer, C. A. (2003). ColorBrewer.org: An online tool for selecting colour schemes for maps. *The Cartographic Journal*, *40*, 27–37. https://doi.org/10.1179/000870403235002042

Herre, E. A. (1993). Population structure and the evolution of virulence in nematode parasites of fig wasps. *Science*, *259*, 1442–1445. https://doi.org/10.1126/science.259.5100.1442

Herre, E. A. (1995). Factors affecting the evolution of virulence: Nematode parasites of fig wasps as a case study. *Parasitology*, *111*, S179–S191. https://doi.org/10.1017/S0031182000075880

Herre, E. A., Knowlton, N., Mueller, U. G., & Rehner, S. A. (1999). The evolution of mutualisms: Exploring the paths between conflict and cooperation. *Trends Ecol. Evol.*, *14*, 49–53. https://doi.org/10.1016/S0169-5347(98)01529-8

Jandér, K. C., & Herre, E. A. (2010). Host sanctions and pollinator cheating in the fig tree–fig wasp mutualism. *Proceedings of the Royal Society B: Biological Sciences*, *277*, 1481–1488. https://doi.org/10.1098/rspb.2009.2157

Johnstone, R. A., & Bshary, R. (2002). From parasitism to mutualism: Partner control in asymmetric interactions. *Ecol. Lett.*, *5*, 634–639. https://doi.org/10.1046/j.1461-0248.2002.00358.x

Karakashian, S. J. (1963). Growth of paramecium bursaria as influenced by the presence of algal symbionts. *Physiol. Zool.*, *36*, 52–68. https://doi.org/10.1086/physzool.36.1.30152738

Kiers, E. T., Rousseau, R. A., West, S. A., & Denison, R. F. (2003). Host sanctions and the legume–rhizobium mutualism. *Nature*, *425*, 78–81. https://doi.org/10.1038/nature01931

Kiers, E. T., Duhamel, M., Beesetty, Y., Mensah, J. A., Franken, O., Verbruggen, E.,et al (2011). Reciprocal rewards stabilize cooperation in the Mycorrhizal symbiosis. *Science*, *333*, 880–882. https://doi.org/10.1126/science.1208473

Lambrechts, L., & Scott, T. W. (2009). Mode of transmission and the evolution of arbovirus virulence in mosquito vectors. *Proceedings of the Royal Society B: Biological Sciences*, *276*, 1369–1378. https://doi.org/10.1098/rspb.2008.1709

Lehmann, L., & Rousset, F. (2014). The genetical theory of social behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*, 20130357. https://doi.org/10.1098/rstb.2013.0357

Lowe, C. D., Minter, E. J., Cameron, D. D., & Brockhurst, M. A. (2016). Shining a light on exploitative host control in a photosynthetic endosymbiosis. *Curr. Biol.*, *26*, 207–211. https://doi.org/10.1016/j.cub.2015.11.052

Maynard Smith, J., & Price, G. R. (1973). The logic of animal conflict. *Nature*, *246*, 15–18. https://doi.org/10.1038/246015a0

Maynard Smith, J., & Szathmary, E. (1995). *The major transitions in evolution*. New York, NY: Oxford University Press.

Messenger, S. L., Molineux, I. J., & Bull, J. J. (1999). Virulence evolution in a virus obeys a trade-off. *Proc. R. Soc. Lond. B Biol. Sci.*, *266*, 397–404. https://doi.org/10.1098/rspb.1999.0651

Moran, N. A., & Sloan, D. B. (2015). The hologenome concept: Helpful or hollow? *PLoS Biol.*, *13*, e1002311. https://doi.org/10.1371/journal.pbio.1002311

El Mouden, C., & Gardner, A. (2008). Nice natives and mean migrants: The evolution of dispersal-dependent social behaviour in viscous populations. *J. Evol. Biol.*, *21*, 1480–1491. https://doi.org/10.1111/j.1420-9101.2008.01614.x

Mullon, C., Keller, L., & Lehmann, L. (2016). Evolutionary stability of jointly evolving traits in subdivided populations. *Am. Nat.*, *188*, 175–195. https://doi.org/10.1086/686900

Mullon, C., Keller, L., & Lehmann, L. (2018). Social polymorphism is favoured by the co-evolution of dispersal with social behaviour. *Nature Ecology & Evolution*, *2*, 132. https://doi.org/10.1038/s41559-017-0397-y

Noë, R., & Hammerstein, P. (1995). Biological markets. *Trends Ecol. Evol.*, *10*, 336–339. https://doi.org/10.1016/S0169-5347(00)89123-5

Nowak, M. A., & May, R. M. (1994). Superinfection and the evolution of parasite virulence. *Proc. R. Soc. Lond. B Biol. Sci.*, *255*, 81–89.

Otto, S. P., & Day, T. (2007). *A biologist's guide to mathematical modeling in ecology and evolution*. Princeton, NJ: Princeton University Press.

Pepper, J. W. (2000). Relatedness in trait group models of social evolution. *J. Theor. Biol.*, *206*, 355–368. https://doi.org/10.1006/jtbi.2000.2132

Richards, A. (2015). *University of Oxford advanced research computing*. Zenodo. https://doi.org/10.5281/zenodo.22558

Sachs, J. L., & Wilcox, T. P. (2006). A shift to parasitism in the jellyfish symbiont Symbiodinium microadriaticum. *Proc. R. Soc. Lond. B Biol. Sci.*, *273*, 425–429.

Speare, L., Cecere, A. G., Guckes, K. R., Smith, S., Wollenberg, M. S., Mandel, M. J., et al (2018). Bacterial symbionts use a type VI secretion system to eliminate competitors in their natural host. *Proceedings of the National Academy of Sciences USA*, *115*(36), E8528–E8537. https://doi.org/10.1073/pnas.1808302115

Stewart, A. D., Logsdon, J. M., & Kelley, S. E. (2005). An empirical study of the evolution of virulence under both horizontal and vertical transmission. *Evolution*, *59*, 730–739. https://doi.org/10.1111/j.0014-3820.2005.tb01749.x

Taylor, P. D. (1992). Altruism in viscous populations — an inclusive fitness model. *Evol Ecol*, *6*, 352–356.

Taylor, P. D., & Frank, S. A. (1996). How to make a kin selection model. *J. Theor. Biol.*, *180*, 27–37. https://doi.org/10.1006/jtbi.1996.0075

Wakano, J. Y., & Lehmann, L. (2014). Evolutionary branching in demestructured populations. *J. Theor. Biol.*, *351*, 83–95. https://doi.org/10.1016/j.jtbi.2014.02.036

Wang, L. (2016). make use of color palettes from http://colorbrewer2.org in Mathematica: https://github.com/wanglongqi/ColorBrewer

West, S. A., & Buckling, A. (2003). Cooperation, virulence and siderophore production in bacterial parasites. *Proc. R. Soc. Lond. B Biol. Sci.*, *270*, 37–44.

West, S. A., Kiers, E. T., Pen, I., & Denison, R. F. (2002). Sanctions and mutualism stability: When should less beneficial mutualists be tolerated? *J. Evol. Biol.*, *15*, 830–837. https://doi.org/10.1046/j.1420-9101.2002.00441.x

West, S. A., Kiers, E. T., Simms, E. L., & Denison, R. F. (2002). Sanctions and mutualism stability: Why do rhizobia fix nitrogen? *Proc. R. Soc. Lond. B Biol. Sci.*, *269*, 685–694. https://doi.org/10.1098/rspb.2001.1878

West, S. A., Fisher, R. M., Gardner, A., & Kiers, E. T. (2015). Major evolutionary transitions in individuality. *Proceedings of the National Academy of Sciences USA*, *112*, 10112–10119. https://doi.org/10.1073/pnas.1421402112

Wickham, H., Chang, W., Henry, L., Pedersen, T.L., Takahashi, K., Wilke, C., et al. (2019). *ggplot2: Create elegant data visualisations using the grammar of graphics*.

Wyatt, G. A. K., Kiers, E. T., Gardner, A., & West, S. A. (2014). A biological market analysis of the plant-mycorrhizal symbiosis. *Evolution*, *68*, 2603–2618. https://doi.org/10.1111/evo.12466

Yamamura, N. (1993). Vertical transmission and evolution of mutualism from parasitism. *Theor. Popul. Biol.*, *44*, 95–109. https://doi.org/10.1006/tpbi.1993.1020

Yamamura, N. (1996). Evolution of mutualistic symbiosis: A differential equation model. *Researches on Population Ecology*, *38*, 211–218. https://doi.org/10.1007/BF02515729

How to cite this article: Leeks A, dos Santos M, Andrew West S. Transmission, relatedness, and the evolution of cooperative symbionts. *J Evol Biol*. 2019;00:1–10. https://doi.org/10.1111/jeb.13505

APPENDIX 1

We are interested in whether relatedness (R) or transmission mode (λ) plays the bigger role in influencing the final level of cooperation in the first analytical model (Equation 3). As a heuristic for this, we determine the effect of marginal changes in either relatedness or transmission mode on the equilibrium level of cooperation. Since the equilibrium level of cooperation depends positively on relatedness, but negatively on increased horizontal transmission, we take the absolute magnitude of the relevant derivatives. Therefore, relatedness plays the bigger role when $\left|\frac{\partial x_0^*}{\partial R}\right| > \left|\frac{\partial x_0^*}{\partial \lambda}\right|$. Given the constraints of our model (i.e. $s > 0$, $f > 0$, $0 \le R \le 1$, $0 \le \lambda \le 1$), this condition simplifies to $\frac{f(1-\lambda)+s}{fR(1-R)} > 1$. This holds provided that $\lambda < \lambda'$, where $\lambda' = \frac{s}{f} + 1 - R\,(1-R)$.

Relatedness always plays the bigger role ($\lambda < \lambda'$) when $\lambda` > 1$, because $\lambda$ is defined between 0 and 1. $\lambda` > 1$ only when fertility accelerates quickly ($f \ge \frac{s}{R(1-R)}$), a condition which becomes prohibitively restrictive as $R \to 0$ or $R \to 1$, and is least restrictive when $R = 0.5$, at which point $f \ge 4s$. Hence, R will always play the bigger role provided that fertility accelerates less than four times as quickly as survival ($f < 4s$).

Furthermore, because $\lambda`$ is increasing with $s$, we can find the lowest value that $\lambda`$ can possibly take, by taking the limit of $\lambda`$ as $s \to 0$: $\lim_{s\to 0} \frac{s}{f} + 1 - R\,(1-R) = 1 - R\,(1-R)$. This expression equals 1 whenever $R$ is 0 or 1, and is lowest when $R = 0.5$, at which point it evaluates to 0.75. Therefore, relatedness is more important ($\lambda < \lambda`$) provided transmission is more than 25% vertical ($\lambda < 0.75$).

To summarize, relatedness (R) always plays a bigger role than transmission mode (λ) in determining the final level of cooperation, unless the three following conditions are all met: (a) horizontal transmission dominates ($\lambda > 0.75$); (b) host fecundity accelerates substantially faster with symbiont cooperation than host survival does ($f < 4s$); (c) relatedness is intermediate ($0 < R < 1$). We plot the area of parameter space in which relatedness has a larger influence than transmission mode in Figure 2.

APPENDIX 2

To "close" the model by expressing relatedness in terms of demographic parameters, this we focus on a focal symbiont and calculate the likelihood that a randomly chosen symbiont infecting the same host, with replacement, is identical by descent (Pepper, 2000). We calculate this for both horizontally transmitting symbionts and vertically transmitting symbionts, and then we combine both expressions,

weighted by the likelihood that each mode of transmission has occurred. Note that this method assumes that evolution in the focal trait does not influence relatedness.

First, with probability λ, all symbionts within the focal symbiont's host were acquired horizontally from a well-mixed pool of symbionts. Because hosts are infected horizontally by exactly k_h symbiont strains, the probability of sampling the same individual twice is $1/k_h$. Because we have assumed a large population of hosts, relatedness to any symbionts which are not identical by descent is zero. Second, with probability $1 - \lambda$, all symbionts were inherited vertically from the host's parent. As before, the probability of picking the same individual twice is $1/k_v$. With probability $(k_v - 1)/k_{v,}$ the second symbiont is different than the first. In this case, the probability that both symbionts are identical by descent is given by the relatedness between pairs of symbionts in the parent's host in the previous generation $R_{(t-1)}$. We therefore derive the relatedness at a given generation t (i.e. $R_{(t)}$) of a focal symbiont to a randomly picked partner in the same host (with replacement) as:

$$R_{(t)} = (1-\lambda)\left(\frac{1}{k_v} + \frac{k_v - 1}{k_v} R_{(t-1)}\right) + \lambda \frac{1}{k_h}. \quad \text{(S1)}$$

At equilibrium, relatedness will no longer change from one generation to the next, and thus, solving $R_{(t)} = R_{(t-1)}$ gives the value of relatedness, R, at equilibrium:

$$R = \frac{k_h(1-\lambda) + \lambda k_v}{k_h\left[1-(1-k_v)\lambda\right]}. \quad \text{(S2)}$$

Since we assume here that the cooperation trait does not influence relatedness, our expression for relatedness (Equation S2) does not depend on the functional forms of fecundity and survival ($s(x)$ and $f(x)$).

If we assume that symbiont bottleneck sizes are the same for both horizontal and vertical transmission ($k_h = k_v$ =k), then our expression for relatedness (Equation S2) simplifies to

$$R = \frac{1}{1+\lambda(k-1)} \quad \text{(S3)}$$

APPENDIX 3

Transmission mode can influence the level of cooperation that evolves through two routes: (a) through influencing cooperation per se and (b) through influencing relatedness, which in turn influences cooperation. To capture the first of these routes, we take the differential of cooperation with respect to transmission mode, $\frac{\partial x_0^*}{\partial \lambda}$. It is difficult to calculate the second route directly, but we instead calculate it indirectly by taking the total effect of transmission mode (which is given by taking $\frac{\partial x_\xi^*}{\partial \lambda}$ and substituting the value for R given in Equation S3) and subtracting the effect of transmission per se. Therefore, transmission mode influences cooperation more via its effect on relatedness, than via its effect per se, when $\frac{\left|\frac{\partial x_\xi^*}{\partial \lambda}\right| - \left|\frac{\partial x_0^*}{\partial \lambda}\right|}{\left|\frac{\partial x_0^*}{\partial \lambda}\right|} > 1$.

This simplifies to $\frac{f+s}{f\lambda} > 2$, which is always true for our constraints ($s > 0$, $f > 0$, $0 \le \lambda \le 1$) provided that $\lambda < \frac{f+s}{2f}$. We know that $\lambda \le 1$, so this condition always holds if $\frac{f+s}{2f} > 1$, which is true whenever $f < s$.

Furthermore, by taking $\lim_{s\to 0} \frac{f+s}{2f}$, we can also see that this condition always holds provided $\lambda \le 0.5$. Therefore, transmission mode always influences cooperation more via its effect on R (indirectly) than via its direct effect per se, if transmission is mostly vertical ($\lambda < 0.5$) () or when host survival increases with symbiont cooperation more quickly than host fecundity does ($f < s$). In Figure 3, we plot the proportion of the influence of transmission mode that is due to its effect on R; this proportion is given by $\frac{\left|\frac{\partial x_\xi^*}{\partial \lambda}\right| - \left|\frac{\partial x_0^*}{\partial \lambda}\right|}{\left|\frac{\partial x_\xi^*}{\partial \lambda}\right|} = 1 - \frac{f\lambda}{f+s}$.

APPENDIX 4

We next wrote a simulation in order to relax our analytical assumption of weak selection, based on the life cycle of our analytical model in the main text (section 2.1).

We initialize our simulation with a constant number of hosts n. At the start of the simulation, each host is infected by k symbionts, with each symbiont's investment into cooperation starting at 0.5. We assume that symbionts reach large population sizes within each host, and so we incorporate mutation by assuming that a constant fraction of the symbiont population (μ = 0.001) mutates to the current value for cooperation ± 0.01 (one mutational step up or down). Transmission of symbionts and interactions between symbionts occur as specified in our first analytical model (section 2.1; Equation 1).

We ran the simulation for 10^6 generations, after determining that this was sufficient to allow evolutionary branching events to occur (Figure 5), by comparing simulation runs for 10^5 and 10^6 generations. We determined whether or not evolutionary branching had occurred by visual inspection and found that we could positively identify branching events when the range of cooperation values (the highest value minus the lowest) in the symbiont population was greater than $0.1 + \bar{x}^2$, where $\bar{x}$ is the mean level of cooperation in the symbiont population. We recorded observed whole-group relatedness in the simulation (Figure 6) by calculating the correlation coefficient between a focal symbiont's trait value for cooperation and the average value for cooperation within its host (Pepper, 2000).

We wrote the simulation in MATLAB version R2016a and obtained results using the University of Oxford Advanced Research Computing (ARC) Facility (Richards, 2015). We analysed the results in R version 3.4 and produced figures with ggplot2 (Wickham et al., 2019). A full simulation life cycle, MATLAB scripts to run the simulation, text files of results produced, and R code with data analysis and figure production are available online at https://osf.io/puwq3/.

6

The Abundance of Defective Viral Genomes in Natural Viral Infections

Introduction

To be successful, viruses must replicate their genome and then package themselves inside capsids to infect new cells. To do this, viruses encode gene products such as replicase enzymes and capsid proteins, which can be shared by multiple viral genomes, not just the genome that produced them (they are 'trans-acting') (Flint *et al.*, 2015; Díaz-Muñoz *et al.*, 2017). From an evolutionary perspective, these trans-acting gene products can be public goods, because they are costly to produce, and provide a benefit to all of the viruses infecting the cell (West *et al.*, 2007). Consequently, they can be open to exploitation by 'cheats', that avoid paying the cost of producing the public good, but still gain its benefits (Ghoul *et al.*, 2013; Leeks *et al.*, 2020).

Defective interfering genomes are a type of viral cheat that exploit these public goods (Huang & Baltimore, 1970). They are formed when large deletions create a mutant that lacks key trans-acting genes, such as those encoding replicase or capsid proteins. Because they lack these key genes, defective interfering genomes cannot successfully infect host cells on their own, but they can spread if they coinfect a host cell alongside a wild-type virus. When this coinfection occurs, defective interfering genomes gain a substantial replication advantage through their shorter length, and so they interfere with the accumulation of the wild-type virus, sometimes replicating thousands of times faster than the wild-type virus (Shirogane *et al.*, 2021).

Defective interfering genomes emerge spontaneously in tissue culture infections of almost all known viruses, and when they spread, this can cause viral populations to collapse (Vignuzzi & López, 2019). Consequently, defective interfering genomes could play an important role in the dynamics of natural viral infections by reducing the accumulation of the wild-type virus, potentially resulting in a less severe infection. Consistent with this, there is some evidence from Influenza A and Ebola viruses that patients infected with a higher abundance of defective viral genomes experienced less severe infections (Vasilijevic *et al.*, 2017; Dong *et al.*, 2020). A further possibility is to use synthetic defective interfering genomes in order to treat viral infections, by introducing them to an infection to reduce the spread of the virus (termed 'therapeutic interfering particles') (Metzger *et al.*, 2011).

However, it is not clear whether defective interfering genomes commonly occur in natural viral infections. On the one hand, they can arise spontaneously in tissue culture infections, and often gain substantial fitness advantages over the wild-type virus, so we might expect them to be common in natural infections. On the other hand, their prevalence in tissue culture could be artefactual, and reflect features of tissue culture that do not carry over into real viral infections. For example, tissue culture infections often contain high ratios of viruses to host cells (a high multiplicity of infection, or MOI), which allows for frequent coinfection. If coinfection is less common in natural viral infections, then defective interfering genomes may not be able to spread, and so they may be found more rarely.

The extent to which defective interfering genomes occur naturally matters, because if they are rare in nature, then they are less likely to be able to explain clinical outcomes, and it may be harder to treat viral infections using therapeutic interfering particles. There are also a number of unknown factors in the biology of defective interfering genomes, that we cannot find out from tissue culture, such as how commonly they spread between hosts, whether the same types of defective interfering genome emerge in different infected hosts, and whether they play a significant role in influencing natural dynamics and clinical outcomes.

Here, we take advantage of public viral sequencing datasets to determine the extent of defective interfering genomes the largest dataset of natural hosts yet studied. Using data from 198 humans infected with influenza, and 1,278 humans infected with SARS-CoV-2 including one longitudinal study, we detect and quantify naturally occurring defective viral genomes. For Influenza, we determine which of these defective genomes are also likely to be interfering by comparing them to known defective interfering genomes from tissue culture. For SARS-CoV-2, we infer that defective viral genomes could be interfering if they persist over multiple timepoints across the longitudinal sample.

Methods

To determine the abundance of defective interfering genomes in natural viral infections, we accessed publicly available deep sequencing datasets from 198 individuals infected with Influenza A between 2015-2017, and 1,277 individuals infected with SARS-CoV-2 during 2020. We also used a longitudinal dataset, consisting of 9 samples taken from a single patient who had been infected with SARS-CoV-2 for more than five months in 2020 (Choi *et al.*, 2020). We analysed these datasets using an existing software package for detecting defective

viral genomes, ViReMa, that has previously been tested in a range of viruses including Influenza A (Fig. 1) (Routh & Johnson, 2014; Jaworski & Routh, 2017; Alnaji *et al.*, 2019). We then extracted information about each candidate defective interfering genome, and estimated its relative abundance. For influenza, we determined which defective genomes were interfering by comparing to known defective interfering genomes from tissue culture infections; for SARS-CoV-2, we inferred the presence of defective interfering genomes by finding candidate defective interfering genomes that persisted over the duration of the longitudinal infection. Code to reproduce the analysis has been made available here: https://unioxfordnexus-my.sharepoint.com/:f:/g/personal/newc4385_ox_ac_uk/EoJmdPBTQzVMnOOqyyN8GUsB74r4GS0lckLqiXzns3eg2A

Influenza

Raw data

On 01/06/2019 we downloaded a list of all complete genome sequences for influenza from fludb where the original fastq files were available. To ensure that any defective genomes we found were really present in the natural host, we filtered this list to contain only samples that had not been subsequently passaged in cell culture. The resulting list contained 449 samples from 198 different individuals, consisting of 145 individuals infected with H3N2, and 53 with H1N1, all collected in New York State between 2015 and 2017. For each sample, we downloaded both the original fastq files (that contain the 'raw' sequencing reads) and the associated consensus sequences using fastq-dump from sra-tools. All samples consisted of paired-end sequencing data, so we downloaded and analysed each set of paired-end reads, recording the output as either 'left' or 'right', before later combining.

Detecting Defective Viral Genomes

We first used ViReMa v0.15 to detect all potential deletion and recombination events within the raw samples, using the consensus sequence for each sample as the reference genome (Fig. 1) (Routh & Johnson, 2014). Based on a previous study that optimised ViReMa for detecting defective viral genomes in influenza A, we appended each reference genome segment with a sequence of 551 'A's, and called ViReMa with the following parameters: --N 1 --X 8 --p 2 (Alnaji *et al.*, 2019).

We then wrote in-house Bash scripts to extract all deletion events occurring on the same genome segment. For each deletion event, we recorded summary information including: the start point; the end point; the gene segment; the number of reads mapping to the deletion

event; the total number of reads mapping to the segment; and whether the read was detected in the forward- or reverse-direction. To distinguish unique defective viral genomes, we then labelled each defective viral genome with the identifier ‘segment_start_end’. For example, a defective viral genome consisting of a deletion between positions 1859 and 1930 in segment PB2 would be labelled ‘PB2_1859_1930’.

Identifying Unique Defective Viral Genomes

To account for the fact that the same defective viral genome could be detected in different ways, we combined some of our abundance metrics for defective viral genomes. When the same defective viral genome, determined by its unique identifier ‘segment_start_end’, was found on both forward- and reverse-strands of the genome, we combined these into the same defective viral genome, summing the read counts and recalculating the relative abundance. To control for bias in whether left or right read-mates were more likely to find each defective viral genome, when the same defective viral genome was found in both left and right read-mates, we combined these and took the mean of the relative abundance. Our final dataset therefore contained a unique set of defective viral genomes for each accession, where each accession represented a sample from a patient. For most of our analyses, we did not combine defective viral genomes that were found in multiple accessions from the same patient, since these could represent distinct samples if the accessions were taken on multiple days or from different locations within the body. When we investigated which defective viral genomes were found across multiple hosts, we pooled the defective viral genomes found in multiple samples from the same host, taking the mean of their relative abundance.

Calculating Relative Abundance

To determine the relative abundance of each defective viral genome, we first recorded the number of reads that contained the deletion event, as given by ViReMa. Then, to account for the fact that read depth fluctuated across different genome segments, we normalised the number of deletion reads by the total number of reads mapping to the genome segment the read was found in. Our relative abundance measure therefore gives the proportion of all reads mapping to that segment that contained the deletion in question. To find the most abundant defective viral genomes within each sample, we then rank-ordered each defective viral genome within each accession according to this abundance metric.

SARS-CoV-2

Raw data

On 09/11/2020 we downloaded 1,277 SARS-CoV-2 fastq files from NCBI, that had been uploaded by the COG-UK sequencing consortium (COG-UK, 2020). For each sample, we downloaded the original fastq files using fastq-dump from sra-tools. All samples were single-end Illumina data.

In collaboration with Dr Manish Choudhary and Dr Jonathan Li of Harvard Medical School, we then accessed a longitudinal dataset that had been obtained at Brigham and Women's Hospital, Boston (Choi *et al.*, 2020). The dataset came from a 45-year-old male patient with an immunosuppressive disorder, who had been infected with SARS-CoV-2 for a period of five months. We had access to nine samples taken over this period, that had been sequenced using two different sets of primers, together with viral load estimates at each timepoint.

For analysing all samples, we used the SARS-CoV-2 reference genome available on NCBI (https://www.ncbi.nlm.nih.gov/nuccore/1798174254).

Detecting and Quantifying Defective Viral Genomes

We followed the same protocol to detect, identify, and quantify defective viral genomes for SARS-CoV-2 as for Influenza, except for three differences. Firstly, instead of using a separate reference genome for each sample, we used the same SARS-CoV-2 reference genome for all samples. Secondly, we used a Python script written by Mariá José Ulmo Oceda at the University of Valéncia to filter out known subgenomic mRNAs from the ViReMa output. Thirdly, we normalised the read count of each defective viral genome to the total number of reads mapped to the whole genome, since the SARS-CoV-2 genome is non-segmented.

Identifying Defective Interfering Genomes

SARS-CoV-2 is a novel virus, so we did not have a body of work on defective viral genomes in tissue culture to compare our natural samples to. Other human coronaviruses do produce defective and defective interfering viral genomes, but their genomes are too distantly related to SARS-CoV-2 to be able to predict which regions of the genome defective viral genomes

are likely to be found in. Therefore, we inferred that defective viral genomes had the potential to be interfering if they persisted over multiple timepoints across the longitudinal sample, as this could indicate a persistent population of defective viral genomes able to be replicated.

To narrow down the list of potential defective interfering genomes, we then searched for polymorphisms in the candidate defective interfering genomes that were not also found in reads that mapped to the wild-type genome. We inferred that if such reads were found over multiple timepoints, this would confirm that the defective viral genomes comprised an independent population that was being replicated, and that they were not being produced *de novo* by the wild-type virus at each timepoint. To do this, we extracted the reads mapping to each candidate defective interfering genome, visualised them using Tablet, recorded any polymorphisms, and determined the effect of these polymorphisms (Milne *et al.*, 2013).

Data and Code Availability

We have made available commented versions of all of the scripts, accession lists, and software environments used, here:
https://unioxfordnexus-my.sharepoint.com/:f:/g/personal/newc4385_ox_ac_uk/EoJmdPBTQzVMnOOqyyN8GUsB74r4GS0lckLqiXzns3eg2A

We used Oxford's ARC high-performance computing cluster to download accessions from NCBI, run ViReMa, and run subsequent Bash scripts to extract the defective viral genomes. We then used R and RStudio locally for further analysis and for generating the figures.

Results

Influenza

Defective Interfering Genomes are Ubiquitous in Natural Influenza Infections

We found that 97.6% of influenza samples (438/449) contained at least one candidate defective interfering genome. We defined candidate defective interfering genomes as those consisting of a deletion of more than 50% of the length of one of the polymerase segments, that was also among the 25% most abundant defective viral genomes within the sample it came from. Across all samples, the vast majority of the most abundant defective viral genomes consisted of large deletions of the polymerase segments (Fig. 2).

Characteristic Patterns of Defective Interfering Genomes in Influenza

We found that large deletions occurred in all genome segments, but that they generally only reached high frequencies in the three polymerase segments (Fig. 3; Fig. 2b). Across most genome segments, small deletion mutants reached relatively high abundance compared to mutants with larger deletions, which is consistent with mutation-selection balance, and the fact that smaller mutations are more likely to be selectively neutral. However, the three polymerase segments appeared to show a different pattern, with both small and large deletions reaching high abundance. This suggests that there could be a selective advantage for large deletions, but only when those large deletions occur on the three polymerase segments.

For the two segments that encode capsid proteins, HA and NA, a small number of large deletions reached high abundance (Fig. 3).

We observed broadly the same qualitative pattern for both H3N2 and H1N1 samples (Fig. S1).

Constrained Structure of Defective Viral Genomes in Influenza

We found that across all segments, the start- and end-points of the deletions were consistent with the known range for viable deletions (Fig. 4). In particular, we found that large deletions reached high frequencies across the three polymerase segments, but there was an upper limit on the size of deletion we found, at around 85% the total length of the genome segment (Fig. 3). This is consistent with there being a minimum viable length for influenza defective interfering genomes, since regions at either end of each genome segment are required for the segment to be replicated and Encapsidated (Liang *et al.*, 2005; Hutchinson *et al.*, 2010; Martin *et al.*, 2019).

The Same Influenza Defective Viral Genomes are Found Across Multiple Individuals

While most defective viral genomes were only found in one patient, we found that a small number were present in many different individuals. 50 defective viral genomes with the same start and end coordinates were found in 30 or more hosts, out of a potential 197 hosts (Fig. 5a). These defective viral genomes primarily contained deletions in the polymerase segments (49 out of the top 50 most shared; Table 1). There was no association between the number of

different individuals that a defective viral genome was found in and its mean relative abundance in each individual (Pearson's correlation coefficient = 0.06, $p < 10^{-15}$; Fig. 5b).

SARS-CoV-2

No Broad-Scale Characteristic Patterns of Defective Interfering Genomes in SARS-CoV-2

In contrast to our results for Influenza, we did not find the same characteristic patterns that would be expected if the same types of defective interfering genomes were arising and spreading in different hosts (Fig. 6). Across the 1,277 latitudinal samples, there was no clear relationship between fraction of genome deleted and relative abundance for any of the combinations of genes that defective viral genomes lost, and we found that defective viral genomes readily formed across all genome coding regions, with no obvious bias towards some areas compared to others (Fig. 7).

Some SARS-CoV-2 Defective Viral Genomes Persisted for Months

We found that in the longitudinal dataset, some SARS-CoV-2 defective viral genomes persisted over many months, with nine defective viral genomes found in at least eight of the nine potential timepoints (Fig. 8). We inferred that the presence of these defective viral genomes across so many timepoints indicated that they came from populations of defective viral genomes that were able to be replicated within the host, and so we designated these as candidate defective interfering genomes. The relative abundances of these candidate defective interfering genomes mostly tended to correspond to the viral load in the patient at each timepoint. However, two candidate defective interfering genomes (1510_8009 and 14206_15647) achieved their maximum relative abundance a few days after the peak viral load recorded in the patient.

None of these nine candidate defective interfering genomes were found in any of the 1,277 latitudinal samples.

Some SARS-CoV-2 Defective Viral Genomes Formed Distinct Viral Populations

Three defective viral genomes (1510_8009, 14206_15647, and 23187_24486) contained polymorphisms that were not found in reads mapping to the wild-type virus (Fig. 9; Table 2). Two of these mutations were non-synonymous, and two interrupted the open reading frame.

This suggests that these defective viral genomes were replicating as independent viral populations, accumulating mutations that were not found in the wild-type virus.

The Most Abundant SARS-CoV-2 Defective Viral Genomes Were Found in Multiple Datasets

To check that our findings were real viral variants and not PCR artefacts, we re-ran our pipeline on the same longitudinal samples, but that had been sequenced using a different set of primers (Artic v1). Only a fraction of the defective viral genomes that had been found in the original sequencing run were also found in the second sequencing run (4382 candidates in the original dataset; 87 of these also found in the new dataset; new dataset found 1771 candidates in total). However, the most abundant defective viral genomes were substantially more likely to be found in the second sequencing run (Fig. 10; glm Z(10.12), P < 10^-16). Of the three candidates that showed polymorphisms in the original data, only one was found in the re-sequenced data, although we did find the same polymorphism in this candidate in the new set of data as well.

Discussion

We explored the defective viral genomes present in publicly available data from 197 patients infected with Influenza A, and 1,278 patients infected with SARS-CoV-2, including one longitudinal dataset spanning more than five months. More than 97% of Influenza samples contained defective viral genomes resembling those that are known to be interfering in tissue culture. Defective viral genomes were most common in the three polymerase segments, reached high abundances across most hosts, and accounted for the vast majority of the reads mapping to defective viral genomes (Fig. 2; Fig. 3). Some defective viral genomes were shared between hosts, with 50 large defective viral genomes, almost exclusively containing deletions in the polymerase segments, being found in at least 30 out of the 197 different individuals in the study (Table 1). These findings confirm, in the largest dataset yet studied, that defective interfering genomes are highly abundant in natural Influenza A infections. These results suggest that defective interfering genomes could play a widespread role in the dynamics of natural Influenza infections, and that therapeutic interfering particles designed against Influenza A are likely to be able to spread in human patients.

We found different patterns of defective viral genome accumulation in the 1,277 SARS-CoV-2 samples, with no clear indication that the same types of defective interfering genomes arose

in different patients (Fig. 6; Fig. 7). However, our longitudinal sample revealed that at least eight different SARS-CoV-2 defective viral genomes persisted over a period of five months within an individual patient (Fig. 8). For three of these defective viral genomes, the presence of polymorphisms confirmed that these were actively replicating viral populations, consistent with being defective interfering genomes (Fig. 9; Table 2). Together, these findings suggest that populations of defective viral genomes can be present in SARS-CoV-2 infections, but that their broader relevance for explaining clinical outcomes, and the possibility for using therapeutic interfering particles in SARS-CoV-2 patients, remain unclear.

Influenza

Defective Interfering Genomes

The patterns of defective viral genomes that we found in our clinical samples of Influenza strongly resembled those of defective interfering genomes from tissue culture infections (Fig. 2; Fig 3). We found that large deletions of the polymerase segments reached high abundance, which is consistent with findings from tissue culture, which find that defective interfering genomes of the polymerase segments gain a replication advantage through being shorter, and can reach high frequencies (Nayak, 1980; Dimmock & Easton, 2014; Alnaji *et al.*, 2019). We found that a small number of large deletion mutants of the capsid protein segments, HA and NA, reached high abundances, which is consistent with the observation that defective interfering genomes of these segments occur in tissue culture, but more rarely than in the polymerase segments (Alnaji *et al.*, 2019). Large deletion mutants of the polymerase segments accounted for the vast majority of the 'biomass' of the viral mutants found, suggesting that the ability of defective interfering genomes to cheat the wild-type virus is the primary factor driving the abundance of defective viral genomes in natural infections (Jaworski & Routh, 2017). Our results also closely agreed with other studies that have looked at defective interfering genomes in natural infections of Influenza (Saira *et al.*, 2013; Martin *et al.*, 2019). Overall, the degree of consistency between our findings and between experimental work on Influenza suggests that defective interfering genomes are not an artefact of tissue culture, and could play an important and widespread role in the dynamics of natural Influenza infections.

Potential Transmission Between Hosts

We found 50 defective viral genomes in 30 or more different individuals, a degree of convergence that is consistent with previous studies on influenza infections of humans (Martin *et al.*, 2019). There are three potential non-exclusive explanations for this: that there are constraints on the types of defective viral genome that can form; that there are selective advantages to certain types of defective viral genome, resulting in convergent evolution; that some defective viral genomes are able to spread between hosts. Tissue culture studies support the first two explanations, with convergent influenza defective viral genome populations emerging in independent tissue culture experiments (Alnaji *et al.*, 2019). However, we found that there was no relationship between the within-host abundance of a defective viral genome and its likelihood of being present in multiple hosts, which could be inconsistent with the second explanation (Fig. 5b). One caveat here is that our 'snapshots' of defective viral genome abundance may not reflect the long-term abundance of each defective viral genome over the course of the infection, especially since defective interfering genomes often show cyclical dynamics (Kirkwood & Bangham, 1994).

There is some support for the third explanation from previous studies on influenza and other viruses such as dengue, that have found indirect evidence that defective viral genomes can transmit between different hosts (Aaskov *et al.*, 2006; Saira *et al.*, 2013; Martin *et al.*, 2019). However, evolutionary theory predicts that this should be relatively rare, since influenza infections are generally only initiated by 1-2 viral genomes, meaning it is unlikely that defective viral genomes can coinfect new hosts alongside a wild-type virus (Leonard *et al.*, 2016; McCrone *et al.*, 2018; Xue & Bloom, 2018). In our study, the 197 patients were infected over a period of two years, across more than thirty different New York counties, and represent only a small fraction of the influenza infections that occurred over that time period. Therefore, if these shared defective viral genomes do reflect transmission, this would suggest that defective viral genomes are circulating among the influenza epidemic over relatively long time periods and wide geographical areas, implying that the same defective viral genome can transmit to new hosts multiple times.

The third possibility could potentially be resolved if we could resequence the original samples and generate long-read sequences of the defective viral genomes. Then, if there was sufficient neutral variation in the defective viral genomes, a phylogenetic analysis could

reveal whether defective viral genomes found across multiple patients originated independently within each patient or resulted from transmission between patients.

Clinical Relevance

Previous studies on small numbers of patients infected with Influenza A (n=12) have found that increased abundance of defective viral genomes is associated with less severe clinical outcomes (Vasilijevic *et al.*, 2017). An obvious next step would be to test whether defective viral genome abundance can also predict clinical outcomes in our dataset of 197 patients. We are currently pursuing this in collaboration with the collectors of the data, however there are two potential methodological issues.

The first issue is that it is unclear whether to expect a positive or a negative relationship between defective genome abundance and more severe clinical outcomes. Both are theoretically possible, and both have been observed in previous studies in different viruses (Vasilijevic *et al.*, 2017; Felt *et al.*, 2021). A further complication is that defective viral genomes can play a disproportionate role in stimulating the host immune system, and in some cases they could even increase the severity of symptoms, by overstimulating host immune responses (Manzoni & López, 2018).

The second potential issue is that our estimates of defective viral genome abundance are noisy, both due to noise from PCR amplification, and from the fact that we have only a snapshot of abundance at a single point in time. In reality, defective interfering genomes may fluctuate in their frequency over time (Kirkwood & Bangham, 1994). An alternative approach could be to see if the presence or absence of specific defective interfering genomes predicts clinical outcome. If we assume that some defective interfering genomes tend to be more effective than others at exploiting the wild-type virus, then the presence or absence of certain defective interfering genomes could be a better predictor of clinical outcome than our noisy estimate of total abundance. We could therefore use a GWAS-like approach to look for associations between specific defective interfering genomes and clinical outcomes.

SARS-CoV-2

SARS-CoV-2 Defective Viral Genomes Occur

Our longitudinal dataset showed that defective viral genomes are possible in SARS-CoV-2 and that they can persist over long timescales (> 5 months) in infected hosts (Fig. 8). We found nine candidate defective interfering viral genomes, and confirmed that three came from viral subpopulations by finding polymorphisms that were not present in the wild-type viral reads (Table 2).

Two of the persistent defective viral genomes, 1510_8009 and 14206_15647, also showed patterns of relative abundance that could be consistent with being defective interfering genomes. These candidates achieved their maximum relative abundances just after the peak viral load (Fig. 8). This pattern would be expected if they were defective interfering genomes, since they would gain their greatest fitness advantage when the wild-type virus is highly abundant, resulting in more frequent coinfection. The subsequent decline of the overall viral load would also be consistent with this, as it could reflect interference from these (and potentially other) defective interfering genomes, although there were confounding factors here, such as treatment with antivirals.

SARS-CoV-2 Defective Viral Genomes Appear to be Rare

However, despite the fact that defective viral genomes are possible in SARS-CoV-2, it appears from our latitudinal dataset of 1,277 patients that SARS-CoV-2 does not produce the same spectrum of predictable defective interfering genomes that exist in Influenza infections, and none of the persist defective viral genomes found in the longitudinal sample were found in any of the 1,277 latitudinal samples.

One explanation for this discrepancy could be if defective viral genomes are more likely to arise and spread in chronic infections. Our first longitudinal sample was taken 18 days after the start of infection, which is much longer than the normal duration of SARS-CoV-2 infections (7-10 days). If SARS-CoV-2 defective interfering genomes arise only rarely, or experience a relatively small advantage over the wild-type, then they might only appear in chronic infections that maintain a high viral load for long periods of time.

An alternative explanation could be if the latitudinal samples were sequenced in a way that doesn't detect defective viral genomes. The latitudinal samples come from COG-UK and were sequenced with Artic primers. When our longitudinal sample was sequenced using similar Artic primers, only 1,771 defective viral genomes were found, compared to 4,382 using the original primers. This would highlight that insights into the natural abundance of defective viral genomes could vary depending on the details of sequencing methodology.

Other Coronaviruses

The extent to which defective interfering genomes are found in other coronaviruses is unclear. On the one hand, defective interfering genomes occur readily in several coronaviruses, such as HCoV-229E and Mouse Hepatitis Virus, and putatively defective interfering genomes have been found in natural infections of MERS (Makino *et al.*, 1985; Xie *et al.*, 2017; Viehweger *et al.*, 2019). On the other hand, this leaves at least ten species of coronavirus for which defective interfering genomes have not been found (Brian & Spaan, 1997). However, it is very difficult to determine whether this absence of evidence is really evidence of absence, or whether it simply reflects differences in sampling effort.

Limitations

Without direct access to the patient samples, we are left to use indirect methods to confirm the presence of defective interfering genomes. This is especially difficult for SARS-CoV-2, because there is no existing body of experimental work to draw from. Our strongest evidence for confirming candidate defective interfering genomes in SARS-CoV-2 was the presence of polymorphisms that were not found in the wild-type reads. Theoretically, this should be a strong indication of an independent viral population, and it is consistent with the observation that defective interfering genomes in other natural infections accumulate genetic variation not found in the wild-type virus (Aaskov *et al.*, 2006; Gelbart *et al.*, 2020). However, our samples used short-read sequencing, with the vast majority of reads ranging from 100 to 150bp. With such short reads, we can only detect polymorphisms if they occur within ~100bp of the junction point. Given that we expect these defective viral genomes to be thousands of base pairs long, our method therefore has an extremely high false negative rate. Long-read sequencing methods could reduce this false negative rate by encompassing more of the genome, but this must be traded off against higher error rates, and lower sequencing depth, both of which would make it more difficult to find polymorphisms.

Defective Interfering Genomes in Other Viruses

We chose SARS-CoV-2 and Influenza because large numbers of sequences were available from natural infections. However, defective interfering genomes are a widespread phenomenon, affecting almost all animal and plant viruses in tissue culture. As sequencing data from natural viral infections becomes increasingly available, it should become possible to investigate the role of defective interfering genomes across a wider range of viruses. It could be productive to combine this kind of sequence analysis with experimental work, to verify whether candidates really are interfering, and to validate the kinds of proxy measure used when experimental work is not possible, such as in this Chapter.

Figures

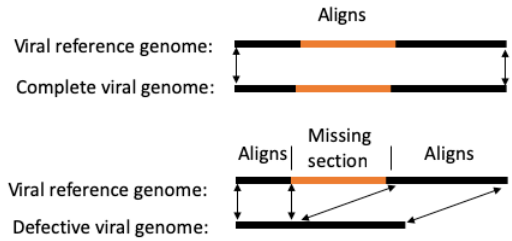
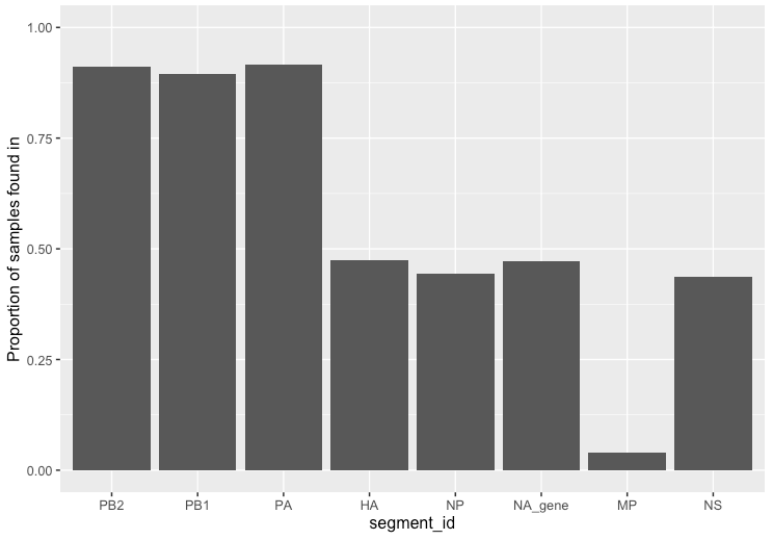


Figure 1: Defective Viral Genomes can be Detected from Deep Sequencing Datasets

We identified defective viral genomes in reads that mapped to the reference genome at either end, with a gap in between indicating a deleted section. We used ViReMa v0.15 to detect defective viral genomes (Routh & Johnson, 2014).

(a)



(b)

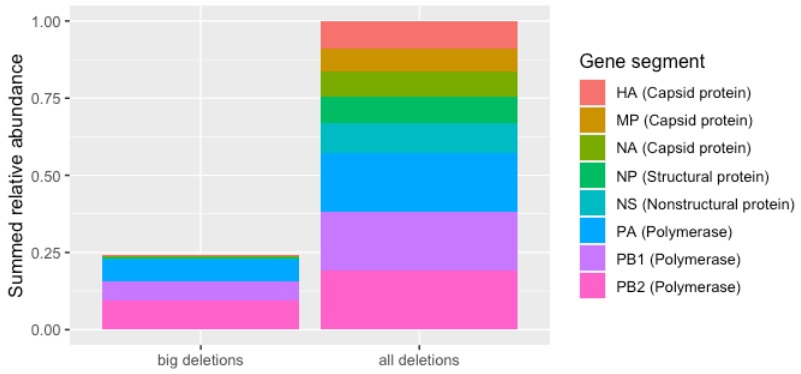


Figure 2: Defective Viral Genomes of the Polymerase Segments are Highly Abundant in Natural Influenza Samples

In (a), we plot the proportion of influenza samples that contained at least one mutant with a large deletion (> 50% segment deleted) in each gene segment. Large-deletion mutants in each of the three segments that encode the polymerase (PA, PB1, and PB2) were found in more than 80% of influenza samples. In (b), we plot the summed relative abundance of all of the large deletion variants (> 50% segment deleted), and all of the deletion variants, across all samples, for each genome segment. y-values are normalised to the total summed read count of all variants. The majority of the biomass of all variants came from mutants of

the three polymerase segments, and the vast majority of the biomass of large-deletion defective viral genomes came from the three polymerase segments.

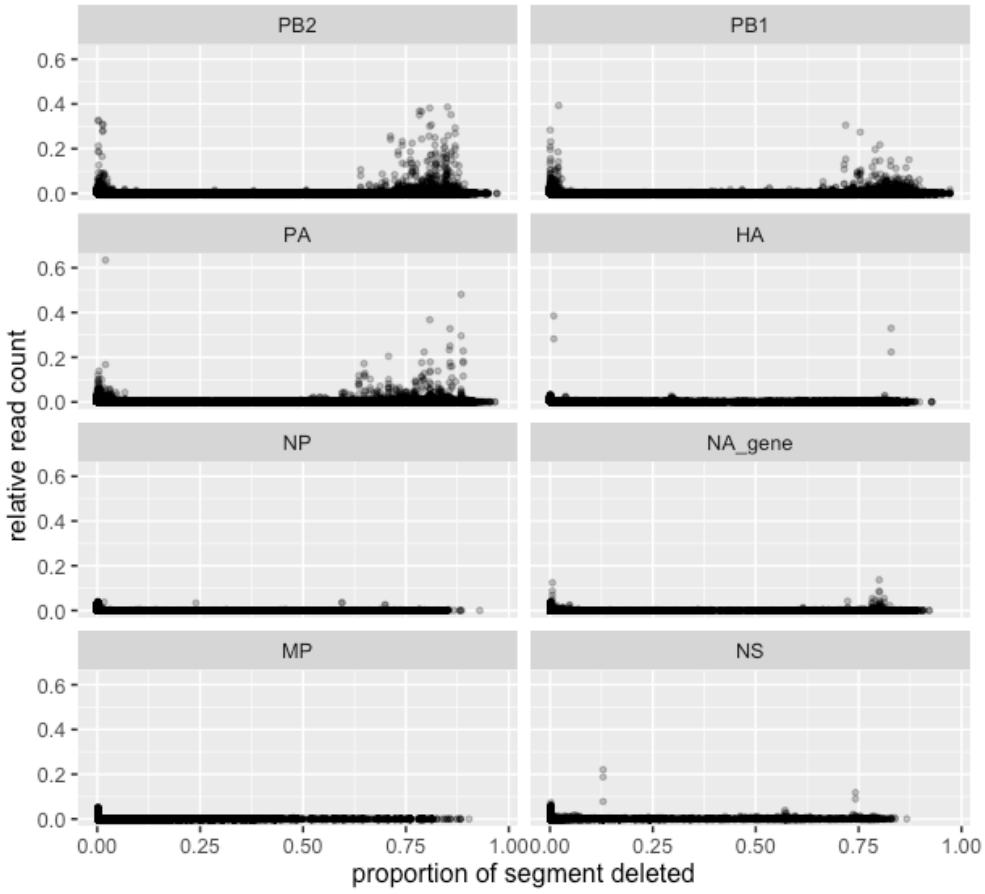


Figure 3: Influenza Polymerase Segments Show Characteristic Patterns of Defective Interfering Genomes

For each deletion variant detected in each influenza genome segment, we plot the proportion of the genome segment that was deleted, against the relative read count of that variant. Both small and large deletions occur in all eight segments, and some small mutants reach relatively high frequencies in all eight segments. Large deletions only reach high frequency in the three polymerase segments (PA, PB1, and PB2). These patterns are consistent with defective interfering genomes of influenza known from tissue culture studies.

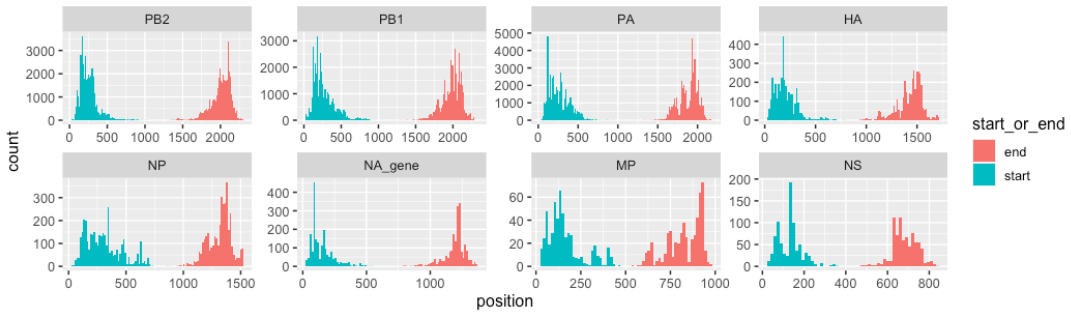


Figure 4: Defective Viral Genomes in Influenza Show Conserved Start- and End-Points

We plot the start- and end-points of every large deletion mutant (> 50% segment deleted) detected in each segment.

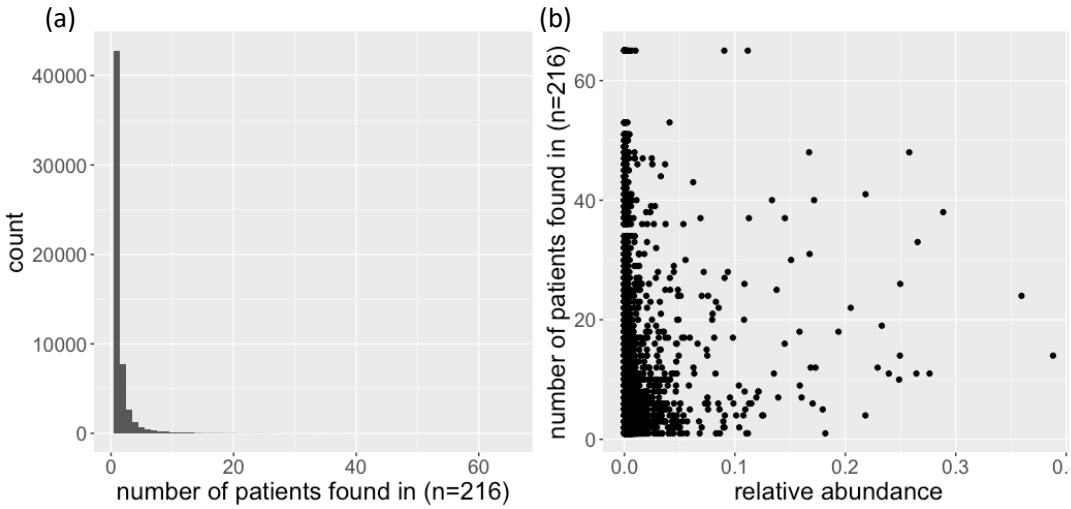


Figure 5: Some Influenza Defective Viral Genomes are Found in Multiple Individual Patients

In (a), we plot a histogram of the number of individual patients that each defective viral genome is found in. The majority are found in just one or a few patients, but a small number of defective viral genomes are found in large numbers of individuals. In (b), we plot the mean relative abundance of each defective viral genome against the number of individuals that defective viral genome is found in. Only large deletions (> 50% genome segment deleted) are plotted in this figure.

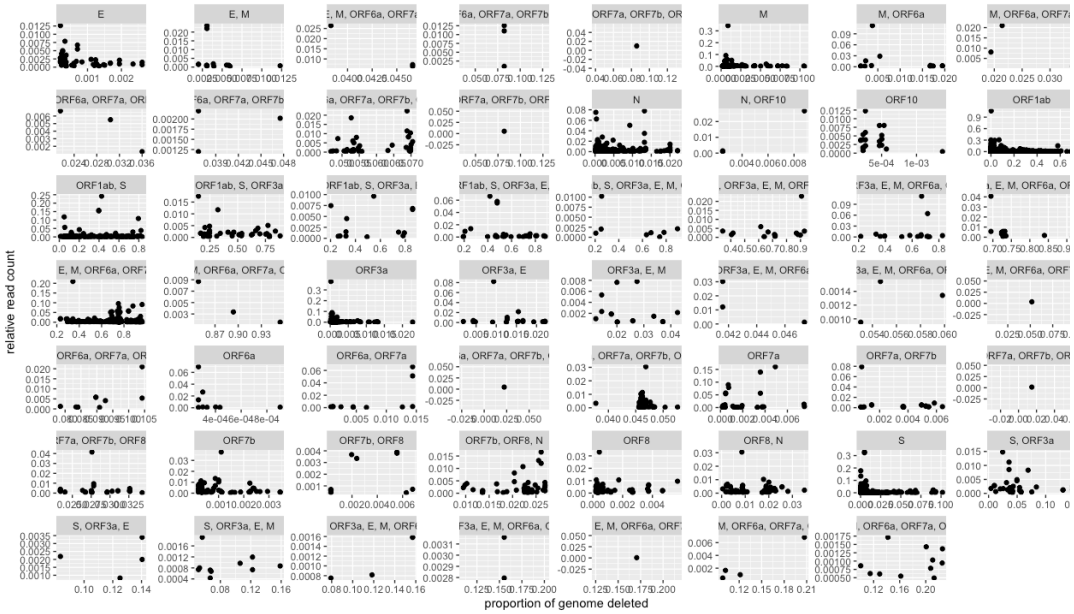
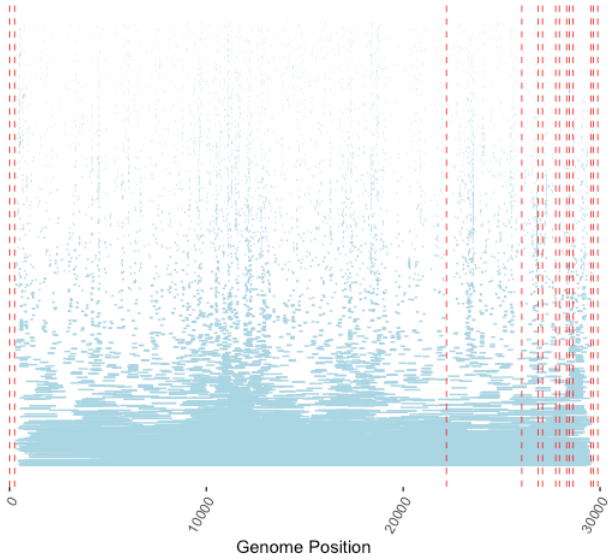


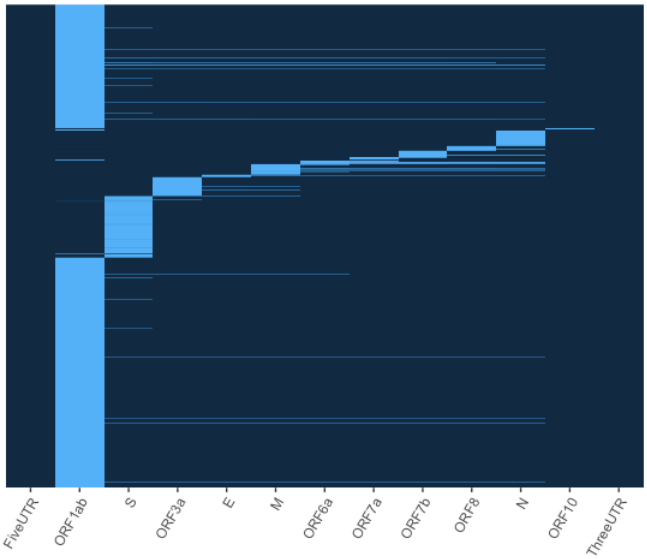
Figure 6: SARS-CoV-2 Variants Did Not Show Characteristic Patterns of Defective Interfering Genomes

For each deletion variant detected in 1,272 samples of SARS-CoV-2, we subset according to the genes that the deletion spanned, and then plotted the proportion of genome deleted against the relative abundance of the variant. In contrast to Influenza (Fig. 3), SARS-CoV-2 variants did not show an increase in abundance with proportion of genome deleted for any of the combinations of genes lost.

(a)


0
10000
20000
30000
Genome Position

(b)


FiveUTR
ORF1ab
S
ORF3a
E
M
ORF6a
ORF7a
ORF7b
ORF8
N
ORF10
ThreeUTR

(c)

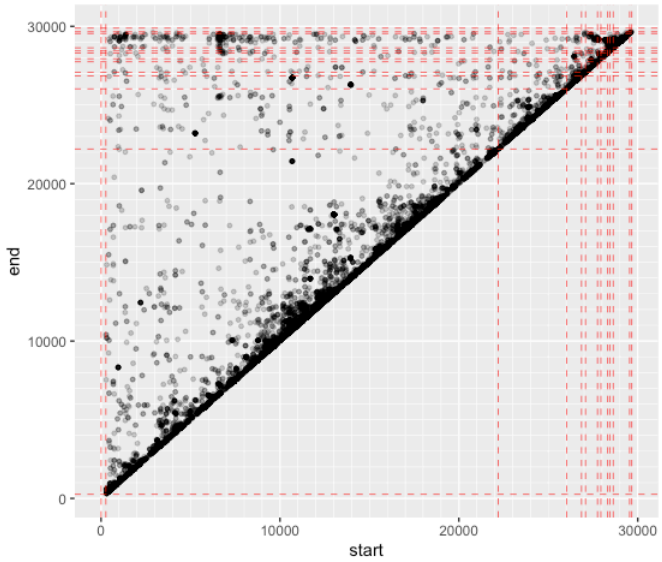


Figure 7: Patterns of SARS-CoV-2 Defective Viral Genomes

We plot the start- and end-positions of every defective viral genome found in 1,272 samples of SARS-CoV-2. In (a), the start- and end-points of each horizontal bar are the start- and end-points of each defective viral genome, with the dashed red lines indicating SARS-CoV-2 gene boundaries. In (b), each horizontal bar represents one defective viral genome, and the genes the defective viral genome disrupts are highlighted in light blue. In (c), we plot the start-position of every defective viral genome on the x-axis, and its end-position on the y-axis, with the dashed red lines indicating SARS-CoV-2 genome boundaries.

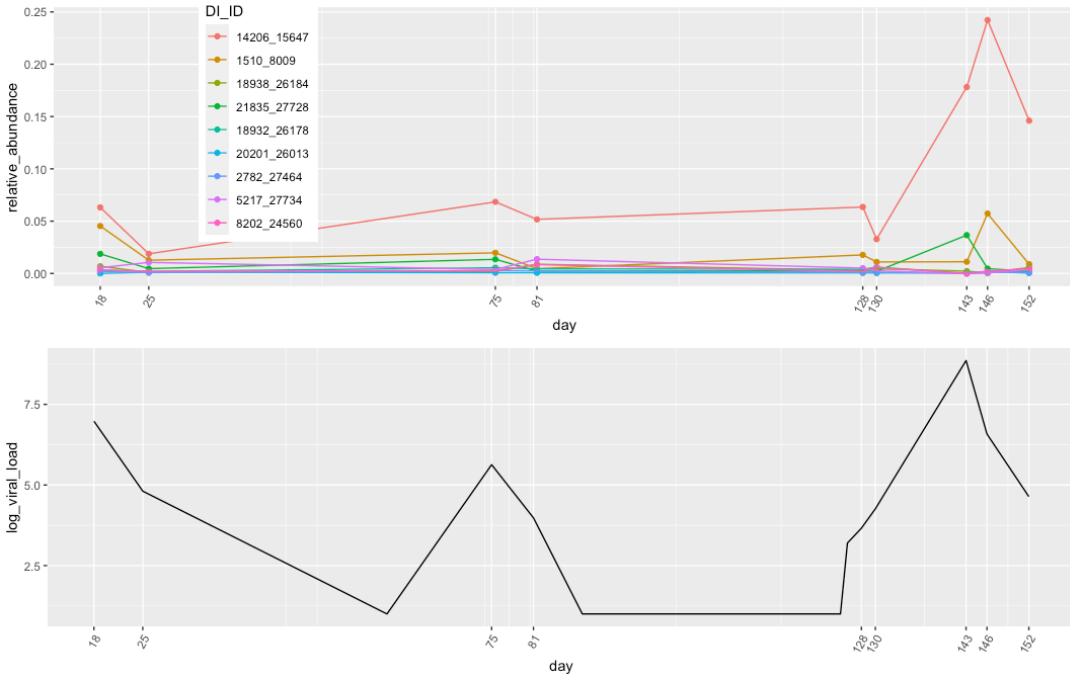


Figure 8: Longitudinal Dynamics of SARS-CoV-2 Defective Viral Genomes

We plot the relative abundance of each of nine SARS-CoV-2 defective viral genomes, each of which were found on at least eight of the nine potential timepoints. In the lower panel, we plot the log10 viral load of the patient at each timepoint.

Figure 9: Detecting Polymorphisms in SARS-CoV-2 Defective Viral Genomes

We show example output from Tablet genome visualiser (Milne *et al.*, 2013). Here, a number of reads mapping to the defective viral genome 1510_8009 are visualised. They contain a U->G polymorphism at genome position 1502, that was not found in the wild-type reads.

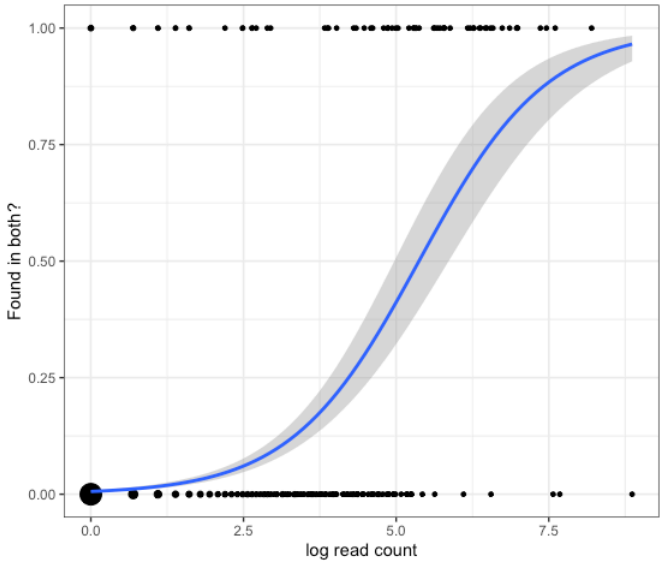


Figure 10: More Abundant SARS-CoV-2 Defective Viral Genomes Were More Likely to be Found by Both Sets of Primers

Each datapoint represents a SARS-CoV-2 defective viral genome, plotted according to whether it is found just in the original sequencing file (0) or in both the original sequencing file and when the same sample was re-sequenced using a different set of primers (1). The x-axis plots the log read count of each defective viral genome, with the curve showing a loess fit to the data. Defective viral genomes that were supported by a larger number of reads in the original sample were more likely to be found when the sample was re-sequenced using a different set of primers.

	DI_ID	number of patients found in (n=216)
1	PA_124_1938	65
2	PA_302_1854	65
3	PA_122_1936	53
4	PA_123_2005	51
5	PA_122_2004	50
6	PB1_280_2159	49
7	PB2_105_2114	48
8	PB2_106_2115	48
9	PA_122_2059	47
10	PA_128_1964	47
11	PA_123_2060	46
12	PA_124_1997	45
13	PB1_176_2031	44
14	PA_300_1852	43
15	PB1_284_2163	43
16	PA_125_1932	42
17	PB2_157_2103	41
18	PA_125_2014	40
19	PB2_155_2101	40
20	PB2_259_2057	40
21	PA_123_1961	39
22	PB2_119_2064	39
23	PA_123_1996	38
24	PA_124_1943	38
25	PA_124_2003	38
26	PA_122_1958	37
27	PA_122_2011	37
28	PA_217_1945	37
29	PB2_286_2053	37
30	PA_285_1856	36
31	PB1_141_2082	36
32	PB1_182_1978	36
33	PB2_176_2104	36
34	PB1_118_2110	34
35	PB2_156_2100	34
36	PB1_108_2156	33
37	PB1_120_2112	33
38	PB2_175_2041	33
39	PB2_201_2125	33
40	PA_124_1972	32
41	PA_280_1854	32
42	PA_302_1818	32
43	PB2_212_2157	32
44	PA_122_1929	31
45	PA_123_2002	31
46	PA_193_1938	31
47	PB2_157_2101	31
48	PB2_161_2088	31
49	PB2_173_2101	31
50	NP_346_1373	30

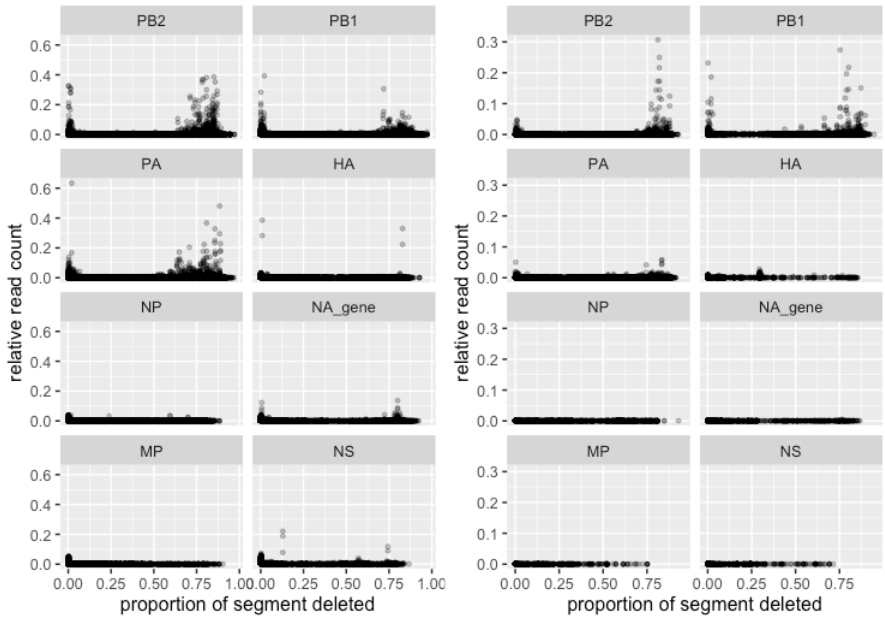
Table 1: Many Defective Viral Genomes Were Shared Between Hosts

50 defective viral genomes (with deletions more than 50% of the length of the genome segment) were found in 30 or more different individuals. 49 of these were deletions in the polymerase segments PA, PB1, or PB2.

	DI_ID	Days Found	day	accession	start	end	raw_read_count	length_abs	length_rel	rel_abundan	strand_types	mean_read_count	maintains ORF?	genes_lost	mutations?
1	14206_15647	9	18	day_18	14206	15647	320	1441	0.04821333	0.06314751	rev	0.096101522	N	ORF1ab; RdRp; Rpol_Nterminus	15657 A->U T->T synonymous (polymorphic) (~25% frequency DI reads) (1 A->T found in WT rea
2	14206_15647	9	25	day_25	14206	15647	46	1441	0.04821333	0.01877551	rev	0.096101522	N	ORF1ab; RdRp; Rpol_Nterminus	15657 A->U T->T synonymous (polymorphic) (~25% frequency DI reads) (1 A->T found in WT rea
3	14206_15647	9	75	day_75	14206	15647	129	1441	0.04821333	0.06845317	rev	0.096101522	N	ORF1ab; RdRp; Rpol_Nterminus	15657 A->U T->T synonymous (polymorphic) (~25% frequency DI reads) (0 A->T found in WT rea
4	14206_15647	9	81	day_81	14206	15647	101	1441	0.04821333	0.05171531	rev	0.096101522	N	ORF1ab; RdRp; Rpol_Nterminus	15657 A->U T->T synonymous (polymorphic) (~25% frequency DI reads) (0 A->T found in WT rea
5	14206_15647	9	128	day_128	14206	15647	198	1441	0.04821333	0.06353281	rev	0.096101522	N	ORF1ab; RdRp; Rpol_Nterminus	15657 A->U T->T synonymous (polymorphic) (~25% frequency DI reads) (1 A->T found in WT rea
6	14206_15647	9	130	day_130	14206	15647	130	1441	0.04821333	0.03280343	rev	0.096101522	N	ORF1ab; RdRp; Rpol_Nterminus	15657 A->U T->T synonymous (polymorphic) (~25% frequency DI reads) (0 A->T found in WT rea
7	14206_15647	9	143	day_143	14206	15647	843	1441	0.04821333	0.17820526	rev	0.096101522	N	ORF1ab; RdRp; Rpol_Nterminus	15657 A->U T->T synonymous (polymorphic) (~25% frequency DI reads) (4 A->T found in WT rea
8	14206_15647	9	146	day_146	14206	15647	698	1441	0.04821333	0.24223495	rev	0.096101522	N	ORF1ab; RdRp; Rpol_Nterminus	15657 A->U T->T synonymous (polymorphic) (~25% frequency DI reads) (0 A->T found in WT rea
9	14206_15647	9	152	day_152	14206	15647	530	1441	0.04821333	0.14604574	rev	0.096101522	N	ORF1ab; RdRp; Rpol_Nterminus	15657 A->U T->T synonymous (polymorphic) (~25% frequency DI reads) (1 A->T found in WT rea
10	1510_8009	9	18	day_18	1510	8009	78	6499	0.21744513	0.04536202	for	0.02092912	Y	ORF1a; nsp2; nsp3; macro; ADP-ribo	1502 U->G Y->D non-synonymous (polymorphic - 20-60%ish of DI reads) (not found in WT reads)
11	1510_8009	9	25	day_25	1510	8009	14	6499	0.21744513	0.01266968	for	0.02092912	Y	ORF1a; nsp2; nsp3; macro; ADP-ribo	1502 U->G Y->D non-synonymous (polymorphic - 20-60%ish of DI reads) (not found in WT reads)
12	1510_8009	9	75	day_75	1510	8009	17	6499	0.21744513	0.01976744	for	0.02092912	Y	ORF1a; nsp2; nsp3; macro; ADP-ribo	1502 U->G Y->D non-synonymous (polymorphic - 20-60%ish of DI reads) (not found in WT reads)
13	1510_8009	9	81	day_81	1510	8009	1	6499	0.21744513	0.00456621	for	0.02092912	Y	ORF1a; nsp2; nsp3; macro; ADP-ribo	1502 U->G Y->D non-synonymous (polymorphic - 20-60%ish of DI reads) (not found in WT reads)
14	1510_8009	9	128	day_128	1510	8009	25	6499	0.21744513	0.01777462	for	0.02092912	Y	ORF1a; nsp2; nsp3; macro; ADP-ribo	1502 U->G Y->D non-synonymous (polymorphic - 20-60%ish of DI reads) (not found in WT reads)
15	1510_8009	9	130	day_130	1510	8009	14	6499	0.21744513	0.01097609	for	0.02092912	Y	ORF1a; nsp2; nsp3; macro; ADP-ribo	1502 U->G Y->D non-synonymous (polymorphic - 20-60%ish of DI reads) (not found in WT reads)
16	1510_8009	9	143	day_143	1510	8009	12	6499	0.21744513	0.01119403	for	0.02092912	Y	ORF1a; nsp2; nsp3; macro; ADP-ribo	1502 U->G Y->D non-synonymous (polymorphic - 20-60%ish of DI reads) (not found in WT reads)
17	1510_8009	9	146	day_146	1510	8009	62	6499	0.21744513	0.05722197	for	0.02092912	Y	ORF1a; nsp2; nsp3; macro; ADP-ribo	1502 U->G Y->D non-synonymous (polymorphic - 20-60%ish of DI reads) (not found in WT reads)
18	1510_8009	9	152	day_152	1510	8009	12	6499	0.21744513	0.00883002	for	0.02092912	Y	ORF1a; nsp2; nsp3; macro; ADP-ribo	1502 U->G Y->D non-synonymous (polymorphic - 20-60%ish of DI reads) (not found in WT reads)
20	23187_24486	2	143	day_143	23187	24486	205	1299	0.04346226	0.06765677	for	0.034235549	N	Spike; Spike receptor binding; S2	24520 A -> deleted (frequency ~1/3 in DI reads; none in WT reads)
21	23187_24486	2	75	day_75	23187	24486	2	1299	0.04346226	0.00081433	for	0.03423555	N	Spike; Spike receptor binding; S2	No obvious polymorphisms (but low read count)

Table 2: Three Defective Viral Genomes Contained Polymorphisms

Three defective viral genomes contained polymorphisms that were maintained over multiple days and not found in the wild-type reads.



Supplementary Figure 1: We Observed Comparable Patterns of Influenza Defective Viral Genome Accumulation in both H3N2 and H1N1 samples.

For each deletion variant detected in each influenza genome segment, we plot the proportion of the genome segment that was deleted, against the relative read count of that variant, for H3N2 (a; n=145) and H1N1 (b; n=53) subtypes. Both small and large deletions occur in all eight segments, and some small mutants reach relatively high frequencies in all eight segments. Large deletions only reach high frequency in the three polymerase segments (PA, PB1, and PB2). These patterns are consistent with defective interfering genomes of influenza known from tissue culture studies.

References

Aaskov, J., Buzacott, K., Thu, H.M., Lowry, K. & Holmes, E.C. 2006. Long-Term Transmission of Defective RNA Viruses in Humans and Aedes Mosquitoes. *Science* **311**: 236–238.

Alnaji, F.G., Holmes, J.R., Rendon, G., Vera, J.C., Fields, C.J., Martin, B.E., *et al.* 2019. Sequencing Framework for the Sensitive Detection and Precise Mapping of Defective Interfering Particle-Associated Deletions across Influenza A and B Viruses. *J. Virol.* **93**.

Brian, D.A. & Spaan, W.J.M. 1997. Recombination and Coronavirus Defective Interfering RNAs. *Semin. Virol.* **8**: 101–111.

Choi, B., Choudhary, M.C., Regan, J., Sparks, J.A., Padera, R.F., Qiu, X., *et al.* 2020. Persistence and Evolution of SARS-CoV-2 in an Immunocompromised Host. *N. Engl. J. Med.*, doi: 10.1056/NEJMc2031364. Massachusetts Medical Society.

COG-UK. 2020. An integrated national scale SARS-CoV-2 genomic surveillance network. *Lancet Microbe* **1**: e99–e100.

Díaz-Muñoz, S.L., Sanjuán, R. & West, S.A. 2017. Sociovirology: Conflict, Cooperation, and Communication among Viruses. *Cell Host Microbe* **22**: 437–441.

Dimmock, N.J. & Easton, A.J. 2014. Defective Interfering Influenza Virus RNAs: Time To Reevaluate Their Clinical Potential as Broad-Spectrum Antivirals? *J. Virol.* **88**: 5217–5227.

Dong, X., Munoz-Basagoiti, J., Rickett, N.Y., Pollakis, G., Paxton, W.A., Günther, S., *et al.* 2020. Variation around the dominant viral genome sequence contributes to viral load and outcome in patients with Ebola virus disease. *Genome Biol.* **21**: 238.

Felt, S.A., Sun, Y., Jozwik, A., Paras, A., Habibi, M.S., Nickle, D., *et al.* 2021. Detection of respiratory syncytial virus defective genomes in nasal secretions is associated with distinct clinical outcomes. *Nat. Microbiol.* 1–10. Nature Publishing Group.

Flint, J., Racaniello, V.R., Rall, G.F. & Skalka, A.M. 2015. *Principles of Virology*, 4th ed. American Society of Microbiology.

Gelbart, M., Harari, S., Ben-Ari, Y., Kustin, T., Wolf, D., Mandelboim, M., *et al.* 2020. Drivers of within-host genetic diversity in acute infections of viruses. *PLOS Pathog.* **16**: e1009029. Public Library of Science.

Ghoul, M., Griffin, A.S. & West, S.A. 2013. Toward an evolutionary definition of cheating. *Evolution* **68**: 318–331.

Huang, A.S. & Baltimore, D. 1970. Defective Viral Particles and Viral Disease Processes. *Nature* **226**: 325–327.

Hutchinson, E.C., von Kirchbach, J.C., Gog, J.R. & Digard, P. 2010. Genome packaging in influenza A virus. *J. Gen. Virol.* **91**: 313–328.

Jaworski, E. & Routh, A. 2017. Parallel ClickSeq and Nanopore sequencing elucidates the rapid evolution of defective-interfering RNAs in Flock House virus. *PLOS Pathog.* **13**: e1006365.

Kirkwood, T.B. & Bangham, C.R. 1994. Cycles, chaos, and evolution in virus cultures: a model of defective interfering particles. *Proc. Natl. Acad. Sci.* **91**: 8685–8689.

Leeks, A., West, S.A. & Ghoul, M. 2020. Cheating in the Viral World. , doi: 10.20944/preprints201906.0106.v2. Preprints.

Leonard, A.S., McClain, M.T., Smith, G.J.D., Wentworth, D.E., Halpin, R.A., Lin, X., *et al.* 2016. Deep Sequencing of Influenza A Virus from a Human Challenge Study Reveals a Selective Bottleneck and Only Limited Intrahost Genetic Diversification. *J. Virol.* **90**: 12.

Liang, Y., Hong, Y. & Parslow, T.G. 2005. cis-Acting packaging signals in the influenza virus PB1, PB2, and PA genomic RNA segments. *J. Virol.* **79**: 10348–10355.

Makino, S., Fujioka, N. & Fujiwara, K. 1985. Structure of the intracellular defective viral RNAs of defective interfering particles of mouse hepatitis virus. *J. Virol.* **54**: 329–336. American Society for Microbiology Journals.

Manzoni, T.B. & López, C.B. 2018. Defective (interfering) viral genomes re-explored: impact on antiviral immunity and virus persistence. *Future Virol.*, doi: 10.2217/fvl-2018-0021.

Martin, M.A., Kaul, D., Tan, G.S., Woods, C.W. & Koelle, K. 2019. *The Dynamics of Influenza A H3N2 Defective Viral Genomes from a Human Challenge Study*. Evolutionary Biology.

McCrone, J.T., Woods, R.J., Martin, E.T., Malosh, R.E., Monto, A.S. & Lauring, A.S. 2018. Stochastic processes constrain the within and between host evolution of influenza virus. *eLife* **7**: e35962.

Metzger, V.T., Lloyd-Smith, J.O. & Weinberger, L.S. 2011. Autonomous Targeting of Infectious Superspreaders Using Engineered Transmissible Therapies. *PLOS Comput. Biol.* **7**: e1002015.

Milne, I., Stephen, G., Bayer, M., Cock, P.J.A., Pritchard, L., Cardle, L., *et al.* 2013. Using Tablet for visual exploration of second-generation sequencing data. *Brief. Bioinform.* **14**: 193–202. Oxford Academic.

Nayak, D.P. 1980. Defective Interfering Influenza Viruses. *Annu. Rev. Microbiol.* **34**: 619–644.

Routh, A. & Johnson, J.E. 2014. Discovery of functional genomic motifs in viruses with ViReMa–a Virus Recombination Mapper–for analysis of next-generation sequencing data. *Nucleic Acids Res.* **42**: e11–e11.

Saira, K., Lin, X., DePasse, J.V., Halpin, R., Twaddle, A., Stockwell, T., *et al.* 2013. Sequence Analysis of In Vivo Defective Interfering-Like RNA of Influenza A H1N1 Pandemic Virus. *J. Virol.* **87**: 8064–8074.

Shirogane, Y., Rousseau, E., Voznica, J., Xiao, Y., Su, W., Catching, A., *et al.* 2021. Experimental and mathematical insights on the interactions between poliovirus and a defective interfering genome. *bioRxiv*, doi: 10.1101/2021.01.11.426198.

Vasilijevic, J., Zamarreño, N., Oliveros, J.C., Rodriguez-Frandsen, A., Gómez, G., Rodriguez, G., *et al.* 2017. Reduced accumulation of defective viral genomes contributes to severe outcome in influenza virus infected patients. *PLOS Pathog.* **13**: e1006650.

Viehweger, A., Krautwurst, S., Lamkiewicz, K., Madhugiri, R., Ziebuhr, J., Hölzer, M., *et al.* 2019. Direct RNA nanopore sequencing of full-length coronavirus genomes provides novel insights into structural variants and enables modification analysis. *Genome Res.*, doi: 10.1101/gr.247064.118.

Vignuzzi, M. & López, C.B. 2019. Defective viral genomes are key drivers of the virus–host interaction. *Nat. Microbiol.* 1.

West, S.A., Griffin, A.S. & Gardner, A. 2007. Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. *J. Evol. Biol.* **20**: 415–432.

Xie, Q., Cao, Y., Su, J., Wu, J., Wu, X., Wan, C., *et al.* 2017. Two deletion variants of Middle East respiratory syndrome coronavirus found in a patient with characteristic symptoms. *Arch. Virol.* **162**: 2445–2449.

Xue, K.S. & Bloom, J.D. 2018. Reconciling disparate estimates of viral genetic diversity during human influenza infections. *bioRxiv* 364430.

7 Discussion

In this thesis, I have studied different aspects of the social lives of viruses, with each chapter drawing its roots from social evolution theory. I suggest that this general approach could be useful going forward for studying other aspects of the social lives of viruses. Each chapter has its own discussion, which I will not repeat here. Instead I will use this chapter to examine some general themes and issues.

I will start by discussing how social evolution differs from existing frameworks within virology, in particular quasispecies theory. Then I will consider some potential future directions, focusing on how social evolution theory could help to provide new answers to existing questions within virology, and could open up new areas of investigation.

7.1 Social evolution and quasispecies theory

Social evolution theory is not the first theoretical framework to provide a way of thinking about virus-virus interactions. Quasispecies theory is an elegant mathematical framework, originally developed for the evolution of early RNA replicators, that deals with the dynamics of large populations of asexual replicators that have extremely high mutation rates (Eigen, 1971; Eigen & Schuster, 1977). Quasispecies theory has historically been popular among theoreticians and empirical

virologists alike, and has introduced a number of key concepts to virology, such as mutational meltdown, and the idea of the error catastrophe (Andino & Domingo, 2015). By focussing attention on the role of virus-virus interactions, quasispecies theory has also driven a large number of seminal empirical studies, without which sociovirology may not have been possible (Holland et al., 1992; Vignuzzi et al., 2006; Lauring & Andino, 2010; Domingo et al., 2012; Andino & Domingo, 2015; Bordería et al., 2015).

However, despite these achievements, the intuition provided by quasispecies theory differs from the logic of social evolution theory in several key ways. Moving forward, I suggest that social evolution theory can provide an alternative framework to quasispecies theory for studying virus-virus interactions.

7.1.1 Survival of the fittest, or survival of the flattest?

The aspect of quasispecies theory that is most relevant for virus-virus interactions is a phenomenon known as 'survival of the flattest'. An early insight from quasispecies theory was that when mutation rates were high enough, some of the standard principles of Darwinian natural selection break down (Eigen, 1971; Eigen & Schuster, 1977). Among these principles is the basic concept of 'survival of the fittest', namely that natural selection will favour of the genotype with the highest fitness. When mutation rates are high enough, rather than acting on individual genotypes, selection may instead on entire mutational neighbourhoods of genotypes, often termed 'mutant clouds' or 'quasispecies'. Consequently, selection can favour mutationally robust groups of genotypes with a high average fitness, even if these neighbourhoods do not contain the genotype that has the highest fitness (Wilke et al., 2001).

One implication of 'survival of the flattest' is that it is not the individual viral genome that is the unit of adaptation, but rather the quasispecies, or mutant cloud. Consequently, adaptations can be selected for if they are beneficial for the whole group of viruses, even if they are not beneficial for the individual viral genome bearing them. If this is really the case, then it offers a drastically different perspective on view virus-virus interactions when compared to social evolution

theory. If adaptation does manifest at the level of the group of viruses, then beneficial virus-virus interactions should be the norm, just as beneficial interactions between cells in the human body are the norm. Rather than seeking to explain when cooperation evolves, we should assume that cooperation is the baseline evolutionary condition, and that it will occur whenever there is a mechanistic opportunity for cooperation. At the same time, apparently negative virus-virus interactions, such as the examples of viral cheating explored in Chapter Two, must be able to be explained as adaptations that ultimately benefit the group of viruses as a whole.

7.1.2 Quasispecies in reality

In reality, 'survival of the flattest' requires mutation rates far higher than those that are found naturally among viruses (Holmes & Moya, 2002; Holmes, 2010). At these lower, more realistic mutation rates, quasispecies theory collapses into standard population genetics, and so there is no disagreement with the logic of social evolution theory (Wilke, 2005). However, the term quasispecies has mutated within virology, and is now commonly used to refer to any genetically diverse viral population, especially if interactions occur between genotypes within that viral population (Domingo et al., 2012).

Just as the term quasispecies lives on virology, so too does the intuition provided by 'survival of the flattest'. Empirical work following in the quasispecies tradition frequently starts from the baseline assumption that the viral population as a whole is being selected, and that this non-Darwinian form of natural selection is the reason why beneficial virus-virus interactions have evolved (Holland et al., 1992; Holmes & Moya, 2002). However, we know that real viral mutation rates are too low for 'survival of the flattest' to really be occurring, and so these interactions are unlikely to actually be group adaptations. Consequently, the assumption that the quasispecies is the unit of adaptation is potentially dangerous, because it can prevent us asking questions about the selective forces that are really operating.

7.1.3 Social evolution as an alternative

The problem is therefore not that quasispecies theory is technically in conflict with social evolution theory, but rather that it provides a conflicting set of intuitions and baseline assumptions. If we start with the baseline assumption that the group of viruses is the unit of adaptation, then if we see beneficial virus-virus interactions, we are unlikely to ask why they have evolved. Meanwhile, if we see negative virus-virus interactions, such as cheating, we may seek answers at the group level that are simply not relevant. We may follow red herrings, while missing fruitful lines of enquiry.

Social evolution theory provides an alternative set of intuitions and baseline assumptions that, I suggest, can be more useful in understanding the evolution of virus-virus interactions. When we see interactions between viruses, social evolution theory prompts us to ask why they have evolved, and to classify them in a formal way based on their fitness consequences. When interactions are cooperative, we have a set of intuitions, based on assumptions that are relevant for viruses, for when the trait can be favoured by natural selection. These intuitions are backed up by a mature body of formal theory that easily lends itself to empirical falsification. When we see negative virus-virus interactions, we do not need to seek convoluted explanations at the level of the group, and can instead offer more parsimonious explanations, that then open new avenues to explore. And even beyond these practical considerations, social evolution theory allows us to consider all social interactions, viral or otherwise, within a broad framework that spans the breadth of life on Earth (Maynard Smith & Szathmary, 1995; Bourke, 2011; Davies et al., 2012).

7.1.4 Quasispecies theory and group selection

There are some parallels between quasispecies theory and the concept of group selection. Within evolutionary biology, cooperation was long seen as a problem in need of explanation, and for a large part of the 20th century, group selection provided that explanation (Darwin, 1859; Williams, 1966). The essential logic was that groups that cooperated more would outcompete groups that did not, and therefore we could explain instances of cooperation as examples of group-level adaptations.

In reality, however, this kind of group adaptationism does not operate in nature. With the advent of inclusive fitness theory, a more parsimonious and empirically successful framework became available for explaining cooperation (Hamilton, 1964). While group selection has since been reformulated to provide a mathematical alternative to kin selection models, the field has made substantial progress by discarding the intuition of group adaptation in favour of the intuition provided by inclusive fitness theory (West et al., 2008; Bourke, 2011; Davies et al., 2012).

There may be formal links between quasispecies theory and group selection. One way of thinking about 'survival of the flattest' is that high mutation rates align the evolutionary interests of genotypes that are likely to mutate into one another. This could be analogous to the concept of relatedness in evolutionary biology; the evolutionary interests of individuals become increasingly aligned as they are increasingly likely to share genes (Hamilton, 1964). In evolutionary biology, group adaptation can occur, but only when relatedness is sufficiently high that the evolutionary interests of the group subsume those of the individual (Gardner & Grafen, 2009); in quasispecies theory, group adaptation can occur when genotypes mutate into one another with sufficient likelihood that selection favours the fittest group of genotypes, rather than the fittest individual genotype (Wilke et al., 2001). It may be possible to draw a formal link between these phenomena, and to express the 'mutational linkage' of quasispecies theory in a way that is mathematically analogous to relatedness, perhaps a form of 'mutational relatedness' (Frank, 1998). Drawing such a formal link could bring together quasispecies theory and social evolution theory into one framework, perhaps resolving their apparent disagreements.

There are also informal links between quasispecies theory and group selection. For both bodies of theory, the intuitions provided by the framework outlived their formal justification in real biological systems. This highlights how the value of a mathematical framework is not just in the formal models that it allows, but also in the informal intuitions that it generates. I suggest that there could be parallels between the empirical progress made when social evolution moved beyond the

intuitions of group selection theory, and the empirical progress that virology could make by moving beyond the intuitions of quasispecies theory.

7.2 New answers

7.2.1 Conflict and Genome Splitting

In contrast with quasispecies theory, social evolution theory places a much larger emphasis on conflict within viral populations. One area where this focus on conflict could be useful is the evolution of genome segmentation and multipartite viruses. Viruses display substantial diversity in the way their genomes are organised, with some containing all genetic information on a single genome, others splitting information across multiple genome segments analogous to chromosomes (segmented viruses), and yet others packaging each segment in its own virion (multipartite viruses) (Nee, 1987; Lucía-Sanz & Manrubia, 2017). Multipartite viruses in particular are puzzling, because the costs of such a lifestyle are obvious and substantial, while the benefits are much less clear.

Most evolutionary explanations for this diversity in genome organisation focus on group-level advantages. In general, models are constructed with the assumption that genome-splitting will evolve if a population of viruses with a split genome can outcompete a population of viruses without a split genome (Iranzo & Manrubia, 2012). Mechanisms that could theoretically allow this include the possibility that segmented genomes could allow for better regulation of gene expression, or that multipartite viruses could allow for a degree of plasticity when infecting new hosts (Holmes, 2009; Zwart & Elena, 2020).

However, an alternative explanation could be that genome splitting is driven by conflicts within viral populations. Viral cheating frequently occurs when mutants lose a section of their genome, and rely on coinfecting cells with a complete wild-type genome in order to make up for the lost gene function. If an initially non-segmented virus encoded multiple gene products that could be complemented in this way, it is plausible that multiple cheats could arise, each cheating a different gene function, and that such a population could persist via these cheats complementing one another

in coinfection. Since this mechanism relies entirely on selection acting on individual viral genomes, it could operate even if the resulting population of segmented viral genomes is overall less productive than the initial, non-segmented population. In principle it could therefore apply in a broader range of cases, and does not require us to search for elusive group-level advantages for genome splitting.

Such a hypothesis would require formal modelling to determine when it is plausible and could be supported by experimental evolution studies in viruses where multiple types of cheat are possible. This kind of mechanism could also apply to cases of genome splitting outside of viruses, such as cicada endosymbionts or mitochondria in Protists (Campbell & McCutcheon, 2015).

However, the key point here is that by focussing on the evolutionary forces shaping individual viral genomes, we are able to think of new kinds of explanation for otherwise puzzling empirical patterns. By placing the emphasis on conflict between individuals, social evolution theory reminds us that natural selection does not always lead to the best outcome for the group.

7.2.2 Adaptive cheating

Another example where breaking the assumption of group optimality could be useful concerns the evolution of defective interfering genomes. As explored in Chapter Two, defective interfering genomes occur in almost all known viruses and can act as cheats, interfering with the accumulation of the wild-type virus and reducing the overall viral load. One consequence of this is that defective interfering genomes could prolong the duration of a viral infection (Vignuzzi & López, 2019). Based on the idea that longer infections could allow for more transmission opportunities, a number of authors have suggested that viruses could evolve in order to produce defective interfering genomes (Li et al., 2011; Poirier et al., 2018; Rezelj et al., 2018). If this were true, defective interfering genomes could be seen as an adaptation that improved the overall fitness of the population.

From a social evolution perspective, this possibility seems counterintuitive, because viruses often incur enormous fitness consequences when interacting with

defective interfering genomes (Shirogane et al., 2021). Consequently, we would expect that viruses should be selected to avoid producing defective interfering genomes.

One way the suggestion could be 'rescued' is if producing defective interfering genomes had a negative effect on the growth rate of all the viral variants within a host, not just the one that produced them. In that case, it is conceivable that viruses could evolve to produce defective interfering genomes as an alternative to slower growth, in order to reduce the level of virulence without sacrificing a competitive advantage against other variants sharing a host. This possibility, which would be analogous to the evolutionary concept of 'policing', would require formal modelling, and may only work under restrictive conditions (Frank, 1995). More broadly, however, social evolution theory highlights when verbal models could be misleading, and provides a toolkit for building formal models that can test the evolutionary plausibility of verbal hypotheses.

7.2.3 The Phenotypic Gambit

I have suggested that a key strength of social evolution theory is the fact that it focuses on the fitness consequences of traits, ignoring mechanistic and genetic detail. On a practical level, this means that social evolution models often deploy a set of assumptions known collectively as the phenotypic gambit (Grafen, 1991). Most of the models in this thesis and proposed in this discussion use the phenotypic gambit; they link phenotypes directly to fitness consequences, ignoring the underlying genetic architecture that produces those phenotypes. This approach is a 'gambit' because it necessarily sacrifices the ability to predict selection on gene frequencies in any given system. In exchange for that sacrifice, the modeller gains a degree of generality, with the ability to predict how traits might evolve across different mechanistic systems and genetic architectures.

The phenotypic gambit has been used to great effect within social evolution and behavioural ecology (Davies et al., 2012). Conceptually, the phenotypic gambit is appealing because it focuses on selective forces that could operate across multiple systems, rather than mechanistic details limited to just a single system. It also brings

empirical advantages, by generating predictions that can be tested comparatively across a range of species. The extent to which these predictions will be successful will depend on features of the trait being investigated. The gambit works best on traits that have substantial fitness consequences and that have been selected upon across many different species for long periods of time. Under these conditions, we can expect that any genetic constraints that could once have prevented the evolutionary optimum being reached will have been eroded by the consistent action of natural selection over time. Many traits studied by behavioural ecologists have these properties, with sex allocation being a prime example, resulting in an unusually precise fit between theory and data (West, 2009).

To what extent can we expect similar empirical successes in viruses? The conceptual advantages of using the phenotypic gambit to study viruses are clear, but there may be limitations when it comes to making empirical predictions. Evolutionary virology often uses experimental methods, notably experimental evolution, to test evolutionary hypotheses (Elena & Sanjuán, 2007). However, these studies are not always suitable for testing predictions made using the phenotypic gambit, because for any given system there may be constraints that prevent the long-term optimum being reached in observable timescales, even in organisms that evolve as quickly as viruses do.

Comparative methods are more appropriate for testing models that use the gambit, but these require us to identify multiple instances of convergent evolution. A number of barriers exist that currently make this difficult in viruses. Firstly, our knowledge of the distribution of traits across viruses is very limited. For example, collective infectious units have been described in some viruses but not others, but does this reflect a real pattern, or just sampling bias? This is in stark contrast to comparative studies in larger organisms, where reliable databases of traits such as bird clutch size or offspring sex ratio are readily available, or at least relatively easy to build. Secondly, there is no overall phylogeny of viruses. Even if viruses do share a common origin, and such a phylogeny is possible, we may be unable to construct it, since viruses' small genomes and fast rates of evolutionary change preclude deep

genetic homology (Holmes, 2009). In the absence of a well-resolved phylogeny, it is difficult to determine the number of independent origins of viral traits. Finally, at present we have only described a tiny fraction of viral species (Zhang et al., 2018). Even if we could conduct a comparative study across viruses, the potential scope may be limited, as we would likely only be testing our hypotheses on a small subset of the viral diversity that really exists.

For viruses, the answer may be to combine models, by first using the gambit to test ideas conceptually, and then using more mechanistically detailed models amenable to experimental falsification. Alternatively, comparative studies may be possible on a more limited scale within viral groups that have well-resolved phylogenies. This latter possibility has been conducted in viruses such as Influenza, which have a fast evolutionary rate that has resulted in many instances of convergent evolution (Escalera-Zamudio et al., 2020).

7.3 New questions

7.3.1 Natural history

As well as offering new insights on existing problems, social evolution also highlights new questions that might not otherwise have come to light. One example is the world of defective interfering genomes. While there is currently a renaissance of interest in these within virology, most attention focuses on their mechanistic consequences, and less on their natural history or comparative biology (Manzoni & López, 2018; Vignuzzi & López, 2019; Yang et al., 2019; Alnaji & Brooke, 2020; Ziegler & Botten, 2020). Some patterns are clear: defective interfering genomes exploit similar gene products across different viruses, and some viruses are exploited by defective interfering genomes more than others. But we are currently very far from being able to quantify these patterns, and even further from being able to explain them.

Chapter Six in this thesis exemplifies this problem: SARS-CoV-2 is a coronavirus, a group that has previously been a model system for defective viral genome research (Brian & Spaan, 1997). Despite this, we have no strong a priori expectations about how common SARS-CoV-2 defective interfering genomes should be, nor what

they should look like. Given the possible role of defective interfering genomes in determining disease outcome, and their potential to be used as antiviral therapeutics, this lack of predictive power is a considerable blind spot (Metzger et al., 2011; Dimmock & Easton, 2014; Vasilijevic et al., 2017; Levi et al., 2021). More broadly, a central aim of evolutionary biology is to explain patterns in natural history, and social evolution has been enormously successful in this regard (Davies et al., 2012; West et al., 2021). As the most diverse and abundant type of cheat in the natural world, a key goal of sociovirology must be to understand the natural history of defective interfering genomes.

However, investigating the natural history of defective interfering genomes is not easy. Many of the generic barriers to comparative biology in viruses outlined in the previous section are relevant, but there are also specific difficulties relating to defective interfering genomes. For one, they are a moving target, since the rate at which a virus produces defective interfering genomes may depend on many factors, such as the cell type, host type, or anything that influences the mutation rate (Alnaji & Brooke, 2020; Levi et al., 2021). Defective interfering genomes must also spread before they can be detected; this depends on the rate of coinfection, which could also vary between cell or host types (Gallagher et al., 2018).

There is also the challenge to combine results from laboratory work with findings from natural viral infections. Laboratory work is more likely to be able to disentangle mechanisms, such as whether the abundance of a defective interfering genome is determined by production by the wild-type virus, or by the defective interfering genome's ability to interfere. However, ultimately we need data on what happens in the environment in which viruses have evolved. This will require far more sampling from natural viral infections than is currently available, and a better understanding of how different sequencing methods influence our ability to detect defective viral genomes (Chapter Six).

There is also an evolutionary complication here, since the rate at which viruses produce defective interfering genomes could itself be a trait that is under selection. In section 7.2.2 of this discussion, I explored the possibility that viruses might

evolve to produce defective interfering genomes more frequently. Perhaps more plausible is the idea that viruses are under selective pressure to produce fewer defective interfering genomes, mirroring other mechanisms of resistance (DePolo et al., 1987). If this is the case, then rates of defective interfering genome production in nature could reflect an ongoing evolutionary arms race rather than a static quantity. This could also produce a 'chicken and egg' problem whereby viruses that are more vulnerable to defective interfering genomes are more likely to evolve resistance mechanisms (Chapter Two; p.42).

7.3.2 Cooperation?

Dealing with ambiguity

In evolutionary biology, a distinction is made between traits that are cooperative, and traits that are mutually beneficial, but not cooperative. Cooperative traits are those that provide a benefit to a recipient, and have evolved at least partially because of these benefits (West et al., 2007). However, some traits provide a mutual benefit to both actor and cooperator, but evolved for a reason that is incidental to these benefits, and are therefore not cooperative.

To illustrate this distinction, consider birds nesting in a tree. The birds may benefit from the shape of the boughs and the cover of the leaves, allowing them to construct a sheltered nest, while the tree may benefit from the presence of the birds, for example if they eat insect herbivores. However, this is not an example of cooperation in the evolutionary sense, because the benefit that the birds gain from the shape of the trees' boughs is purely incidental to the evolution of the tree's boughs; in the absence of the birds, the tree would look no different. A counterexample that is cooperative would be the 'domatia' of acacia trees, which are specialised structures produced by the plant in order to house ants, that then defend the tree from herbivores. In this case, constructing the domatia is a cooperative trait; without the ants, the tree would be selected to stop constructing domatia, and so we can say that the domatia have evolved because of the fitness benefit they provide to the ants.

In viruses, it can be difficult to draw a sharp distinction between these two types of trait. One example of this kind of ambiguity is the beneficial coinfection in Influenza modelled in Chapter Three. Here, the putatively cooperative trait is encoding a modified 'G version' of Neuraminidase. When wild-type genomes and G-variant genomes infect the same host cells, virions are produced that display both versions of Neuraminidase, and which consequently have a greater ability to infect new cells than virions produced from singly infected cells. This interaction provides a large enough fitness benefit that it allows the G-variant to persist in the population, even though the G-variant is an order of magnitude less fit than the wild-type in the absence of coinfection. At the same time, the beneficial interaction means that both wild-type and G-variant increase in absolute frequency, bringing mutual benefits to both variants.

In this example, encoding the G-variant of Neuraminidase provides a mutual benefit, but does that mean that the G-variant is selected for because of this mutual benefit? Is it more like the boughs of a tree, or more like domatia, and how would we tell? In this case, the trait could plausibly be cooperative or non-cooperative, and it is difficult to distinguish between these possibilities.

The mutual benefits that define this interaction could provide a basis for cooperation to favour encoding the G-version of Neuraminidase. The G-variant does better in cells that are also infected by the wild-type, and so the G-variant benefits if the wild-type is at a higher frequency in the population. Therefore, selection on the G-variant could be driven at least partially by this benefit to the wild-type, in a way that is loosely analogous to reciprocity; if the wild-type does well, this provides more opportunity for beneficial interactions in the next generation, that will benefit the next generation of G-variant genomes.

Alternatively, the trait could be a non-cooperative adaptation to the social environment, that is selected for entirely because of the benefit the G-variant gains in coinfected cells. To illustrate this possibility, consider a hypothetical experiment in which the experimenter adds the wild-type version of Neuraminidase to each host cell. In this experiment, the G-variant would increase in frequency, as it adapted to

the presence of the wild-type Neuraminidase. However, this would not be driven by cooperation, since there is no recipient to receive a benefit.

This case highlights the potential difficulties that can arise when trying to incorporate new biological examples within an existing evolutionary framework. Here, both non-cooperative and cooperative explanations for the trait result in the same qualitative outcome: the G-variant increases in frequency. Therefore, distinguishing between them is practically very difficult, even if theoretically the two mechanisms would lead to quantitatively different outcomes. In this case, we can gain insight into the evolutionary dynamics of the interaction without needing to label it as cooperative or not, as in Chapter Three of this thesis. But will this always be the case? When is it important to distinguish cooperation from non-cooperation?

When do semantics matter?

The G-variant of Influenza provides one example, but more broadly, this kind of ambiguity is widespread within virology. Most notably, the field of minor variant interactions has revealed a plethora of cases where viral variants interact in mutually beneficial ways. These range from simple interactions such as complementing missing gene products, all the way through to purportedly complex adaptations where different variants of a virus 'distract' different components of adaptive immune systems (so-called 'antigenic cooperation') (Skums et al., 2015; Brooke 2017). Virologists usually refer to these kinds of traits as cooperative, even though in an evolutionary sense it is ambiguous whether they really are. Given that in many of these cases it will be practically very difficult to distinguish cooperation from non-cooperation, does it matter, and should we try?

I suggest that in cases where it is ambiguous, it is useful to be cautious and avoid describing a trait as cooperative. The reason is that cooperative traits require specific conditions to evolve and be maintained, that depend on the reasons why cooperation is favoured. For example, if a trait is altruistic, then it is maintained by kin selection, and so the trait may be selected against when relatedness is low, even if the trait is relatively essential. When traits really are unambiguously cooperative,

then investigating the evolutionary mechanisms at play can provide valuable insight into the selective forces that maintain the trait.

A good example of this is interferon-blocking in mammalian viruses. A careful mixture of experiments and modelling in vesicular stomatitis virus have dissected the fitness consequences of this trait, demonstrating that it is unambiguously a case of altruistic cooperation (Domingo-Calap et al., 2019). In the process, this work has shown that interferon-blocking is maintained by kin selection, that cheats can evolve, and has even quantified the minimum genetic relatedness required to maintain the trait.

This kind of work then opens the door to many useful avenues going forward. Although it seems counterintuitive that a trait as essential interferon-blocking can be selected against, we now know that it can be, because it is cooperative. When relatedness varies throughout time during an infection, or across different infections, we can now link this variation to the evolutionary dynamics of interferon-blocking, and in turn to viral dynamics at the population level. This offers an insight into the pathogenicity and dynamics of viral infections that would not have been possible if we did not know that the trait was altruistic.

We also gain new ways of looking at viral data, that could explain otherwise puzzling trends. Interferon-blocking is a common trait found across mammalian viruses (Flint et al., 2015). It is also accompanied by much genetic variability; viral mutants that lack the ability to suppress interferon are common, including in clinically important viruses such as SARS-CoV-2 and Influenza (Konno et al., 2020; Martin et al., 2020; Sun et al., 2020; Yuen et al., 2020). Could these variants be cheats, exploiting the cooperative nature of interferon-suppression? If they are, does this suggest that interferon-suppression breaks down commonly in natural infections? Given the drastic fitness consequences for viral populations when interferon-suppression does break down, these social dynamics have the potential to play a substantial role in shaping viral population dynamics and pathogenicity.

The example of interferon-blocking shows us that by understanding the selective forces that underlie cooperative traits, we gain insights into viral dynamics and

evolution that would not otherwise be possible. This field is currently in its infancy, and interferon-blocking is just one of so many viral traits where social dynamics are likely to be important. There is enormous potential here for improving our understanding of virus evolution and viral population dynamics.

However, if we are to capitalise on this potential, it is critical to use the formal definitions for cooperation and altruism that social evolution theory provides. Evolutionary definitions of altruism and cooperation are not arbitrary, but are defined in terms of fitness consequences, the ultimate currency of natural selection (West et al., 2007). When these definitions are combined with careful virology, as in interferon-blocking, they can transform the way that we investigate even the most fundamental of viral traits. In contrast, if the fitness consequences of interferon-blocking had not been carefully dissected, and if these consequences had not been linked to the maintenance of the trait, then none of these insights would be possible. The key first step was to ask whether interferon-blocking really is cooperation in the evolutionary sense, and then to pin down the selective forces that maintain it. The opportunity to do this is lost when evolutionary definitions are applied too liberally, such as when every beneficial virus-virus interaction is described as cooperative.

Transitions to cooperation

Viruses are not the only organisms where it is difficult to distinguish between cooperative and non-cooperative explanations for mutually beneficial traits. This issue also comes up when studying cross-feeding interactions in bacteria. Many bacterial species have evolved to digest one another's waste products, forming mutually beneficial partnerships. In most cases, these interactions are not examples of evolutionary cooperation, because each species would produce the waste product regardless of whether the product is digested by another species. However, there are cases where such initially non-cooperative interactions can transition into cooperative mutualisms, such as in experimental populations of Salmonella enterica and Escherichia coli. Here, each species evolves to secrete a costly resource, gaining a fitness advantage through increasing the frequency of its partner species and

therefore having more of the other species' waste product to digest (Harcombe et al., 2018). When these non-cooperative interactions become cooperative in this way, each species can then be invaded by cheats, that do not produce the costly cooperative product; such cheats do not exist when in non-cooperative cross-feeding interactions.

Could similar transitions occur in viruses? Are there cases where viruses have evolved to cooperate, in order to capitalise upon an existing mutually beneficial interaction? One example could be the evolution of collective infectious units. In Chapter Four, we considered how the evolution of collective infectious units could be driven by beneficial interactions between viral genomes. However, we did not consider that there is like to be a cooperative component to constructing a collective infectious unit. For example, constructing a larger viral capsid will require costly gene products that act as public goods, and that could be exploited by public goods cheats. More broadly, given how common virus-virus interactions are, it is plausible that many examples of viral cooperation could be driven by the benefits of capitalising upon existing, non-cooperative, mutually beneficial interactions.

7.3.3 Double-edged sword

It is clear that both positive and negative interactions are important for viral evolution. Beneficial interactions can be critical for viral population dynamics, as in Chapters Three and Four of this thesis, and the rapidly growing subfield studying minority variant interactions. Even the basic dynamics of viral genome replication depend on positive feedbacks between viral genomes and genome products such as replicase enzymes (Andreu-Moreno et al., 2020).

However, these kinds of beneficial interactions can often be exploited by viral cheats, with potentially drastic fitness consequences, as highlighted by Chapter Two in this thesis, and the subfields of virology that study defective interfering genomes, satellites, and other viral cheats. Many of these cheats are common and can arise spontaneously, meaning that virus-virus interactions are potentially double-edged; viruses face the challenge of balancing the benefits of interactions with the potential for exploitation.

Balancing these benefits may be easiest for interactions between genetically identical viruses. For example, the kinds of collective benefit that underpin the evolution of collective infectious units in Chapter Four, or the positive feedbacks involved in viral replication (Andreu-Moreno et al., 2020). In these cases, viruses may be able to restrict interactions to genetically identical sibling genomes, achieving the benefits of interaction while minimising the risks of exploitation by cheats. The need to control relatedness in this way could shape the types of collective infectious unit that evolve, and could be one reason for the evolution of superinfection exclusion (Bondy-Denomy et al., 2016; Andreu-Moreno & Sanjuán, 2020).

However, it would appear to be a more difficult balancing act when the benefits of interaction depend on interactions between different viral genotypes. Examples of this kind of interaction include the coinfection benefits modelled in Chapter Three, and likely the majority of minority variant interactions (Vignuzzi et al., 2006; Bordería et al., 2015; Xue et al., 2016). In these cases, controlling relatedness might avoid the risk of cheats, but it would also preclude the benefits of these interactions. It is possible that viruses have mechanisms that we have not yet thought of, that allow them to achieve the best of both worlds here. Alternatively, perhaps the threat of cheating is an ever-present constraint that prevents viruses from taking advantage of some of the benefits of sociality.

7.4 Conclusion

Viruses are social organisms, and social evolution is the study of social organisms. Using social evolution theory to study viruses can offer new answers for both fundamental and applied virology, while also posing new questions and opening up new fields of enquiry. At the same time, incorporating viruses within the broader field of social evolution provides an excellent opportunity to test and expand a mature and successful body of theory. Incorporating viruses is not only beneficial, but I would argue it is necessary for any theory that offers general answers about the evolution of life on Earth.

7.5 References

- Alnaji, F.G. & Brooke, C.B. 2020. Influenza virus DI particles: Defective interfering or delightfully interesting? PLOS Pathog. 16: e1008436. Public Library of Science.

- Andino, R. & Domingo, E. 2015. Viral quasispecies. Virology 479–480: 46–51.

- Andreu-Moreno, I., Bou, J.-V. & Sanjuán, R. 2020. Cooperative nature of viral replication. Sci. Adv. 6: eabd4942. American Association for the Advancement of Science.

- Andreu-Moreno, I. & Sanjuán, R. 2020. Collective Viral Spread Mediated by Virion Aggregates Promotes the Evolution of Defective Interfering Particles. mBio 11.

- Bondy-Denomy, J., Qian, J., Westra, E.R., Buckling, A., Guttman, D.S., Davidson, A.R., et al. 2016. Prophages mediate defense against phage infection through diverse mechanisms. ISME J. 10: 2854–2866.

- Bordería, A.V., Isakov, O., Moratorio, G., Henningsson, R., Agüera-González, S., Organtini, L., et al. 2015. Group Selection and Contribution of Minority Variants during Virus Adaptation Determines Virus Fitness and Phenotype. PLOS Pathog. 11: e1004838.

- Bourke, A.F.G. 2011. Principles of Social Evolution. Oxford University Press, Oxford, New York.

- Brian, D.A. & Spaan, W.J.M. 1997. Recombination and Coronavirus Defective Interfering RNAs. Semin. Virol. 8: 101–111.

- Brooke, C.B. 2017. Population diversity and collective interactions during influenza virus replication and evolution. J. Virol. JVI.01164-17.

- Campbell, M.A. & McCutcheon, J.P. 2015. Genome expansion via lineage splitting and genome reduction in the cicada endosymbiont Hodgkinia. Proc Natl Acad Sci U S A. 112(33):10192-9.

- Darwin, C. 1859. On the Origin of Species. John Murray and sons, London.

- Davies, N.B., Krebs, J.R. & West, S.A. 2012. An Introduction to Behavioural Ecology, 4th ed. John Wiley & Sons.

- DePolo, N.J., Giachetti, C. & Holland, J.J. 1987. Continuing coevolution of virus and defective interfering particles and of viral genome sequences during undiluted passages: virus mutants exhibiting nearly complete resistance to formerly dominant defective interfering particles. J. Virol. 61: 454–464.

- Dimmock, N.J. & Easton, A.J. 2014. Defective Interfering Influenza Virus RNAs: Time To Reevaluate Their Clinical Potential as Broad-Spectrum Antivirals? J. Virol. 88: 5217–5227.

- Domingo, E., Sheldon, J. & Perales, C. 2012. Viral Quasispecies Evolution. Microbiol. Mol. Biol. Rev. 76: 159–216.

- Domingo-Calap, P., Segredo-Otero, E., Durán-Moreno, M. & Sanjuán, R. 2019. Social evolution of innate immunity evasion in a virus. Nat. Microbiol. 1.

- Eigen, M. 1971. Selforganization of matter and the evolution of biological macromolecules. Naturwissenschaften 58: 465–523.

- Eigen, M. & Schuster, P. 1977. The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle. Naturwissenschaften 64: 541–565.

- Elena, S.F. & Sanjuán, R. 2007. Virus Evolution: Insights from an Experimental Approach. Annu. Rev. Ecol. Evol. Syst. 38: 27–52.

- Escalera-Zamudio, M., Golden, M., Gutiérrez, B., Thézé, J., Keown, J.R., Carrique, L., et al. 2020. Parallel evolution in the emergence of highly pathogenic avian influenza A viruses. Nat. Commun. 11: 5511. Nature Publishing Group.

- Flint, J., Racaniello, V.R., Rall, G.F. & Skalka, A.M. 2015. Principles of Virology, 4th ed. American Society of Microbiology.

- Frank, S.A. 1998. Foundations of Social Evolution. Princeton University Press, Princeton, NJ.

- Frank, S.A. 1995. Mutual policing and repression of competition in the evolution of cooperative groups. Nature 377: 520–522.

- Gallagher, M.E., Brooke, C.B., Ke, R. & Koelle, K. 2018. Causes and Consequences of Spatial Within-Host Viral Spread. Viruses 10: 627.

- Gardner, A. & Grafen, A. 2009. Capturing the superorganism: a formal theory of group adaptation. J. Evol. Biol. 22: 659–671.

- Grafen, A. 1991. Modelling in Behavioural Ecology. In: Behavioural Ecology, pp. 5–31. Blackwell Scientific Publications, Oxford.

- Hamilton, W.D. 1964. The genetical evolution of social behaviour. I. J. Theor. Biol. 7: 1–16.

- Harcombe, W.R., Chacón, J.M., Adamowicz, E.M., Chubiz, L.M. & Marx, C.J. 2018. Evolution of bidirectional costly mutualism from byproduct consumption. Proc. Natl. Acad. Sci. 201810949.

- Holland, J.J., De La Torre, J.C. & Steinhauer, D.A. 1992. RNA virus populations as quasispecies. Curr. Top. Microbiol. Immunol. 176: 1–20.

- Holmes, E.C. 2009. The Evolution and Emergence of RNA Viruses. Oxford University Press, Oxford, New York.

- Holmes, E.C. 2010. The RNA virus quasispecies: fact or fiction? J. Mol. Biol. 400: 271–273.

- Holmes, E.C. & Moya, A. 2002. Is the quasispecies concept relevant to RNA viruses? J. Virol. 76: 460–465.

- Iranzo, J. & Manrubia, S.C. 2012. Evolutionary dynamics of genome segmentation in multipartite viruses. Proc. R. Soc. Lond. B Biol. Sci. 279: 3812–3819.

- Konno, Y., Kimura, I., Uriu, K., Fukushi, M., Irie, T., Koyanagi, Y., et al. 2020. SARS-CoV-2 ORF3b Is a Potent Interferon Antagonist Whose Activity Is Increased by a Naturally Occurring Elongation Variant. Cell Rep. 32. Elsevier.

- Lauring, A.S. & Andino, R. 2010. Quasispecies Theory and the Behavior of RNA Viruses. PLOS Pathog. 6: e1001005.

- Levi, L.I., Rezelj, V.V., Henrion-Lacritick, A., Erazo, D., Boussier, J., Vallet, T., et al. 2021. Defective viral genomes from chikungunya virus are broad-spectrum antivirals and prevent virus dissemination in mosquitoes. PLOS Pathog. 17: e1009110. Public Library of Science.

- Li, D., Lott, W.B., Lowry, K., Jones, A., Thu, H.M. & Aaskov, J. 2011. Defective Interfering Viral Particles in Acute Dengue Infections. PLOS ONE 6: e19447.

- Lucía-Sanz, A. & Manrubia, S. 2017. Multipartite viruses: adaptive trick or evolutionary treat? Npj Syst. Biol. Appl. 3: 34.

- Manzoni, T.B. & López, C.B. 2018. Defective (interfering) viral genomes re-explored: impact on antiviral immunity and virus persistence. Future Virol., doi: 10.2217/fvl-2018-0021.

- Martin, B.E., Harris, J.D., Sun, J., Koelle, K. & Brooke, C.B. 2020. Cellular co-infection can modulate the efficiency of influenza A virus production and shape the interferon response. PLOS Pathog. 16: e1008974. Public Library of Science.

- Maynard Smith, J. & Szathmary, E. 1995. The Major Transitions in Evolution. Oxford University Press, New York.

- Metzger, V.T., Lloyd-Smith, J.O. & Weinberger, L.S. 2011. Autonomous Targeting of Infectious Superspreaders Using Engineered Transmissible Therapies. PLOS Comput. Biol. 7: e1002015.

- Nee, S. 1987. The evolution of multicompartmental genomes in viruses. J. Mol. Evol. 25: 277–281.

- Poirier, E.Z., Goic, B., Tomé-Poderti, L., Frangeul, L., Boussier, J., Gausson, V., et al. 2018. Dicer-2-Dependent Generation of Viral DNA from Defective Genomes of RNA Viruses Modulates Antiviral Immunity in Insects. Cell Host Microbe 23: 353-365.e8.

- Rezelj, V.V., Levi, L.I. & Vignuzzi, M. 2018. The defective component of viral populations. Curr. Opin. Virol. 33: 74–80.

- Shirogane, Y., Rousseau, E., Voznica, J., Xiao, Y., Su, W., Catching, A., et al. 2021. Experimental and mathematical insights on the interactions between poliovirus and a defective interfering genome. bioRxiv. doi: 10.1101/2021.01.11.426198

- Skums, P., Bunimovich, L. & Khudyakov, Y. 2015. Antigenic cooperation among intrahost HCV variants organized into a complex network of cross-immunoreactivity. Proc. Natl. Acad. Sci. 112: 6653–6658.

- Sun, J., Vera, J.C., Drnevich, J., Lin, Y.T., Ke, R. & Brooke, C.B. 2020. Single cell heterogeneity in influenza A virus gene expression shapes the innate

antiviral response to infection. PLOS Pathog. 16: e1008671. Public Library of Science.

- Vasilijevic, J., Zamarreño, N., Oliveros, J.C., Rodriguez-Frandsen, A., Gómez, G., Rodriguez, G., et al. 2017. Reduced accumulation of defective viral genomes contributes to severe outcome in influenza virus infected patients. PLOS Pathog. 13: e1006650.

- Vignuzzi, M. & López, C.B. 2019. Defective viral genomes are key drivers of the virus–host interaction. Nat. Microbiol. 1.

- Vignuzzi, M., Stone, J.K., Arnold, J.J., Cameron, C.E. & Andino, R. 2006. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. Nature 439: 344–348.

- West, S.A. 2009. Sex Allocation. Princeton University Press (Monographs in Population Biology Series).

- West, S.A., Cooper, G.A., Ghoul, M.B. & Griffin, A.S. 2021. Ten recent insights for our understanding of cooperation. Nat. Ecol. Evol. 1–12. Nature Publishing Group.

- West, S.A., Griffin, A.S. & Gardner, A. 2007. Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. J. Evol. Biol. 20: 415–432.

- West, S.A., Griffin, A.S. & Gardner, A. 2008. Social semantics: how useful has group selection been? J. Evol. Biol. 21: 374–385.

- Wilke, C.O. 2005. Quasispecies theory in the context of population genetics. BMC Evol. Biol. 5: 44.

- Wilke, C.O., Wang, J.L., Ofria, C., Lenski, R.E. & Adami, C. 2001. Evolution of digital organisms at high mutation rates leads to survival of the flattest. Nature 412: 331–333.

- Williams, G.C. 1966. Adaptation and Natural Selection. Princeton University Press, Princeton, NJ.
- Xue, K.S., Hooper, K.A., Ollodart, A.R., Dingens, A.S. & Bloom, J.D. 2016. Cooperation between distinct viral variants promotes growth of H3N2 influenza in cell culture. eLife 5: e13974.
- Yang, Y., Lyu, T., Zhou, R., He, X., Ye, K., Xie, Q., et al. 2019. The antiviral and antitumor effects of defective interfering particles/genomes and their mechanisms. Front. Microbiol. 10.
- Yuen, C.-K., Lam, J.-Y., Wong, W.-M., Mak, L.-F., Wang, X., Chu, H., et al. 2020. SARS-CoV-2 nsp13, nsp14, nsp15 and orf6 function as potent interferon antagonists. Emerg. Microbes Infect. 0: 1–29. Taylor & Francis.
- Zhang, Y.-Z., Shi, M. & Holmes, E.C. 2018. Using Metagenomics to Characterize an Expanding Virosphere. Cell 172: 1168–1172.
- Ziegler, C.M. & Botten, J.W. 2020. Defective Interfering Particles of Negative-Strand RNA Viruses. Trends Microbiol., doi: 10.1016/j.tim.2020.02.006.
- Zwart, M.P. & Elena, S.F. 2020. Modeling multipartite virus evolution: the genome formula facilitates rapid adaptation to heterogeneous environments†. Virus Evol. 6.

Appendices

A

Altruism in a Virus

KIN SELECTION IN VIRUSES

Altruism in a virus

A recent study finds that viruses cooperate altruistically to overcome innate host immunity and that this can be explained in the same way we explain altruism between animals.

Asher Leeks and Stuart West

In many organisms, the first line of defence against viruses is to quarantine the infected area. Mammalian cells achieve this by releasing interferons, which are signalling molecules that spread to neighbouring cells and place them on 'red alert', enhancing their antiviral defences. Many viruses fight back by blocking interferon signalling, keeping nearby cells susceptible to infection. Domingo-Calap and colleagues now show that in vesicular stomatitis virus (VSV), blocking interferon signalling is a form of altruistic cooperation that is favoured because it provides a benefit to the local population of viruses[1] (Fig. 1).

A behaviour or trait is altruistic when it is costly to perform and provides a benefit to others. Cost and benefit here are defined in terms of evolutionary fitness — the number of offspring produced. Explaining altruism is a fundamental problem for evolutionary biology. Given the Darwinian idea of 'survival of the fittest', how can we explain individuals carrying out a behaviour that reduces their own fitness, while increasing that of others?

Bill Hamilton's kin selection theory provides an explanation for altruism[2]. Hamilton showed that altruism could be favoured when it is directed towards relatives who share the gene for altruism. This theory is encapsulated in a pleasingly simple form by Hamilton's rule, which states that altruism is favoured when $rb - c > 0$; where c is the fitness cost to the altruist, b is the fitness benefit to the recipient of the altruism and r is the relatedness between these individuals[3]. Relatedness is a measure of genetic similarity (kinship) that essentially captures the likelihood that they would share a gene for altruism.

Altruism is usually discussed within the context of the animal world; for example, the sterile workers in social insect colonies, or helpers at the nest in cooperatively breeding birds such as long-tailed tits[4–6]. Domingo-Calap and colleagues show that altruism also occurs in viruses. To do this, they derived a version of Hamilton's rule that was appropriate to interferon blocking in VSV, and then estimated its parameters.

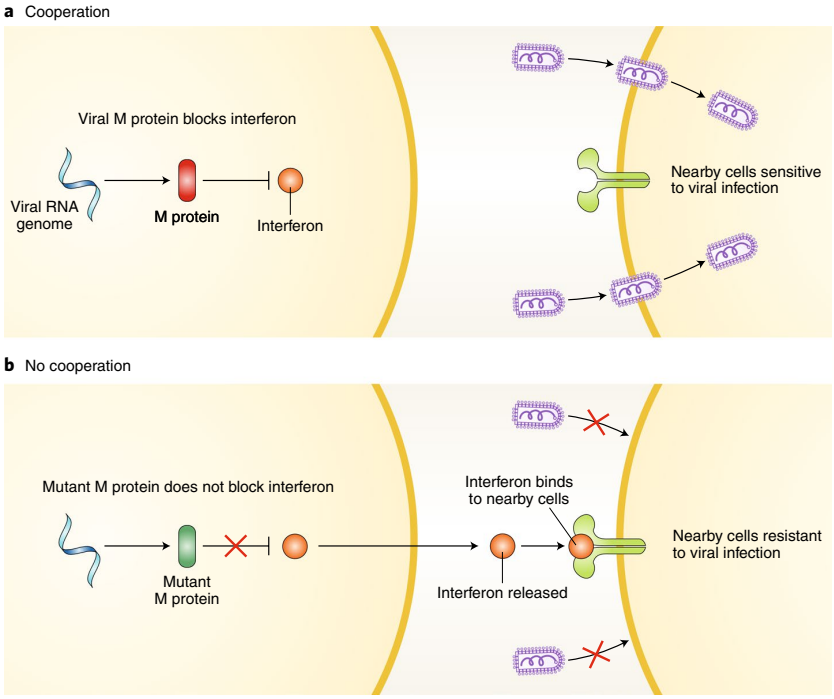


Fig. 1 | Blocking interferon release indicates altruistic cooperation in a virus. Following viral infection, mammalian cells release interferon, a signalling molecule that spreads to nearby cells and enhances antiviral defences. **a**, Wild-type VSV encodes a protein (the M protein) that blocks the release of interferon from infected cells. This provides a benefit to the local group of viruses because it keeps nearby cells susceptible to infection. **b**, In contrast, the non-cooperative Δ51 mutant of VSV encodes a different version of the M protein that does not block the release of interferon. This provides a growth advantage to Δ51, but it comes at a cost to the local group of viruses because nearby host cells activate their antiviral defences and become resistant to viral infection.

They did this through a series of elegant experiments using two mutants: a wild-type VSV strain that blocked interferon, and a genetically engineered mutant of VSV (Δ51) that differed from the wild type only in its inability to block interferon[7] (Fig. 1).

To isolate the cost of blocking interferon, the authors compared the growth rates of Δ51 and wild-type VSV when the host cells' interferon response was reduced. In these conditions, viruses would pay the cost of blocking interferon but not receive the benefit. They showed that Δ51 grew faster than wild-type VSV in the early stages of infections (before the cells' interferon response had kicked in). This showed that blocking interferon comes at a cost to an individual virus because it reduces the virus's growth rate. The authors used multiple lines of evidence to confirm this result, including competition assays over time and an additional experiment in which viruses were grown on cells that were not able to produce interferon.

Next, the authors determined the extent to which blocking interferon provided a shared benefit to local viruses. They did this by comparing the overall growth rates of pure wild-type and pure Δ51 infections, accounting for Δ51's faster inherent growth rate. Blocking interferon provided a group benefit, ultimately allowing wild-type VSV infections to reach more than 16 times higher viral densities than Δ51 infections.

Using values calculated 43 hours after the start of the infection, the authors obtained approximate estimates of $c = 0.4$ and $b = 1.6$. These can be used in Hamilton's rule to obtain $r \times 1.6 > 0.4$, giving $r = 0.25$ as the critical value of relatedness above which cooperation is favoured. This is not a very high level of relatedness — it roughly requires that benefits go to clone mates one-fourth of the time.

Domingo-Calap and colleagues did not directly measure relatedness, but they tested whether manipulating relatedness influenced the advantage of blocking interferon. Comparing across different relatedness treatments, they found that the wild type did better when relatedness was higher. Specifically, in a fully structured condition ($r = 1$), the wild type out-competed Δ51; in an intermediate condition, where the authors estimated that $r = 0.3$, the wild type won with a relatively small fitness advantage; and in a fully mixed condition ($r = 0$), Δ51 out-competed the wild type. The authors also performed in vivo work and showed that, in mixed infections, Δ51 out-competed wild-type VSV. A possible explanation for how cooperation could be maintained in natural infections is if infections often start with a small number of viral particles, keeping relatedness high and mixed infections relatively rare.

Overall, this study provides an elegant demonstration of how social interactions can be critical for the evolutionary success of viruses. In particular, it shows how cooperation can occur between viruses infecting different cells, as well as those infecting the same cell[8]. This raises a number of new questions about the social nature of viruses. Are other viral traits maintained by kin selection? Can we disrupt relatedness to select against cooperation for therapeutic purposes? Does cooperation break down in natural viral infections, and does this affect disease dynamics? Over the last 20 years, we have discovered that cooperation plays a fundamental role in the success and virulence of bacteria — are we about to experience a similar revolution in viruses[9]?

More broadly, this study shows how the idea of kin selection, though originally developed to explain altruism in animals, can also be applied to viruses[3]. Since its development, kin selection has been used to explain cooperation and conflict at all levels of biology, including animal behaviour, sterility in social insects, genes inside a genome, fruiting bodies in slime moulds and the production of virulence factors by bacteria[10]. It now seems that we can add viruses to this list. ❐

Asher Leeks and Stuart West*

Department of Zoology, University of Oxford, Oxford, UK.

**e-mail: stuart.west@zoo.ox.ac.uk*

Published online: 22 May 2019
https://doi.org/10.1038/s41564-019-0463-0

References

1. Domingo-Calap, P., Segredo-Otero, E., Durán-Moreno, M. & Sanjuán, R. *Nat. Microbiol.* https://doi.org/10.1038/s41564-019-0379-8 (2019).
2. Hamilton, W. D. *J. Theor. Biol.* **7**, 1–16 (1964).
3. Hamilton, W. D. *Am. Nat.* **97**, 354–356 (1963).
4. Boomsma, J. J. *Curr. Biol.* **17**, R673–R683 (2007).
5. Cornwallis, C. K., West, S. A., Davis, K. E. & Griffin, A. S. *Nature* **466**, 969–972 (2010).
6. Hatchwell, B. J., Gullett, P. R. & Adams, M. *J. Philos. T. R. Soc. B* **369**, 20130565 (2014).
7. Stojdl, D. F. et al. *Cancer Cell* **4**, 263–275 (2003).
8. Turner, P. E. & Chao, L. *Nature* **398**, 441–443 (1999).
9. West, S. A., Diggle, S. P., Buckling, A., Gardner, A. & Griffin, A. S. *Annu. Rev. Ecol. Evol. S.* **38**, 53–77 (2007).
10. Bourke, A. F. G. *Principles of Social Evolution* (Oxford University Press, 2001).

Competing interests

The authors declare no competing interests.