

Adaptation and genetic conflict

Thomas Scott

Wadham College

University of Oxford



A thesis submitted for the degree of
Doctor of Philosophy

Trinity 2019

Declaration

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. The work has not been submitted for any degree or professional qualification except as specified.

Thomas W. Scott

Acknowledgements

First and foremost, thanks to Stu for supervision – couldn't have really asked for a better experience as a graduate student. Thanks to Ashleigh also and to everyone in the West / Griffin lab groups I've worked alongside – it's a great place to do science.

The work presented in this thesis has benefitted greatly from conversations and feedback from lots of people, including: Duur Aanen, Koos Boomsma, John Bruce, Max Burton-Chellew, Shana Caro, Guy Cooper, David Crosse, Anna Dewar, Philip Downing, Jared Field, Berti Fisher, Mel Ghoul, Rebecca Goldberg, Alan Grafen, Ashleigh Griffin, Tom Hitchcock, Steffi Kapsetaki, Toby Kiers, Chucky Lee, Asher Leeks, Egbert Leigh, Sam Levin, Ming Liu, Alper Mutlu, Mati Patel, Dave Queller, Ian Sanders, Miguel dos Santos, Joan Strassman, Josh Thomas, Daniel Unterweger, Gijsbert Werner, Stu West, Geoff Wild & Dan Wilson. Thanks to all!

Publications and Contributions

The following papers have arisen from this thesis, and are presented in Chapter 2, Chapter 3 & the appendix.

▪ Chapter 2

- **Scott, T. W.**, Kiers, E. T., Cooper, G. A., Santos, dos, M., & West, S. A. (2019). Evolutionary maintenance of genomic diversity within arbuscular mycorrhizal fungi. *Ecology and Evolution*, *9*(5), 2425–2435.
 - I conceived of the manuscript with feedback from SAW and ETK. I carried out the modelling with help from GAC and MdS. I wrote the first draft of the manuscript and SAW, ETK and I contributed equally to the preparation of the manuscript in its final form.

▪ Chapter 3

- **Scott, T. W.**, & West, S. A. (2019). Adaptation is maintained by the parliament of genes. *Nature Communications*, *10*, 5163.
 - I conceived of the manuscript with feedback from SAW. I carried out the modelling and wrote the first draft of the manuscript. SAW and I contributed equally to the preparation of the manuscript in its final form.

▪ Appendix

- Levin, S.R., **Scott, T.W.**, Cooper, H.S. & West, S.A. (2019) Darwin's aliens. *International Journal of Astrobiology* *18*(1), 1-9
 - SAW and SRL conceived of the manuscript. SRL wrote the first draft, and SAW, SRL and I contributed equally to preparing the

final draft. HSC created the illustrations. SRL created Figure 3 and I created Figure 1.

The following chapter contains unpublished work.

- **Chapter 4**

- **Scott, T. W., & West, S. A.** The evolution of genetic kin discrimination.
 - I conceived of the project with feedback from SAW. I carried out the modelling and wrote the first draft of the chapter. I prepared the final draft with input from SAW.

Abstract

Genes that increase organism fitness can come to prominence as a result of natural selection, leading to the appearance of organismal design, or 'adaptation'. However, genes that compromise organism fitness can also come to prominence if they are able to secure a selfish propagation advantage from doing so. Such genes are called 'selfish genetic elements'. I consider the consequences of selfish genetic elements for organismal design (adaptation). First, I consider a fungus in which – strikingly – different nuclei in the same individual may be genetically different from each other ('chimera'). I show how such diversity could be maintained by natural selection, and what consequences this has on the organism. Second, I consider, in general and in a range of specific biological scenarios, whether selfish genetic elements might be expected to gain control of organism traits. I show that the 'parliament of genes' is generally effective in suppressing selfish genetic elements, meaning organism design is generally preserved. Third, I ask whether animals can, in principle, evolve to recognise kin via genetic cues. I show they often can, as long as genetic kin discrimination is favoured, over indiscriminate cooperation and indiscriminate defection, at the individual level (it maximises individual fitness).

Contents

1. Introduction	p1
2. Evolutionary maintenance of genomic diversity within arbuscular mycorrhizal fungi.	p16
3. Adaptation is maintained by the parliament of genes	p37
4. The evolution of genetic kin discrimination	p147
5. Discussion	p202
<i>Appendix</i>	
6. Darwin's aliens	p219

Introduction

One of the most striking features of the natural world is the extent to which organisms appear designed or adapted (Paley, 1829). Darwin (1869) provided the solution to this problem with his theory of natural selection. Our modern understanding of this theory is that genes that raise organism fitness will increase in frequency, leading to organisms that appear as if they have been designed to maximise their fitness (Fisher, 1930; Gardner, 2009; Grafen, 2006; 2014; Hamilton, 1964).

A problem is that there is also considerable evidence for selfish genetic elements, which increase their own contribution to future generations at the expense of other genes in the same organism (Ågren & Clark, 2018; Burt & Trivers, 2006; Gardner & Úbeda, 2017). Selfish genetic elements create conflict within the genome. My thesis concerns how our understanding of organism design (adaptation) is affected by genetic conflict.

Each research chapter in this thesis has its own introduction. Therefore, in this chapter, I do not provide an exhaustive literature review of all issues raised in subsequent chapters. Instead, I provide a brief overview of what I take *adaptation* to mean, and how adaptation might be compromised by genetic conflict, whilst pointing out some unanswered questions in this area of research.

Adaptation

An entity is 'adapted' if it 'appears designed'. We follow Paley (1829) in breaking this down further (Gardner, 2009; Lewens, 2019). For an entity to 'appear designed', it must satisfy two criteria. Firstly, the entity must be goal-directed. That is, it must operate in a way that is not random or unpredictable. Rather, the entity must act with purpose – it must be pursuing some 'maximand'. Secondly, the entity must be integrated. That is, its constituent parts must interact with each other – they must work together, towards a shared goal, rather than independently, towards individual goals.

Biological organisms are said to be 'adapted' because they satisfy the two required criteria. They are goal directed because, as pointed out by Darwin (1869), they strive to maximise their fitness (Grafen, 2006; Hamilton, 1964). They are integrated because, despite being hierarchical in nature (bodies comprise cells, which comprise organelles, which comprise genes), the component parts work together rather than independently (Maynard Smith & Szathmáry, 1995; West, Fisher, Gardner, & Kiers, 2015).

In recent years, evolutionary biologists have tightened this two-criterion definition of adaptation (Gardner, 2009; Gardner & Grafen, 2009; Queller & Strassmann, 2009; Strassmann & Queller, 2010). When we observe that a biological organism is 'goal directed', what we are really observing is that: its constituent parts (cells, organelles, genes) are not in disagreement with each other over organism trait values; or, where there is disagreement, this isn't readily observable at the organism level. We can

therefore say that biological organisms are characterised by ‘low conflict’ amongst their constituent parts (single maximand).

When we observe that a biological organism is ‘integrated’, what we are really observing is that its constituent parts (cells, organelles, genes) cooperate with each other – for example, they share, rather than compete over, resources. We can therefore say that biological organisms are characterised by ‘high cooperation’ amongst their constituent parts (integration; West et al., 2015).

Therefore, the dual characteristics of ‘low conflict’ and ‘high cooperation’ amongst constituent parts is what characterises biological organisms. Strassmann & Queller have even gone as far as to suggest that we should *define* an organism as any entity with these dual characteristics, which means that groups of individuals, such as some social insect colonies or between-species mutualisms, may count as organisms (Queller & Strassmann, 2009; Strassmann & Queller, 2007; 2010).

However, for our present purposes, we take the traditional (if vague) definition of an organism as (in general) a single physiologically continuous unit, such as an individual animal, plant or fungal network (Pepper & Herron, 2008).

Organisms are adapted to differing extents. Organisms with constituent parts that are strongly cooperating and weakly conflicting will be strongly adapted. Conversely, organisms with constituent parts that are weakly cooperating and strongly conflicting will only be weakly adapted. This raises an interesting question – why are some organisms highly adapted (high cooperation and low conflict) and others less so? Or

to re-frame the question – what biological features do highly-adapted organisms have that weakly-adapted organisms lack?

Modular versus unitary organisms

The extent of ‘cooperation’ between the component parts of a multicellular organism (integration) is likely to depend on whether most cells retain reproductive capability after the organism matures (Strassmann & Queller, 2004). Modular organisms grow through iterations of modules, like hyphae or stems, that each retain reproductive capability (Pineda-Krch & Lehtila, 2004). By contrast, unitary organisms grow from a single genome (single-celled bottleneck), and most cells lose reproductive capability after organism maturation (germ-soma divide).

The modules (cells) of modular organisms are faced with an evolutionary trade-off (Aanen, Debets, de Visser, & Hoekstra, 2008; Bastiaans, Debets, & Aanen, 2016; Frank, 1995; 2003). They could act cooperatively, sharing their resources with other modules in pursuit of the shared goal of maximising organism fitness. They could alternatively act selfishly, replicating quickly, at a potential fitness cost to the organism, to increase their propagation within the modular organism and its offspring (which are generated by module fragmentation). Modular organisms include many filamentous fungi, colonial invertebrates like sponges, and plants that grow from underground connected stems called rhizomes (Herron, Rashidi, Shelton, & Driscoll, 2013; Pineda-Krch & Lehtila, 2004).

In contrast, owing to the single cell bottleneck, it is very difficult for a gene in a unitary organism to increase its propagation to future generations via selfish replication within an organism (Strassmann & Queller, 2004). In other words, the option to replicate selfishly within organisms is less feasible for genes in unitary organisms relative to genes in modular organisms. As a result, the evolutionary trade-off faced by genes or cells within a unitary organism, regarding whether to selfishly replicate within the body, or to contribute cooperatively to the shared goal of organism fitness maximisation, is less stark. The cooperative strategy is more mutationally accessible than the selfish strategy, and therefore is more likely to be chosen (Ågren, Davies, & Foster, 2019; Frank, 1995; 2003). Counter-exceptions are occasionally found, like transposons, meiotic drivers and cancers, but these cases are rare compared to the amount of genes and cells within unitary organisms that do not adopt these selfish strategies (Burt & Trivers, 2006; Grafen, 2006).

Therefore, we would expect modular organisms to have 'low cooperation' between component parts (low integration) and unitary organisms to have 'high cooperation' between component parts (high integration). Modular organisms should be less 'adapted' (Gardner & Grafen, 2009). An open question is, if modular organisms are poorly adapted, how are they able to persist stably across evolutionary time? How are they able to acquire the specific, fine-tuned organism forms that allow them to fit into narrow environmental niches?

Conflicting coreplicons

The extent of 'conflict' between the component parts of an organism may depend on the extent that its genome is factionalised into different 'coreplicons'. Coreplicons are collections of loci within a genome that are inherited in the same way (Cosmides & Tooby, 1981). For instance, autosomal loci and Y chromosome loci comprise different coreplicons, because the former are transmitted equally through males and females and the latter are transmitted only through males.

Coreplicons may disagree over organism trait values, generating 'conflict' within the organism (Haig, 2014; Úbeda & Haig, 2003). They disagree because different coreplicons may be propagated more efficiently at different organism trait values. For instance, in a large, randomly mating (panmictic) population, autosomes are propagated most efficiently when the offspring sex ratio is equal (equal production of males and female offspring). An equal offspring sex ratio is favoured by autosomes because a bias in favour of one sex means that offspring of the common sex suffer from increased mate competition (Fisher, 1930). In contrast to this, Y chromosomes prefer a completely male-biased offspring sex ratio, because they are only transmitted through males (Hamilton, 1967).

The disagreement amongst coreplicons, over the value of traits like offspring sex ratio, has the potential to undermine the goal directedness (fitness maximisation) of organisms. If organism trait values were to fall in between the preferred values of different coreplicons (e.g. a slightly male-biased sex ratio), the organism would not be maximising any one single quantity – it would be more like a 'compromise' between divergent interests, lacking a clear purpose (Haig, 2014). Furthermore, the

factionalisation of organisms into rival coreplicons is ubiquitous across the tree of life, being driven by phenomena like genomic imprinting and horizontal gene transfer (Haig, 2002; Nogueira et al., 2009; Úbeda & Gardner, 2010; 2011; 2012).

Given then that organisms are highly factionalised, a salient question is: how do they retain 'low conflict' amongst their constituent parts (coreplicons), leading to adaptation? To put this another way: why does the *potential* for conflict amongst coreplicons rarely translate into *actual* conflict at the organism level, compromising organism design (adaptation; Queller & Strassmann, 2018)?

Thesis outline

In this thesis, I consider some questions relating to organism design (adaptation) in the light of different kinds of genetic conflict. Specifically:

In **Chapter 2**, I consider a highly abundant and ecologically significant modular organism: Arbuscular Mycorrhizal fungi. Strikingly, different nuclei in the same organism (fungal network) may be genetically different from each other (chimera), and nuclear diversity is inherited via multi-nucleate spores (no single-celled bottleneck; Sanders & Croll, 2010). I show that nuclear diversity could provide a benefit at the level of the organism (fungal network), by improving growth in variable environments, and that this can stabilise nuclear diversity over evolutionary timescales (Scott, Kiers, Cooper, Santos, & West, 2019). These results have wider significance in helping to explain how modular organisms are able persist over

evolutionary timescales despite the potential for evolutionary maladaptation (Pineda-Krch & Lehtila, 2004).

In **Chapter 3**, I address an unresolved contradiction in evolutionary biology. Fields such as behavioural and evolutionary ecology are built on the assumption that natural selection leads to organisms that behave as if they are trying to maximise their fitness (Davies, Krebs, & West, 2012). And yet selfish genetic elements, which may change the behaviour of individuals to increase their own transmission, are common (Burt & Trivers, 2006). I reconcile this contradiction by showing that the ‘parliament of genes’ is generally effective in suppressing ‘cabals’ of selfish genetic elements (Leigh, 1971). This means that, even when there is the potential for considerable genetic conflict, there will often be negligible effect at the organism level. These results have wider significance in helping to explain why organisms usually remain adapted (low internal conflict) despite the factionalisation of genomes into different coreplicons.

In **Chapter 4**, I consider the evolution of genetic kin discrimination. Kin discrimination has the potential to promote the evolution of altruism via kin selection. However, genetic kin discrimination, in which individuals recognise kin based on a shared genetically-encoded signal (‘tag’), is thought to be inherently unstable, because individuals with common tags will find social partners at a faster rate than individuals with rare tags (common-tag advantage), meaning rare tags are purged from the population (‘Crozier’s paradox’; Crozier, 1986). However, I show that the common-tag advantage is often reversed because ‘cheaters’ (non-altruists) build up within

groups of individuals using common tags (Grafen, 1990). I show that, for genetic kin discrimination to evolve, it must be favoured, over indiscriminate cooperation and indiscriminate defection, at the individual level (it must maximise individual fitness).

In **Chapter 5**, I discuss the broader implications of this research for our understanding of adaptation in the light of genetic conflict, and future directions.

Finally, the **appendix** contains a review paper, where we argued that, even on other planets, adapted organisms (aliens) will necessarily be characterised by high internal cooperation and low internal conflict (Levin, Scott, Cooper, & West, 2017).

References

Aanen, D. K., Debets, A. J. M., de Visser, J. A. G. M., & Hoekstra, R. F. (2008). The social evolution of somatic fusion. *BioEssays*, *30*(11-12), 1193–1203.

<http://doi.org/10.1002/bies.20840>

Ågren, J. A., & Clark, A. G. (2018). Selfish genetic elements. *PLoS Genetics*, *14*(11), e1007700–20. <http://doi.org/10.1371/journal.pgen.1007700>

Ågren, J. A., Davies, N. G., & Foster, K. R. (2019). Enforcement is central to the evolution of cooperation. *Nature Ecology & Evolution*, *3*(7), 1–12.

<http://doi.org/10.1038/s41559-019-0907-1>

Bastiaans, E., Debets, A. J. M., & Aanen, D. K. (2016). Experimental evolution reveals that high relatedness protects multicellular cooperation from cheaters.

Nature Communications, *7*, 11435. <http://doi.org/10.1038/ncomms11435>

Burt, A., & Trivers, R. (2006). *Genes in Conflict*. Cambridge, MA and London,

- England: Harvard University Press. <http://doi.org/10.4159/9780674029118>
- Cosmides, L. M., & Tooby, J. (1981). Cytoplasmic inheritance and intragenomic conflict. *Journal of Theoretical Biology*, *89*(1), 83–129.
[http://doi.org/10.1016/0022-5193\(81\)90181-8](http://doi.org/10.1016/0022-5193(81)90181-8)
- Crozier, R. H. (1986). Genetic clonal recognition abilities in marine invertebrates must be maintained by selection for something else. *Evolution*, *40*(5), 1100–1101. <http://doi.org/10.1111/j.1558-5646.1986.tb00578.x>
- Darwin, C. (1869). *On the Origin of Species by Means of Natural Selection*.
- Davies, N. B., Krebs, J. R., & West, S. A. (2012). *An Introduction to Behavioural Ecology*. John Wiley & Sons.
- Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*. Oxford University Press.
- Frank, S. A. (1995). Mutual policing and repression of competition in the evolution of cooperative groups. *Nature*, *377*(6549), 520–522.
<http://doi.org/10.1038/377520a0>
- Frank, S. A. (2003). Perspective: repression of competition and the evolution of cooperation. *Evolution*, *57*(4), 693–705. <http://doi.org/10.1111/j.0014-3820.2003.tb00283.x>
- Gardner, A. (2009). Adaptation as organism design. *Biology Letters*, *5*(6), 861–864.
<http://doi.org/10.1098/rsbl.2009.0674>
- Gardner, A., & Grafen, A. (2009). Capturing the superorganism: a formal theory of group adaptation. *Journal of Evolutionary Biology*, *22*(4), 659–671.
<http://doi.org/10.1111/j.1420-9101.2008.01681.x>
- Gardner, A., & Úbeda, F. (2017). The meaning of intragenomic conflict. *Nature*

- Ecology & Evolution*, 1(12), 1–9. <http://doi.org/10.1038/s41559-017-0354-9>
- Grafen, A. (1990). Do animals really recognize kin? *Animal Behaviour*, 39(1), 42–54.
- Grafen, A. (2006). Optimization of inclusive fitness. *Journal of Theoretical Biology*, 238(3), 541–563. <http://doi.org/10.1016/j.jtbi.2005.06.009>
- Grafen, A. (2014). The formal darwinism project in outline. *Biology & Philosophy*, 29(2), 155–174. <http://doi.org/10.1007/s10539-013-9414-y>
- Haig, D. (2002). *Genomic Imprinting and Kinship*. Rutgers University Press.
- Haig, D. (2014). Genetic dissent and individual compromise. *Biology & Philosophy*, 29(2), 233–239. <http://doi.org/10.1007/s10539-013-9418-7>
- Hamilton, W. D. (1964). The genetical evolution of social behaviour. I. *Journal of Theoretical Biology*, 7(1), 1–16. [http://doi.org/10.1016/0022-5193\(64\)90038-4](http://doi.org/10.1016/0022-5193(64)90038-4)
- Hamilton, W. D. (1967). Extraordinary sex ratios. A sex-ratio theory for sex linkage and inbreeding has new implications in cytogenetics and entomology. *Science*, 156(3774), 477–488. <http://doi.org/10.1126/science.156.3774.477>
- Herron, M. D., Rashidi, A., Shelton, D. E., & Driscoll, W. W. (2013). Cellular differentiation and individuality in the “minor” multicellular taxa. *Biological Reviews*, 88(4), 844–861. <http://doi.org/10.1111/brv.12031>
- Leigh, E. G. (1971). *Adaptation and diversity*.
- Levin, S. R., Scott, T. W., Cooper, H. S., & West, S. A. (2017). Darwin's aliens. *International Journal of Astrobiology*, 18(1), 1–9. <http://doi.org/10.1017/S1473550417000362>
- Lewens, T. (2019). Neo-Paleyan biology. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 76, 101185. <http://doi.org/10.1016/j.shpsc.2019.101185>

- Maynard Smith, J., & Szathmáry, E. (1995). *The Major Transitions in Evolution*. Oxford University Press. <http://doi.org/10.2307/2410462>
- Nogueira, T., Rankin, D. J., Touchon, M., Taddei, F., Brown, S. P., & Rocha, E. P. C. (2009). Horizontal Gene Transfer of the Secretome Drives the Evolution of Bacterial Cooperation and Virulence. *Current Biology*, *19*(20), 1683–1691. <http://doi.org/10.1016/j.cub.2009.08.056>
- Paley, W. (1829). *Natural Theology*. New York,: S. King. <http://doi.org/10.5962/bhl.title.52208>
- Pepper, J. W., & Herron, M. D. (2008). Does Biology Need an Organism Concept? *Biological Reviews*, *83*(4), 621–627. <http://doi.org/10.1111/j.1469-185X.2008.00057.x>
- Pineda-Krch, M., & Lehtila, K. (2004). Challenging the genetically homogeneous individual. *Journal of Evolutionary Biology*, *17*(6), 1192–1194. <http://doi.org/10.1111/j.1420-9101.2004.00815.x>
- Queller, D. C., & Strassmann, J. E. (2009). Beyond society: the evolution of organismality. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1533), 3143–3155. <http://doi.org/10.1098/rstb.2009.0095>
- Queller, D. C., & Strassmann, J. E. (2018). Evolutionary Conflict. *Annual Review of Ecology, Evolution, and Systematics*, *49*(1), annurev–ecolsys–110617–062527–21. <http://doi.org/10.1146/annurev-ecolsys-110617-062527>
- Sanders, I. R., & Croll, D. (2010). Arbuscular Mycorrhiza: The Challenge to Understand the Genetics of the Fungal Partner. *Annual Review of Genetics*, *44*(1), 271–292. <http://doi.org/10.1146/annurev-genet-102108-134239>
- Scott, T. W., Kiers, E. T., Cooper, G. A., Santos, dos, M., & West, S. A. (2019).

- Evolutionary maintenance of genomic diversity within arbuscular mycorrhizal fungi. *Ecology and Evolution*, 9(5), 2425–2435. <http://doi.org/10.1002/ece3.4834>
- Strassmann, J. E., & Queller, D. C. (2004). Genetic conflicts and intercellular heterogeneity. *Journal of Evolutionary Biology*, 17(6), 1189–1191. <http://doi.org/10.1111/j.1420-9101.2004.00814.x>
- Strassmann, J. E., & Queller, D. C. (2007). Insect societies as divided organisms: The complexities of purpose and cross-purpose. *Proceedings of the National Academy of Sciences*, 104(Supplement 1), 8619–8626. <http://doi.org/10.1073/pnas.0701285104>
- Strassmann, J. E., & Queller, D. C. (2010). The Social Organism: Congresses, Parties, and Committees. *Evolution*, 64(3), 605–616. <http://doi.org/10.1111/j.1558-5646.2009.00929.x>
- Úbeda, F., & Gardner, A. (2010). A model for genomic imprinting in the social brain: juveniles. *Evolution*, 64(9), 2587–2600. <http://doi.org/10.1111/j.1558-5646.2010.01015.x>
- Úbeda, F., & Gardner, A. (2011). A model for genomic imprinting in the social brain: adults. *Evolution*, 65(2), 462–475. <http://doi.org/10.1111/j.1558-5646.2010.01115.x>
- Úbeda, F., & Gardner, A. (2012). A model for genomic imprinting in the social brain: elders. *Evolution*, 66(5), 1567–1581. <http://doi.org/10.1111/j.1558-5646.2011.01517.x>
- Úbeda, F., & Haig, D. (2003). Dividing the child. *Genetica*, 117(1), 103–110. <http://doi.org/10.2307/3479394>
- West, S. A., Fisher, R. M., Gardner, A., & Kiers, E. T. (2015). Major evolutionary

transitions in individuality. *Proceedings of the National Academy of Sciences*,
112(33), 10112–10119. <http://doi.org/10.1073/pnas.1421402112>

2

**Evolutionary maintenance of
genomic diversity within arbuscular
mycorrhizal fungi**



ORIGINAL RESEARCH

Evolutionary maintenance of genomic diversity within arbuscular mycorrhizal fungi

Thomas W. Scott¹ | E. Toby Kiers² | Guy A. Cooper¹ | Miguel dos Santos^{1,3} | Stuart A. West^{1,4}¹Department of Zoology, University of Oxford, Oxford, UK²Institute of Ecological Sciences, Faculty of Earth and Life Sciences, Vrije Universiteit, Amsterdam, The Netherlands³Department of Social Psychology and Social Neuroscience, Institute of Psychology, University of Bern, Bern, Switzerland⁴Magdalen College, Oxford, UK**Correspondence**Thomas W. Scott, Department of Zoology, University of Oxford, Oxford, UK.
Email: thomas.scott@zoo.ox.ac.uk**Funding information**

Natural Environment Research Council, Grant/Award Number: Doctoral Training Partnership

Abstract

Most organisms are built from a single genome. In striking contrast, arbuscular mycorrhizal fungi appear to maintain genomic variation within an individual fungal network. Arbuscular mycorrhizal fungi dwell in the soil, form mutualistic networks with plants, and bear multiple, potentially genetically diverse nuclei within a network. We explore, from a theoretical perspective, why such genetic diversity might be maintained within individuals. We consider selection acting within and between individual fungal networks. We show that genetic diversity could provide a benefit at the level of the individual, by improving growth in variable environments, and that this can stabilize genetic diversity even in the presence of nuclear conflict. Arbuscular mycorrhizal fungi complicate our understanding of organismality, but our findings offer a way of understanding such biological anomalies.

KEYWORDS

arbuscular mycorrhizal fungi, chimera, genetic conflict, individuality, intraorganismal genetic heterogeneity, levels of selection, modular organisms, mosaic, mycorrhizal networks, organismality

1 | INTRODUCTION

Most organisms are built from a single collection of genes (*genome*), copied into all nuclei, across all cells. Genomic homogeneity means that the cells and nuclei within organisms have the same evolutionary interest, to transmit that genome to the next generation (Buss, 1988; Maynard Smith & Szathmary, 1997; Strassmann & Queller, 2004). The components of organisms therefore work together, cooperatively, to increase reproductive success. From an evolutionary perspective, this cooperation and lack of conflict define organisms (Maynard Smith & Szathmary, 1997; Queller & Strassmann, 2009, 2016; West, Fisher, Gardner, & Kiers, 2015).

Arbuscular mycorrhizal (AM) fungi appear to be a striking exception to this rule of genomic homogeneity within organisms (Angelard, Colard, Niculita-Hirzel, Croll, & Sanders, 2010; Angelard et al., 2013;

Ehinger, Croll, Koch, & Sanders, 2012; Wyss, Masclaux, Rosikiewicz, Pagni, & Sanders, 2016). AM fungi form large branching networks composed of filaments called hyphae. These hyphal networks (*individuals*), which germinate from spores, live in soil and colonize plant roots, exchanging mineral resources for host carbon (Bonfante & Genre, 2010). A hyphal network can potentially bear thousands of coexisting nuclei at once (*heterokaryotic*) (Sanders & Croll, 2010), and connect multiple plants simultaneously (Rosendahl & Stukenbrock, 2004). There are no internal septal walls within the hyphal networks (*coenocytic*), and so nuclei can potentially move across entire networks. Individual networks of closely related fungal strains can fuse (*anastomose*) and share nuclei (Giovannetti, Avio, & Sbrana, 2015), potentially generating individuals bearing two genomes (Corradi & Brachmann, 2017; Ropars et al., 2016) or possibly many more (Croll et al., 2008; de Novais, Sbrana, Junior, Siqueira, & Giovannetti, 2013;

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

Hijri & Sanders, 2005; Kuhn, Hijri, & Sanders, 2001; Sanders & Croll, 2010; Wyss et al., 2016). Small levels of genomic variation might also arise through different de novo mutations occurring in different nuclei within an individual (Tisserant et al., 2013). When individuals sporulate, hundreds of nuclei flow into the emerging spore, allowing a large portion of the genomic variation to be maintained (Jany & Pawlowska, 2010).

From an evolutionary perspective, the potential for genomic variation within individuals, and the apparent absence of any mechanism to regulate it, poses problems (Frank, 1995, 2003; Strassmann & Queller, 2007). First, it is likely that nuclei replicate at different rates within hyphal networks (Jany & Pawlowska, 2010; Roberts & Gladfelter, 2015), so we would expect the most competitive and fast-growing nucleus lineage to outcompete the rest. In other words, we would expect *within-individual selection* to lead to genomic purity (Gilbert, Foster, Mehdiabadi, Strassmann, & Queller, 2007; Inglis, Ryu, Asikhia, Strassmann, & Queller, 2017; Kooij, Aanen, Schiøtt, & Boomsma, 2015; Meunier, Hosseini, Heidari, Maryush, & Johannesson, 2018; Vreeburg, Nygren, & Aanen, 2016). Within-individual evolution would eventually lead to genomic purity even if nuclei are equally competitive, through drift, because not all nuclei migrate from parent hyphal networks into daughter cells (Angelard et al., 2010; Boon, Zimmerman, St-Arnaud, & Hijri, 2013; Marleau, Dalpé, St-Arnaud, & Hijri, 2011; Masclaux, Wyss, Mateus-Gonzalez, Aletti, & Sanders, 2018). Secondly, we would expect genomic variation within individuals to lead to conflict among different genomic (nuclear) lineages and hence reduce the fitness of that individual. Consequently, individuals with high genomic variation could be outcompeted by individuals with genomic homogeneity. In other words, we would expect *between-individual selection* to also lead to genomic purity (Bastiaans, Debets, & Aanen, 2016; Meunier et al., 2018).

We address the theoretical problem of why genomic diversity would be maintained in AM fungi. We develop theoretical models to address two questions. First, can genomic diversity provide a benefit at the individual level that gives individuals with genomic diversity a competitive advantage over those with genomic homogeneity, despite potential conflict between genomes? Second, how can genomic diversity be maintained within individuals, if one nucleus lineage is more competitive and able to reproduce faster? Our hypothesis is that different fungal genotypes are better at colonizing different plant species, and so fungal individuals with genomic diversity are better able to better colonize multiple plants. If fungal individuals encounter sufficiently different plant species, then this could maintain genomic diversity.

We develop simple analytical models, building upon previous theory, to illustrate the general points. We then develop a more detailed individual-based simulation, to better match the biology of AM fungi. To emphasize applicability to other organisms, we use the general terms “individual” and “genomic diversity,” rather than the AM-specific terms “hyphal network” and “nuclear diversity.” Conversely, although we often talk specifically about competing nucleus lineages, our theory applies more generally to *genomic lineages* of a modular organism that may in fact be cell lines as opposed to

nucleus lines (Pineda-Krch & Lehtila, 2004; Strassmann & Queller, 2004). The extent of genomic diversity in AM fungi is a matter of considerable debate, which is beyond the scope of our paper (Lin et al., 2014; Maeda et al., 2018; Ropars & Corradi, 2015; Tisserant et al., 2013; Wyss et al., 2016). Our aim is to examine how, if diversity exists, it could plausibly be maintained (Bruns, Corradi, Redecker, Taylor, & Öpik, 2017; Sanders, 2018).

2 | MODELS

2.1 | Competing individuals

Our first question is whether genomic diversity can provide a benefit at the level of the individual, allowing individuals with genomic diversity to outcompete those without. Our hypothesis is that genomic diversity provides a way of acquiring a generalist phenotype, which is better able to cope with an unpredictable environment. We take an ESS approach, based on previous theory (Levins, 1962), to find the level of genomic diversity that maximizes individual fitness.

We assume that there are two different plant species, which we term plant 1 (P_1) and plant 2 (P_2). Individual hyphal networks associate with and grow on multiple plants simultaneously. We assume that all individuals are in the same environment, with a proportion p of their interactions being with plant 1 (P_1), and the remaining proportion $(1-p)$ with plant 2 (P_2). The overall fitness of an individual (W) depends on its fitness (how well it grows) on type 1 plants (w_1), weighted by the extent to which it is growing on type 1 plants (p), and its fitness on type 2 plants (w_2), weighted by the extent to which it is growing on type two plants ($1-p$), with $W = pw_1 + (1-p)w_2$. This equation was originally formulated as a general way to represent fitness under simultaneous exposure to two different environments (Levins, 1962). For our purposes, the two plant hosts provide the two environments.

We make the fitness terms in Levins' equation (w_1 and w_2) explicit, so that the fitness of an individual can be written:

$$W(x) = p(\kappa + (1 - \kappa)x^\alpha) + (1 - p)(\kappa + (1 - \kappa)(1 - x)^\alpha). \quad (1)$$

Individuals contain two types of nuclei (N_1 and N_2), which are genetically distinct, nonrecombining, and each specialized on one plant type, N_1 on P_1 , and N_2 on P_2 (Chen et al., 2018b). Fitness on each plant depends on the parameter x , which is the individual's proportion of type 1 nuclei (N_1) relative to type 2 nuclei (N_2). There is a trade-off, meaning as the type 1 nuclear proportion x is increased, fitness on P_1 (w_1) increases from κ to 1, but fitness on P_2 (w_2) decreases, symmetrically, from 1 to κ . The slope of fitness (w_1, w_2) against nucleus proportion (x) may be concave ($0 < \alpha < 1$), corresponding to diminishing fitness returns to plant specialization, or convex ($\alpha > 1$), corresponding to accelerating returns.

The curvature parameter α encapsulates multiple biological phenomena. If the size of the hyphal network (individual) is large relative to the number of plant associations it has, there may be an

overabundance of nuclei in the network (Shoji, Kikuma, Arioka, & Kitamoto, 2010). This would make specialized nuclei less effective at high proportions, where they are not being fully utilized, causing diminishing returns to specialization ($0 < \alpha < 1$). Conversely, small networks with relatively many plant associations may be insufficiently productive to engage each of their host plants in a mutually beneficial relationship, given that host plants divert their resources away from poorly cooperating AM fungi (Kiers et al., 2011). This would render specialized nuclei ineffective at low proportions, causing increasing returns to specialization ($\alpha > 1$). Conflict and interference between nuclei would also lead to increasing returns from specialization. Nuclear conflict could render specialized nuclei ineffective at low proportions where their relatedness to other nuclei is low. Interference among nuclei may mean low proportions of specialized nuclei are swamped and unable to contribute to network-level functionality.

We now ask when genomic diversity ($0 < x < 1$), as opposed to purity ($x = 0$ or $x = 1$), is favored at the individual level. This will be the case when the fitness of an individual (W ; Equation 1) is maximized at some intermediate nuclear proportion, which requires the mathematical conditions: $\frac{dW}{dx} = 0$, $\frac{d^2W}{dx^2} < 0$, $0 < x^* < 1$ (Maynard Smith & Price, 1973; Taylor, 1996). These conditions are satisfied when there is a mixture of the two plant species in the environment ($0 < p < 1$), and the returns to specialization are diminishing ($0 < \alpha < 1$) (Appendix 1). Given this, genomic diversity is favored, and the specific nuclear proportion (x) that is favored is as follows:

$$x^* = \frac{1}{1 + \left(\frac{1-p}{p}\right)^{\frac{1}{1-\alpha}}} \quad (2)$$

We can convert the equilibrium nuclear proportion (x^*) to a measure of genomic diversity (z^*), which ranges from zero to one, and is

maximal when there is an equal proportion of type 1 and type 2 nuclei ($z^* = 1 - 2|x^* - 0.5|$). More extreme genomic diversity is favored by between-individual selection ($z^* \rightarrow 1$) as returns become more diminishing ($\alpha \rightarrow 0$) and the environment becomes more mixed ($p \rightarrow 0.5$) (Figure 1a). As returns become more diminishing, the relative benefit of having a small fraction of each nucleus is increased, favoring diversity. Our result illustrates, for the specific case of genomic diversity in an individual, how life history and ecology can select for “generalist” phenotypes (Hedrick, Ginevan, & Ewing, 1976; Levins, 1962, 1966; Levins & MacArthur, 1966). Furthermore, our model implies that genomic diversity might be favored in some, but not all environments (Sanders, 2018). Discrepancies between different empirical estimates of genomic diversity in natural AM fungi populations might reflect environmental differences in either: (a) the density of plants (which may affect the returns on nucleus specialization); or (b) the mixture of different plant types.

2.2 | Competing nuclei

Our above model examined why individuals with genomic diversity might outcompete individuals with genomic homogeneity. A potential problem here is that nucleus (genome) lineages might be more competitive or selfish, replicating faster within individuals and eliminating genomic diversity as they come to dominance (Frank, 1998). Consequently, we now examine whether such within-individual competition could be balanced by the benefits of being in an individual with genomic diversity (between-individual selection). We are therefore taking the result from the Competing Individuals (Levins, 1962) model that individuals with genomic diversity have a higher fitness, and examining the consequences for the maintenance of within-individual genomic diversity. Our aim here is to analyze an abstract, heuristic case—in the following

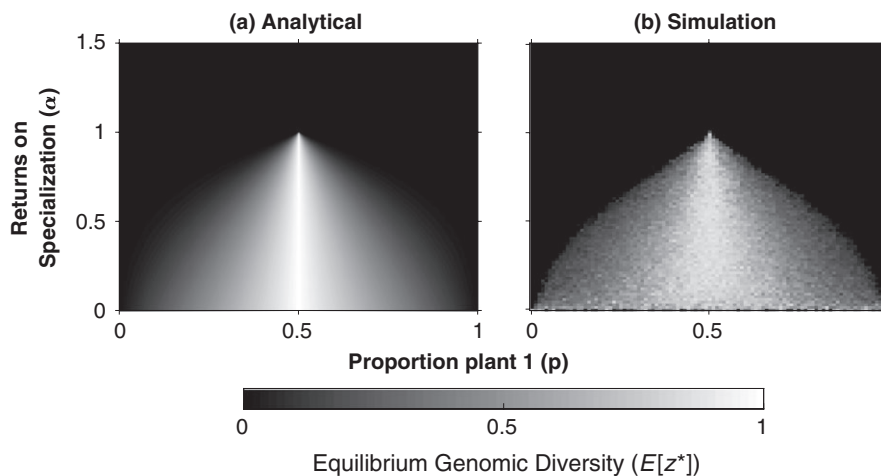


FIGURE 1 Effect of environmental variability (p) and the curvature of specialization returns (α) on genomic diversity. Both parts show the level of genomic diversity at evolutionary equilibrium ($E[z^*]$) in the absence of nuclear replicative differences. The y-axis is the shape of the relationship between fitness and nucleus proportion (α), where $\alpha > 1$ reflects accelerating returns to specialization and $\alpha < 1$ reflects diminishing returns. The x-axis is the proportion of plant species one (p), relative to plant species two ($1-p$). Part (a) shows the analytically derived ESS of the Competing Individuals model, and part (b) shows the results of our individual-based simulation ($n = 2000$, $f = 0.005$, $d = 0.5$, $m = 0$). The results of our ESS model and our simulation are quantitatively equivalent, showing that genomic diversity is stabilized, for diminishing returns to specialization ($\alpha \rightarrow 0$) and mixed environments ($p \rightarrow 0.5$), in the absence of replicative differences between nuclei

section, we use a simulation approach to analyze a more biologically realistic scenario.

We model a population of individuals assuming different proportions of type 1 (N_1) relative to type 2 (N_2) nuclei, x . We model the population as a distribution with a mean nuclear proportion $E[X]$. Every generation, individuals undergo nucleus replication, where *within-individual* selection can occur, then asexual reproduction (sporulation), where *between-individual* selection can occur (Supporting information Figure S3). There is no sharing of nuclei between individuals; individuals die at an arbitrary rate independent of nuclear proportion (x); offspring have the same nuclear proportion (x) as their asexual parent (perfect inheritance).

In the nucleus replication phase, type 1 and type 2 nuclei replicate and compete within individuals, with type 1 nuclei gaining a propagative advantage. We assume a competitive regime within individuals in which the population average nuclear proportion increases by some constant value (θ , where $\theta > 0$). Individuals then reproduce (sporulate) asexually in proportion to their (individual) fitness. The fitness of an individual increases as its genomic diversity approaches some environmentally determined optimal value (μ , where $0 < \mu < 1$). We assume an abstract competitive regime, contingent on the exact form of the distribution of individuals, and of fitness, across different nuclear proportions, in which the response of the population to between-individual selection is constant and given by s ($0 < s < 1$). This will be higher if nuclei strongly affect fitness, and if there is high variation between individuals. Combining our assumptions, the generational change in mean nuclear proportion is as follows:

$$E[X]_{t+1} = s\mu + (1-s)(E[X]_t + \theta). \quad (3)$$

We set $E[X]_t = E[X]_{t+1} = E[X^*]$, and find that the equilibrium (*absorption*) state of the distribution occurs at a mean genomic diversity of $E[X^*] = \mu + \frac{1-s}{s}\theta$. We show in Appendix 2 that this state corresponds to genomic diversity ($0 < E[X^*] < 1$) when:

$$s(1-\mu) > (1-s)\theta. \quad (4)$$

The left-hand side $s(1-\mu)$ represents the stabilizing force of between-individual selection, effective when between-individual selection strongly disfavors fast-replicating nuclei (high s ; low μ). The right-hand side $(1-s)\theta$ represents the destabilizing, directional force of within-individual selection, effective when competitive differences between nuclei within individuals are large relative to the competitive differences between individuals (high θ ; low s). Genomic diversity is evolutionarily stabilized if between-individual selection for genomic diversity exceeds within-individual selection for competitive genomes (nuclei).

This condition is analogous to mutation-selection balance in population genetics (Haldane, 1927; Lande, 1975), and group versus individual selection in social evolution theory (Hamilton, 1975; Price, 1972). In these cases, a given evolutionary outcome is dependent on how two opposing evolutionary forces are

resolved (Frank, 2011). This perspective provides a framework for understanding why genomic diversity is common in organisms that enforce synchronous nuclear replication ($\theta = 0$), and why nonfunctional “cheating” nuclei are sometimes evolutionarily stable (Appendix 3). Our qualitative conclusions hold when the order of within- and between-individual selection is reversed (Supporting information Data S1), when within-individual selection and between-individual selection are modeled in a more general, less abstracted, framework (Supporting information Data S2), and when an explicit form of the distribution of individuals is assumed (*unpublished*).

2.3 | AM fungi simulation

In the Competing Individuals model, we showed that between-individual selection can favor within-individual genomic diversity. In the Competing Nuclei model, we took this result and showed that diversity can be stably maintained even if genomes compete within individuals. However, to make our analysis general and analytically tractable, we made several simplifying assumptions with regard to: within-individual selection (nuclear replication was not explicitly modeled); between-individual selection (distribution of individuals, and of fitness, across different nuclear proportions, was not explicitly modeled); unstructured populations (no dispersal); no fusion of individuals (*anastomosis*); no stochasticity regarding which nuclei enter asexual spores (perfect inheritance of the nuclear proportion, x).

We built a simulation model that allowed us to relax these simplifying assumptions, resulting in a closer representation of the biology of AM fungi and many other modular organisms (Figure 2). We have two broad aims with our simulation. First, we examine whether the predictions of our simple analytical models hold when more biological realism is incorporated, in a fully dynamical model. Second, we examine the influence of a number of additional factors, including differential rates of replication between strains, the fusion of individuals (*anastomosis*), dispersal, and spore size.

2.3.1 | Simulation details

We implement a population of n individuals in an individual-based computer simulation model. The population is split into j patches with n/j individuals per patch. Individuals bear some proportion of type 1 (N_1) relative to type 2 (N_2) nuclei (x , as in previous models). An individual's initial nuclear proportion is drawn at random from a uniform distribution bound between zero and one. We assume the following lifecycle. First, individuals grow from a single spore and their nuclei grow exponentially, with type 1 nuclei replicating faster than type 2 nuclei ($r_1 > r_2$). Next, individuals temporarily fuse with a random patch-mate with some probability (m), share nuclei, and acquire new nuclear proportions (x) that are a mean of their nuclear proportions prior to fusion. The actual probability of nonself fusion between AM fungi networks in nature is unclear,

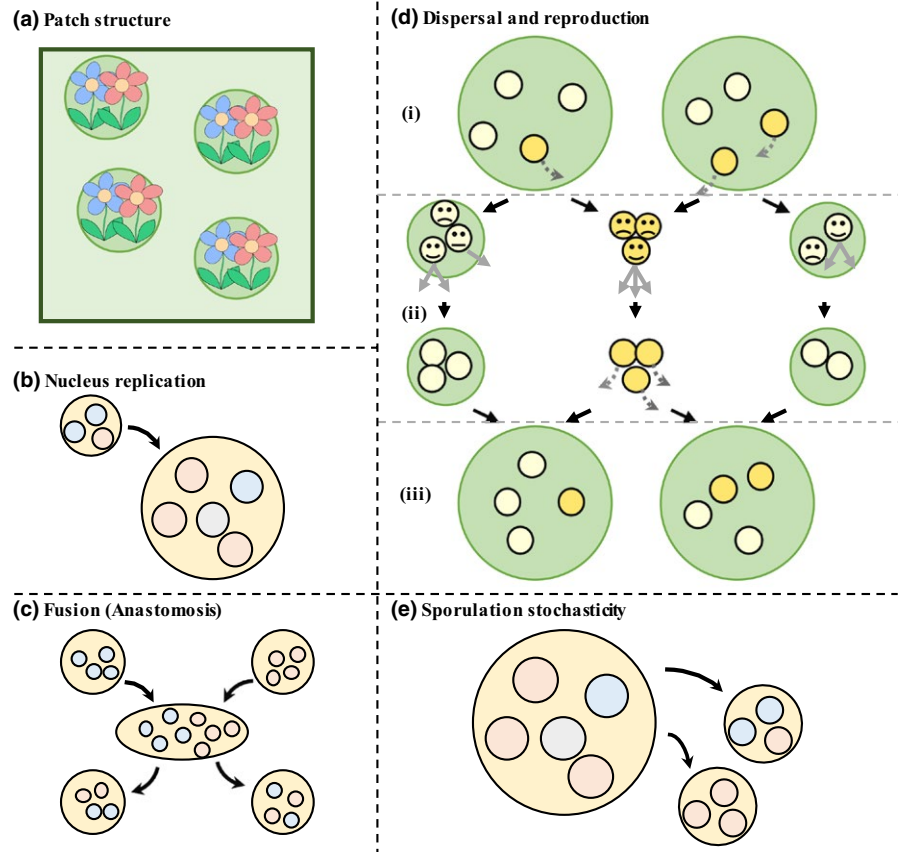


FIGURE 2 Simulation lifecycle. (a) The population of individuals (green box) is patch structured (circles containing plants). (b) Type 1 nuclei (red) replicate faster than type 2 nuclei (blue). (c) Fusion (*anastomosis*) is pairwise, with nuclei shared evenly between individuals via the formation then lesion of a large fused individual. (d) Individuals with dispersing offspring are orange, and compete with each other globally. Individuals with non-dispersing offspring are beige, and compete with each other locally on their native patch (green circles). (dii) Individuals with higher fitness (smile) are more likely to reproduce (gray solid lines) into free spots. (diii) Offspring that have dispersed (orange) are sorted at random back into patches (green circles). (e) An offspring's genotype deviates stochastically from its asexual parent's genotype

with experimental estimates ranging from 6% to 90% (Giovannetti et al., 2015).

Next, individuals reproduce with a probability proportional to their fitness, which is given by Equation (1). As shown in the Competing Individuals model, this fitness equation favors genomic diversity if there is a mixture of host plants ($0 < p < 1$) and functional synergy between type 1 and type 2 nuclei ($0 < \alpha < 1$); it favors purity of one nucleus strain otherwise. Fitness is judged relative to patchmates if an individual's offspring are not dispersed; fitness is judged relative to global dispersers if an individual's offspring are dispersed. Offspring dispersal occurs with some probability (d), and in AM fungi, it is likely to occur via soil-disrupting vertebrates that transfer spores between otherwise-isolated clusters of plants (Savary, Masclaux, et al., 2018; Vályi, Mardhiah, Rillig, & Hempel, 2016).

Offspring inherit a random sample of nuclei from their asexual parent. Offspring nuclear proportion deviates from their asexual parent by some number drawn randomly from a truncated normal distribution with a standard deviation (f) reflecting the level of sporulation stochasticity. The parameter f captures spore size—spores that inherit a small proportion of parental nuclei will be subject to higher stochasticity in nuclear inheritance (f). Parents die after reproducing. Though generational death (nonoverlapping generations) does not strictly apply, this is a standard modeling assumption to simplify analysis. More precise simulation details are given in Appendix 4.

We track nuclear proportion in each individual (x), over many generations, until the system equilibrates, to see if genomic diversity

is stable. An intermediate mean nuclear proportion ($0 < E[x^*] < 1$) is not sufficient to show that diversity is present within individuals, because this condition is also satisfied by populations comprising genomically pure individuals, some bearing type 1 nuclei and others type 2. Therefore, for each individual, we convert the nuclear proportion (x) to a genomic diversity score (z), which ranges from zero to one ($z = 1 - 2|x - 0.5|$). Genomic diversity is stable if the population average level of diversity is greater than zero at equilibrium ($E[z^*] > 0$).

2.3.2 | Simulation results

We found broad support between our analytical models and our simulation—when there is replicative synchrony between nuclei ($r_1 = r_2$), genomic diversity can be favored (Figure 1). As the replicative advantage of type 1 nuclei ($(r_1 - r_2)/r_2$) is increased, the diversity at equilibrium ($E[z^*]$) is reduced and tends toward zero (Figure 3a; solid line). This result holds regardless of the nature of between-individual selection ($\alpha > 0$, $0 \leq p \leq 1$) (Figure 4a).

Examining the extra factors in our simulation, we found that, as the replicative advantage of type 1 nuclei is increased ($(r_1 - r_2)/r_2$), the corresponding reduction in equilibrium genomic diversity ($E[z^*]$) is exaggerated by fusion between individuals (*anastomosis*) (Figure 3a; dashed line), and attenuated by sporulation stochasticity (f) (Figure 3a; dotted line). The exaggerating force of fusion and the attenuating force of sporulation stochasticity are observable across the full range of between-individual selection ($\alpha > 0$, $0 \leq p \leq 1$) (Figure 4b and c).

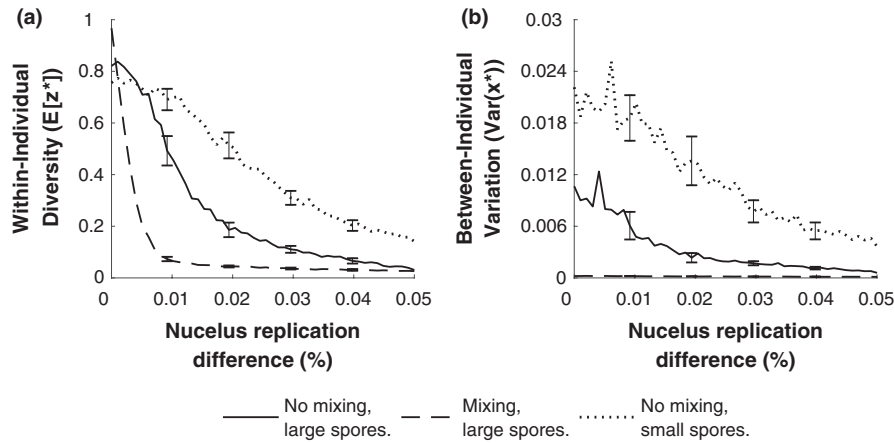


FIGURE 3 Nuclear diversity within and between individuals. The within-individual genomic diversity (a), and between-individual variation in nuclear proportion (b), is plotted against the nuclear replicative advantage of type 1 nuclei ($r_1 - r_2 / r_2$) ($\alpha = 0.8$, $p = 0.5$, $d = 0.5$, $r_2 = 0.3$, r_1 is varied). The different lines represent different degrees of fusion (no fusion $m = 0$; fusion: $m = 0.05$) and different spore sizes (large: $f = 0.005$; small: $f = 0.01$). Fusion between lines (higher m) leads to an effectively complete loss of variation between individuals, which reduces the strength of between-individual selection, and hence leads to a faster rate of loss of within-individual genomic diversity. Smaller spores (higher $f = 0.01$) lead to an increased sporulation stochasticity, which increases between-individual variation, resulting in a slower rate of loss of within-individual genomic diversity. The plots represent the average results taken across 10 trials. Error bars, where plotted, show one standard deviation above and below the mean across these 10 trials

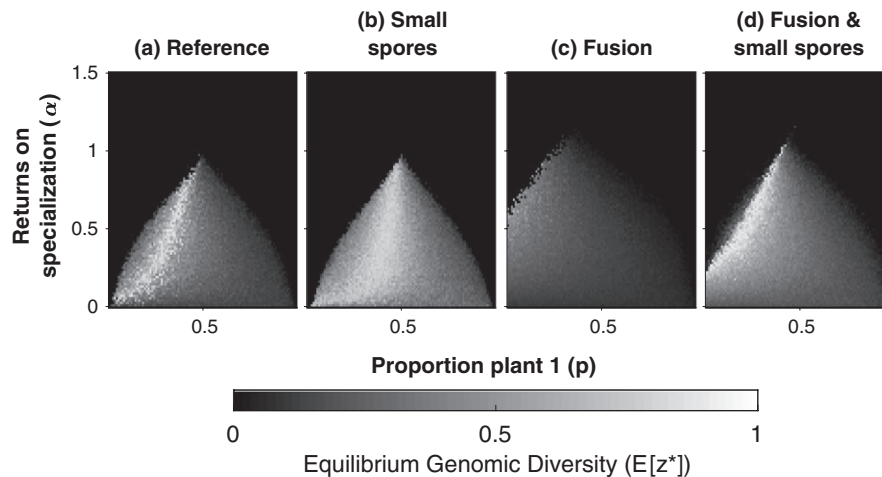


FIGURE 4 Maintenance of genomic diversity for different between-individual selection pressures. The results of the AM Fungi Simulation model are plotted, showing the level of genomic diversity maintained within individuals at equilibrium ($E[z^*]$). The heat maps plot the full range of between-individual selection, from decelerating to accelerating returns on plant specialization (α , y-axis), and from a plant 2 to a plant 1 dominated environment (p , x-axis). Nucleus 1 has a replicative advantage ($r_1 = 0.305$, $r_2 = 0.3$), meaning (a) genomic diversity is favored in environments that are slightly dominated by plant 2, which the slower replicating nucleus is specialized on ($m = 0$, $f = 0.005$). (b) As sporulation stochasticity is increased (small spores), more genomic diversity is stable across the between-individual selection parameter space ($f = 0.01$). (c) Fusion of individuals destabilizes genomic diversity over most of the parameter space at equilibrium ($m = 0.05$). (d) The counteracting effects of fusion and sporulation stochasticity can cancel each other out ($f = 0.01$, $m = 0.05$). These results assumed $n = 2,000$ (population size), $d = 0.5$ (dispersal)

These effects arise because fusion reduces (Figure 3b; dashed line) between-individual variation ($\text{Var}(x)$) and sporulation stochasticity increases it (Figure 3b; dotted line), correspondingly decreasing, and respectively, increasing, the efficacy of (stabilizing) between-individual selection relative to (destabilizing) within-individual selection.

We find that if genomic diversity is neutral at the within-individual ($r_1 = r_2$) and not favored at the individual level ($\alpha \geq 1$), fusion

(anastomosis) can prolong the maintenance of genomic diversity in a nonequilibrium state, by attenuating the loss of genomic diversity through individual-level drift (Supporting information Figure S4; Bever & Wang, 2005; Pawlowska & Taylor, 2004). We find that dispersal does not significantly increase between-individual variation (Supporting information Figure S5b), but increases the effective population size by connecting patches, in turn increasing the

efficacy of between-individual selection, slightly stabilizing genomic diversity (Supporting information Figure S5).

3 | DISCUSSION

We provide an evolutionary explanation for the maintenance of genomic diversity in AM hyphal networks that may apply more broadly to other modular organisms. If nuclei, or specifically, *particular genes on nuclei*, are functionally specialized on different plant hosts, the cost of genome conflict borne by individuals with genomic diversity may be outweighed by the benefit of being a good generalist in a variable environment. If this between-individual selection for genomic diversity exceeds within-individual selection for the single fastest replicating nucleus genome, genomic diversity can be evolutionarily stable.

A key assumption in our models is that genomes (nuclei) are functionally specialized on aspects of their environment (host plants) (Strassmann & Queller, 2004). Consistent with this, the fitness of AM fungal individuals (hyphal networks) has been empirically shown to depend on an interaction between the strain of the hyphal network (genotype) and its host plant species (environment), implying nucleus specialization (Angelard et al., 2010, 2013; Ehinger, Koch, & Sanders, 2009; Savary, Masclaux, et al., 2018; Savary, Villard, & Sanders, 2018). Our model could be extended in numerous ways, to explore other factors, potentially important to AM fungi, or other organisms. For example, more nucleus types could be considered, or replication rates could be allowed to evolve (Czárán, Hoekstra, & Aanen, 2014; Frank, 1994; Wyss et al., 2016).

There are organisms other than AM fungi capable of genomic diversity, mostly restricted to those that grow through iterations of modules, like hyphae or stems, that each retains reproductive capability. These *modular* organisms include many filamentous fungi, colonial invertebrates like sponges, and plants that grow from underground connected stems called rhizomes (Herron, Rashidi, Shelton, & Driscoll, 2013; Pineda-Krch & Lehtila, 2004). Our theory is that genomic diversity allows modular organisms to adapt to heterogeneous environments. Although a benefit to genomic diversity has been demonstrated in some other organisms, including ascidians, red algae, and other fungi, it is unclear whether environmental specialization of genomes contributes to these benefits (Rinkevich & Shapira, 1999; Santelices et al., 1999). Other hypotheses for the benefit of genomic diversity include the following: the simultaneous generation of multiple predator-defense phenotypes (genetic mosaic hypothesis), a mechanism for screening and selecting the best mutations, and the increased size that can result from fusing individuals (Aanen, Debets, Visser, & Hoekstra, 2008; Bastiaans, Debets, & Aanen, 2015; Castillo, Switz, Foster, Queller, & Strassmann, 2005; Foster, Fortunato, Strassmann, & Queller, 2002; Gill, Chao, Perkins, & Wolf, 1995; Jany & Pawlowska, 2010; Otto & Hastings, 1998; Otto & Orive, 1995; Whitham & Slobodchikoff, 1981).

We have focused on long-term evolutionary (*ultimate*) causes of genomic diversity, which complement previous studies of immediate (*proximate*) generators of genomic diversity. Fusion (*anastomosis*) promotes the (proximate) spread of new genomes through populations of individuals (Croll et al., 2009; de Novais et al., 2013), but destabilizes genomic diversity in evolutionary time by reducing variation between individuals for selection to act on. Stochasticity in sporulation can result in the (proximate) loss of genomic diversity over a generation (Angelard et al., 2010; Boon et al., 2013; Marleau et al., 2011; Masclaux et al., 2018), though it stabilizes genomic diversity in evolutionary time by increasing variation between individuals. Other possible proximate influencers of genomic diversity include the following: de novo mutations; the restriction of fusion to close kin (*allorecognition*) (Czárán et al., 2014); and genetic exchange between nuclei (Chen et al., 2018a; Croll & Sanders, 2009).

To conclude, throughout this paper, we have referred to AM fungi and other modular organisms exhibiting genomic diversity as “individuals.” However, from an evolutionary perspective, “individuality” or “organismality” requires cooperation and lack of conflict between component parts (Buss, 1988; Gardner & Grafen, 2009; Maynard Smith & Szathmáry, 1997; Queller & Strassmann, 2009, 2016; West et al., 2015). Genomic conflict pulls entities away from optimal trait values (Competing Nuclei model), limiting adaptation (Strassmann & Queller, 2007). Despite this, we have shown that entities with genomic diversity can be selected and come to dominate populations. For this reason, although we may not wish to call them “organisms” (Folse & Roughgarden, 2010; Queller & Strassmann, 2009), such entities are capable of lasting evolutionary stability—hundreds of millions of years in the case of AM fungi (Heckman, 2001).

ACKNOWLEDGMENTS

We thank Duur Aanen, Ian Sanders, Daniel Wilson, and Jared Field for useful discussion and comments on the manuscript; Natural Environment Research Council, Engineering and Physical Sciences Research Council, Swiss National Science Foundation, European Research Council ERC Grant Agreement 335542 (to E.T.K.) for funding; Magdalen College for emergency housing.

CONFLICT OF INTEREST

None declared.

AUTHOR CONTRIBUTION

T.W.S., E.T.K., and S.A.W. designed the study and wrote the article; T.W.S., M.d.S., and G.A.C. contributed to mathematical modeling.

DATA ACCESSIBILITY

We agree to deposit our data to a public repository.

ORCID

Thomas W. Scott  <https://orcid.org/0000-0002-5941-3905>

REFERENCES

- Aanen, D. K., Debets, A. J. M., de Visser, J. A. G. M., & Hoekstra, R. F. (2008). The social evolution of somatic fusion. *BioEssays*, 30, 1193–1203. <https://doi.org/10.1002/bies.20840>
- Angelard, C., Colard, A., Niculita-Hirzel, H., Croll, D., & Sanders, I. R. (2010). Segregation in a mycorrhizal fungus alters rice growth and symbiosis-specific gene transcription. *Current Biology*, 20, 1216–1221. <https://doi.org/10.1016/j.cub.2010.05.031>
- Angelard, C., Tanner, C. J., Fontanillas, P., Niculita-Hirzel, H., Masclaux, F., & Sanders, I. R. (2013). Rapid genotypic change and plasticity in arbuscular mycorrhizal fungi is caused by a host shift and enhanced by segregation. *ISME Journal*, 8, 284–294. <https://doi.org/10.1038/ismej.2013.154>
- Bastiaans, E., Debets, A. J. M., & Aanen, D. K. (2015). Experimental demonstration of the benefits of somatic fusion and the consequences for allorecognition. *Evolution*, 69, 1091–1099. <https://doi.org/10.1111/evo.12626>
- Bastiaans, E., Debets, A. J. M., & Aanen, D. K. (2016). Experimental evolution reveals that high relatedness protects multicellular cooperation from cheaters. *Nature Communications*, 7, 11435. <https://doi.org/10.1038/ncomms11435>
- Bever, J. D., & Wang, M. (2005). Arbuscular mycorrhizal fungi: Hyphal fusion and multigenomic structure. *Nature*, 433, E3–E4. <https://doi.org/10.1038/nature03294>
- Bonfante, P., & Genre, A. (2010). Mechanisms underlying beneficial plant–fungus interactions in mycorrhizal symbiosis. *Nature Communications*, 1, 1–11. <https://doi.org/10.1038/ncomms1046>
- Boon, E., Zimmerman, E., St-Arnaud, M., & Hijri, M. (2013). Allelic differences within and among sister spores of the Arbuscular Mycorrhizal fungus *glomus etunicatum* suggest segregation at sporulation. *PLoS One*, 8, e83301. <https://doi.org/10.1371/journal.pone.0083301>
- Bruns, T. D., Corradi, N., Redecker, D., Taylor, J. W., & Öpik, M. (2017). Glomeromycotina: What is a species and why should we care? *New Phytologist*, 8, 274. <https://doi.org/10.1111/nph.14913>
- Buss, L. W. (1988). *The evolution of individuality*. Princeton: Princeton University Press.
- Castillo, D. I., Switz, G. T., Foster, K. R., Queller, D. C., & Strassmann, J. E. (2005). A cost to chimerism in *Dictyostelium discoideum* on natural substrates. *Evolutionary Ecology Research*, 7, 263–271.
- Chen, E. C., Mathieu, S., Hoffrichter, A., Sedzielewska-Toro, K., Peart, M., Pelin, A., ... Corradi, N. (2018a). Single nucleus sequencing reveals evidence of internucleus recombination in arbuscular mycorrhizal fungi. *eLife*, 7, 692.
- Chen, E. C. H., Morin, E., Beaudet, D., Noel, J., Yildirim, G., Ndikumana, S., ... Corradi, N. (2018b). High intraspecific genome diversity in the model arbuscular mycorrhizal symbiont *Rhizophagus irregularis*. *New Phytologist*, 19, 455.
- Corradi, N., & Brachmann, A. (2017). Fungal mating in the most widespread plant symbionts? *Trends in Plant Science*, 22, 175–183. <https://doi.org/10.1016/j.tplants.2016.10.010>
- Croll, D., Giovannetti, M., Koch, A. M., Sbrana, C., Ehinger, M., Lammers, P. J., & Sanders, I. R. (2009). Nonself vegetative fusion and genetic exchange in the arbuscular mycorrhizal fungus *Glomus intraradices*. *New Phytologist*, 181, 924–937.
- Croll, D., & Sanders, I. R. (2009). Recombination in *Glomus intraradices*, a supposed ancient asexual arbuscular mycorrhizal fungus. *BMC Evolutionary Biology*, 9, 13.
- Croll, D., Wille, L., Gamper, H. A., Mathimaran, N., Lammers, P. J., Corradi, N., & Sanders, I. R. (2008). Genetic diversity and host plant preferences revealed by simple sequence repeat and mitochondrial markers in a population of the arbuscular mycorrhizal fungus *Glomus intraradices*. *New Phytologist*, 178, 672–687.
- Czárán, T., Hoekstra, R. F., & Aanen, D. K. (2014). Selection against somatic parasitism can maintain allorecognition in fungi. *Fungal Genetics and Biology*, 73, 128–137. <https://doi.org/10.1016/j.fgb.2014.09.010>
- de Novais, C. B., Sbrana, C., Júnior, O. J. S., Siqueira, J. O., & Giovannetti, M. (2013). Vegetative compatibility and anastomosis formation within and among individual germlings of tropical isolates of arbuscular mycorrhizal fungi (Glomeromycota). *Mycorrhiza*, 23, 325–331. <https://doi.org/10.1007/s00572-013-0478-y>
- Ehinger, M. O., Croll, D., Koch, A. M., & Sanders, I. R. (2012). Significant genetic and phenotypic changes arising from clonal growth of a single spore of an arbuscular mycorrhizal fungus over multiple generations. *New Phytologist*, 196, 853–861. <https://doi.org/10.1111/j.1469-8137.2012.04278.x>
- Ehinger, M., Koch, A. M., & Sanders, I. R. (2009). Changes in arbuscular mycorrhizal fungal phenotypes and genotypes in response to plant species identity and phosphorus concentration. *New Phytologist*, 184, 412–423. <https://doi.org/10.1111/j.1469-8137.2009.02983.x>
- Folse, H. J. III, & Roughgarden, J. (2010). What is an individual organism? A multilevel selection perspective. *The Quarterly Review of Biology*, 85, 447–472. <https://doi.org/10.1086/656905>
- Foster, K. R., Fortunato, A., Strassmann, J. E., & Queller, D. C. (2002). The costs and benefits of being a chimera. *Proceedings of the Royal Society B-Biological Sciences*, 269, 2357–2362. <https://doi.org/10.1098/rspb.2002.2163>
- 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. <https://doi.org/10.1098/rspb.1994.0156>
- Frank, S. A. (1995). Mutual policing and repression of competition in the evolution of cooperative groups. *Nature*, 377, 520–522.
- Frank, S. A. (1998). *Foundations of social evolution*. Princeton: Princeton University Press.
- Frank, S. A. (2003). Perspective: Repression of competition and the evolution of cooperation. *Evolution*, 57, 693. [https://doi.org/10.1554/0014-3820\(2003\)057\[0693:PROCAT\]2.0.CO;2](https://doi.org/10.1554/0014-3820(2003)057[0693:PROCAT]2.0.CO;2)
- Frank, S. A. (2011). Natural selection. III. Selection versus transmission and the levels of selection. *Journal of Evolutionary Biology*, 25, 227–243. <https://doi.org/10.1111/j.1420-9101.2011.02431.x>
- Gardner, A., & Grafen, A. (2009). Capturing the superorganism: A formal theory of group adaptation. *Journal of Evolutionary Biology*, 22, 659–671. <https://doi.org/10.1111/j.1420-9101.2008.01681.x>
- Gilbert, O. M., Foster, K. R., Mehdiabadi, N. J., Strassmann, J. E., & Queller, D. C. (2007). High relatedness maintains multicellular cooperation in a social amoeba by controlling cheater mutants. *Proceedings of the National Academy of Sciences*, 104, 8913–8917. <https://doi.org/10.1073/pnas.0702723104>
- Gill, D. E., Chao, L., Perkins, S. L., & Wolf, J. B. (1995). Genetic mosaicism in plants and clonal animals. *Annual Review of Ecology and Systematics*, 26, 423–444. <https://doi.org/10.1146/annurev.es.26.110195.002231>
- Giovannetti, M., Avio, L., & Sbrana, C. (2015). Functional significance of anastomosis in arbuscular mycorrhizal networks. In T.R. Horton (Ed.), *Mycorrhizal networks* (3rd ed., pp. 41–67). Dordrecht: Springer Netherlands.
- Haldane, J. B. S. (1927). A mathematical theory of natural and artificial selection, Part V: Selection and mutation. *Mathematical Proceedings of the Cambridge Philosophical Society*, 23, 838–844. <https://doi.org/10.1017/S0305004100011750>
- Hamilton, W. D. (1975). Innate social aptitudes of man: An approach from evolutionary genetics. In R. Fox (Ed.), *Biosocial anthropology* (pp. 133–155). New York, NY: Wiley.
- Heckman, D. S. (2001). Molecular Evidence for the Early Colonization of Land by Fungi and Plants. *Science*, 293, 1129–1133. <https://doi.org/10.1126/science.1061457>

- Hedrick, P. W., Ginevan, M. E., & Ewing, E. P. (1976). Genetic polymorphism in heterogeneous environments. *Annual Review of Ecology and Systematics*, 7, 1–32. <https://doi.org/10.1146/annurev.es.07.110176.000245>
- Herron, M. D., Rashidi, A., Shelton, D. E., & Driscoll, W. W. (2013). Cellular differentiation and individuality in the “minor” multicellular taxa. *Biological Reviews*, 88, 844–861. <https://doi.org/10.1111/brv.12031>
- Hijri, M., & Sanders, I. R. (2005). Low gene copy number shows that arbuscular mycorrhizal fungi inherit genetically different nuclei. *Nature*, 433, 160–163. <https://doi.org/10.1038/nature03069>
- Inglis, R. F., Ryu, E., Asikhia, O., Strassmann, J. E., & Queller, D. C. (2017). Does high relatedness promote cheater-free multicellularity in synthetic lifecycles? *Journal of Evolutionary Biology*, 30, 985–993. <https://doi.org/10.1111/jeb.13067>
- Jany, J. L., & Pawlowska, T. E. (2010). Multinucleate spores contribute to evolutionary longevity of asexual glomeromycota. *The American Naturalist*, 175, 424–435. <https://doi.org/10.1086/650725>
- Kiers, E. T., Duhamel, M., Beesetty, Y., Mensah, J. A., Franken, O., Verbruggen, E., ... Bucking, H. (2011). Reciprocal rewards stabilize cooperation in the Mycorrhizal symbiosis. *Science*, 333, 880–882. <https://doi.org/10.1126/science.1208473>
- Kooij, P. W., Aanen, D. K., Schiøtt, M., & Boomsma, J. J. (2015). Evolutionarily advanced ant farmers rear polyploid fungal crops. *Journal of Evolutionary Biology*, 28, 1911–1924. <https://doi.org/10.1111/jeb.12718>
- Kuhn, G., Hijri, M., & Sanders, I. R. (2001). Evidence for the evolution of multiple genomes in arbuscular mycorrhizal fungi. *Nature*, 414, 745–748. <https://doi.org/10.1038/414745a>
- Lande, R. (1975). The maintenance of genetic variability by mutation in a polygenic character with linked loci. *Genetical Research*, 26, 221–235. <https://doi.org/10.1017/S0016672300016037>
- Levins, R. (1962). Theory of fitness in a heterogeneous environment. I. The fitness set and adaptive function. *The American Naturalist*, 96, 361–373. <https://doi.org/10.1086/282245>
- Levins, R. (1966). The strategy of model building in population. *American Science*, 54, 421–431.
- Levins, R., & MacArthur, R. (1966). The maintenance of genetic polymorphism in a spatially heterogeneous environment: Variations on a theme by Howard Levene. *The American Naturalist*, 100, 585–589. <https://doi.org/10.1086/282452>
- Lin, K., Limpens, E., Zhang, Z., Ivanov, S., Saunders, D. G. O., Mu, D., ... Huang, S. (2014). Single nucleus genome sequencing reveals high similarity among nuclei of an Endomycorrhizal Fungus. *PLoS Genetics*, 10, e1004078. <https://doi.org/10.1371/journal.pgen.1004078>
- Maeda, T., Kobayashi, Y., Kameoka, H., Okuma, N., Takeda, N., Yamaguchi, K., ... Kawaguchi, M. (2018). Evidence of non-tandemly repeated rDNAs and their intragenomic heterogeneity in *Rhizophagus irregularis*. *Communications Biology*, 1, 1–13. <https://doi.org/10.1038/s42003-018-0094-7>
- Marleau, J., Dalpé, Y., St-Arnaud, M., & Hijri, M. (2011). Spore development and nuclear inheritance in arbuscular mycorrhizal fungi. *BMC Evolutionary Biology*, 11, 96. <https://doi.org/10.1186/1471-2148-11-51>
- Masclaux, F. G., Wyss, T., Mateus-Gonzalez, I. D., Aletti, C., & Sanders, I. R. (2018). Variation in allele frequencies at the bg112 locus reveals unequal inheritance of nuclei in a dikaryotic isolate of the fungus *Rhizophagus irregularis*. *Mycorrhiza*, 28, 369–377. <https://doi.org/10.1007/s00572-018-0834-z>
- 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., & Szathmáry, E. (1997). *The major transitions in evolution*. Oxford: Oxford University Press.
- Meunier, C., Hosseini, S., Heidari, N., Maryush, Z., & Johannesson, H. (2018). Multilevel selection in the filamentous ascomycete *Neurospora tetrasperma*. *American Naturalist*, 191, 290–305.
- Otto, S. P., & Hastings, I. M. (1998). Mutation and selection within the individual. In R. Woodruff & J. Thompson (Eds.), *Mutation and evolution* (Vol. 7, pp. 507–524). Dordrecht: Springer Netherlands.
- Otto, S. P., & Orive, M. E. (1995). Evolutionary consequences of mutation and selection within an individual. *Genetics*, 141, 1173–1187.
- Pawlowska, T. E., & Taylor, J. W. (2004). Organization of genetic variation in individuals of arbuscular mycorrhizal fungi. *Nature*, 427, 733–737. <https://doi.org/10.1038/nature02290>
- Pineda-Krch, M., & Lehtila, K. (2004). Costs and benefits of genetic heterogeneity within organisms. *Journal of Evolutionary Biology*, 17, 1167–1177. <https://doi.org/10.1111/j.1420-9101.2004.00808.x>
- Price, G. R. (1972). Extension of covariance selection mathematics. *Annals of Human Genetics*, 35, 485–490. <https://doi.org/10.1111/j.1469-1809.1957.tb01874.x>
- Queller, D. C., & Strassmann, J. E. (2009). Beyond society: The evolution of organismality. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364, 3143–3155. <https://doi.org/10.1098/rstb.2009.0095>
- Queller, D. C., & Strassmann, J. E. (2016). Problems of multi-species organisms: Endosymbionts to holobionts. *Biology and Philosophy*, 31, 855–873. <https://doi.org/10.1007/s10539-016-9547-x>
- Rinkevich, B., & Shapira, M. (1999). Multi-partner urochordate chimeras outperform two-partner chimerical entities. *Oikos*, 87, 315. <https://doi.org/10.2307/3546746>
- Roberts, S. E., & Gladfelter, A. S. (2015). Nuclear autonomy in multinucleate fungi. *Current Opinion in Microbiology*, 28, 60–65. <https://doi.org/10.1016/j.mib.2015.08.009>
- Ropars, J., & Corradi, N. (2015). Homokaryotic vs heterokaryotic mycelium in arbuscular mycorrhizal fungi: Different techniques, different results? *New Phytologist*, 208, 638–641. <https://doi.org/10.1111/nph.13448>
- Ropars, J., Toro, K. S., Noel, J., Pelin, A., Charron, P., Farinelli, L., ... Corradi, N. (2016). Evidence for the sexual origin of heterokaryosis in arbuscular mycorrhizal fungi. *Nature Microbiology*, 1, 16033. <https://doi.org/10.1038/nmicrobiol.2016.33>
- Rosendahl, S., & Stukenbrock, E. H. (2004). Community structure of arbuscular mycorrhizal fungi in undisturbed vegetation revealed by analyses of LSU rDNA sequences. *Molecular Ecology*, 13, 3179–3186. <https://doi.org/10.1111/j.1365-294X.2004.02295.x>
- Sanders, I. R. (2018). Sex, plasticity, and biologically significant variation in one Glomeromycotina species. *New Phytologist*, 20, 1216.
- Sanders, I. R., & Croll, D. (2010). Arbuscular Mycorrhiza: The Challenge to Understand the Genetics of the Fungal Partner. *Annual Review of Genetics*, 44, 271–292. <https://doi.org/10.1146/annurev-genet-102108-134239>
- Santelices, B., Correa, J. A., Aedo, D., Flores, V., Hormazábal, M., & Sánchez, P. (1999). Convergent biological processes in coalescing Rhodophyta. *Journal of Phycology*, 35, 1127–1149. <https://doi.org/10.1046/j.1529-8817.1999.3561127.x>
- Savary, R., Masclaux, F. G., Wyss, T., Droh, G., Cruz Corella, J., Machado, A. P., ... Sanders, I. R. (2018). A population genomics approach shows widespread geographical distribution of cryptic genomic forms of the symbiotic fungus *Rhizophagus irregularis*. *ISME Journal*, 12, 17–30.
- Savary, R., Villard, L., & Sanders, I. R. (2018). Within-species phylogenetic relatedness of a common mycorrhizal fungus affects evenness in plant communities through effects on dominant species. *PLoS ONE* 13(11): e0198537. <https://doi.org/10.1371/journal.pone.0198537>
- Shoji, J.-Y., Kikuma, T., Arioka, M., & Kitamoto, K. (2010). Macroautophagy-mediated degradation of whole nuclei in the filamentous fungus *Aspergillus oryzae*. *PLoS ONE*, 5, e15650. <https://doi.org/10.1371/journal.pone.0015650>
- Strassmann, J. E., & Queller, D. C. (2004). Genetic conflicts and intercellular heterogeneity. *Journal of Evolutionary Biology*, 17, 1189–1191. <https://doi.org/10.1111/j.1420-9101.2004.00814.x>

- Strassmann, J. E., & Queller, D. C. (2007). Insect societies as divided organisms: The complexities of purpose and cross-purpose. *Proceedings of the National Academy of Sciences*, 104, 8619–8626. <https://doi.org/10.1073/pnas.0701285104>
- Taylor, P. D. (1996). The selection differential in quantitative genetics and ESS models. *Evolution*, 50, 2106. <https://doi.org/10.2307/2410769>
- Tisserant, E., Malbreil, M., Kuo, A., Kohler, A., Symeonidi, A., Balestrini, R., ... Martin, F. (2013). Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. *Proceedings of the National Academy of Sciences*, 110, 20117–20122. <https://doi.org/10.1073/pnas.1313452110>
- Vályi, K., Mardhiah, U., Rillig, M. C., & Hempel, S. (2016). Community assembly and coexistence in communities of arbuscular mycorrhizal fungi. *The ISME Journal*, 10, 2341–2351. <https://doi.org/10.1038/ismej.2016.46>
- Vreeburg, S., Nygren, K., & Aanen, D. K. (2016). Unholy marriages and eternal triangles: How competition in the mushroom life cycle can lead to genomic conflict. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371, 20150533. <https://doi.org/10.1098/rstb.2015.0533>
- West, S. A., Fisher, R. M., Gardner, A., & Kiers, E. T. (2015). Major evolutionary transitions in individuality. *Proceedings of the National Academy of Sciences*, 112, 10112–10119. <https://doi.org/10.1073/pnas.1421402112>
- Whitham, T. G., & Slobodchikoff, C. N. (1981). Evolution by individuals, plant-herbivore interactions, and mosaics of genetic variability: The adaptive significance of somatic mutations in plants. *Oecologia*, 49, 287–292.
- Wyss, T., Masclaux, F. G., Roskiewicz, P., Pagni, M., & Sanders, I. R. (2016). Population genomics reveals that within-fungus polymorphism is common and maintained in populations of the mycorrhizal fungus *Rhizophagus irregularis*. *ISME Journal*, 10, 2514–2526. <https://doi.org/10.1038/ismej.2016.29>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Scott TW, Kiers ET, Cooper GA, dos Santos M, West SA. Evolutionary maintenance of genomic diversity within arbuscular mycorrhizal fungi. *Ecol Evol*. 2019;9:2425–2435. <https://doi.org/10.1002/ece3.4834>

APPENDIX

APPENDIX 1: BETWEEN-INDIVIDUAL SELECTION

Equation 2 gives the nuclear proportion corresponding to a stationary point (x^*). It can be seen by inspection of Equation 2 that there is always, and only, one sensical x^* value (one that lies in the range $0 \leq x \leq 1$) for any combination of parameter values. The stationary point (x^*) could represent a maximum (if $\frac{d^2W}{dx^2} < 0$), minimum ($\frac{d^2W}{dx^2} > 0$) or inflection point (if $\frac{d^2W}{dx^2} = 0$). If it is a maximum, then x^* represents the individual-favoured nuclear proportion (the ESS). If it is a minimum or inflection point, then the individual-favoured nuclear proportion (the ESS) will be found at a boundary of $x = 0$ or $x = 1$.

We examine the form of the stationary point for the different ranges of the shape parameter α and baseline fitness κ . When there are increasing returns to specialisation ($\alpha > 1$) and nuclear proportions affect fitness ($0 \leq \kappa < 1$), substituting $\alpha > 1$ into $\frac{d^2W}{dx^2} > 0$ (the condition for x^* to be a minimum) gives $px^{\alpha-2} + (1-p)(1-x)^{\alpha-2} > 0$, which, given that $0 \leq p \leq 1$, is always true. x^* therefore always represents a minimum when returns are increasing. When returns to specialisation are linear ($\alpha = 1, 0 \leq \kappa < 1$) we find that $\frac{d^2W}{dx^2} = 0$, and so x^* always represents an inflection point.

We ask what ESS will arise when x^* represents a minimum or inflection point. Given that there is only one equilibrium solution x^* for each set of parameter values, it must be the case that $W(x)$ is maximal at either $x = 0$ or $x = 1$. It is maximal at $x = 1$ if $W(x=0) < W(x=1)$ is satisfied. Evaluating this shows that this is true for $p > 0.5$. Conversely, $W(x)$ is maximised at $x = 0$ when $p < 0.5$. When $W(x=0) = W(x=1)$, which is the case when $p = 0.5$, individuals can maximize fitness with either of two strategies, and individuals may assume either $x = 0$ or $x = 1$ at equilibrium. So, when returns to specialisation are increasing or linear ($\alpha \geq 1$), the ESS is positioned at nuclear purity, and the nucleus type that is chosen is the one that grows better with the most common plant host. In the special case where nuclear proportions have no effect on fitness ($\kappa = 1$), substitution of $\kappa = 1$ gives $\frac{d^2W}{dx^2} = 0$ and $W(x=0) = W(x=1)$, meaning nuclear purity of type one or type two nuclei will evolve with equal likelihood.

For diminishing returns to specialisation ($0 < \alpha < 1, 0 \leq \kappa < 1$), substituting $0 < \alpha < 1$ into $\frac{d^2W}{dx^2} < 0$ (the condition for x^* to be a maximum) gives $px^{\alpha-2} + (1-p)(1-x)^{\alpha-2} > 0$, which, given that $0 \leq p \leq 1$, is always satisfied, meaning x^* always represents a maximum. Because there is one maximum, x^* confers the global optimum fitness (W), and so represents an ESS. The maximum corresponds to genomic diversity ($0 < x^* < 1$) when the host plant environment is mixed ($0 < p < 1$). Between-individual selection therefore favours genomic diversity if there are diminishing returns to specialisation ($0 < \alpha < 1$) and a mixed host plant environment ($0 < p < 1$).

APPENDIX 2: STABLE GENOMIC DIVERSITY

Equation 3 gives the change in the population mean nuclear proportion ($E[X]$) over one generation. The population mean nuclear proportion will not undergo further evolution if $E[X]_{t+1} = E[X]_t = E[X^*]$. By equating $E[X]_{t+1} = E[X]_t$ and solving, we find this position (the stationary distribution) to be $E[X^*] = \mu + \frac{1-s}{s}\theta$. A population with this average nuclear proportion ($E[X^*]$) will not evolve, but we now ask whether populations will evolve to this position from elsewhere (whether the stationary distribution is absorbing).

We perturb the equilibrium by a small positive value ϵ and see that rightward perturbations are restored if $E[X^*] + \epsilon > (E[X^*] + \epsilon + \theta)(1-s) + s\mu$, and leftward perturbations are restored if $E[X^*] - \epsilon < (E[X^*] - \epsilon + \theta)(1-s) + s\mu$. Substituting the equilibrium condition $E[X^*] = \mu + \frac{1-s}{s}\theta$ and simplifying generates $\epsilon > 0$ in both cases, and so the population of individuals will evolve to this position

($E[X^*]$) regardless of its initial mean nucleus proportion ($E[X]$); it is an evolutionary end point.

We are interested in cases where populations maintain nuclear diversity within individuals. In principle, an intermediate mean population nuclear proportion ($0 < E[X^*] < 1$) could correspond to a mixture of genomically pure individuals, some with type one nuclei and others with type two. However, there is no diversifying selection in this model, and so nuclear diversity within the population corresponds to nuclear diversity within individuals ($0 < E[X^*] < 1$). $0 < E[X^*]$ always holds, because type two nucleus purity is never selected for. However, $E[X^*] < 1$ only holds for the condition given in Equation 4, which is the condition for stable genomic diversity.

APPENDIX 3: COMPETING NUCLEI

In AM fungi, replicative differences between nuclei (θ) may be high, but in other organisms with multiple genomes, replicative synchrony ($\theta \rightarrow 0$) might be well enforced. For example, other filamentous fungi (Basidiomycetes and Ascomycetes) can form dikaryons, in which replicative synchrony is often well enforced (by structures called *clamp connections* and *croziers*, respectively). Stable genomic diversity in these cases requires only that it provides some benefit to the individual ($s > 0$).

Large genomic deletions may generate nuclei that are faster replicating as a result of their smaller genome, but non-functional or deleterious to the individual. Between-individual selection disfavors such nuclei ($\mu = 0$), but we see that they can still coexist alongside functional nuclei if the between-individual selection to purge the deleterious nuclei is (a) stronger than their replicative advantage within individuals ($(1 - s)\theta > s$; this means that the equilibrium is a stable absorption point), and (b) not maximal, corresponding to lethal

nuclei ($s < 1$; this means that the absorption point is $E[X^*] > 0$). As predicted by this, deleterious 'cheating' nuclei have been observed in heterokaryotic fungi (Meunier *et al.* 2018; Bastiaans *et al.* 2016). A theoretical treatment of when such cheating nuclei will arise in the first place is a question for future study; here we are content to show that such nuclei, if they arise, can be maintained stably.

APPENDIX 4: SIMULATION

We give further details regarding how nuclear replication, and individual dispersal, was modelled.

- (i) Nucleus Replication Phase. Type 1 (N_1) and type 2 (N_2) nuclei replicate repeatedly, increasing exponentially: $N_1(t + 1) = (1 + r_1) N_1$; $N_2(t + 1) = (1 + r_2) N_2$, where the generational growth rate of type one nuclei (r_1) exceeds that of type two nuclei ($r_1 > r_2$). An individual's generational change in nuclear proportion (x) is therefore given by: $x(t + 1) = \frac{x(1 + r_1)}{x(r_1 - r_2) + 1 + r_2}$
- (ii) Sporulation & Dispersal Phase. With probability d , an individual's offspring disperse and compete on a population scale with other dispersing offspring. There are $d(n/j)$ spots available on each patch for dispersing offspring, and an individual with dispersing offspring reproduces into each of these spots with the probability given by their fitness (Equation 1) divided by the total fitness of all individuals with dispersing offspring. With probability $(1 - d)$, an individual's offspring do not disperse and compete on the local patch with other non-dispersing offspring for the $(1 - d)(n/j)$ free spots. An individual with non-dispersing offspring reproduces into each of these spots with the probability given by their fitness divided by the total fitness of all individuals with non-dispersing offspring on the native patch.

Supplementary Information

Data S1: alternative lifecycle in the Competing Nuclei model

We reformulate the Competing Nuclei model of the main text, assuming an alternative lifecycle where selection between individuals occurs at the spore stage, with nucleus replication following this as individuals mature. The order of within- and between-individual selection is therefore reversed relative to the model presented in the main text. The generational change in the population mean nuclear proportion is given by: $E[X]_{t+1} = (s\mu + (1-s) E[X]_t) + \theta$. The absorption point becomes $E[X^*] = \theta/s + \mu$, which corresponds to genomic diversity when $s(1-\mu) > \theta$. This qualitatively resembles the condition of the model presented in the main text (equation 4), capturing the opposing pulls of selection within and between individuals, except that the destabilising force of within-individual selection is here un-tempered by the strength of between-individual selection, as it occurs afterwards in the lifecycle (the right-hand side of the condition is θ rather than $\theta(1-s)$).

Data S2: more general capture of nucleus replication in the Competing Nuclei model

We extend the Competing Nuclei model of the main text to capture nucleus replication in a more general framework. Firstly, like in the original model, we assume a lifecycle in which within-individual selection precedes between-individual selection. In the original model, within-individual selection was captured by assuming that the nuclear proportion (which we denote here as X') in an individual after a bout of within-individual selection can be given by $X' = X + \theta$, where X is the nuclear proportion beforehand. This corresponds to a hypothetical lifecycle in which, every generation, a group of less competitive nuclei (N_2) are replaced by a group more competitive nuclei (N_1), with this group size constant with respect to the proportion of the two nuclei in individuals. This might occur if modular turnover is constrained, such that modules do not replicate freely.

However, the action of within-individual selection may depend on the current mean nuclear proportion (X). For example, if type one nuclei are nearly at fixation ($X \rightarrow 1$), further proportional increases in type one nuclei might be slight. We can model within-individual selection more generally: $X' = X + \theta_1 + \theta_2 * X$. Our formula for generational changes in nuclear proportion then becomes: $s\mu + (1-s) (E[X]_t + \theta_1 + E[X]_t \theta_2) = E[X]_{t+1}$. Solving for $E[X]_t = E[X]_{t+1} = E[X^*]$ and checking for stability, we find one absorption point at $E[X^*] = (\theta_1 - s\theta_1 + s\mu) / (s + (s-1)\theta_2)$, where μ and s respectively denote the optimum nuclear proportion and strength of between-individual selection. The mean corresponds to genomic diversity when $(1-s)(\theta_1 + \theta_2) < s(1-\mu)$. This condition qualitatively resembles the condition of the model presented in the main text (equation 4), capturing the opposing pulls of selection within and between individuals. The previous coefficient θ has been replaced the two coefficients, θ_1 and θ_2 , which respectively capture X -independent and X -dependent generational shifts in nuclear proportion.

Higher orders of X are still required to capture other modes of nuclear replication (e.g. $X' = X + \theta_1 + \theta_2 * X + \theta_3 * X^2 + \dots$), and the qualitative results of our abstract model are retained with arbitrary generalisations of nuclear replication.

Similarly, we assumed that the response to between-individual selection (s) is constant and independent of nuclear proportion (X). This is contingent on the shape of the distribution of individuals and fitness values across different nuclear proportions. However, relaxing these assumptions, to allow the response to selection (s) to vary with the current nuclear proportion (X), does not change our qualitative findings. Furthermore, the results of the Competing Model are corroborated in the simulation model of the main text, in which within- and between-individual selection are modelled in a more biologically realistic scenario.

Data S3: Additional Figures

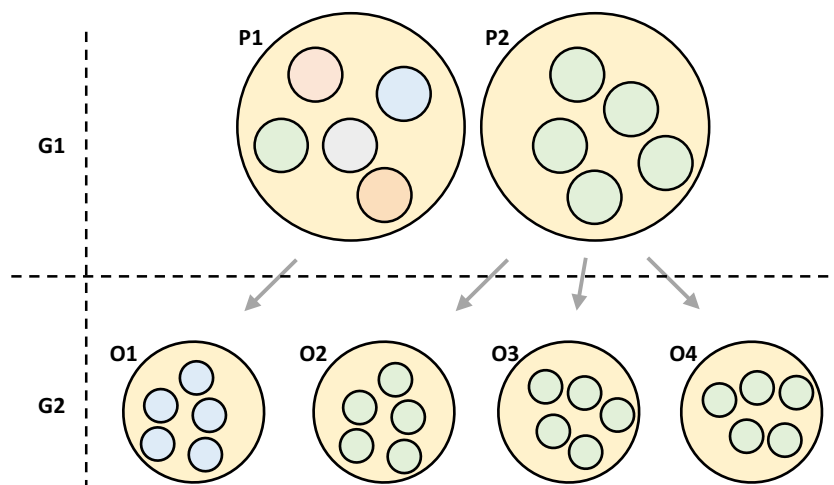


Figure S1 | The theoretical problem of genomic diversity. Two generations of AM fungi, G1 and G2, are represented. Hyphal networks (individuals) are represented by beige circles. Individuals bear nuclei, which can be different strains, as indicated by colour. Generation 1 (G1) consists of two parent individuals: P1, which has genomic diversity, and P2, which has genomic homogeneity. Genomic diversity leads to conflict and so P1 bears an evolutionary cost; P2 doesn't bear this cost. P2 therefore has more asexual offspring (O2, O3 & O4) than P1 (O1), and so G2 is shifted towards genome purity as a result of between-individual selection. Due to competition within individuals (within-individual selection) or drift, O1 has a lower genomic diversity than its parent (P1), shifting G2 towards genome purity.

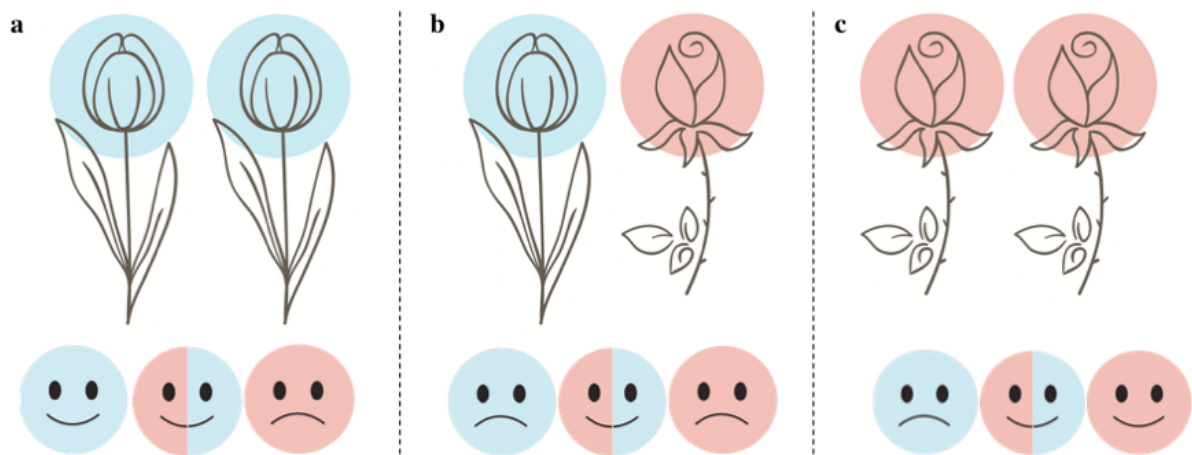


Figure S2 I Hypothesis for the maintenance of genomic diversity. The red and blue flowers represent two species of host plant. In environment (a) all host plants are the blue species; in (c) all are red; in (b) there is a mixture of blue and red. Host plants are connected to AM fungal networks (individuals), represented by faces, which contain multiple nuclei of two types. Individuals can contain solely nuclei that are specialised on red plants (red faces), solely nuclei that are specialised on blue plants (blue faces), or a mixture of nucleus types (red / blue faces). Mixed individuals grow well in all host plant environments (a-c) because they always contain some specialised nuclei. Pure individuals can only grow well when connected to their specialised host plant (blue face in (a); red face in (c)) but don't grow well when there is a mixture of host plants (b). Genomic diversity may therefore be favoured in mixed host plant environments.

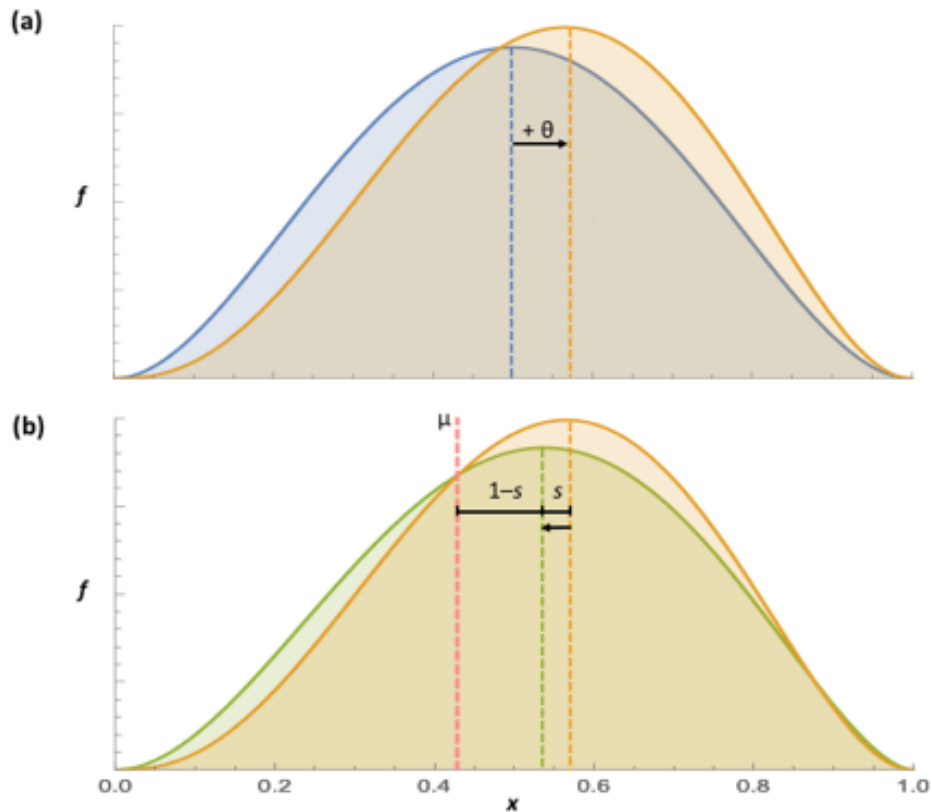


Figure S3 I Competing Nuclei model set-up. The distributions plot the frequency of AM fungal networks (individuals) with differing proportions of type one relative to type two nuclei (x) in a population. Each generation, the population undergoes within-individual selection (a), which increases the mean nuclear proportion ($E[X]$) by θ . The population then undergoes between-individual selection (b), which pulls the mean nuclear proportion ($E[X]$) towards the individual optimum nuclear proportion μ by some proportion given by the strength of between-individual selection (s). Individuals then reproduce asexually, which doesn't change the mean nuclear proportion ($E[X]$), and the process is iterated. We ask if the two processes can lead to an equilibrium with genomic diversity ($0 < E[X] < 1$).

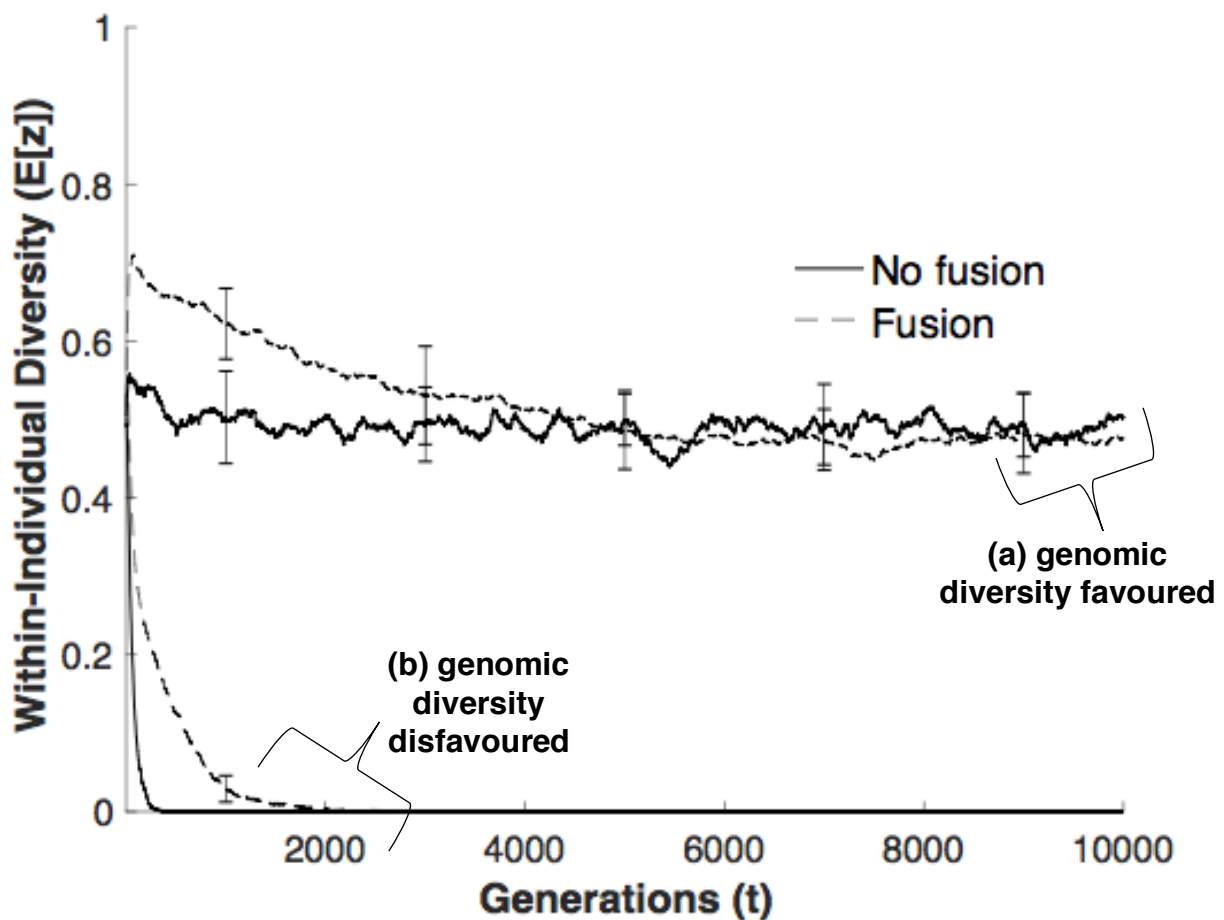


Figure S4 | Effect of fusion on genomic diversity in the absence of within-individual selection. Within-individual genomic diversity (z^*), in the absence of replicative differences between nuclei ($r_1=r_2$), is plotted as a function of time (t). Two scenarios are considered, when genomic diversity is: (a) favoured ($\alpha=0.3$, $p=0.3$), and (b) disfavoured ($\alpha=1.1$, $p=0.6$), by between-individual selection. The dashed lines assume fusion between individuals ($m=0.2$), and the solid lines assume no fusion ($m=0$). Under fusion, equilibrium is reached more slowly. This means that genomic diversity can be maintained in a non-equilibrium state over a longer time period, as is the case in (b). These results assumed $f=0.005$ (sporulation stochasticity), $d=0.5$ (dispersal). The plots represent the average results taken across 10 trials. Error bars, where plotted, show one standard deviation above and below the mean across these 10 trials.

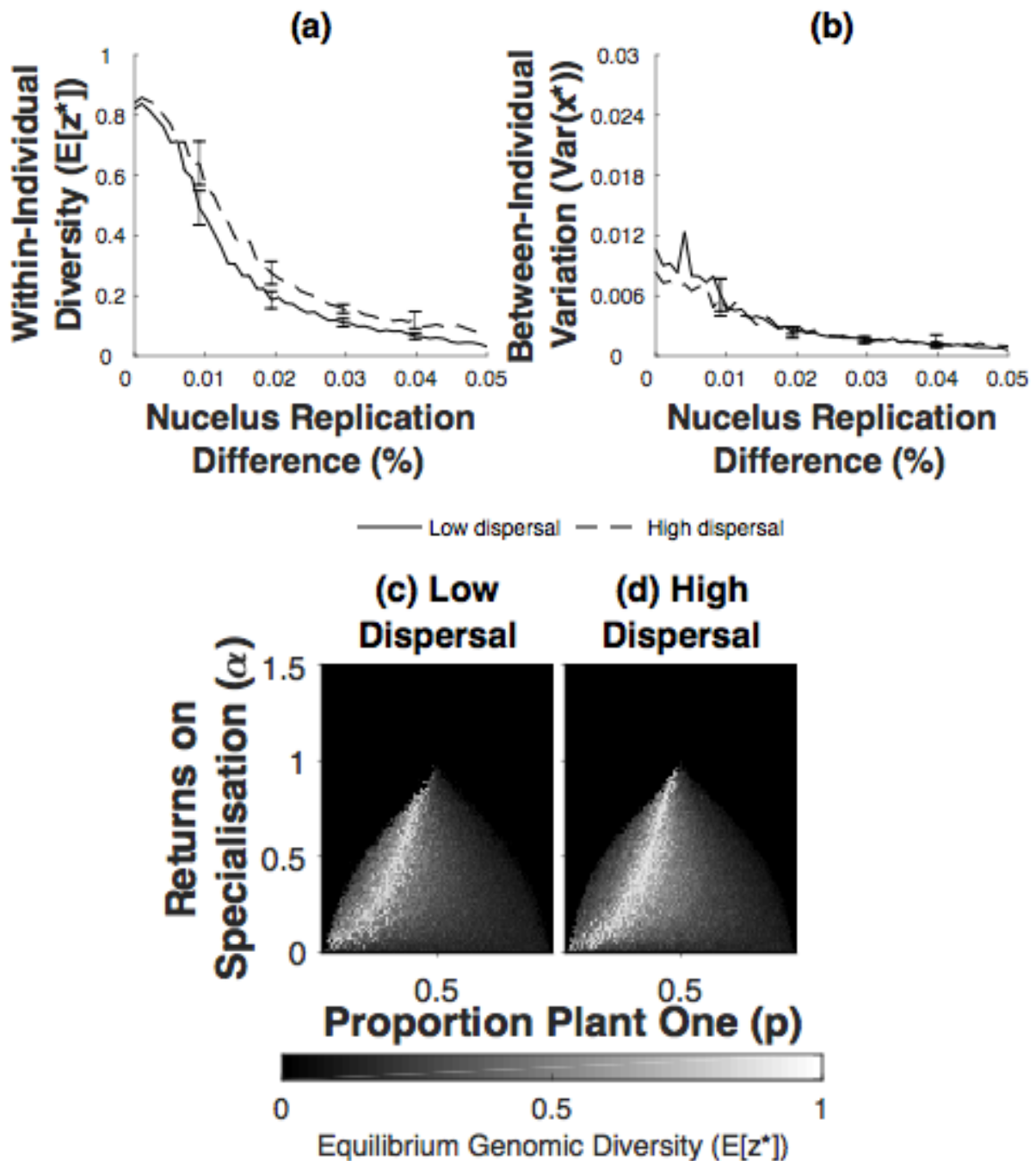


Figure S5 | Effect of dispersal on genomic diversity. The within-individual genomic diversity (a), and between-individual variation in nuclear proportion (b), is plotted against the nuclear replicative advantage of type one nuclei ($r_1 - r_2/r_2$) ($\alpha=0.8$, $p=0.5$, $d=0.5$, $r_2=0.3$, r_1 is varied). The different lines represent different degrees of dispersal (low: $d=0.5$; high: $d=1$). Dispersal does not significantly affect between-individual variation, but can nevertheless slightly increase within-individual diversity, because it increases the effective population size of the population, and hence the efficacy of between-individual selection. The plots represent the average results taken across 10 trials. Error bars, where plotted, show one standard deviation above and below the mean across these 10 trials. (c) and (d) plot the full range of between-individual selection, from decelerating to accelerating returns on plant specialisation

(α , y axis), and from a plant two to a plant one dominated environment (ρ , x axis). Equilibrium genomic diversity is slightly greater when dispersal is high, across a large range of between-individual selection. These results assumed $f=0.005$ (sporulation stochasticity) and $m=0$ (fusion).

3

**Adaptation is maintained by the
parliament of genes**

ARTICLE

<https://doi.org/10.1038/s41467-019-13169-3>

OPEN

Adaptation is maintained by the parliament of genes

Thomas W. Scott ^{1*} & Stuart A. West¹

Fields such as behavioural and evolutionary ecology are built on the assumption that natural selection leads to organisms that behave as if they are trying to maximise their fitness. However, there is considerable evidence for selfish genetic elements that change the behaviour of individuals to increase their own transmission. How can we reconcile this contradiction? Here we show that: (1) when selfish genetic elements have a greater impact at the individual level, they are more likely to be suppressed, and suppression spreads more quickly; (2) selection on selfish genetic elements leads them towards a greater impact at the individual level, making them more likely to be suppressed; (3) the majority interest within the genome generally prevails over 'cabals' of a few genes, irrespective of genome size, mutation rate and the sophistication of trait distorters. Overall, our results suggest that even when there is the potential for considerable genetic conflict, this will often have negligible impact at the individual level.

¹Department of Zoology, University of Oxford, Zoology Research and Administration Building, 11a Mansfield Road, Oxford OX1 3SZ, UK.
*email: thomas.scott@zoo.ox.ac.uk

There is a contradiction between major branches of modern evolutionary biology. On the one hand, fields such as behavioural and evolutionary ecology are based on the assumption that organisms will behave as if they are trying to maximise their fitness^{1–4}. Models based on fitness maximisation are used to make predictions about the selective forces (reasons) for adaptation, and these are then tested empirically^{5,6}. This approach has been phenomenally successful, explaining many aspects of behaviour, life history and morphology. For example, fitness maximisation underpins our evolutionary explanations of: foraging behaviour, resource competition, sexual selection, parental care, sex allocation, signalling and cooperation^{7–12}.

On the other hand, there is considerable evidence for selfish genetic elements, which increase their own contribution to future generations at the expense of other genes in the same organism^{13–17}. These selfish genetic elements may distort traits away from the values that would maximise individual fitness, to increase their own transmission^{14,18–22}. Evidence for such genetic conflict has been found across the tree of life, from simple prokaryotes to complex animals. The contradiction is that selfish genetic elements mess up individual fitness maximisation, and appear to be common, but individual fitness maximisation still appears to occur^{17,23,24}. This contradiction is especially apparent in the study of sex allocation: theoretical models based on individual fitness maximisation have explained a wide range of natural variation in sex ratio, and yet there have been many reported cases of selfish sex ratio distorters^{9,14,25–27}.

Leigh²⁸ provided a potential solution to this contradiction by suggesting that selfish genetic elements would be suppressed by the ‘parliament of genes’. Leigh’s argument was that, because selfish genetic elements reduce the fitness of most of the other genes in the organism, these other genes will have a united interest in suppressing selfish genetic elements. Furthermore, because these other genes are far more numerous, they will be likely to win the conflict. Consequently, even when there is considerable potential for conflict within individuals, we would still expect fitness maximisation at the individual level^{29–34}. Leigh²⁸ demonstrated the plausibility of his argument by showing theoretically how a suppressor of a sex ratio distorter could be favoured. Since then, numerous suppressors have been studied from a theoretical and an empirical perspective^{14,35,36}.

However, several issues may affect the validity of the parliament of genes hypothesis. First, whether a suppressor spreads can depend upon biological details such as the extent to which a selfish genetic element is distorting a trait, the population frequency of that element and the cost of suppression^{14,37–43}. Are certain types of selfish genetic elements, which cause substantial distortion, less likely to be suppressed? Second, if the spread of suppressors through populations is slow, and if selfish genetic elements arise continuously over evolutionary time, non-equilibrium trait distortion may be possible³⁵. Third, selfish genetic elements are themselves also under evolutionary pressure to cause a level of trait distortion that would maximise their transmission to the next generation¹⁵. Could the evolution of selfish genetic elements lead to trait distortion that is less likely to be suppressed?³² Fourth, if a suppressor does not reach fixation in a population, or a selfish genetic element is not purged from a population, subsequent mating may decouple selfish genetic elements and suppressors to expose previously suppressed trait distortion³⁸. How important is this problem of polymorphism likely to be?

We address these issues, by investigating the parliament of genes hypothesis theoretically. Our aim is to investigate the extent to which genetic conflict distorts traits away from the value that would maximise individual fitness. We find that: (i) the greater the level of trait distortion caused by a selfish genetic element, the

more likely and the quicker it is suppressed; (ii) selection on selfish genetic elements leads towards greater trait distortion, making them more likely to be suppressed; (iii) in genome-wide arms races to gain control of organism traits, the majority interest within the genome generally prevails over ‘cabals of a few’, regardless of genome size, mutation rate, and the strength and sophistication of trait distorters. We find the same patterns with an illustrative model, and when examining three specific scenarios: selfish trait distortion of the sex ratio by an X chromosome driver; an altruistic helping behaviour encoded by an imprinted gene; and production of a cooperative public good encoded on a horizontally transmitted bacterial plasmid. Furthermore, we find close agreement when analysing scenarios with population genetic analyses and individual-based simulations. Our results suggest that even when there is potential for considerable genetic conflict, it has relatively little impact on traits at the individual level.

Results

Modelling approach. We examine conflict between two groups of genes within the genome. We assume a selfish genetic element that can gain a propagation advantage through distorting some trait of the organism (‘trait distorter’). This trait distortion only benefits alleles at a subset of loci within the genome—Leigh termed this subset of loci a ‘cabal’³⁰. The rest of the genome, which does not gain the propagation advantage from the trait distortion, will be selected to suppress the trait distorter. Leigh termed this collection of genes, which will comprise most of the genome, and so will constitute the majority within the parliament of genes, the ‘commonwealth’³⁰.

We used two complementary theoretical approaches. First, we developed ‘Equilibrium models’, where we assume that the trait distorter and their cabal are only a very small fraction of the genome. We allow for this by assuming that it is highly likely that a potential suppressor of a trait distorter can arise by mutation. Consequently, in these models, we focus our analyses on when a trait distorter and its suppressor can spread. We use this approach to examine, given the potential for suppression, what direction would we expect natural selection to take on average.

We then developed ‘Dynamics models’, where we relaxed the assumption that the trait distorter and its cabal are a negligible fraction of the genome. In this case, rather than focus on the equilibrium state, we allowed trait distorters and their suppressors to arise continuously, at different loci across the genome. This approach allows us to investigate the influence of factors such as genome size, mutation rate and cabal size. We use this approach to determine the outcome of an evolutionary conflict that embroils the whole genome, to elucidate how far an organism trait is likely to be distorted at any given point in evolutionary time.

Equilibrium models. We assessed, given the potential for suppression, the extent to which a trait distorter will distort an organism trait away from the optimum for individuals. In order to elucidate the selective forces, we ask four questions in a step-wise manner, with increasing complexity:

- (1) In the absence of a suppressor, when can a trait distorter invade?
- (2) When can a costly suppressor of the trait distorter invade?
- (3) What are the overall consequences of trait distorter-suppressor dynamics for trait values, at the individual and population level, at evolutionary equilibrium and before equilibrium has been reached?
- (4) If the extent to which the trait distorter manipulates the organism trait can evolve, how will this influence the

likelihood that it is suppressed, and hence the individual and population trait values?

We assume an arbitrary trait that influences organism fitness. In the absence of trait distorters, all individuals have the trait value that maximises their individual fitness. The trait distorter manipulates the trait away from the individual optimum, to increase their own transmission to offspring. We assume a large population of diploid, randomly mating individuals. The aim of this model is to establish key aspects of the population genetics governing trait distorters and their suppressors, in an abstract setting. In Supplementary Notes 3, 4 and 5, we address the same issues in three specific biological scenarios.

(1) Spread of a trait distorter: We consider a trait distorter, which we denote by D_1 , that is dominant and distorts an organism trait value by some positive amount k ($k > 0$). This trait distortion increases the transmission of the trait distorter to offspring. Specifically, the trait distorter (D_1) drives at meiosis, in heterozygotes, against a trait non-distorter (D_0), being passed into the proportion $(1 + t(k))/2$ of offspring. $t(k)$ denotes the transmission bias ($0 \leq t(k) \leq 1$) and is a monotonically increasing function of trait distortion ($\frac{dt}{dk} \geq 0$).

We emphasise that, in nature, trait distorters need not be meiotic drivers—the key point here is that we are considering when trait distortion increases the propagation of that trait distorter. We chose meiotic drive in this model for simplicity, and model different mechanisms in the biologically specific models (Supplementary Notes 3, 4 and 5). Indeed, in many natural cases, meiotic drivers would not gain their advantage by distorting a trait, in which case they would not enter any conflict with the rest of the genome over organism trait values, and therefore would not have any lasting influence on whether trait values are those that maximise individual fitness. For example, the segregation distorter (SD) meiotic driver in *Drosophila melanogaster* gains its advantage in heterozygous males by disrupting the proper development of rival sperm, and not by trait distortion⁴⁴. Any organism-level fitness costs associated with SD would be opposed by SD as well as across the rest of the genome⁴⁵. Our focus in this paper is on selfish genetic elements that gain an advantage by trait distortion, and therefore disagree with the majority of genes over trait values.

Trait distortion leads to a fitness (viability) cost ($c_{\text{trait}}(k)$) at the individual level, reducing an individual’s number of offspring from 1 to $1 - c_{\text{trait}}(k)$ ($0 \leq c_{\text{trait}}(k) \leq 1$). Owing to trait distorter dominance, the fitness cost of trait distortion is borne by heterozygous as well as trait distorter-homozygous individuals. The fitness cost is a monotonically increasing function of trait distortion ($\frac{dc_{\text{trait}}}{dk} \geq 0$). We assume that $t(k)$ and $c_{\text{trait}}(k)$ do not change with population allele frequencies, but relax this assumption in our specific models.

We first ask what frequency the trait distorter will reach in the population in the absence of suppression. If we take p and p' as the population frequency of the trait distorter in two consecutive generations, then the population frequency of the trait distorter in the latter generation is:

$$\bar{w} p' = (1 - c_{\text{trait}}(k)) (p^2 + (1 - p)p(t(k) + 1)), \quad (1)$$

where \bar{w} is the average fitness of individuals in the population in the current generation, and can be written in full as: $\bar{w} = (1 - c_{\text{trait}}(k))(p^2 + 2p(1 - p)) + (1 - p)^2$. In ‘Trait distorter population frequency’ in the Methods, we show, with a population genetic analysis of Eq. 1, that the trait distorter will spread from rarity and reach fixation when $c_{\text{trait}}(k) < t(k)(1 - c_{\text{trait}}(k))$. This shows that trait distortion will evolve when the number of offspring that the trait distorter gains as a result of trait distortion ($t(k)(1 -$

$c_{\text{trait}}(k)$) is greater than the number of offspring bearing the trait distorter that are lost as a result of reduced individual fitness ($c_{\text{trait}}(k)$).

(2) Spread of an autosomal suppressor: We assume that the trait distorter (D_1) can be suppressed by an unlinked autosomal allele (suppressor), denoted by S_1 . We assume that this suppressor (S_1) is dominant and only expressed in the presence of the trait distorter (facultative), but found similar results when the suppressor is constitutively expressed (obligate; Supplementary Note 6). Expression of the suppressor incurs a fitness cost to the individual, c_{sup} ($0 \leq c_{\text{sup}} \leq 1$), which could arise for multiple reasons, including energy expenditure, or errors relating to the use of gene silencing machinery^{46,47}. Gene silencing generally precedes the translation of the targeted gene, and so we assume that the cost of suppression (c_{sup}) is independent of the amount of trait distortion caused by the trait distorter (k).

We can write recursions detailing the generational change in the frequencies of the four possible gametes, D_0/S_0 , D_0/S_1 , D_1/S_0 and D_1/S_1 , with the respective frequencies in the current generation denoted by x_{00} , x_{01} , x_{10} and x_{11} , and the frequencies in the subsequent generation denoted by an appended dash (‘)’:

$$\bar{w} x'_{00} = x_{00}^2 + x_{00}x_{01} + (1 - t)(1 - c_{\text{trait}})x_{00}x_{10} + \left(\frac{1 - c_{\text{sup}}}{2}\right)x_{00}x_{11} + \left(\frac{1 - c_{\text{sup}}}{2}\right)x_{01}x_{10}, \quad (2)$$

$$\bar{w} x'_{01} = x_{00}x_{01} + \left(\frac{1 - c_{\text{sup}}}{2}\right)x_{00}x_{11} + x_{01}^2 + \left(\frac{1 - c_{\text{sup}}}{2}\right)x_{01}x_{10} + (1 - c_{\text{sup}})x_{01}x_{11}, \quad (3)$$

$$\bar{w} x'_{10} = (1 + t)(1 - c_{\text{trait}})x_{00}x_{10} + \left(\frac{1 - c_{\text{sup}}}{2}\right)x_{00}x_{11} + \left(\frac{1 - c_{\text{sup}}}{2}\right)x_{01}x_{10} + (1 - c_{\text{trait}})x_{10}^2 + (1 - c_{\text{sup}})x_{10}x_{11}, \quad (4)$$

$$\bar{w} x'_{11} = \left(\frac{1 - c_{\text{sup}}}{2}\right)x_{00}x_{11} + \left(\frac{1 - c_{\text{sup}}}{2}\right)x_{01}x_{10} + (1 - c_{\text{sup}})x_{01}x_{11} + (1 - c_{\text{sup}})x_{10}x_{11} + (1 - c_{\text{sup}})x_{11}^2, \quad (5)$$

where \bar{w} is the average fitness of individuals in the current generation, and equals the sum of the equations’ right-hand sides. In ‘Suppressor invasion condition’ in the Methods, we show, with a population genetic analysis of these equations, that a suppressor will spread from rarity if trait distortion (k) is greater than some threshold value, at which the cost of suppression (c_{sup}) is less than the cost of being subjected to trait distortion, $c_{\text{sup}} < c_{\text{trait}}(k)$. A threshold with respect to the level of trait distortion (k) arises because the cost of trait distortion ($c_{\text{trait}}(k)$) increases with greater trait distortion, but the cost of suppression (c_{sup}) is constant. Given that the individual cost of pre-translational suppression at a single locus is likely to be low^{46,47}, trait distortion conferred by unsuppressed trait distorters is likely to be negligible.

(3) Consequences for organism trait values: The extent of trait distortion at the individual level shows a discontinuous relationship with the strength of the trait distorter (Fig. 1a). When trait distortion is low, a suppressor will not spread ($c_{\text{sup}} > c_{\text{trait}}(k)$) and so the level of trait distortion at the individual level will increase with the level of trait distortion induced by the trait distorter (k). However, once a threshold is reached ($c_{\text{sup}} < c_{\text{trait}}(k)$), the suppressor spreads. We show in ‘Equilibrium trait distorter and suppressor frequencies’ in the Methods that the spread of the suppressor (S_1) causes the trait distorter (D_1) to lose its selective advantage and be eliminated from the population, leading to an absence of trait distortion at the individual level. In contrast, we

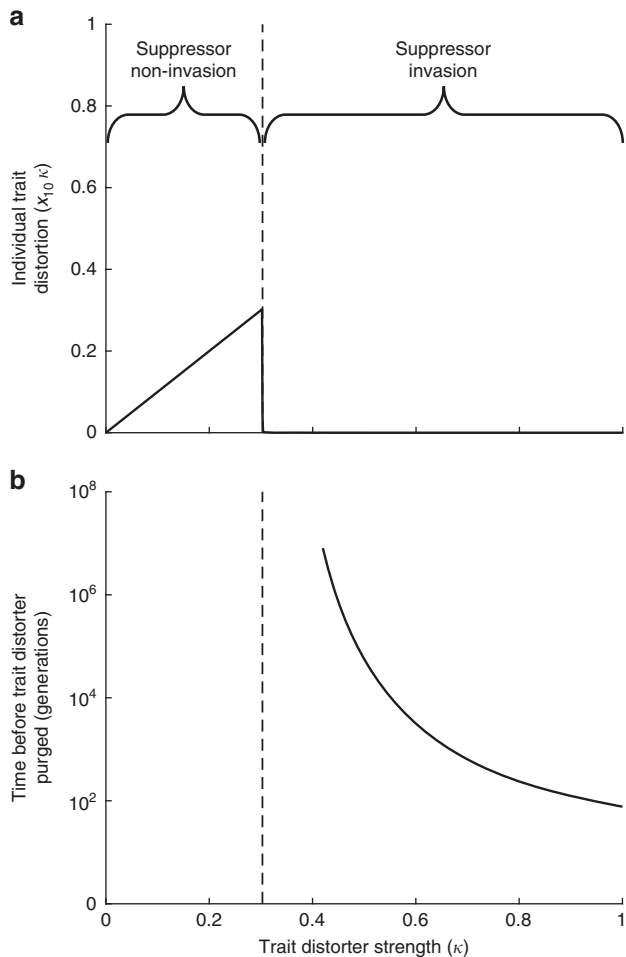


Fig. 1 Trait distorter-suppressor dynamics and consequences for the organism. The trait distorter (D_1) and its suppressor (S_1) are introduced from rarity. In **a**, the resulting average trait distortion ($x_{10, k}$) is plotted at equilibrium, against the extent to which the trait distorter causes trait values to deviate from the individual optimum (k). Below a certain threshold strength ($c_{sup} > c_{trait}(k)$; left of dashed line), the suppressor does not invade, and so the resulting trait distortion increases with the strength of the trait distorter (k). Above this threshold, the suppressor invades, and the trait distorter is purged, restoring the trait to the individual optimum. In **b**, the number of generations between trait distorter introduction and loss is plotted, on a log₁₀ scale, for trait distorters that are purged at equilibrium (having been suppressed), which lie to the right of the dashed line. Stronger trait distorters are purged more quickly than weaker trait distorters ($c_{sup} = 0.15$; $t = 0.87k$ and $c_{trait} = 0.9k^{1.5}$)

show in Supplementary Note 6 that if the suppressor is constitutively expressed (obligate), the spread of the suppressor (S_1) to fixation in the population causes the trait distorter (D_1) to become neutral, meaning the trait distorter (D_1) can be maintained in the population without being expressed.

Overall, these results suggest that, given a relatively low cost of suppression (c_{sup}), the level of trait distortion observed at the individual level will either be low or absent. When a trait distorter is weak (low k), it will not be suppressed, but it will only have a small influence at the level of the individual. When a trait distorter is strong (high k), it will be suppressed and so there will be no influence at the level of the individual (Fig. 1a).

In addition, we found that stronger trait distorters are suppressed more quickly (Fig. 1b). In ‘Non-equilibrium trait distortion’ in the Methods, we numerically iterated our recursions

to determine how many generations it takes for suppressors to reach equilibrium. As long as trait distortion continues to reduce individual fitness non-negligibly after suppression is favoured (such that $\frac{df}{dk} / \frac{dc_{trait}}{dk}$ is not excessively high after $c_{sup} < c_{trait}(k)$), stronger trait distorters (higher k) are suppressed and purged more rapidly than weaker trait distorters, limiting the potential for non-equilibrium trait distortion (Fig. 1b).

(4) Evolution of trait distortion: We then considered the consequence of allowing the level of trait distortion (k) to evolve. We assume a trait distorter (D_1) that distorts by k , and then introduce a rare mutant (D_2) that distorts by a different amount \hat{k} ($\hat{k} \neq k$). This mutant (D_2) is propagated into the proportion $(1 + t(\hat{k}) - t(k))/2$ of the offspring of D_2D_1 heterozygotes, and into the proportion $(1 + t(\hat{k}))/2$ of the offspring of D_2D_0 heterozygotes. We assume that the stronger of the two trait distorters is dominant, but found similar results when assuming additivity (‘Invasion of a mutant trait distorter’ in the Methods). We assume that the similarity in coding sequence and regulatory control means that the original trait distorter and the mutant are both suppressed by the same suppressor allele, at the same cost (c_{sup})^{46,47}. In ‘Invasion of a mutant trait distorter’ in the Methods, we write the recursions that detail the generational frequency changes in the different possible gametes (D_0/S_0 , D_0/S_1 , D_1/S_0 , D_1/S_1 , D_2/S_0 and D_2/S_1).

We found that stronger mutant trait distorters ($\hat{k} > k$) will invade from rarity when the marginal increase in offspring they are propagated into exceeds the marginal increase in offspring they are lost from as a result of reduced fitness ($\Delta t(1 - c_{trait}(\hat{k})) > \Delta c_{trait}$, where Δ denotes marginal change ($\Delta t = t(\hat{k}) - t(k)$; $\Delta c_{trait} = c_{trait}(\hat{k}) - c_{trait}(k)$). Consequently, if trait distortion is initially low, and successive mutant trait distorters are introduced, each deviating only slightly from the trait distorters from which they are derived (‘ δ -weak selection’⁴⁸), invading trait distorters will approach a ‘target’ strength, denoted by k_{target} . This target strength corresponds to the level of trait distortion that would maximise the fitness of the gene¹⁵, and is when the marginal benefit of transmission is exactly counterbalanced by the marginal individual cost of reduced offspring, $\frac{df}{dk}(1 - c_{trait}) = \frac{dc_{trait}}{dk}$. The target strength of trait distortion (k_{target}) will therefore be greater if increased trait distortion (k) leads to a low rate of decrease in marginal transmission benefit ($-\frac{d^2t}{dk^2}$) relative to the rate of increase in marginal individual cost ($\frac{d^2c_{trait}}{dk^2}$) (Fig. 2b). If mutations are larger (strong selection), invading trait distorters may overshoot the target strength of trait distortion ($\hat{k} > k_{target}$). Weaker mutant trait distorters ($\hat{k} < k$) are recessive so cannot invade from rarity.

As evolution on the trait distorter increases the level of trait distortion, it makes it more likely that the trait distorter goes above the critical level of trait distortion where suppression will be favoured. When this is the case ($c_{sup} < c_{trait}(k_{target})$), the trait distorter spreads to high frequency, which then causes the suppressor to increase in frequency, reversing the direction of selection on the trait distorter, towards non-trait distortion (D_0), resulting in 0 trait distortion at equilibrium ($k^* = 0$) (Fig. 2a; ‘Equilibrium allele frequencies after mutant invasion’ in the Methods). Suppression only fails to spread if the individual fitness cost associated with suppression is greater than the individual fitness cost associated with the target trait distortion ($c_{sup} > c_{trait}(k_{target})$; Fig. 2a). Given that the individual fitness cost of pre-translational suppression at a single locus is likely to be low, then any non-negligible trait distorter is likely to be suppressed.

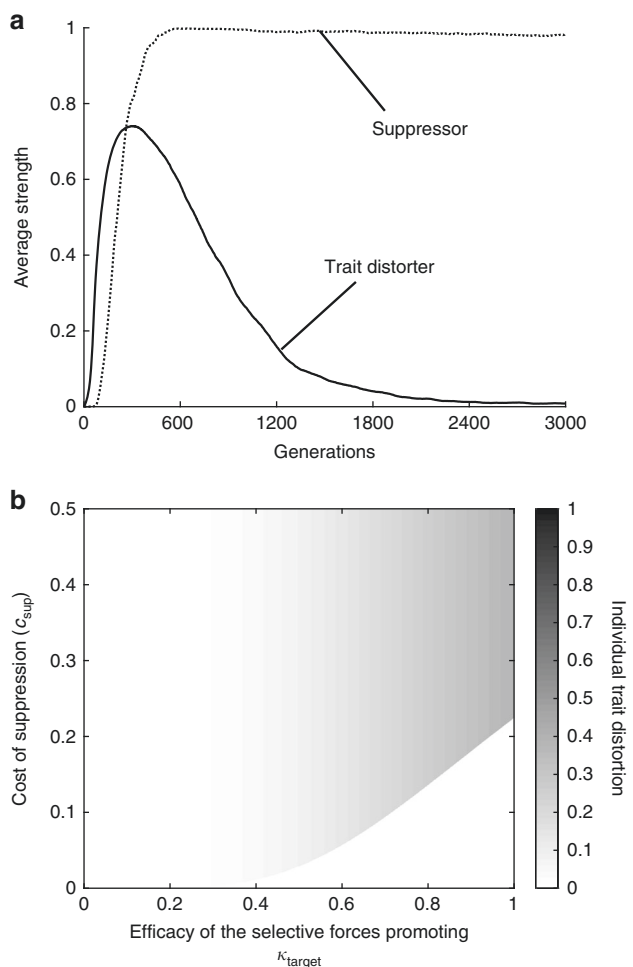


Fig. 2 Evolution of trait distortion. In **a**, a trait distorter and suppressor are introduced in our agent-based simulation model (Methods: ‘Agent-based simulation (single trait distorter locus)’), with $c_{\text{sup}} = 0.1$, $t = k$ and $c_{\text{trait}} = \max(k_a, k_b)/2$. The population average trait distorter and suppressor strengths over 100 simulation runs are plotted for successive generations. Initially, both trait distorter and suppressor strength increases, but then the trait distorters are purged from the population. **b** shows how the trait distortion at equilibrium is influenced by the cost of suppression (c_{sup}), and the target level of trait distortion (k_{target}), which is determined by the rate of increase in the marginal individual cost of trait distortion ($\frac{d^2 c_{\text{trait}}}{dk^2}$) relative to the rate of decrease in the marginal transmission benefit ($-\frac{d^2 t}{dk^2}$) (Supplementary Note 1). Trait distortion is low, unless there is both a high target level of trait distortion and a relatively high cost of suppression (top right of heat map)

Overall, our results suggest that selection on trait distorters will tend to lead to the eventual suppression of those trait distorters. In ‘Agent-based simulation (single trait distorter locus)’ in the Methods, we developed an agent-based simulation, which allowed us to continuously vary the level of both trait distortion and suppression, and obtained results in close agreement (Fig. 2a; Supplementary Note 2, Supplementary Fig. 2).

Specific biological scenarios. In Supplementary Notes 3, 4 and 5, we tested the robustness of our above conclusions by developing models for three different biological scenarios: a sex ratio distorter on an X chromosome (X driver); an imprinted gene that is only expressed when maternally inherited; and a gene for the

production of a public good by bacteria, which is encoded on a mobile genetic element^{14,26,36,49–52}. We examined these cases because they are different types of trait distortion, involving different selection pressures, in very different organisms. In all three specific models, we obtained the same qualitative results as with our above illustrative model for an arbitrary trait (Fig. 3).

Dynamics models. Our Equilibrium models assumed that the suppressor of any given trait distorter will arise quickly by mutation. This assumption becomes less likely if suppressors are complex and hard to evolve, or favoured across a reduced portion of the genome (smaller commonwealth). Also, multiple trait distorters and their suppressors may arise continually in populations, through evolutionary time, at different loci within the cabal and commonwealth respectively. Organisms may therefore never rest at equilibria where all trait distorters are suppressed or of negligible strength.

We address these issues by relaxing our assumption that the commonwealth is very large relative to the cabal, assuming instead that the commonwealth encompasses some majority of loci within the genome, with the remaining loci comprising the cabal. We examined the average and extremes of trait distortion produced by trait distorters and suppressors, by asking three further questions, of increasing complexity, in a step-wise manner:

- (5) To what extent are organism traits distorted when populations of individuals are only ever subjected to one segregating trait distorter at a time (no trait distorter co-segregation)?
- (6) To what extent are organism traits distorted when populations of individuals may be exposed to multiple, co-segregating, interacting trait distorters?
- (7) To what extent are organism traits distorted when the strength of each trait distorter may evolve?

(5) Trait distortion when no trait distorter co-segregation: We model a population of individuals, each with a genome size of γ loci. Within this genome, the cabal constitutes a fraction θ of all loci, and the commonwealth constitutes the remaining fraction $1 - \theta$ of all loci. If a fraction of the genome is inherited in the same way, such that it favours the same trait values (same maximand), it is termed a ‘coreplicon’^{20,22}. The cabal comprises all coreplicons that favour the distortion of a particular trait, along a particular axis, in a particular direction, away from individual fitness maximisation. The commonwealth comprises the remaining replicons. Cabals and commonwealths are therefore trait-specific. It is useful, when analysing a specific trait, to partition the genome along these lines, because it is this conflict—between the cabal and commonwealth—that drives the evolution of the trait value.

Cabals and commonwealths are defined a priori, by partitioning and summing up the coreplicons that, respectively, disfavour and favour the trait distortion under study. The ‘individual’ is the majority interest within the genome, and so the cabal size can never exceed more than half of the genome, because then it would be the majority ($\theta \leq 0.5$)⁵³. In Supplementary Note 8, we calculate some real-world proportional cabal sizes (θ) by dividing the number of genes in a cabal by the total number of genes in a genome. In *Drosophila melanogaster*, a Y chromosome cabal, which favours male biased sex ratio distortion, has a proportional size of $\sim \theta \approx 0.001$ ^{54,55}. In human females, a cabal comprising cytoplasmic elements as well as the X chromosomes, which favours female-biased sex ratio distortion, has a proportional size of $\sim \theta \approx 0.04$ ^{56–58}. In *Escherichia coli*, a cabal made up of horizontally transferrable plasmids, which could favour

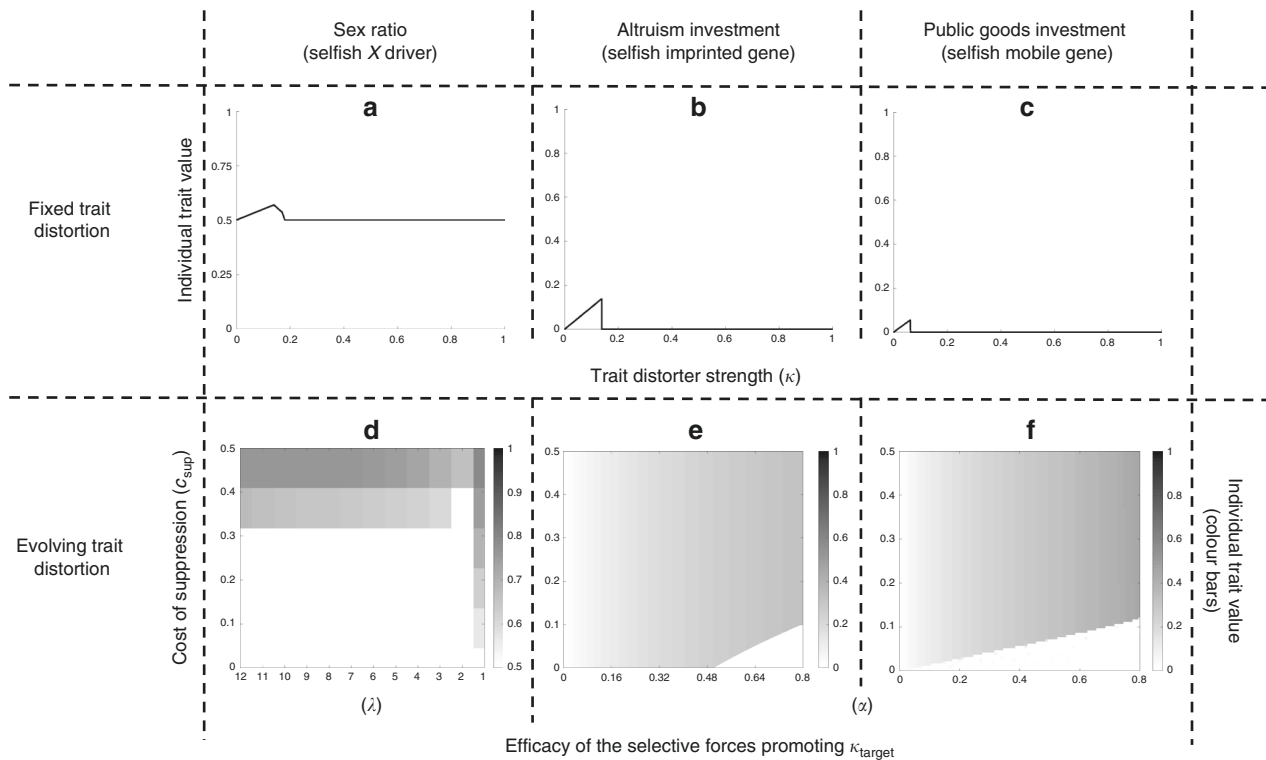


Fig. 3 Specific biological scenarios. We consider three biological scenarios: **a, d** sex ratio distortion by an X driver; **b, e** cooperative investment by an imprinted gene (cost and benefit of cooperation are $c = k$ and $b = k^\alpha$, respectively); **c, f** cooperative public goods investment by a mobile gene (cost and benefit of cooperation are $c = k$ and $b = 8k^\alpha$, respectively). In all three scenarios, we obtain the same pattern as our illustrative model, that trait distorters have either a minor impact at the individual level or are suppressed ($c_{sup} = 0.05$; **a**: $\lambda = 2$; **b, c**: $\alpha = 0.9$). In **d-f**, we allowed the trait distorters to evolve. We show how the equilibrium level of trait distortion is determined by the cost of suppression (c_{sup}), and parameters that determine the target level of trait distortion (k_{target} : **d**: λ ; **e, f**: α). Trait distortion is low unless there is both a high cost of suppression (c_{sup}) and a high target level of trait distortion (top right of heat maps)

upregulated public goods production⁴⁹, varies in size across strains, but has an average of $\sim \theta \approx 0.036$.

For analytical tractability, we start by assuming that new trait distorters and suppressors are introduced at a fixed rate (deterministic). Biologically, new trait distorters and suppressors are likely to arise via some combination of de novo mutation and the acquisition, via gene conversion or transposition, of pre-existing sequences contributing to trait distortion or suppression^{35,59,60}. We assume that a trait distorter arises at a new locus within the cabal every $1/(\theta\gamma\rho_{D_1})$ generations, and its dedicated suppressor arises at a locus inside the commonwealth $1/((1-\theta)\gamma\rho_{S_1})$ generations afterwards. ρ_{D_1} and ρ_{S_1} , respectively, give the generational per-locus probabilities of generating new trait distorters and suppressors. These probabilities (ρ_{D_1}, ρ_{S_1}) increase linearly, according to the same gradient, as the baseline mutation rate in the genome, denoted by ρ , is increased.

As in our equilibrium models, we assume that unsuppressed trait distorters distort organism traits by the fixed amount k , at an individual cost $c_{trait}(k)$, gaining a meiotic transmission advantage in heterozygotes of $(1+t(k))/2$. Similarly, we again assume that suppressors are dominant and completely suppress their target trait distorters at the cost c_{sup} , and are facultatively expressed in the presence of their target trait distorter⁵⁻⁸. We assume that the trait distortion experienced by an organism is given by the strength of its strongest unsuppressed trait distorter (inter-locus dominance).

We emphasise again that the mechanism by which the trait distorter gains its advantage (meiotic drive) is chosen here purely

for illustrative purposes (see Supplementary Notes 3, 4 and 5 for different mechanisms). We are interested in the subset of selfish genetic elements that gain their selfish benefit by distorting a trait away from the value that maximises individual fitness. The same trait distortion would be favoured across the coreplicon/cabal of which these selfish genetic elements are a part. This contrasts with selfish genetic elements that gain a selfish benefit through their ability to be meiotic drivers, without distorting a trait—such drivers could conceivably arise at any locus in a genome. The key difference here is between meiotic drive (could be favoured at any locus; selfish benefit does not arise via distorting a trait) and selfish genetic elements that gain a benefit by distorting a trait (the specific examples that we consider and model in this paper)^{14,15}.

We calculate the average and extremes of trait distortion faced by organisms in the population across evolutionary time, for different trait distorter strengths (k), and different proportional cabal sizes (θ). Considering trait distorters that do not trigger suppressor invasion ($c_{sup} > c_{trait}(k)$), the average trait distortion is trivially given by the strength of the trait distorters available to the cabal (k). Considering trait distorters that are suppressed and purged at equilibrium ($c_{sup} < c_{trait}(k)$), for analytical tractability, we first consider parameter regimes in which trait distorters are introduced at new loci more slowly than they are purged at old loci, meaning they do not co-segregate.

In ‘Long-term trait distortion (exact numerical solution)’ in the Methods, we develop a population genetic model based on these assumptions, and solve it numerically to show that individual trait distortion increases and decreases cyclically over evolutionary time, ranging between peaks of k and troughs of 0, as new

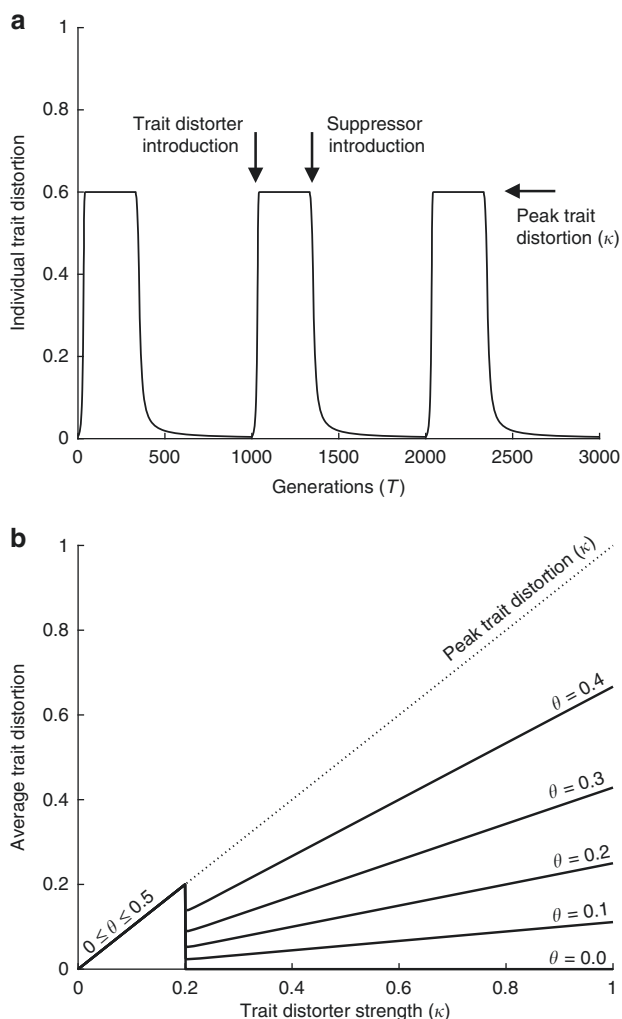


Fig. 4 The dynamics of conflict when trait distorters do not co-segregate. **a** plots individual trait distortion over evolutionary time, for trait distorters of strength $k = 0.6$. We introduced trait distorters (D_1) deterministically at new loci every $1/(\theta\gamma\rho_{D_1})$ generations, and their dedicated suppressors after a lag of $1/((1 - \theta)\gamma\rho_{S_1})$ generations. Individual trait distortion increases and decreases cyclically over evolutionary time, between peaks of k and troughs of 0. **b** plots average trait distortion for different proportional cabal sizes (θ), against the strength of trait distorters available to the cabal (k). Below a certain threshold strength ($c_{sup} > c_{trait}(k)$), suppressors do not invade, and so the resulting trait distortion increases with the strength of the trait distorter (k), and is unaffected by proportional cabal size (θ) (flatter lines). Above this threshold ($c_{sup} < c_{trait}(k)$), suppressors are favoured, and average trait distortion is approximately given by Eq. 6, increasing with trait distorter strength (k), although less appreciably for decreased proportional cabal size (θ) (flatter lines) ($c_{sup} = 0.1$; $t = k$, $c_{trait} = k/2$, $\rho_{S_1} = 10^{-11}$, $\rho_{D_1} = 10^{-11}$, $\gamma = 10^6$)

trait distorters and suppressors advance and retreat through the population (Fig. 4a). In ‘Long-term trait distortion (analytical approximation)’ in the Methods, we show that the average trait distortion over these cycles is given by

$$\frac{k\theta\rho_{D_1}}{(1-\theta)\rho_{S_1}}, \tag{6}$$

by making the assumption that the rate of gene frequency equilibration after trait distorter/suppressor introduction is very fast relative to the rate of trait distorter/suppressor introduction

(separation of timescales). For our three specific biological scenarios (Supplementary Notes 3, 4 and 5), the rate of gene frequency equilibration after trait distorter/suppressor introduction varies in each scenario, but these details are inconsequential when the separation of timescales assumption is made, meaning average trait distortion is given by Eq. 6 in each of the three specific biological scenarios. Furthermore, we also found with numerical analysis that Eq. 6 is a good approximation, even when the separation of timescales is relaxed (Fig. 4b).

Smaller proportional cabal sizes (θ) lead to a slower rate of trait distorter introduction relative to suppressor introduction, and so both: (i) an absolute reduction in average trait distortion; and (ii) a reduced effect of distorter strength (k) on average trait distortion ($k - \theta$ interaction) (Fig. 4b). In the limit of negligible proportional cabal size ($\theta \rightarrow 0$), we recover the result from our Equilibrium models that the proportion of evolutionary time in which a trait distorter is present approaches 0, leading to an average trait distortion of 0 for trait distorters above the threshold of suppression ($c_{sup} < c_{trait}(k)$).

Both genome size (γ) and baseline mutation rate (ρ) have no influence on the average trait distortion. Increases in both of these factors leads to a proportional increase in trait distorter introduction rate, and the same proportional increase in suppressor introduction rate, which exactly cancel (Supplementary Note 7, Supplementary Fig. 11).

(6) Trait distortion when trait distorters may co-segregate: We then considered the possibility that different trait distorters may co-segregate for some periods of evolutionary time^{59,60}. In ‘Agent-based simulation (multiple loci; discrete)’ in the Methods, we developed an agent-based simulation that allowed us to investigate the scenario where mutations appear stochastically rather than deterministically. When an individual contains multiple trait distorters, we assume that extent of trait distortion is determined by the strongest trait distorter (inter-locus dominance).

The consequence of allowing trait distorters to co-segregate will depend on mechanistic assumptions about how trait distorters and suppressors act and interact. To capture different ends of the continuum of possibilities, we model two different types of trait distorter, which we term low-sophistication (D_{1L}) and high-sophistication (D_{1H}) (Supplementary Note 7, Supplementary Fig. 12). High-sophistication trait distorters are only suppressed by dedicated suppressors that evolved to suppress that specific trait distorter, and incur a low cost when inter-locus recessive. In contrast, low-sophistication trait distorters can be suppressed to some extent by any suppressor (background or generalist suppression)^{35,59,60}, and incur a high cost when inter-locus recessive. High-sophistication trait distorters are more functionally complex, and so are likely to be less mutationally accessible than low-sophistication trait distorters.

We found that, for a sufficiently small proportional cabal size ($\theta \rightarrow 0$), trait distorters scarcely co-segregate, and Eq. 6 is recovered. Consequently, for sufficiently small proportional cabal sizes, the average level of trait distortion is again not influenced by genome size (γ), mutation rate (ρ), or the mechanics of trait distorter interaction (D_{1L}/D_{1H}).

In contrast, with larger cabals ($\theta \rightarrow 0.5$), trait distorters often co-segregate. In this case, the details of genome size (γ), mutation rate (ρ), and trait distorter sophistication (D_{1L}/D_{1H}) matter. Specifically, trait distortion may be: (i) greater than Eq. 6 if trait distorters are high sophistication (D_{1H}); (ii) lower than Eq. 6 if trait distorters are low sophistication (D_{1L}). The deviation from Eq. 6 is exaggerated for increased trait distorter co-segregation, which is promoted by: (i) high genome size (γ)/mutation rate (ρ) (Fig. 5); (iii) low trait distorter strength (k), which causes trait distorters to be purged more slowly (Supplementary Note 7,

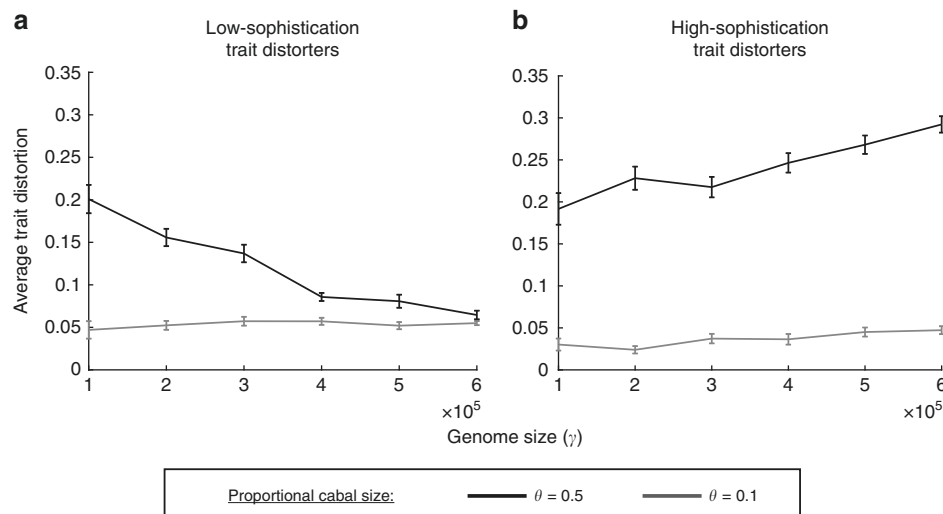


Fig. 5 Effect of genome size (γ) on average trait distortion when trait distorters can co-segregate. The average trait distortion is plotted against genome size (γ), for two different proportional cabal sizes (black: $\theta = 0.5$; grey: $\theta = 0.1$). **a** shows low-sophistication trait distorters (D_{1L}), which are partially suppressed by non-dedicated suppressors (background or generalist suppression), and incur a high cost even when inter-locus recessive. **b** shows high sophistication distorters (D_{1H}), which are only suppressed by dedicated suppressors, and incur a low cost when inter-locus recessive. The results plotted are an average of 36 runs of our agent-based simulation (error bars represent 1 standard error in each direction), each over $T_{\text{end}} = 30,000$ generations, where trait distorters and their dedicated suppressors are introduced stochastically. When proportional cabal size (θ) is smaller, the level of trait distortion is small, and genome size has little influence. When proportional cabal size (θ) is larger, the trait distortion is larger and depends upon genome size. Low-sophistication trait distorters interact counter-productively when co-segregating, meaning trait distortion decreases with genome size, while high-sophistication trait distorters interact productively when co-segregating, meaning trait distortion increases with genome size ($c_{\text{sup}} = 0.01$, $t = k$, $c_{\text{trait}} = \text{Dist}/2$, $\rho_{S_1} = 4 \times 10^{-9}$, $\rho_{D_{1L}} = 4 \times 10^{-9}$, $\rho_{D_{1H}} = 2 \times 10^{-9}$, $k = 0.5$)

Supplementary Fig. 14); (iv) low trait distorter sophistication (D_{1L}), which increases the mutational accessibility of trait distorters. The proportional cabal sizes that make these different factors matter are, however, much larger than we generally find in nature.

(7) Evolution of trait distortion and suppression: We then examined the consequences of allowing the level of trait distortion and suppression to evolve freely at each locus¹⁵. In ‘Agent-based simulation (multiple loci; continuous)’ in the Methods, we generalised our agent-based simulation to allow for this, and found that trait distorters evolve increased trait distortion (approaching k_{target}) while unsuppressed (Supplementary Note 7, Supplementary Fig. 15). Stronger trait distorters are suppressed and purged more quickly than weaker ones, and are less likely to co-segregate as a result. Consequently, when evolution is permitted at trait distorter loci, average trait distortion again approaches that predicted by Eq. 6, so is less influenced by genome size (γ), mutation rate (ρ), and the mechanics of trait distorter interaction (D_{1L}/D_{1H}).

Discussion

We obtained three main results: First, larger trait distortions are more likely to be suppressed. Consequently, trait distorters will either lead to small trait distortions, with minor fitness consequences, or be suppressed (Figs. 1a and 3a–c). Second, selection on trait distorters favours the evolution of higher levels of trait distortion, which will favour their suppression. Consequently, trait distorters will evolve to bring about their own demise (Figs. 2, 3d–f and 6). Third, if trait distortion is favoured at only a small proportion of the genome (proportionally small cabals), the extent of trait deviation away from the individual level optima is low and unaffected by factors, such as genome size, mutation rate and mechanism of trait distortion (Figs. 4 and 5). The reason for this result is that the influence of all of these factors is determined by proportional cabal size. Overall, these results suggest that even

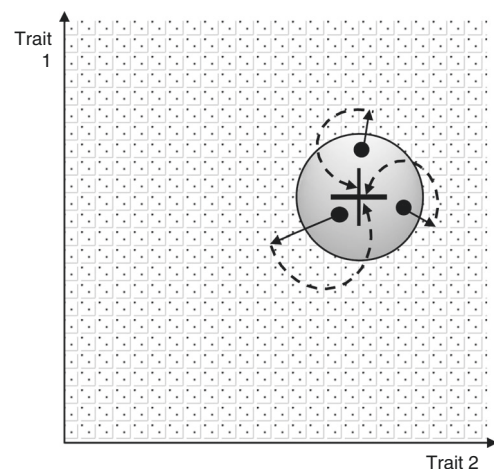


Fig. 6 Selfish genetic elements evolve to be suppressed by the parliament of genes. The cross represents the position in phenotype space, here defined with respect to two traits, 1 and 2, that maximises the fitness of an individual. The circle surrounding the cross represents the phenotype space where suppression of selfish genetic elements, that have distorted traits 1 or 2, would not be selected for. The surrounding area represents the phenotype space in which the parliament of genes is selected to suppress selfish genetic elements. The three dots represent three possible individuals, each with differently weakly selfish genetic elements, which incur a small fitness cost. Because these deviations from individual fitness maximisation are only slight, costly suppression of the weakly selfish genetic elements does not evolve. However, the selfish genetic elements will evolve to become more distorting (solid arrows), bringing individuals into the area of phenotype space where they will be suppressed and individual fitness maximisation (the black cross) is regained (dashed arrows)

if there is substantial potential for genetic conflict, trait distorters will have relatively little influence at the individual level, in support of Leigh's²⁸ parliament of genes hypothesis.

Suppressing trait distorters: We have shown that suppressors spread when the cost of suppression is lower than the fitness cost imposed by trait distortion ($c_{\text{trait}}(k) > c_{\text{sup}}$). The individual fitness cost of pre-translational suppression at a single locus is likely to be low. For example, a molecularly characterised suppressor (*nmy*) destroys the messenger RNA transcripts of a sex ratio distorter (*Dox*) via RNA interference (RNAi), the costs of which are likely to be negligible at the individual level^{46,47,60,61}. Consequently, in order to not be suppressed, a trait distorter would have to have relatively negligible influence on a trait, or influence a trait that has a negligible influence on fitness. Furthermore, we also showed that selection on trait distorters will often favour higher level trait distortion, bringing trait distorters into the region where $c_{\text{trait}}(k) > c_{\text{sup}}$, and hence where suppression is favoured (Figs. 2, 3 and 6).

Our analyses have focused on selfish genetic elements that increase their own transmission by manipulating some organism trait in a specific direction^{15,17}. Examples include the sex ratio distorters and public goods genes considered in our specific models. We focused on such 'trait distorters' because they can have substantial influences on the traits of organisms, even when at fixation. In contrast, we have not considered selfish genetic elements, such as transposons and meiotic drivers, that do not need to manipulate organism traits in order to give themselves a selfish propagation advantage⁴³. We have not considered such selfish genetic elements because: (i) they do not distort traits away from individual maxima; and (ii) the cost of such drivers makes them disfavoured across the entire genome, leading to selection to attenuate that cost.

Our Dynamics models have validated various verbal arguments that have previously been made for the parliament of genes hypothesis. We found that, if trait distortion is only favoured across a small proportion of the genome (proportionally small cabal), the trait distortion experienced by individuals is likely to be low, and unaffected by details such as genome size, mutation rate and mechanism of trait distortion. Empirically, cabals typically comprise small proportions of genomes^{54,56}. Furthermore, more sophisticated trait distorters, with the potential to interact synergistically with each other, are likely to have a lower mutational accessibility, and so are more likely to be suppressed and purged before they have a chance to co-segregate. Real-world examples of trait distortion are typically caused by lone genes, or genes that do not interact synergistically^{14,60}. In contrast, complex adaptations are typically underpinned by multitudes of synergistically interacting genes residing in the parliamentary majority (commonwealth)²³.

We are not claiming that appreciable trait distortion will never evolve, or that biological details will never matter^{14,32,59,60}. Instead, our results suggest that the modal outcome will be a relative lack of trait distortion. This conclusion is supported empirically by cases where appreciable distortion is only revealed in hybrid crosses, implying that trait distorters are generally suppressed⁶². Furthermore, we find that, after suppression has evolved, trait distorters are generally purged from the population at equilibrium. If suppressors are constitutively expressed (obligate), trait distorters are not purged from the population, but in these cases, suppressors spread to fixation (Supplementary Note 6). Regardless of the extent to which suppressors are constitutive, there is negligible polymorphism in at least one locus, meaning trait distortion is unlikely to be revealed by mating within a population³⁸. When trait distorters are not purged from the population, trait distortion will be revealed by matings between populations/species⁶².

Sex ratio distorters as a case study: The relatively large literature on sex ratio distorters offers a chance for us to assess the validity of our models, and their predictions. In Supplementary Note 3, we detail how our assumptions are consistent with the biology of sex ratio distorters and their suppressors. For example, X drivers increase their own transmission by killing Y bearing sperm, and hence producing a female-biased offspring sex ratio. This comes at a cost to the rest of the genome through both a reduction in sperm number, and through Fisherian selection disfavoured the more common sex (females). The scope of the parliament of genes to act against such drivers is shown by the fact that, in most species in which an X driver is present, suppressors have been found on both the autosomes and the Y chromosome³⁶. Our assumptions about how suppressors act, and the cost of suppression, are analogous to those in a molecularly characterised suppressor (*Nmy*) of a sex ratio distorter (*Dox*)^{46,60,61}; and more generally to suppressors that act pre-translationally^{63,64}.

Our model predictions are consistent with the available data on X drivers in *Drosophila*. As predicted by our model: (1) Across natural populations of *Drosophila simulans*, there is a positive correlation between the extent of sex ratio distortion and the extent of suppression⁶⁵. (2) In both *Drosophila mediopunctata* and *D. simulans* the presence of an X-linked driver led to the experimental evolution of suppression^{66,67}. In addition, consistent with our model: (3) In natural populations of *D. simulans*, the prevalence of an X driver has been shown to sometimes decrease under complete suppression⁶⁸. (4) Crossing different species of *Drosophila* has been shown to lead to appreciable sex ratio deviation, by unlinking trait distorters from their suppressors, and hence revealing previously hidden trait distorters⁶². Work on other sex ratio distorters has also shown that suppressors can spread extremely quickly from rarity, reaching fixation in as little as ~5 generations⁶⁹.

Individual fitness maximisation: We emphasise that when the assumption of individual fitness maximisation is made in behavioural and evolutionary ecology, it is not being assumed that natural selection produces perfect fitness maximisers⁵. Many factors could constrain adaptation, such as genetic architecture, mutation and phylogenetic constraints^{70,71}. Instead, the assumption of fitness maximisation is used as a basis to investigate the selective forces that have favoured particular traits (adaptations). The aim is not to test if organisms maximise fitness, or behave 'optimally', but rather to try to understand the selective forces favouring particular traits or behaviours². We have examined how the parliament of genes prevents selfish genetic elements from constraining adaptation, focusing on the maintenance, rather than the emergence, of traits (Supplementary Discussion).

To conclude, debate over the validity of assuming individual level fitness maximisation has usually revolved around whether selfish genetic elements are common or rare^{4,20,21,24,72}. We have shown that that even if selfish genetic elements are common, they will tend to be either weak and negligible, or suppressed. This suggests that even if there is the potential for appreciable genetic conflict, individual level fitness maximisation will still often be a reasonable assumption. This allows us to explain why certain traits, especially the sex ratio, have been able to provide such clear support for both individual level fitness maximisation and genetic conflict⁹.

Methods

Trait distorter population frequency. We ask when a rare trait distorter (D_1) can invade a population fixed for the trait non-distorter (D_0). We take Eq. 1, set $p' = p = p^*$, and solve to find two possible equilibria: $p^* = 0$ (trait non-distorter fixation) and $p^* = 1$ (trait distorter fixation). The trait distorter (D_1) can invade from rarity when the $p^* = 0$ equilibrium is unstable, which occurs when the differential of p'

with respect to p , at $p^* = 0$, is >1 . The trait distorter invasion criterion is therefore $c_{\text{trait}}(k) < t(k)(1 - c_{\text{trait}}(k))$.

We now ask what frequency the trait distorter (D_1) will reach after invasion. The trait distorter (D_1) can spread to fixation if the $p^* = 1$ equilibrium is stable, which requires that the differential of p' with respect to p , at $p^* = 1$, is <1 . This requirement always holds true, demonstrating that there is no negative frequency dependence on the trait distorter, and that it will always spread to fixation after its initial invasion.

Suppressor invasion condition. We ask when the suppressor (S_1) can spread from rarity in a population in which the trait distorter (D_1) and non-suppressor (S_0) are fixed at equilibrium. We derive the Jacobian stability matrix for this equilibrium, which is a matrix of each genotype frequency (x_{00}' , x_{01}' , x_{10}' , x_{11}') differentiated by each genotype frequency in the prior generation (x_{00} , x_{01} , x_{10} , x_{11}), at the equilibrium position given by $x_{00}^* = 0$, $x_{01}^* = 0$, $x_{10}^* = 1$, $x_{11}^* = 0$:

$$J = \begin{pmatrix} 1-t & \frac{1-c_{\text{sup}}}{2(1-c_{\text{trait}})} & 0 & 0 \\ 0 & \frac{1-c_{\text{sup}}}{2(1-c_{\text{trait}})} & 0 & 0 \\ t-1 & \frac{-3(1-c_{\text{sup}})}{2(1-c_{\text{trait}})} & 0 & \frac{-(1-c_{\text{sup}})}{1-c_{\text{trait}}} \\ 0 & \frac{1-c_{\text{sup}}}{2(1-c_{\text{trait}})} & 0 & \frac{1-c_{\text{sup}}}{1-c_{\text{trait}}} \end{pmatrix}, \quad (7)$$

The suppressor can invade when the equilibrium is unstable, which occurs when the leading eigenvalue is greater than one. The leading eigenvalue is $(1 - c_{\text{sup}})/(1 - c_{\text{trait}})$, meaning the suppressor invasion criterion is $c_{\text{trait}} > c_{\text{sup}}$.

Equilibrium trait distorter and suppressor frequencies. We ask what frequency the trait distorter (D_1) and suppressor (S_1) will reach after initial suppressor (S_1) invasion. We assume that the suppressor is introduced from rarity when the trait distorter has reached the population frequency given by $f(x_{00} \rightarrow f, x_{10} \rightarrow 1-f, \{x_{01}, x_{11}\} \rightarrow 0)$. We numerically iterate Eqs. 2–5, over successive generations, until equilibrium has been reached. At equilibrium, for all parameter combinations ($f, t, c_{\text{sup}}, c_{\text{trait}}$), the suppressor reaches an internal equilibrium and the trait distorter is lost from the population ($x_{00}^* + x_{01}^* = 1, x_{10}^* = 0, x_{11}^* = 0$). This equilibrium arises because trait distorter presence gives the suppressor (S_1) a selective advantage, leading to high suppressor frequency, which in turn reverses the selective advantage of the trait distorter (D_1), leading to trait distorter loss and suppressor equilibration.

Non-equilibrium trait distortion. We consider a trait distorter that is suppressed and therefore purged at equilibrium ($c_{\text{trait}} > c_{\text{sup}}$), and ask to what extent it can contribute to individual trait distortion in the period after its initial invasion, but before its eventual loss (non-equilibrium). We introduce the trait distorter (D_1) and suppressor (S_1) from rarity and numerically iterate our recursions until the trait distorter has been purged from the population (or a cap of 20,000,000 generations has been reached). We vary parameters between $0 \leq t \leq 1, c_{\text{sup}} < c_{\text{trait}} \leq 1, 0 \leq c_{\text{sup}} \leq 1$.

We find that a higher cost of trait distortion (c_{trait}) relative to suppression (c_{sup}) leads to shorter non-equilibrium maintenance of the trait distorter in the population. This is because the cost of trait distortion relative to suppression mediates selection on the suppressor (Methods: ‘Suppressor invasion condition’). We find that a higher transmission bias (t) leads to longer non-equilibrium maintenance of the trait distorter in the population, but this effect is diluted as the cost of trait distortion (c_{trait}) is increased relative to suppression (c_{sup}) (Supplementary Note 2, Supplementary Fig. 1). Stronger trait distorters (with higher k , leading to higher c_{trait} and t) are therefore generally suppressed and purged more rapidly than weaker trait distorters (Fig. 1b). Exceptions are trait distorters that reduce individual fitness relatively negligibly after the point (k) at which suppression is favoured, such that $\frac{dt}{dk} / \frac{dc_{\text{trait}}}{dk}$ is very high for values of k satisfying $c_{\text{sup}} < c_{\text{trait}}(k)$.

Invasion of a mutant trait distorter. We ask when a mutant trait distorter (D_2) will invade against a resident trait distorter (D_1) that is unsuppressed and at fixation ($k \neq \hat{k}$). We write recursions detailing the generational frequency changes in the six possible gametes, $D_0/S_0, D_0/S_1, D_1/S_0, D_1/S_1, D_2/S_0, D_2/S_1$, with current generation frequencies denoted, respectively by $x_{00}, x_{01}, x_{10}, x_{11}, x_{20}, x_{21}$, and next-generation frequencies denoted with an appended dash ($'$):

$$\begin{aligned} \bar{w}x'_{00} &= x_{00}x_{00} + x_{00}x_{01} + (1-t(k))(1-c_{\text{trait}}(k))x_{00}x_{10} \\ &+ \left(\frac{1-c_{\text{sup}}}{2}\right)x_{00}x_{11} + (1-t(\hat{k}))(1-c_{\text{trait}}(\hat{k})) \\ &x_{00}x_{20} + \left(\frac{1-c_{\text{sup}}}{2}\right)x_{00}x_{21} + \left(\frac{1-c_{\text{sup}}}{2}\right) \\ &x_{01}x_{10} + \left(\frac{1-c_{\text{sup}}}{2}\right)x_{01}x_{20}, \end{aligned} \quad (8)$$

$$\begin{aligned} \bar{w}x'_{01} &= x_{00}x_{01} + \left(\frac{1-c_{\text{sup}}}{2}\right)x_{00}x_{11} + \left(\frac{1-c_{\text{sup}}}{2}\right) \\ &x_{00}x_{21} + x_{01}x_{01} + \left(\frac{1-c_{\text{sup}}}{2}\right)x_{01}x_{10} + (1-c_{\text{sup}})x_{01}x_{11} \\ &+ \left(\frac{1-c_{\text{sup}}}{2}\right)x_{01}x_{20} + (1-c_{\text{sup}})x_{01}x_{21}, \end{aligned} \quad (9)$$

$$\begin{aligned} \bar{w}x'_{10} &= (1+t(k))(1-c_{\text{trait}}(k))x_{00}x_{10} + \left(\frac{1-c_{\text{sup}}}{2}\right) \\ &x_{00}x_{11} + \left(\frac{1-c_{\text{sup}}}{2}\right)x_{01}x_{10} + (1-c_{\text{trait}}(k)) \\ &x_{10}x_{10} + (1-c_{\text{sup}})x_{10}x_{11} + (1+t(k)-t(\hat{k})) \\ &(1-c_{\text{trait}}(\max(k, \hat{k})))x_{10}x_{20} + \left(\frac{1-c_{\text{sup}}}{2}\right)x_{10}x_{21} \\ &+ \left(\frac{1-c_{\text{sup}}}{2}\right)x_{11}x_{20}, \end{aligned} \quad (10)$$

$$\begin{aligned} \bar{w}x'_{11} &= \left(\frac{1-c_{\text{sup}}}{2}\right)x_{00}x_{11} + \left(\frac{1-c_{\text{sup}}}{2}\right) \\ &x_{01}x_{10} + (1-c_{\text{sup}})x_{01}x_{11} + (1-c_{\text{sup}}) \\ &x_{10}x_{11} + \left(\frac{1-c_{\text{sup}}}{2}\right)x_{10}x_{21} \\ &+ (1-c_{\text{sup}})x_{11}x_{11} + \left(\frac{1-c_{\text{sup}}}{2}\right) \\ &x_{11}x_{20} + (1-c_{\text{sup}})x_{11}x_{21}, \end{aligned} \quad (11)$$

$$\begin{aligned} \bar{w}x'_{20} &= (1+t(\hat{k}))(1-c_{\text{trait}}(\hat{k}))x_{00}x_{20} + \left(\frac{1-c_{\text{sup}}}{2}\right) \\ &x_{00}x_{21} + \left(\frac{1-c_{\text{sup}}}{2}\right)x_{01}x_{20} + (1-t(k)+t(\hat{k})) \\ &(1-c_{\text{trait}}(\max(k, \hat{k})))x_{10}x_{20} + \left(\frac{1-c_{\text{sup}}}{2}\right) \\ &x_{10}x_{21} + \left(\frac{1-c_{\text{sup}}}{2}\right)x_{11}x_{20} + (1-c_{\text{trait}}(\hat{k})) \\ &x_{20}x_{20} + (1-c_{\text{sup}})x_{20}x_{21}, \end{aligned} \quad (12)$$

$$\begin{aligned} \bar{w}x'_{21} &= \left(\frac{1-c_{\text{sup}}}{2}\right)x_{00}x_{21} + \left(\frac{1-c_{\text{sup}}}{2}\right) \\ &x_{01}x_{20} + (1-c_{\text{sup}})x_{01}x_{21} + \left(\frac{1-c_{\text{sup}}}{2}\right)x_{10}x_{21} + \left(\frac{1-c_{\text{sup}}}{2}\right) \\ &x_{11}x_{20} + (1-c_{\text{sup}})x_{11}x_{21} + (1-c_{\text{sup}})x_{20}x_{21} + (1-c_{\text{sup}})x_{21}x_{21}, \end{aligned} \quad (13)$$

where \bar{w} is the average fitness of individuals in the current generation, and equals the sum of the right-hand side of the system of equations. The mutant trait distorter can invade when the equilibrium given by $x_{00}^* = 0, x_{01}^* = 0, x_{10}^* = 1, x_{11}^* = 0, x_{20}^* = 0, x_{21}^* = 0$ is unstable, which occurs when the leading eigenvalue of the Jacobian stability matrix for this equilibrium is >1 . Testing for stability in this way, we find that, if the mutant trait distorter is weaker than the resident, it can never invade. If the mutant trait distorter is stronger than the resident, it invades from rarity when $\Delta(1 - c_{\text{trait}}(\hat{k})) > \Delta c_{\text{trait}}$, where $\Delta t = t(\hat{k}) - t(k)$, $\Delta c_{\text{trait}} = c_{\text{trait}}(\hat{k}) - c_{\text{trait}}(k)$.

The implication is that, if trait distortion is initially low, and mutant trait distorters are successively introduced, each deviating only very slightly from the resident trait distorter from which they are derived, such that $\hat{k} = k \pm \delta$, where δ is very small (δ -weak selection⁴⁸), then trait distorters will approach a ‘target’ strength at which $\frac{dt}{dk}(1 - c_{\text{trait}}) = \frac{dc_{\text{trait}}}{dk}$. In the absence of suppression, this target (k_{target}) is the equilibrium level of trait distortion ($k^* = k_{\text{target}}$). However, if mutant trait distorters (D_2) are allowed to deviate appreciably from residents (D_1) (strong selection), then trait distorters may invade even if they overshoot the target ($\hat{k} > k_{\text{target}}$). In the absence of suppression, k_{target} is then not the equilibrium level of trait distortion, but rather, the minimum equilibrium level of trait distortion ($k^* \geq k_{\text{target}}$) (Supplementary Note 2, Supplementary Fig. 2b).

We could alternatively have assumed that an individual’s trait is distorted according to the average strength of its alleles (additive gene interactions), rather than according to the stronger (higher k) allele (dominance). Such an assumption leads to a single invasion criterion for a mutant trait distorter, regardless of whether the mutant trait distorter is stronger or weaker than the resident trait distorter, given by: $\Delta(2 - c_{\text{trait}}(k) - c_{\text{trait}}(\hat{k})) > \Delta c_{\text{trait}}$. In the absence of suppression, this leads to an equilibrium level of trait distortion (k^*), which holds even under strong selection, and satisfies $2 \frac{dt}{dk}(1 - c_{\text{trait}}) = \frac{dc_{\text{trait}}}{dk}$.

Equilibrium allele frequencies after mutant invasion. We ask what equilibrium state will arise after the invasion of a mutant trait distorter. We assume that the mutant trait distorter (D_2) is introduced from rarity when the resident trait distorter (D_1) has reached the population frequency given by q . We numerically iterate Eq. 8–13, over successive generations, until equilibrium has been reached. At equilibrium, for all parameter combinations ($q, t(k), t(\hat{k}), c_{\text{sup}}, c_{\text{trait}}(k), c_{\text{trait}}(\hat{k})$), the resident trait distorter (D_1) is lost from the population ($x_{10}, x_{11} = 0$), with either the mutant trait distorter (D_2) and non-suppressor (S_0) at fixation ($x_{20}^* = 1$), or the trait non-distorter (D_0) at fixation alongside the suppressor (S_1) at an internal equilibrium ($x_{00}^* + x_{01}^* = 1$). The latter scenario arises if the mutant trait distorter triggers suppressor invasion ($c_{\text{sup}} < c_{\text{trait}}(\hat{k})$). This equilibrium arises because mutant trait distorter presence gives the suppressor (S_1) a selective advantage, leading to high

suppressor frequency, which in turn reverses the selective advantage of trait distortion, leading to trait distorter (D_1, D_2) loss and suppressor equilibration.

Agent-based simulation (single trait distorter locus). We construct an agent-based simulation to ask what level of trait distortion evolves when continuous variation is permitted at trait distorter and suppressor loci. We model a population of $N = 2000$ individuals and track evolution at two autosomal loci: a trait distorter locus and a suppressor locus. Each individual has two alleles at the trait distorter locus, with strengths denoted by k_a and k_b , and two alleles at the suppressor locus, with strengths denoted by m_a and m_b (diploid). Strengths can take any continuous value between 0 and 1. We assume that, for both loci, the strongest (highest value) allele within an individual is dominant. The absolute fitness of an individual with at least one active meiotic driver ($\max(k_a, k_b) > 0$) is: $1 - c_{\text{trait}}(\max(k_a, k_b))(1 - \max(m_a, m_b)) - c_{\text{sup}}\max(m_a, m_b)$, and the absolute fitness of an individual lacking an active trait distorter ($\max(k_a, k_b) = 0$) is 1. The function $c_{\text{trait}}(\max(k_a, k_b))$ is given an explicit form in simulations (Supplementary Note 2, Supplementary Fig. 2).

In each generation, there are N breeding pairs. To fill each position in each breeding pair, individuals are drawn from the population, with replacement, with probabilities given by their fitness (hermaphrodites). Breeding pairs then reproduce to produce one offspring, before dying (non-overlapping generations). Alleles at the suppressor locus are inherited in Mendelian fashion. Alleles at the trait distorter locus may drive, meaning the parental allele of strength k_a is inherited, rather than the allele of strength k_b , with the probability $(1 + (t(k_a) - t(k_b))(1 - \max(m_a, m_b))) / 2$. The transmission bias function, t , is given an explicit form in simulations (Supplementary Note 2, Supplementary Fig. 2). Each generation, trait distorter and suppressor alleles have a 0.01 chance of mutating to a new value, which is drawn from a normal distribution centred around the pre-mutation value, with variance 0.2, and truncated between 0 and 1. We track the population average trait distorter strength, denoted by $E[k]$, and suppressor strength, denoted by $E[m]$, over 20,000 generations. We see that, allowing for continuous variation at the trait distorter and suppressor loci, if the cost of suppression (c_{sup}) is not excessively high, trait distortion at equilibrium is either low or nothing (Fig. 2a; Supplementary Note 2, Supplementary Fig. 2b).

Long-term trait distortion (exact numerical solution). We ask how the trait distortion experienced by organisms changes across evolutionary time as new trait distorters and suppressors are continuously introduced and lost from a population. We construct a population genetic model and solve it numerically and exactly. We introduce a trait distorter from rarity and iterate our recursion for an unsuppressed trait distorter (Eq. 1) from $T = 1$ to $T = 1 / ((1 - \theta)\gamma\rho_{S_i})$ generations. During this period, the trait distortion experienced by individuals rises to a peak of k , corresponding to the strength of trait distorters available to the cabal. We then introduce a suppressor from rarity and iterate our recursions for trait distorter-suppressor co-segregation (Eqs. 2–5), from $T = 1 / ((1 - \theta)\gamma\rho_{S_i})$ until the trait distorter has been purged ($T = X$). During this period, the trait distortion experienced by individuals falls to a trough of 0.

Average trait distortion over evolutionary time is given by weighting average trait distortion during the interval $T = \{1, 2, \dots, X\}$ by the proportion of evolutionary time in which a trait distorter is segregating in the population ($X(\theta\gamma\rho_{D_i})$). This methodology provides exact, numerical values for average trait distortion. These values correspond closely to the analytical approximation for average trait distortion (Eq. 6), which is derived under a separation of timescales assumption (Methods: ‘Long-term trait distortion (analytical approximation)’; Fig. 4).

Long-term trait distortion (analytical approximation). When a trait distorter is initially introduced into the population, it will spread, and the population will equilibrate when the trait distorter reaches fixation (Methods: ‘Long-term trait distortion (exact numerical solution)’). Similarly, when a suppressor is initially introduced into the population, it will spread if its target trait distorter is sufficiently costly ($c_{\text{sup}} < c_{\text{trait}}(k)$), and the population will equilibrate when the suppressor’s target trait distorter is purged from the population (Methods: ‘Long-term trait distortion (exact numerical solution)’). We assume that, after the introduction of a new trait distorter or suppressor, the rate at which gene frequencies equilibrate is very fast relative to the rate at which new trait distorters and suppressors are introduced at new loci (separation of timescales).

On this assumption, we can partition evolutionary time into two repeating periods. In the first period, comprising the $1 / ((1 - \theta)\gamma\rho_{S_i})$ generations in between trait distorter and suppressor introduction, individual trait distortion is k . In the second period, comprising the following $1 / (\theta\gamma\rho_{D_i}) - 1 / ((1 - \theta)\gamma\rho_{S_i})$ generations, and ending when the next trait distorter is introduced at a new locus, individual trait distortion is 0. We average over these two time periods to calculate the average trait distortion experienced by individuals across evolutionary time (Eq. 6).

Agent-based simulation (multiple loci; discrete). We build on the agent-based model detailed in Methods: ‘Agent-based simulation (single trait distorter locus)’ to capture the evolutionary dynamics of arbitrarily large numbers of co-segregating trait distorters and suppressors across the genome. The specific details of how mate partners are attributed (e.g. panmictic; hermaphrodite), and how the population is

sampled to implement fitness effects (e.g. non-overlapping generations), are fully described in Methods: ‘Agent-based simulation (single trait distorter locus)’. We model a diploid population of $N = 2000$ individuals, each with $\gamma = 10^6$ loci, $\theta\gamma$ of which constituting the cabal and $(1 - \theta)\gamma$ of which constituting the commonwealth.

We assume that each locus across the genome is initially ‘dormant’. The alleles segregating in the population at dormant loci are neutral with respect to trait distortion and suppression. Loci are activated when the alleles segregating there have drifted to lie one mutational step away from distortion or suppression. For a given dormant locus in the cabal and in the commonwealth, the generational activation probability is given, respectively, by ρ_{D_i} and ρ_{S_i} . Each successively activated cabal and commonwealth locus is indexed with a consecutive integer within the respective sets $I_{\text{cabal}} = \{1, 2, \dots, n_{\text{cabal}}\}$ and $I_{\text{commonwealth}} = \{1, 2, \dots, n_{\text{commonwealth}}\}$, where n_{cabal} and $n_{\text{commonwealth}}$ give respectively the total number of activated cabal and commonwealth loci, which increase as generations (T) pass. After locus activation, alleles mutate between functional and neutral forms with a generational probability of 0.001. If, at any time, all trait distorters ($i \in I_{\text{cabal}}$) have dedicated suppressors ($i \in I_{\text{commonwealth}}$), such that $n_{\text{cabal}} = n_{\text{commonwealth}}$, further commonwealth loci cannot be activated until new trait distorters arise ($n_{\text{cabal}} > n_{\text{commonwealth}}$). If trait distorters are low-sophistication as opposed to high-sophistication, the generational cabal locus activation probability (ρ_{D_i}) is increased by a factor two (such that $\rho_{D_{iL}} = 2^* \rho_{D_{iH}}$).

For each individual, the set $I_{\text{distorter}} \subseteq I_{\text{cabal}}$ comprises every locus within the cabal where one (heterozygous) or two (homozygous) trait distorters are present. A given suppressor at a locus within the commonwealth ($i \in I_{\text{commonwealth}}$) is only expressed if its target trait distorter ($i \in I_{\text{distorter}}$) is also present in the individual. However, if expressed, a given suppressor ($i \in I_{\text{commonwealth}}$) may also contribute to the ‘background’ suppression of unsuppressed non-target trait distorters ($I_{\text{distorter}} \setminus i$), at a fraction z of its usual strength. We assume that, for low-sophistication trait distorters (D_{1L}), $z = 0.5$, and for high-sophistication trait distorters (D_{1H}), $z = 0$.

The total suppression faced by a trait distorter ($i \in I_{\text{distorter}}$) is therefore $\text{TotSup}_i = 1$ if its dedicated suppressor is present in the individual, or $\text{TotSup}_i = \min(zq, 1)$ if its dedicated suppressor is absent, where q is the number of expressed suppressors present in the individual, and where the ‘min’ notation indicates that the total suppression cannot exceed 1 (complete suppression). The total cost of suppression for an individual is $c_{\text{sup}} \sum_{i \in I_{\text{distorter}}} \text{TotSup}_i$. The least suppressed trait distorter in each individual ($i_{\text{dom}} \in I_{\text{distorter}}$) exerts inter-locus dominance, and causes a trait distortion of $\text{Dist} = \max_{i \in I_{\text{distorter}}} ((1 - \text{TotSup}_i)k)$. The individual cost of trait distortion, which is given by $c_{\text{trait}}(\text{Dist})$, increases monotonically with the extent that the trait is distorted ($\frac{dc_{\text{trait}}}{d\text{Dist}} \geq 0$).

Expression of the remaining ‘inter-locus recessive’ trait distorters ($I_{\text{distorter}} \setminus i_{\text{dom}}$) leads to a pool of gene products with an abundance that is proportional to: $\text{Waste} = \sum_{i \in I_{\text{distorter}}} \max_{i \neq i_{\text{dom}}} ((1 - \text{TotSup}_i)k)$. The individual cost arising from inter-locus recessive trait distorters, which is given by c_{rec} , increases monotonically with the size of the pool of redundant gene products ($\frac{dc_{\text{rec}}}{d\text{Waste}} \geq 0$). We assume that, for low-sophistication trait distorters (D_{1L}), the individual cost arising from any one inter-locus recessive trait distorter is equal to the cost of trait distortion itself ($c_{\text{trait}}(\text{Dist}) = \frac{c_{\text{rec}}(\text{Waste})}{|I_{\text{distorter}}| - 1} \geq 0$). For high-sophistication trait distorters (D_{1H}), this cost is lower relative to the cost of trait distortion ($c_{\text{trait}}(\text{Dist}) = \frac{5c_{\text{rec}}(\text{Waste})}{3(|I_{\text{distorter}}| - 1)} \geq 0$). The total fitness (viability) of an individual is then given by: $1 - c_{\text{trait}}(\text{Dist}) - c_{\text{rec}}(\text{Waste}) - c_{\text{sup}} \sum_{i \in I_{\text{distorter}}} \text{TotSup}_i$.

We define the set $I_{\text{het}} \subseteq I_{\text{distorter}} \subseteq I_{\text{cabal}}$ as the collection of loci in an individual at which one (heterozygous) trait distorter, as opposed to two (homozygous) trait distorters, are present. The trait distorters at these loci (I_{het}) drive at meiosis, as a unit. The least suppressed trait distorter in the group pulls the unit through meiosis, meaning the group of trait distorters (at loci I_{het}) is inherited by each offspring with the probability $(1 + \max_{i \in I_{\text{het}}} (1 - \text{TotSup}_i)k) / 2$.

Agent-based simulation (multiple loci; continuous). We adapt the simulation model detailed in Methods: ‘Agent-based simulation (multiple loci; discrete)’ so that trait distorters and suppressors are not of fixed strength (of k and 1, respectively), but are free to evolve continuously between 0 and 1.

Homologous alleles at activated cabal loci ($i \in I_{\text{cabal}}$) have strengths k_{ai} and k_{bi} , and homologous alleles at activated commonwealth loci ($i \in I_{\text{commonwealth}}$) have strengths m_{ai} and m_{bi} . Within an individual, the loci bearing trait distorters ($I_{\text{distorter}} \subseteq I_{\text{cabal}}$) each satisfy $\max(k_{ai}, k_{bi}) > 0$. Each trait distorter (at locus $i \in I_{\text{distorter}}$) is suppressed to the following extent:

$$\text{TotSup}_i = \min \left(\max(m_{ai}, m_{bi}) + z \sum_{j \in I_{\text{distorter}}} \max_{j \neq i} (m_{aj}, m_{bj}), 1 \right).$$

Within an individual, the strongest trait distorter (after suppression) is inter-locus dominant ($i_{\text{dom}} \in I_{\text{distorter}}$), and distorts the individual trait

by: $\text{Dist} = \max_{i \in I_{\text{distorter}}} ((1 - \text{TotSup}_i) \max(k_{ai}, k_{bi}))$. The inter-locus recessive trait distorters ($I_{\text{distorter}}^{\text{Vdom}}$) bring about an additional individual level cost of $c_{\text{rec}}(\text{Waste})$, which is a monotonically increasing function of $\text{Waste} = \sum_{i \in I_{\text{distorter}}^{i \neq \text{dom}}} ((1 - \text{TotSup}_i) \max(k_{ai}, k_{bi}))$.

If an allele is more trait-distorting than its homologue (k_{ai} vs. k_{bi}), it can drive at meiosis. The strongest alleles across each homologous pair drive together as a single unit. The unit is inherited by each offspring with the probability $\left(1 + \frac{\max_{i \in I_{\text{distorter}}} (1 - \text{TotSup}_i) \text{abs}(k_{ai} - k_{bi})}{2}\right)$. Every generation, each allele at an activated locus has a 0.01 chance of mutating to a new strength, which is drawn from a normal distribution centred around the pre-mutation strength, with variance 0.2, and truncated between 0 and 1.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data that support the findings of this study are available upon request.

Code availability

Simulation code is available upon request.

Received: 8 March 2019; Accepted: 24 October 2019;

Published online: 14 November 2019

References

- Fisher, R. A. *The Genetical Theory of Natural Selection* (Oxford Univ. Press, 1930).
- Gardner, A. Adaptation as organism design. *Biol. Lett.* **5**, 861–864 (2009).
- Hamilton, W. D. The genetical evolution of social behaviour. I. *J. Theor. Biol.* **7**, 1–16 (1964).
- Grafen, A. The formal darwinism project in outline. *Biol. Philos.* **29**, 155–174 (2014).
- Parker, G. A. & Maynard Smith, J. Optimality theory in evolutionary biology. *Nat. Publ. Group* **348**, 27–33 (1990).
- Frank, S. A. *Foundations of Social Evolution*. *Heredity* **82**, 268–344 (1998).
- Stephens, D. W. & Krebs, J. R. *Foraging Theory* (Princeton Univ. Press, 1986).
- Stearns, S. C. *The Evolution of Life Histories*. (Oxford Univ. Press, 1992).
- West, S. *Sex Allocation* (Princeton Univ. Press, 2009).
- Davies, N. B., Krebs, J. R. & West, S. A. *An Introduction to Behavioural Ecology* (Wiley, 2012).
- Westneat, D. & Fox, C. W. *Evolutionary Behavioral Ecology* (Oxford Univ. Press, 2010).
- Bourke, A. F. G. *Principles of Social Evolution* (Oxford Univ. Press, 2011).
- Werren, J. H., Nur, U. & Wu, C. I. Selfish genetic elements. *Trends Ecol. Evol.* **3**, 297–302 (1988).
- Burt, A. & Trivers, R. *Genes in Conflict* (Harvard Univ. Press, 2006).
- Gardner, A. & Úbeda, F. The meaning of intragenomic conflict. *Nat. Ecol. Evol.* **1**, 1–9 (2017).
- Ågren, J. A. & Clark, A. G. Selfish genetic elements. *PLoS Genet.* **14**, e1007700–e1007720 (2018).
- Eshel, I. Are intragenomic conflicts common in nature? Do they represent an important factor in evolution? *J. Theor. Biol.* **108**, 159–162 (1984).
- Hamilton, W. D. Extraordinary sex ratios. A sex-ratio theory for sex linkage and inbreeding has new implications in cytogenetics and entomology. *Science* **156**, 477–488 (1967).
- Haig, D. *Genomic Imprinting and Kinship* (Rutgers Univ. Press, 2002).
- Haig, D. Genetic dissent and individual compromise. *Biol. Philos.* **29**, 233–239 (2014).
- Bourke, A. F. G. The gene's-eye view, major transitions and the formal darwinism project. *Biol. Philos.* **29**, 241–248 (2014).
- Cosmides, L. M. & Tooby, J. Cytoplasmic inheritance and intragenomic conflict. *J. Theor. Biol.* **89**, 83–129 (1981).
- West, S. A. & Gardner, A. Adaptation and inclusive fitness. *Curr. Biol.* **23**, R577–R584 (2013).
- Ågren, J. A. Selfish genetic elements and the gene's-eye view of evolution. *Curr. Zool.* **62**, 659–665 (2016).
- Charnov, E. L. *The Theory of Sex Allocation* (Princeton Univ. Press, 1982).
- Jaenike, J. Sex chromosome meiotic drive. *Annu. Rev. Ecol. Syst.* **32**, 25–49 (2001).
- Leigh, E. G., Herre, E. A. & Fischer, E. A. Sex allocation in animals. *Experientia* **41**, 1265–1276 (1985).
- Leigh, E. G. *Adaptation and Diversity* (Freeman, 1971).
- Leigh, E. G. When does the good of the group override the advantage of the individual? *Proc. Natl Acad. Sci. USA* **80**, 2985–2989 (1983).
- Leigh, E. G. How does selection reconcile individual advantage with the good of the group? *Proc. Natl Acad. Sci. USA* **74**, 4542–4546 (1977).
- Strassmann, J. E. & Queller, D. C. The social organism: congresses, parties, and committees. *Evolution* **64**, 605–616 (2010).
- Queller, D. C. & Strassmann, J. E. Evolutionary conflict. *Annu. Rev. Ecol. Evol. Syst.* **49**, 73–93 (2018).
- Gardner, A. & Ross, L. Mating ecology explains patterns of genome elimination. *Ecol. Lett.* **17**, 1602–1612 (2014).
- Úbeda, F. Why Mendelian segregation? *Biochem. Soc. Trans.* **34**, 566–568 (2006).
- Lindholm, A. K. et al. The ecology and evolutionary dynamics of meiotic drive. *Trends Ecol. Evol.* **31**, 315–326 (2016).
- Helleu, Q., Gérard, P. R. & Montchamp-Moreau, C. Sex chromosome drive. *Cold Spring Harb. Perspect. Biol.* **7**, a017616–a017616 (2015).
- Crow, J. F. Why is Mendelian segregation so exact? *BioEssays* **13**, 305–312 (1991).
- Ridley, M. *Mendel's Demon* 176–177 (Weidenfeld & Nicolson, 2000).
- Prout, T., Bundgaard, J. & Bryant, S. Population genetics of modifiers of meiotic drive I. The solution of a special case and some general implications. *Theor. Popul. Biol.* **4**, 446–465 (1973).
- Caubet, Y., Hatcher, M. J., Mocquard, J. P. & Rigaud, T. Genetic conflict and changes in heterogametic mechanisms of sex determination. *J. Evol. Biol.* **13**, 766–777 (2000).
- Randerson, J. P. The evolutionary dynamics of male-killers and their hosts. *Heredity* **84**, 1–9 (2000).
- Haig, D. Intragenomic politics. *Cytogenet. Genome Res.* **113**, 68–74 (2006).
- Eshel, I. Evolutionary genetic stability of mendelian segregation and the role of free recombination in the chromosomal system. *Am. Naturalist* **125**, 412–420 (1985).
- Larracuente, A. M. & Presgraves, D. C. The selfish segregation distorter gene complex of *Drosophila melanogaster*. *Genetics* **192**, 33–53 (2012).
- Zanders, S. E. & Unckless, R. L. Fertility costs of meiotic drivers. *Curr. Biol.* **29**, R512–R520 (2019).
- Tao, Y., Masly, J. P., Ararape, L., Ke, Y. & Hartl, D. L. A sex-ratio meiotic drive system in *Drosophila simulans*. I: an autosomal suppressor. *PLoS Biol.* **5**, e292 (2007).
- Qiu, S. A computational study of off-target effects of RNA interference. *Nucleic Acids Res.* **33**, 1834–1847 (2005).
- Wild, G. & Traulsen, A. The different limits of weak selection and the evolutionary dynamics of finite populations. *J. Theor. Biol.* **247**, 382–390 (2007).
- Nogueira, T. et al. Horizontal gene transfer of the secretome drives the evolution of bacterial cooperation and virulence. *Curr. Biol.* **19**, 1683–1691 (2009).
- Mc Ginty, S. E., Lehmann, L., Brown, S. P. & Rankin, D. J. The interplay between relatedness and horizontal gene transfer drives the evolution of plasmid-carried public goods. *Proc. R. Soc. Ser. B* **280**, 20130400–20130400 (2013).
- Rood, E. S. & Freedberg, S. Intragenomic conflict produces sex ratio dynamics that favor maternal sex ratio distorters. *Ecol. Evol.* **6**, 8085–8093 (2016).
- Vaz, S. C. & Carvalho, A. B. Evolution of autosomal suppression of the sex-ratio trait in *Drosophila*. *Genetics* **166**, 265–277 (2004).
- Grafen, A. Optimization of inclusive fitness. *J. Theor. Biol.* **238**, 541–563 (2006).
- Bernardo Carvalho, A., Koerich, L. B. & Clark, A. G. Origin and evolution of Y chromosomes: *Drosophila* tales. *Trends Genet.* **25**, 270–277 (2009).
- Matthews, B. B. et al. Gene model annotations for *Drosophila melanogaster*: impact of high-throughput data. *G3 (Bethesda)* **5**, 1721–1736 (2015).
- Taanman, J. W. The mitochondrial genome: structure, transcription, translation and replication. *Biochim. Biophys. Acta* **1410**, 103–123 (1999).
- Perete, M. et al. CHES: a new human gene catalog curated from thousands of large-scale RNA sequencing experiments reveals extensive transcriptional noise. *Genome Biol.* **19**, 1–14 (2018).
- Zerbino, D. R. et al. Ensembl 2018. *Nucleic Acids Res.* **46**, D754–D761 (2017).
- Eickbush, M. T., Young, J. M. & Zanders, S. E. Killer meiotic drive and dynamic evolution of the wtf gene family. *Mol. Biol. Evol.* **36**, 1201–1214 (2019).
- Lin, C.-J. et al. The hpRNA/RNAi pathway is essential to resolve intragenomic conflict in the *Drosophila* male germline. *Dev. Cell* **46**, 316–326.e5 (2018).
- Tao, Y. et al. A sex-ratio meiotic drive system in *Drosophila simulans*. II: an X-linked distorter. *PLoS Biol.* **5**, e293 (2007).
- Blows, M. W., Berrigan, D. & Gilchrist, G. W. Rapid evolution towards equal sex ratios in a system with heterogamety. *Evol. Ecol. Res.* **1**, 277–283 (1999).
- Aravin, A. A., Hannon, G. J. & Brennecke, J. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* **318**, 761–764 (2007).

64. Doron, S. et al. Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* **359**, eaar4120–13 (2018).
65. Atlan, A., Mercot, H., Landre, C. & Montchamp-Moreau, C. The sex-ratio trait in *Drosophila simulans*: geographical distribution of distortion and resistance. *Evolution* **51**, 1886 (1997).
66. Carvalho, A. B., Sampaio, M. C., Varandas, F. R. & Klaczko, L. B. An experimental demonstration of Fisher's principle: evolution of sexual proportion by natural selection. *Genetics* **148**, 719–731 (1998).
67. Capillon, C. & Atlan, A. Evolution of driving X chromosomes and resistance factors in experimental populations of *Drosophila simulans*. *Evolution* **53**, 506 (1999).
68. Bastide, H. et al. Rapid rise and fall of selfish sex-ratio X chromosomes in *Drosophila simulans*: spatiotemporal analysis of phenotypic and molecular data. *Mol. Biol. Evol.* **28**, 2461–2470 (2011).
69. Hornett, E. A. et al. Evolution of male-killer suppression in a natural population. *PLoS Biol.* **4**, e283–e286 (2006).
70. Maynard Smith, J. et al. Developmental constraints and evolution: a perspective from the mountain lake conference on development and evolution. *Q. Rev. Biol.* **60**, 265–287 (1985).
71. Lehmann, L., Alger, I. & Weibull, J. Does evolution lead to maximizing behavior? *Evolution* **69**, 1858–1873 (2015).
72. Gardner, A. & Grafen, A. Capturing the superorganism: a formal theory of group adaptation. *J. Evol. Biol.* **22**, 659–671 (2009).

Acknowledgements

We thank Geoff Wild, Alan Grafen, Egbert Giles Leigh, Jr., Guy Cooper, Sam Levin, Asher Leeks, Matishalin Patel and Tom Hitchcock for comments on the manuscript; Anna Dewar for sharing data on the ratio of plasmid to chromosomal genes in *E. coli*.

Author contributions

T.W.S. and S.A.W. designed the study and wrote the paper. T.W.S. carried out mathematical analysis.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-019-13169-3>.

Correspondence and requests for materials should be addressed to T.W.S.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019

Methods (additional entries)

The following entries to the Methods section were added following initial thesis examination, and do not feature in the published version of the paper (Scott & West, Nat Commun 10, 5163 (2019)).

Fitness maximisation. We adopt the definition of fitness given by Rousset (2013). Fitness is the number of adult offspring of an adult. In other words, fitness is the number of descendants of an individual after one full iteration of the life cycle. At a long term evolutionary equilibrium, a given individual is ‘fitness maximising’, with respect to a given trait, if the individual adopts a trait value that, relative to the population average, maximises its production of adult offspring.

Intragenomic conflict. We adopt the definition of intragenomic conflict given by Gardner and Úbeda (2017). Intragenomic conflict occurs when the “inclusive fitness interests” of different genes disagree.

Intragenomic conflict and fitness maximisation. We are only concerned with the subset of intragenomic conflicts that compromise individual fitness maximisation (adaptation). To illustrate the type of intragenomic conflict that we are *not* concerned with, imagine the joint phenotype: “drive at meiosis”. This phenotypic change will be favoured by the single gene that has the potential to drive, but, if there is an associated organism-level (fertility) cost, will be opposed by other genes residing elsewhere in the genome.

This example of intragenomic conflict (meiotic drive) does not undermine organism design, because “compensatory mutations” will be selected, across the whole genome, to minimise organism-level consequences of the gene-drive. For instance, in a stalk-eyed fly, males bearing a meiotic driver (SR) suffer from sperm destruction as a direct consequence of drive, but compensate for this by having enlarged testes, with the result that their fertility is not compromised relative to males lacking the driver, even when challenged with fertilising large numbers of females (Meade et al. 2019). Therefore, conflicts over whether a gene should be able to drive or not (which also includes conflicts over gene transposition) are not the focus of our study.

We are concerned with the subset of intra-genomic conflicts that cannot be resolved by compensatory mutations across the genome. This includes conflicts over sex ratio and public goods production. These conflicts pose potential problems for organism design in a way that conflicts over gene-drive do not.

Modifier theory. Modifier theory analyses co-evolutionary dynamics at (i) a primary locus (or primary loci) under selection, and (ii) a secondary (modifier) locus that affects (“modifies”) gene transmission at the primary locus (or loci) in some specified way. Due to the complexity of multi-locus population genetic models, a given model typically considers only one type of modification in isolation. Mathematical treatment has been given, but is not limited, to the following types of modification: mutation rate at the primary locus (Karlin and McGregor 1974); dominance at the primary locus (Fisher 1930); segregation distortion at the primary locus (Eshel 1985; Feldman and

Otto 1991); recombination between primary loci (Nei 1967); migration rate (Balkau and Feldman 1973); life history (Charlesworth 1980); mate choice (Kirkpatrick 1982).

Different types of modifier (e.g. mutation rate modifier; dominance modifier) are governed by different population genetics, and are likely to suffer different evolutionary fates as a result. Despite this, in many different scenarios, it has been found that unlinked modifiers that maximise the population mean fitness are favoured by natural selection (Karlin and McGregor 1974). This result is recovered in our *Equilibrium Model 2*. Specifically, if a suppressor invades a population fixed for a trait distorter, which occurs when $c_{sup} < c_{trait}$, this will result in an increase in population mean fitness.

Modifiers of segregation. Within the field of modifier theory, some general points have been made regarding models of: (i) a primary locus subject to segregation distortion, and (ii) a secondary (modifier) locus that determines the rate of segregation distortion at the primary locus. It has been found that, if a costly meiotic driver is stably maintained at equilibrium alongside a non-driving allele, a unlinked neutral modifier will always invade from rarity if it reduces segregation distortion at the primary locus (Eshel 1985). The result of successive allele invasions at the secondary (modifier) locus is therefore that Mendelian segregation (no segregation distortion) is instated at the primary locus. This result has been used to explain why Mendelian segregation is so exact in general (meiotic drivers are rare or transient) (Eshel 1985; Crow 1991).

In order capture the biological phenomena of organismal adaptation and trait distortion, our model departs from classical modifier theory. Specifically, we assume that a modifier of segregation distortion (i.e. “suppressor”) is costly. We also assume that the fitness cost of meiotic drive is recovered under suppression. These features violate an assumption of the classical models of meiotic drive modification – that there is no direct selection at the secondary (modifier) locus (Eshel 1985; Crow 1991).

Suppressors of trait distortion (our model) are favoured far more readily than suppressors of meiotic drive (classical modifier theory). Suppressors of meiotic drive are only favoured if there is a stable polymorphism (allelic diversity) at the primary locus, which is biologically unlikely, requiring either that: (i) the non-driving allele is recurrently re-introduced by mutation, or (ii) individuals bearing both driving and non-driving alleles (heterozygotes) have an appreciable fitness advantage (heterozygote advantage) (Eshel 1985; Crow 1991). It is more likely that a meiotic driver will reach population fixation at equilibrium, removing selection for suppressor spread (Charlesworth and Hartl 1978; Eshel 1985; Zanders and Unckless 2019). By contrast, suppressors of trait distortion (our model) are favoured even if there is no polymorphism at the primary locus (trait distorter is at fixation), owing to the lingering fitness cost of trait distortion, which persists even after the trait distorter has gone to fixation.

Generalising c_{sup} . We consider the consequences of generalising the functional form for the cost of suppression (c_{sup}). We assume that the cost of suppression (c_{sup})

is monotonically increasing with the strength of the trait distorter being suppressed ($\frac{dc_{sup}}{dk} \geq 0$) and is positive when the strength of the trait distorter is zero, owing to the “overhead costs” of utilising gene-silencing machinery ($c_{sup}(k=0) > 0$). This general functional form for the cost of suppression reduces to our previous (specific) functional form when $\frac{dc_{sup}}{dk} = 0$.

We first ask if the results of the trait distorter-suppressor coevolution model (Equations 2-5) are changed when the cost of suppression is generalised ($\frac{dc_{sup}}{dk} \geq 0; c_{sup}(k = 0) > 0$). The suppressor (S_1) invasion condition ($c_{trait} > c_{sup}$), and the mutant distorter (D_2) invasion condition ($\Delta t(1 - c_{trait}) > \Delta c_{trait}$), are derived without assuming a functional form for c_{sup} , so are unchanged. Furthermore, the following result does not depend on the functional form of c_{sup} , so is unchanged: if a suppressor invades, the trait distorter will be ultimately purged at equilibrium.

Given that these results still hold when c_{sup} is generalised, the shapes of Figures 1 & 4, characterised by a discontinuous relationship between trait distorter strength and organismal trait distortion, will be unchanged. The key point is that a suppressor will spread only once a threshold has been passed, after which, the cost trait distortion exceeds the cost of suppressing trait distortion.

We now ask if the results of the trait distorter mutant model ($D_1/D_2/S_1$; Equations 8-13) are changed when the cost of suppression is generalised ($\frac{dc_{sup}}{dk} \geq 0; c_{sup}(k = 0) > 0$). We find that, if the cost of suppression (c_{sup}) increases sharply

with trait distortion ($\frac{dc_{sup}}{dk} \gg 0$), the invasion and spread of a suppressor to suppress a stronger mutant trait distorter (D_2) does not necessarily lead to the simultaneous suppression of weaker trait distorters (D_1), because the weaker trait distorters (D_1) are now subject to a relatively lower fitness cost of being suppressed ($c_{sup,D1} < c_{sup,D2}$).

As a result of this, selection on trait distorters to increase their strength does not always result (via suppressor spread) in the ultimate prominence of non-distorters (D_0 ; $k=0$). Instead, when $\frac{dc_{sup}}{dk} \gg 0$, selection on trait distorters to increase their strength may result in unsuppressed intermediate-strength (k) trait distorters being ultimately prominent. Therefore, if the cost of suppression (c_{sup}) increases sharply with trait distortion ($\frac{dc_{sup}}{dk} \gg 0$), trait distortion can be more pronounced than previously realised. However, equilibrium trait distortion is not *appreciable*, because high-strength (k) trait distorters still trigger suppression, even when c_{sup} is generalised.

References (for additional Methods entries)

Balkau, B. J., and M. W. Feldman. 1973. Selection for Migration Modification.

Genetics 74:171–174. Genetics Society of America.

Charlesworth, B. 1980. Evolution in Age-Structured Populations. Cambridge University Press.

Charlesworth, B., and D. L. Hartl. 1978. Population Dynamics of the Segregation

- Distorter Polymorphism of *Drosophila Melanogaster*. *Genetics* 89:171–192.
- Crow, J. F. 1991. Why is mendelian segregation so exact? *Bioessays* 13:305–312.
- Eshel, I. 1985. Evolutionary Genetic Stability of Mendelian Segregation and the Role of Free Recombination in the Chromosomal System. *The American Naturalist* 125:412–420.
- Feldman, M. W., and S. P. Otto. 1991. A Comparative Approach to the Population-Genetics Theory of Segregation Distortion. *The American Naturalist* 137:443–456.
- Fisher, R. A. 1930. *The Genetical Theory of Natural Selection*. Oxford University Press.
- Gardner, A., and F. Úbeda. 2017. The meaning of intragenomic conflict. *Nature Ecology & Evolution* 1:1–9. Springer US.
- Karlin, S., and J. McGregor. 1974. Towards a Theory of Evolution of Modifier Genes. *Theoretical Population Biology* 5:59–103.
- Kirkpatrick, M. 1982. Sexual Selection and the Evolution of Female Choice. *Evolution* 36:1–12.
- Meade, L., S. Finnegan, R. Kad, K. Fowler, and A. Pomiankowski. 2019. Maintenance of fertility in the face of meiotic drive. *The American Naturalist* 707372–49.
- Nei, M. 1967. Modification of Linkage Intensity by Natural Selection. *Genetics* 57:625–&. Genetics Society of America.

Rousset, F. 2013. Genetic Structure and Selection in Subdivided Populations (MPB-40). Princeton University Press.

Zanders, S. E., and R. L. Unckless. 2019. Fertility Costs of Meiotic Drivers. *Current Biology* 29:R512–R520. Elsevier Ltd.

Supplementary Information

**Adaptation is maintained by the
parliament of genes**

Scott, T.W., & West, S.A.

Supplementary Notes:

- (1) Functional Forms Assumed in Figure 2b (Main Text) p64
- (2) Equilibrium Models (Additional Figures) p65
- (3) Sex Ratio Distortion p67
- (4) Genomic Imprinting and Altruism p91
- (5) Horizontal Gene Transfer and Public Goods p103
- (6) Suppressor Conditionality p111
- (7) Dynamics Models (Additional Figures) p114
- (8) Real-World Estimates of Proportional Cabal Size p121

Supplementary Discussion:

- (1) Relation to Gardner & Úbeda (2017) and Grafen's Formal Darwinism p123
Individual fitness maximisation: emergence versus maintenance
- (2) Simple Selfish Genetic Elements vs Trait Distorters p126
Relation to Eshel (1984) and Eshel (1985)
- (3) Relation to Cosmides & Tooby (1981): Coreplicons, Cabal & Commonwealth p131

Supplementary References p137

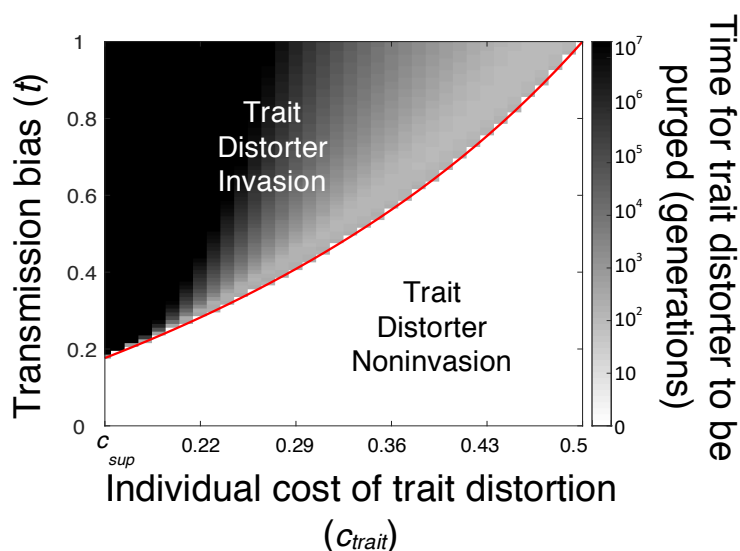
Supplementary Note 1

Functional Forms Assumed in Figure 2b (Main Text)

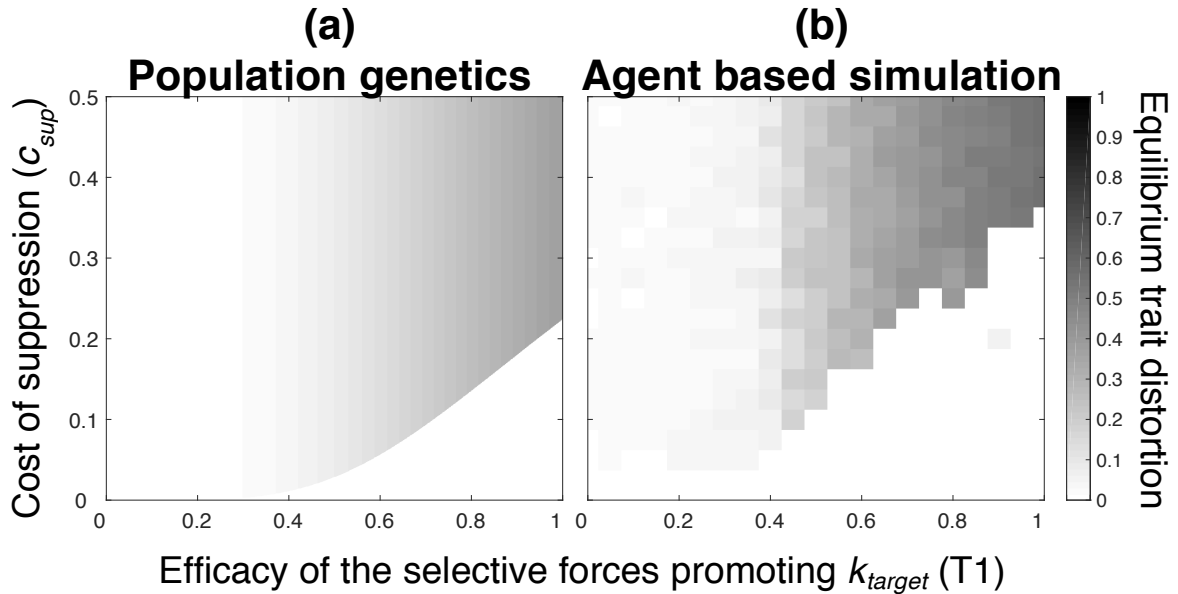
To generate Figure 2b (Main Text), we assumed the functional forms $c_{trait}=0.8k^{1.3}$ and $t=T1k$, where $0\leq T1\leq 1$. The parameter T1 mediates the rate at which the marginal transmission advantage dissipates, relative to the marginal individual cost of trait distortion, as the trait becomes increasingly distorted (k). This parameter (T1) is plotted along the x axis of Figure 2b (Main Text). We analytically derived the target trait distortion (k_{target}), for different values of c_{sup} and T1, by substituting our specific functional forms into the condition that specifies k_{target} : $\frac{dt}{dk}(1 - c_{trait}) = \frac{dc_{trait}}{dk}$. This gave: $\frac{d(T1k)}{dk}(1 - 0.8k^{1.3}) = \frac{d(0.8k^{1.3})}{dk}$, which simplifies to $T1(1-0.8k^{1.3})=1.04k^{0.3}$, which was solved for k to give $k = k_{target}$. We derived the equilibrium trait distortion (k^*) by substituting each value of k_{target} into the condition for suppressor spread: $c_{sup} < c_{trait}(k_{target})$. Satisfaction of this condition implies that $k^*=0$; else, $k^*=k_{target}$.

Supplementary Note 2

Equilibrium Models (Additional Figures)



Supplementary Figure 1. Non-equilibrium trait distortion. A trait distorter (D_1) is introduced from rarity alongside its suppressor (S_1) (initial genotype frequencies: $x_{00}=0.97$, $\{x_{01}, x_{10}, x_{11}\}=0.01$). The trait distorter (D_1) is associated with some individual cost when expressed (c_{trait}) which is varied along the x axis (the cost of suppression is fixed at $c_{sup}=0.15$). The trait distorter (D_1) is also associated with a transmission bias at meiosis (t) which is varied along the y axis. We consider trait distorters that induce suppressor spread ($c_{sup} < c_{trait}$) and ask whether such trait distorters can cause appreciable trait distortion before they are ultimately suppressed and purged from the population. The red line plots the formula $t = c_{trait} / (1 - c_{trait})$; above this line, trait distorters can spread from rarity. We plot the number of generations (on a \log_{10} scale) until equilibrium is reached (trials that did not equilibrate by 20,000,000 generations were capped). We see that less costly trait distorters (c_{trait} only slightly greater than c_{sup}) can invade even with a relatively low transmission bias (t), and are purged at a very slow rate, causing extended non-equilibrium trait distortion. More costly trait distorters (c_{trait} large compared to c_{sup}) require a high transmission bias (t) to invade, and if they can invade, they are purged relatively quickly, causing shorter non-equilibrium trait distortion. Therefore, non-equilibrium trait distortion is either not-so-costly and extended, or costly and ephemeral, and so has limited impact on individual fitness maximisation in either case.



Supplementary Figure 2. Comparison of population genetic and agent based simulation

results of the illustrative model. Trait distorter strength evolves in the presence of a suppressor of distortion (S_T) and the resulting equilibrium trait distortion ($0 \leq k^* \leq 1$) is plotted, using the following functional forms for trait distorter cost (c_{trait}) and transmission advantage (t): $c_{trait}(k) = 0.8k^{1.3}$ and $t(k) = T1k$, where $0 \leq T1 \leq 1$, and where $T1$ and c_{sup} are varied. $T1$ mediates the rate at which the marginal transmission advantage ($\frac{dt}{dk}$) dissipates, relative to the marginal individual cost ($\frac{dc_{trait}}{dk}$) of trait distortion, as the trait becomes increasingly distorted (k). Part (a) plots the equilibrium trait distortion under weak selection, as analytically derived from the population genetic model in the main text. Part (b) plots the equilibrium trait distortion as obtained by the agent based simulation detailed in *Methods: Agent-based simulation (single trait distorter locus)*, which does not assume weak selection, and which allows continuous variation at the trait and suppressor loci. There is good correspondence between the models; however, the region in which trait distorters are suppressed ($k^* = 0$; bottom right) is larger in the agent based simulation, and outside of this region, the equilibrium trait distortion (k^*) is slightly greater in the agent based simulation. Discrepancy arises from the different assumptions about the strength of selection.

Supplementary Note 3

Sex Ratio Distortion

We examine sex ratio evolution in a diploid species, in a large outbreeding (panmictic) population, with non-overlapping generations, and where males and females are equally costly to produce. Fisher¹ and many others have shown that, in this scenario, individuals would be selected to invest equally in male and female offspring^{2,3}. We assume genetic sex determination, with males as XY and females as XX⁴.

We consider a selfish genetic element residing on an X chromosome, that may gain a propagation advantage by distorting the offspring sex ratio towards a greater production of females. The genes that do not gain a propagation advantage from female sex ratio bias reside on both the autosomes and the Y chromosome⁵. We focus on suppressors in the autosomes, for simplicity, and because this is the larger group of genes, constituting the majority within the parliament of genes⁶. Consequently, we focus our analyses on when an X driver and an autosomal suppressor can spread.

Our overall aim is to assess, given the potential for suppression, the extent that an X chromosome driver can distort the sex ratio away from the individual optimum. The individual optimum is taken to be the evolutionarily stable strategy (ESS) adopted by individuals in the absence of selfish trait distortion, which is an equal investment in offspring of both sexes.

We build our model in a step-wise manner, as described in the “Equilibrium Models” section of the main text. Aspects of questions 1-3 have been analysed before with respect to sex ratio, but we go over them here for the specific case of our model, and to elucidate the underlying selective forces. There is available data on the fitness consequences of sex ratio distortion and suppression, and so, in this case, we aim for a biologically realistic model that can be parameterised.

(1) Spread of a Trait distorter

We considered the spread, in the absence of suppression, of a selfish sex ratio distorter that skews offspring sex ratio towards females. In the literature, selfish X drivers are often denoted by *SR* (for *sex-ratio* distortion), with non-distorting rival alleles denoted by *ST* (for *standard*)⁷. However, we denote the trait distorting and non-distorting alleles respectively by D_1 and D_0 for consistency across our models. We assume that normal (D_0/Y) males produce X and Y sperm equally. The trait distorter (D_1) causes D_1/Y males to kill Y-bearing sperm, leading to a female-biased sex ratio⁸⁻¹³. In males with an unsuppressed trait distorter, its proportion of X-bearing sperm, and correspondingly, the proportion of its offspring that are female, is given by $0.5(1+k)$, where k denotes the proportion of Y-bearing sperm that are killed ($0 < k \leq 1$).

We assume that males with an unsuppressed trait distorter (D_1) suffer a fertility cost as a result of sperm death, and have a reduced ejaculate size of $1-k/2$, relative to 1 in all other males¹⁴⁻¹⁹. We assume that, each generation, each female copulates with λ random males, and that each sperm cell is equally competitive in the female’s

internal store. The likelihood of a male's sperm fertilising an egg (*fertility, F*) is given by his ejaculate size relative to the total amount of ejaculate that female has received. Letting *l* be the proportion of males in the present generation with an unsuppressed trait distorter, the fertility of those males with an unsuppressed trait distorter (F_{drive}), and the fertility of those without (F_{normal}), is given by:

$$F_{drive} = \sum_{i=1}^{\lambda} \frac{\left(\frac{1-k}{2}\right)^i}{\left(\frac{1-k}{2}\right)^{i+(\lambda-i)}} l^{i-1} (1-l)^{\lambda-i}, \quad (1)$$

$$F_{normal} = \sum_{i=1}^{\lambda} \frac{1}{\left(\frac{1-k}{2}\right)^{(i-1)+(\lambda-i+1)}} l^{i-1} (1-l)^{\lambda-i}. \quad (2)$$

There is no sperm competition, and therefore no fertility cost of sex ratio distortion, when females are singly mated ($F_{normal} = F_{drive}$ when $\lambda=1$). There is increased sperm competition at higher female mating rates, meaning the relative fertility cost of sex ratio distortion (F_{normal}/F_{drive}) increases and plateaus for high λ at $F_{normal}/F_{drive} = 1/(1-k/2)$.

The trait distorter has no fitness consequences for females, and so the condition for the spread of the trait distorter (D_1) allele is that D_1/Y males sire more female offspring than D_0/Y males. In the absence of suppression, D_1/Y males have $F_{drive}(1+k)/2$ female offspring, and D_0/Y males have $F_{normal}/2$ female offspring, meaning the trait distorter (D_1) is selected when:

$$F_{normal} / F_{drive} < (1+k). \quad (3)$$

The left-hand side of Supplementary Equation 3 gives the between-individual relative fertility cost of trait distortion and the right-hand side gives the within-individual relative transmission advantage of trait distortion. When we substitute our explicit fertility functions (Supplementary Equations 1 & 2) into Supplementary Equation 3, we find that Supplementary Equation 3 is always satisfied. Consequently, analogous to previous arguments, the distorting D_1 chromosome will always spread to fixation, irrespective of female mating frequency (λ)^{20,21}. This distorts the offspring sex ratio, defined as the proportion of females, to $(1+k)/2$.

Previous models have relaxed some of our simplifying assumptions, allowing a fixed (mating rate (λ)-independent) cost of distortion, and allowing the female mating rate (λ) to change as the trait distorter spreads^{5,8,14,17,22-27}. We have explored these factors and found that our general conclusions: (i) are not altered; and (ii) do not depend on the trait distorter (D_1) allele spreading all the way to fixation (Scott, unpublished).

(2) Spread of an autosomal suppressor

We assume that the sex ratio distorter can be suppressed by an autosomal allele (suppressor), as has been found in many *Drosophila* species²⁸⁻³¹. We base our model upon the biology of *Nmy*, which suppresses the X chromosome trait distorter *Dox* in *Drosophila simulans*. *Nmy* works by RNAi-mediated destruction of the trait distorter's mRNA transcripts. *Nmy* is dominant, and only expressed in the presence of the trait distorter (*Dox*)³²⁻³⁴.

For consistency across models, we denote the autosomal suppressor allele as S_1 , and the wild type non-suppressor allele as S_0 . We assume that the suppressor (S_1) is dominant, meaning individuals bearing at least one suppressor (S_1) allele suffer no sperm death and consequentially no fertility loss or sex ratio distortion. We assume that the suppressor (S_1) is only expressed in the presence of an active trait distorter (in D_1/Y males). When the suppressor is expressed it leads to a cost, which reduces the probability (V) that an individual survives from zygote to adult³⁵⁻³⁸, from $V_{normal}=1$ in individuals without an active suppressor, to $V_{suppression}=1-c_{sup}$ in individuals with one. The cost of suppression is a fixed cost (c_{sup}) of activating an RNAi pathway. Assuming alternatively that the suppression cost affects fertility rather than viability does not qualitatively change our results (Scott, unpublished).

We ask when an autosomal suppressor (S_1) will spread from rarity, given that an X chromosome trait distorter (D_1) is at fixation. Given that the suppressor only has phenotypic effects in D_1/Y males, it will spread from rarity if D_1/Y males bearing a suppressor (S_1) have more mated offspring than D_1/Y males lacking a suppressor (S_0/S_0). Assuming that the trait distorter and non-suppressor alleles are at fixation, and random mating, D_1/Y males with a suppressor will have

$V_{suppression} * F_{normal} * (1/2) * ((1+k) / 2)$ mated female offspring, and $V_{suppression} * F_{normal} * (1/2) * ((1-k)/2)$ mated male offspring, leading to a total of $V_{suppression} * F_{normal} * (1/4)$ mated offspring. D_1/Y males lacking a suppressor will have a total of $2 * V_{normal} * F_{drive} * ((1-k)/2) * ((1+k)/2)$ mated offspring. Suppressed D_1/Y males will therefore have more offspring, and the suppressor allele (S_1) will spread from rarity, when the following condition is satisfied:

$$(F_{normal}/F_{drive}) * (1/(1-k^2)) > (V_{normal}/V_{suppression}). \quad (4)$$

The overall cost of letting the trait distorter (D_1) go unsuppressed is a product of the costs to fertility (F_{normal}/F_{drive}) and offspring mating success ($1/(1-k^2)$). For a suppressor to spread, this must be greater than the viability cost of suppression ($V_{normal}/V_{suppression}$). Consequently, analogous to previous results, the suppressor (S_1) will only spread when the trait distorter (D_1) leads to appreciable trait distortion^{36,39-42}.

A previous model asked whether female-biased sex ratio distortion can select for compensatory evolution on autosomes, such that the autosomes evolve to encode a male-biased sex ratio in the absence of the trait distorter⁴³. It found that compensatory evolution does not evolve when the female-biased sex ratio distorter is transmitted into female offspring with 100% certainty, as is the case for X drivers acting in males. This is why we did not allow compensatory strategies to evolve on autosomes in our model, and only allowed autosomes to suppress the trait distorter.

(3) Consequences for organism trait values

We turn to the question of how trait distorter-suppressor dynamics affect sex ratio.

When both the trait distorter (D_1) and suppressor (S_1) are in a population, the genotypes they are in matters (epistasis), and so we explicitly track the frequencies of all 15 possible genotypes, with 15 recursions. The 15 equations represent the generational changes in each of the 15 possible genotypes. We let p_{fi} and q_{mi} be the proportion of the i th female genotype and the i th male genotype, respectively, in the

current generation (Supplementary Table 1). We let p_{fi} and q_{mi} be the frequencies of female and male genotypes in the next generation. The population sex ratio is given by the population proportion of females, $\sum p_{fi}$. The equations are listed in Supplementary Table 2. We note that, in the absence of the trait distorter (D_1), population sex ratio evolves to 0.5, and after this, genotype frequencies remain constant over time (Hardy-Weinberg equilibrium).

Supplementary Table 1: Selection coefficients, drive values, and genotype frequency notation.

For each male and female genotype, its proportion in the population at generation t , and its probability of maturing from a zygote to an adult (viability, V) is given. For each male genotype, the proportion of X chromosomes in its sperm store (drive), and its probability of successfully fertilising the female's egg cell after copulation (fertility, F), is given. Male fertility (F) depends on the number of mates each female has per generation (λ), and is written in full in Supplementary Equations 1 & 2. k gives the proportion of a male's Y bearing sperm that are killed, and c_{sup} gives the viability cost of trait distorter suppression.

		Females			Males	
		D_0/D_0	D_0/D_1	D_1/D_1	D_0/Y	D_1/Y
S_1 / S_1	Proportion	p_{f1}	p_{f4}	p_{f7}	p_{m1}	p_{m4}
	Fertility, F	/	/	/	F_{normal}	F_{normal}
	Viability, V	1	1	1	1	$1 - c_{sup}$
	Drive	/	/	/	0.5	0.5
S_1 / S_0	Proportion	p_{f2}	p_{f5}	p_{f8}	p_{m2}	p_{m5}
	Fertility, F	/	/	/	F_{normal}	F_{normal}
	Viability, V	1	1	1	1	$1 - c_{sup}$
	Drive	/	/	/	0.5	0.5
S_0 / S_0	Proportion	p_{f3}	p_{f6}	p_{f9}	p_{m3}	p_{m6}
	Fertility, F	/	/	/	F_{normal}	F_{drive}
	Viability, V	1	1	1	1	1
	Drive	/	/	/	0.5	$(1+k)/2$

Supplementary Table 2: Recursions detailing the change in proportion of each genotype

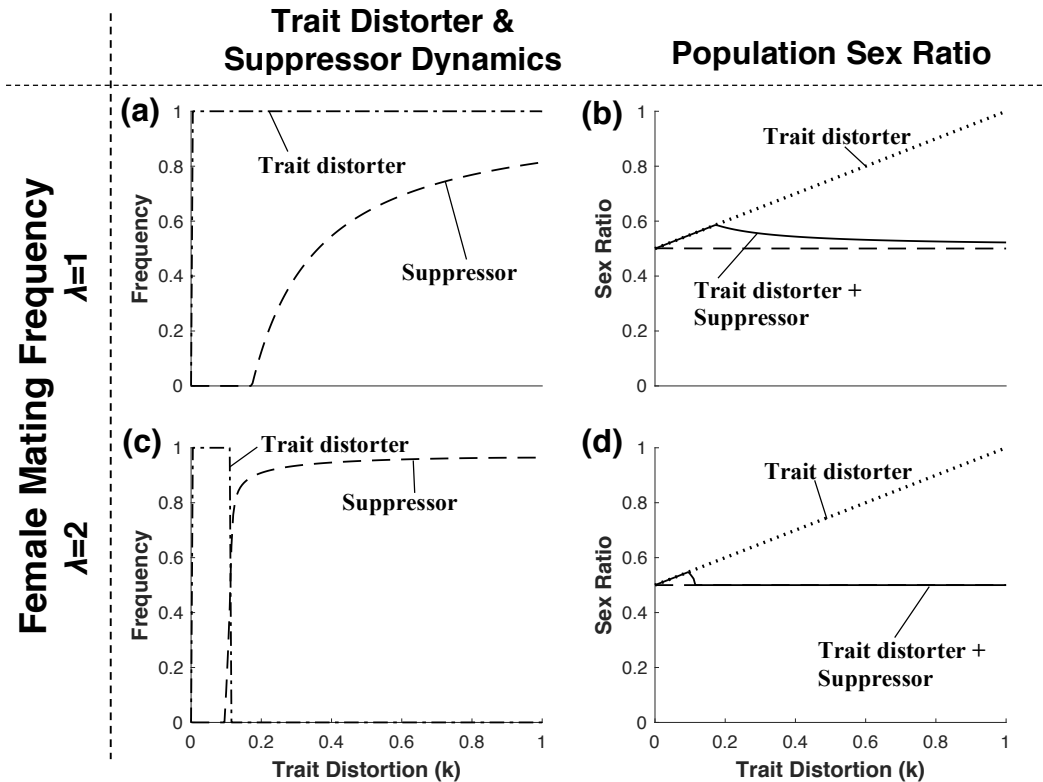
across one generation (D_0 and D_1 segregating at trait locus). Notation is defined in

Supplementary Table 1. T is the sum of the right sides of the system of equations such that $\sum p=1$. It normalises the recursions to ensure that gene frequency changes reflect proportions.

$T p_{f1}' =$	$(p_{f1} + 0.5 p_{f2} + 0.5 p_{f4} + 0.25 p_{f5}) (0.5 p_{m1} + 0.25 p_{m2}) F_{normal}$
$T p_{f2}' =$	$((0.5 p_{f2} + p_{f3} + 0.25 p_{f5} + 0.5 p_{f6}) (0.5 p_{m1} + 0.25 p_{m2}) + (p_{f1} + 0.5 p_{f2} + 0.5 p_{f4} + 0.25 p_{f5}) (0.25 p_{m2} + 0.5 p_{m3})) F_{normal}$
$T p_{f3}' =$	$(0.5 p_{f2} + p_{f3} + 0.25 p_{f5} + 0.5 p_{f6}) (0.25 p_{m2} + 0.5 p_{m3}) F_{normal}$
$T p_{f4}' =$	$(0.5 p_{f4} + 0.25 p_{f5} + p_{f7} + 0.5 p_{f8}) (0.5 p_{m1} + 0.25 p_{m2}) + (p_{f1} + 0.5 p_{f2} + 0.5 p_{f4} + 0.25 p_{f5}) (p_{m4}/2 + p_{m5}/4) F_{normal}$
$T p_{f5}' =$	$((0.25 p_{f5} + 0.5 p_{f6} + 0.5 p_{f8} + p_{f9}) (0.5 p_{m1} + 0.25 p_{m2}) + (0.5 p_{f4} + 0.25 p_{f5} + p_{f7} + 0.5 p_{f8}) (0.25 p_{m2} + 0.5 p_{m3})) F_{normal} + (p_{f1} + 0.5 p_{f2} + 0.5 p_{f4} + 0.25 p_{f5}) (1/2 (1 + k) F_{drive} p_{m6} + 1/4 p_{m5} F_{normal}) + (0.5 p_{f2} + p_{f3} + 0.25 p_{f5} + 0.5 p_{f6}) (p_{m4}/2 + p_{m5}/4) F_{normal}$
$T p_{f6}' =$	$(0.25 p_{f5} + 0.5 p_{f6} + 0.5 p_{f8} + p_{f9}) (0.25 p_{m2} + 0.5 p_{m3}) F_{normal} + (0.5 p_{f2} + p_{f3} + 0.25 p_{f5} + 0.5 p_{f6}) (1/2 (1 + k) F_{drive} p_{m6} + 1/4 p_{m5} F_{normal})$
$T p_{f7}' =$	$(0.5 p_{f4} + 0.25 p_{f5} + p_{f7} + 0.5 p_{f8}) (p_{m4}/2 + p_{m5}/4) F_{normal}$
$T p_{f8}' =$	$((0.5 p_{f4} + 0.25 p_{f5} + p_{f7} + 0.5 p_{f8}) (1/2 (1 + k) F_{drive} p_{m6} + 1/4 p_{m5} F_{normal}) + (0.25 p_{f5} + 0.5 p_{f6} + 0.5 p_{f8} + p_{f9}) (p_{m4}/2 + p_{m5}/4) F_{normal})$
$T p_{f9}' =$	$(0.25 p_{f5} + 0.5 p_{f6} + 0.5 p_{f8} + p_{f9}) (1/2 (1 + k) F_{drive} p_{m6} + 1/4 p_{m5} F_{normal})$
$T p_{m1}' =$	$(p_{f1} + 0.5 p_{f2} + 0.5 p_{f4} + 0.25 p_{f5}) ((0.5 p_{m1} + 0.25 p_{m2}) F_{normal} + 1/2 p_{m4} F_{normal} + 1/4 p_{m5} F_{normal})$
$T p_{m2}' =$	$(p_{f1} + 0.5 p_{f2} + 0.5 p_{f4} + 0.25 p_{f5}) ((0.25 p_{m2} + 0.5 p_{m3}) F_{normal} + 1/2 (1 - k) F_{drive} p_{m6} + 1/4 p_{m5} F_{normal}) + (0.5 p_{f2} + p_{f3} + 0.25 p_{f5} + 0.5 p_{f6}) ((0.5 p_{m1} + 0.25 p_{m2}) F_{normal} + 1/2 p_{m4} F_{normal} + 1/4 p_{m5} F_{normal})$
$T p_{m3}' =$	$(0.5 p_{f2} + p_{f3} + 0.25 p_{f5} + 0.5 p_{f6}) ((0.25 p_{m2} + 0.5 p_{m3}) F_{normal} + 1/2 (1 - k) F_{drive} p_{m6} + 1/4 p_{m5} F_{normal})$
$T p_{m4}' =$	$V_{suppression} (0.5 p_{f4} + 0.25 p_{f5} + p_{f7} + 0.5 p_{f8}) ((0.5 p_{m1} + 0.25 p_{m2}) F_{normal} + 1/2 p_{m4} F_{normal} + 1/4 p_{m5} F_{normal})$
$T p_{m5}' =$	$V_{suppression} ((0.5 p_{f4} + 0.25 p_{f5} + p_{f7} + 0.5 p_{f8}) (0.25 F_{normal} p_{m2} + 0.5 F_{normal} p_{m3} + 1/2 (1 - k) F_{drive} p_{m6} + 1/4 p_{m5} F_{normal}) + (0.25 p_{f5} + 0.5 p_{f6} + 0.5 p_{f8} + p_{f9}) (0.5 p_{m1} + 0.25 p_{m2} + 1/2 p_{m4} F_{normal} + 1/4 p_{m5} F_{normal}))$
$T p_{m6}' =$	$(0.25 p_{f5} + 0.5 p_{f6} + 0.5 p_{f8} + p_{f9}) ((0.25 p_{m2} + 0.5 p_{m3}) F_{normal} + 1/2 (1 - k) F_{drive} p_{m6} + 1/4 p_{m5} F_{normal})$

To illustrate the logic of the equations, we derive one recursion explicitly. We derive the recursion for p_{m5}' , which gives the frequency, in the next generation, of males bearing the trait distorter (D_1/Y) and one suppressor allele (S_1/S_0). We denote the

current generation as G1 and the next generation as G2. A mating between D_1/Y , S_0/S_0 males (at frequency p_{m6}) and D_1/D_1 , S_0/S_1 females (at frequency p_{f8}) can give rise to individuals in G2 with our focal genotype D_1/Y , S_1/S_0 . Of all matings between males and females in G1 ($\sum_{i=1}^6 \sum_{j=1}^9 p_{mi} * p_{fj}$), these matings occur in the proportion $(p_{m6} * p_{f8}) / (\sum_{i=1}^6 \sum_{j=1}^9 p_{mi} * p_{fj})$ of cases. Copulation success, in which the egg is successfully fertilised to form a zygote, depends on the fertility of the male, which in our case is F_{drive} (Supplementary Equation 1). Of all the zygotes produced by the population in G1, our parents will contribute the proportion $(p_{m6} * p_{f8}) / (\sum_{i=1}^6 \sum_{j=1}^9 p_{mi} * p_{fj}) * (F_{drive} / \sum_{i=1}^6 p_{mi} F_{mi})$ of them, where $\sum_{i=1}^6 p_{mi} F_{mi}$ gives average male fertility. These zygotes will have the focal offspring genotype (D_1/Y , S_1/S_0) if they inherit a YS_0 gamete from the father (with probability $(1-k)/2$) and a D_1S_1 gamete from the mother (with probability $1/2$), meaning the proportion of zygotes in the population with the focal genotype, (D_1/Y , S_1/S_0), stemming from copulations between our focal parents, is $(p_{m6} * p_{f8}) / (\sum_{i=1}^6 \sum_{j=1}^9 p_{mi} * p_{fj}) * (F_{drive} / \sum_{i=1}^6 p_{mi} F_{mi}) * (1+k)/4$. Finally, only the proportion $V_{suppression}$ of these focal zygotes (D_1/Y , S_1/S_0) will successfully mature to adulthood in G2, meaning the proportion of mature adults in G2 that have the focal genotype (D_1/Y , S_1/S_0) and arose from copulations between D_1/Y , S_0/S_0 males and D_1/D_1 , S_0/S_1 females in G1 is given by $(p_{m6} * p_{f8}) / (\sum_{i=1}^6 \sum_{j=1}^9 p_{mi} * p_{fj}) * (F_{drive} / \sum_{i=1}^6 p_{mi} F_{mi}) * (1+k)/4 * (V_{suppression} / \sum p^* V)$, where $\sum p^* V$ is the average viability of individuals. We simplify the expression by making the substitution $T = (\sum_{i=1}^6 \sum_{j=1}^9 p_{mi} * p_{fj}) (\sum_{i=1}^6 p_{mi} F_{mi}) (\sum p^* V)$, our normalisation factor. We now need to sum over all possible parental copulations that can give rise to D_1/Y , S_0/S_1 offspring. Doing so gives the frequency of the D_1/Y , S_0/S_1 genotype in the next generation, p_{m5}' , written in full in Supplementary Table 2.



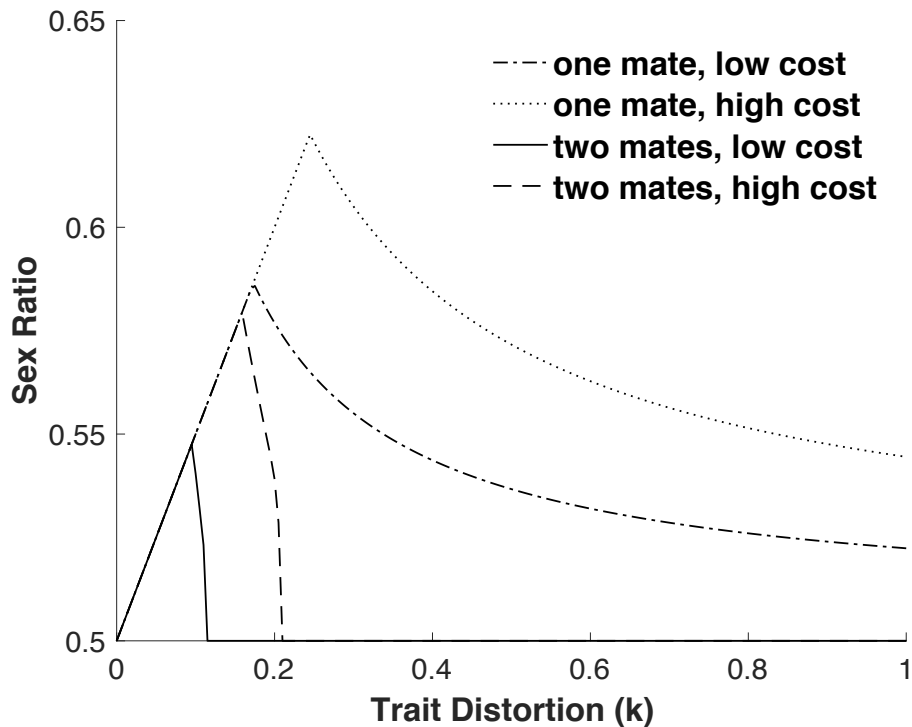
Supplementary Figure 3. Drive-suppressor coevolution and resulting sex ratio. A sex ratio distorter (D_1) and its suppressor (S_1) are introduced from rarity. The left-hand column shows the resulting equilibrium trait distorter (D_1 ; dot-dash line) and suppressor (S_1 ; dashed line) frequencies, for different trait distorter strengths ($0 < k \leq 1$). The right-hand column shows the resulting population average sex ratio. The dotted and dashed lines are plotted for reference, to show, respectively, the sex ratio $((1+k)/2)$ that arises in the absence of suppression, and the sex ratio $(1/2)$ that arises in the absence of the trait distorter (D_1). The top row shows results for when females are singly mated ($\lambda=1$) and the bottom row shows results for when females are doubly mated ($\lambda=2$). The numerical solutions assume that the cost of suppression is $c_{\text{sup}}=0.03$. We see that equilibrium suppressor (S_1) frequency is greater for strong (higher- k) drivers, resulting in full ($\lambda=2$) or partial ($\lambda=1$) restoration of the individual optimal sex ratio (0.5) for strong (higher- k) drivers.

We iterated these recursions to find the trait distorter (D_1) and suppressor (S_1) frequencies, and the population sex ratio ($\sum p_i$), at equilibrium (Supplementary Figure 3). When we introduced both the trait distorter and suppressor at low frequencies,

we confirmed our above results that the trait distorter (D_1) initially spreads to fixation, and that the suppressor allele (S_1) only invades and reaches high frequencies if it is suppressing a strong trait distorter (high k).

We used our recursions to examine whether the spread of the suppressor led to the subsequent loss of the trait distorter. As the suppressor increases in frequency, the population sex ratio becomes less biased, and the fitness benefit of further trait distorter suppression is reduced (negative frequency dependence). This means that, when females are singly mated ($\lambda=1$), the rise of the suppressor allele towards some nonzero equilibrium frequency does not cause subsequent loss of the trait distorter (D_1), which remains at fixation (Supplementary Figure 3a). When females are multiply mated ($\lambda>1$), there is an additional fertility cost of distortion, and so the suppressor continues to spread, to a higher equilibrium frequency, until the trait distorter (D_1) is lost completely from the population (Supplementary Figure 3c).

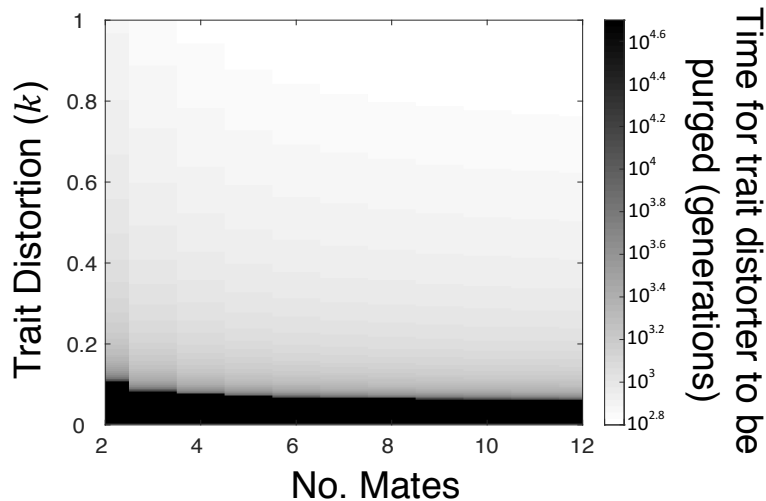
We also considered the overall consequences of the trait distorter-suppressor dynamics for the sex ratio. For weak trait distorters (low k), suppressors do not spread. Consequently, there is sex ratio distortion, but it is negligible. For trait distorters of intermediate strength (intermediate k , e.g. $k \approx 0.2$), suppressors are still at low population frequency, and so there can be greater sex ratio distortion. For strong trait distorters (high k), suppressors spread to high population frequency, and so there is little (when $\lambda=1$) or no (when $\lambda>1$) sex ratio distortion. Consequently, the extent that the sex ratio deviates from the individual optimum of equal investment in the sexes: (a) shows a domed relationship with the extent of distortion (k); and (b)



Supplementary Figure 4. Mating frequency and cost of suppression. A sex ratio distorter (D_1) and its suppressor (S_1) are introduced from rarity. Equilibrium sex ratio is plotted, across multiple trials where D_1 has different levels of drive ($0 < k \leq 1$), and where different assumptions are made about the cost of suppression (c_{sup}) and female mating rate (λ). The four parameter regimes plotted assume single ($\lambda=1$) or double ($\lambda=2$) female mating; and, low ($c_{sup}=0.03$) or high ($c_{sup}=0.06$) viability cost of trait distorter suppression. The sex ratio is more easily distorted when the suppressor is costlier (c_{sup}) and when female mating rate is lower (λ).

will often be negligible⁴¹ (Supplementary Figure 3b & Supplementary Figure 3d). It should be noted that, in reality, the population is a mixture of two types of individual, one adopting a sex ratio of $\frac{1}{2}$ and the other adopting a distorted sex ratio of $(1+k)/2$, and here we are capturing the population average deviation of individuals from the optimal sex ratio.

The case of singly mated females ($\lambda=1$) is of special interest because there is no fertility cost of trait distortion ($F_{normal}=F_{drive}$), meaning the individual level cost of



Supplementary Figure 5. Non-equilibrium sex ratio distortion. An X driver (D_1) is introduced from rarity alongside its suppressor (S_1). The proportion of Y-bearing sperm killed by the X driver (k) is varied alongside number of times each female mates per generation (λ). The cost of suppression is fixed at $c_{sup}=0.03$. We consider trait distorters that are purged from the population, after being suppressed, at equilibrium ($\lambda>1$). We plot the number of generations (on a \log_{10} scale) until equilibrium is reached (trials that did not equilibrate by 50,000 generations were capped). We see that stronger trait distorters (high k) are purged at a faster rate, reducing the potential for non-equilibrium sex ratio distortion. Increased female mating (high λ) increases the fertility cost of distortion, meaning trait distorters are purged at a faster rate.

bearing the selfish genetic element (D_1) arises *solely* because an individual level trait (sex ratio) is suboptimal (not $\frac{1}{2}$). Sex ratio distortion is often negligible even in this special case ($\lambda=1$), indicating that the parliament of genes can act for the sole purpose of trait (sex ratio) restoration, without the additional incentive of fertility recovery.

Additionally, we considered the effects of model parameters on sex ratio. We found that increasing the rate of female mating (λ) and decreasing the cost of suppression

(c_{sup}) both led to a reduced tolerance of drive, and a correspondingly reduced level of sex ratio distortion (Supplementary Figure 4).

Finally, we considered how far sex ratio can be distorted in the time period after the trait distorter initially invades and before the trait distorter is suppressed and purged from the population. We iterated our recursions and timed how many generations it took to reach equilibrium. We found that stronger trait distorters (higher k) are suppressed and purged from the population more quickly, especially at higher female mating rates (λ) where the fertility cost of sex ratio distortion is greater (Supplementary Figure 5).

4) Evolution of trait distortion

In the above analyses, we assumed that the strength of the trait distorter (D_1) was a fixed parameter (k). We now consider the consequence of allowing the level of trait distortion to evolve²⁰. We first consider the scenario in which there is no suppressor. We take a game theoretical approach to find the evolutionarily stable strength of X chromosome sex ratio distortion (k^*) in the absence of suppression. We assume a population where all males have an X chromosome with the same strength of distortion, denoted by a capital K . We then assume that a mutation arises in the X chromosome of one male in the population that causes it to assume a new strength of distortion, denoted by \check{k} . We wish to find the strength of distortion that, when adopted by every X chromosome in the population, cannot be invaded by the mutant X chromosome adopting a different strength of distortion. This strength of distortion (k^*) represents the evolutionarily stable strategy (ESS)⁴⁴.

Trait distorters have no effect in females, so the fitness of the mutant trait distorter depends only on its action in males. The male bearing the mutant trait distorter has fertility given by its proportional sperm contribution to a female mate's sperm store:

$\left(\frac{(1-\check{k}/2)}{(1-\check{k}/2)+(\lambda-1)(1-K/2)} \right)$. The mutant trait distorter is passed into $(1+\check{k})/2$ offspring, and

so the fitness of the mutant X chromosome is proportional to: $w =$

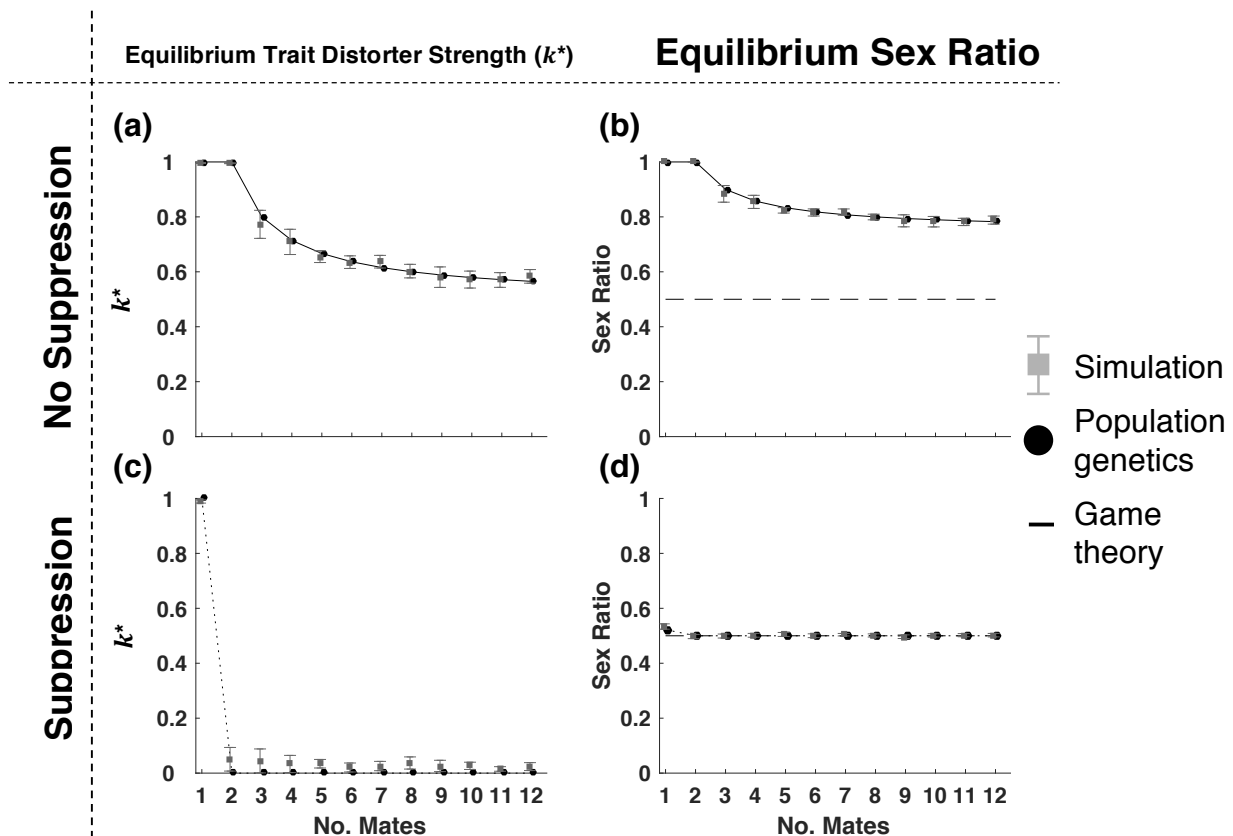
$\frac{(1+\check{k})}{2} \left(\frac{(1-\check{k}/2)}{(1-\check{k}/2)+(\lambda-1)(1-K/2)} \right)$. The ESS strength of X chromosome distortion is the

value of k^* that satisfies $\frac{dw}{d\check{k}} = 0$ and $\frac{d^2w}{d\check{k}^2} < 0$ when $\check{k}=K=k^*$, and is given by:

$$k^*=(\lambda+1)/(2\lambda-1). \tag{5}$$

When females mate singly ($\lambda=1$) or doubly ($\lambda=2$), maximal trait distorter strength is favoured ($k^*=1$), resulting in population collapse due to lack of males. As female mating frequency increases to $\lambda \geq 3$, the increased fertility cost of distortion means that the equilibrium strength of distortion (k^*) decreases²⁰, until it plateaus at the minimum of $k^*=0.5$ as $\lambda \rightarrow \infty$ (Supplementary Figure 6a). The game theoretic equilibrium is verified in fully dynamical population genetic and agent-based simulation models, as described below (Supplementary Figure 6a).

We now consider what sex ratio will evolve in the presence of a suppressor (S_1). We assume that a mutant X chromosome trait distorter (D_2) arises from a mutation on the old sex ratio distorter (D_1), and kills a different proportion of sperm when unsuppressed ($\hat{k} \neq k$), biasing individual sex ratio by $(1+\hat{k})/2$ and reducing ejaculate



Supplementary Figure 6. Equilibrium trait distorter strength and sex ratio. (a) In the absence of a suppressor (S_T), equilibrium trait distorter strength (k^*) decreases with the number of mates each female has per generation (λ). (b) Trait distorters causes more significantly distorted equilibrium sex ratio at lower female mating rates. A dashed line shows the sex ratio that would evolve in the absence of selfish genetic elements ($1/2$). (c) In the presence of suppression, maximally distorting (but largely suppressed) trait distorters ($k^*=1$) evolve when females are singly mated ($\lambda=1$); otherwise, non-trait distorters ($k^*=0$) evolve. (d) Owing to the spread of suppressors, sex ratio is completely ($\lambda>1$) or partially ($\lambda=1$) recovered at equilibrium. (c) and (d) assumed a small cost of suppression ($c_{sup}=0.03$). For all graphs, the results of the simulation (grey boxes) are plotted alongside the population genetics result (black circles). For (a) and (b), the result of a game theoretic analysis is also plotted (solid line). All methods give the same equilibrium trait distorter strength and sex ratio. The error bars show one standard deviation from the mean over 10 trials of the simulation.

size to $1-\hat{k}/2$. We assume that D_2 and D_1 share a similar genetic and mechanistic basis of drive, such that the mutant distorting X chromosome (D_2) is suppressed by

the same suppressor allele (S_1)⁴⁵⁻⁴⁸. In Supplementary Table 4, we display 27 recursions to describe the generational changes in genotype frequencies when the alleles D_1 , D_0 , D_2 , Y , S_0 and S_1 are segregating in a population (notation defined in Supplementary Tables 1 & 3). We note that, in the absence of the trait distorters (D_1 and D_2), population sex ratio evolves to 0.5, and after this, genotype frequencies remain constant over time (Hardy-Weinberg equilibrium). These equations reduce to those in Supplementary Table 2 when genotypes bearing D_2 are set to zero.

We work out the evolved level of sex ratio distortion, under the assumption that trait distorter strength is initially low, and the additional assumption of weak selection. We assume the mutant trait distorter (D_2) is only slightly stronger than the trait distorter from which it is derived (D_1), so that $\hat{k}=k+\delta$, where δ is positive and very small (*δ -weak selection*⁴⁹). We see if a mutant trait distorter can spread by iterating our recursions in Supplementary Table 4 until equilibrium is reached. If the stronger trait distorter (D_2) displaces the weaker one (D_1), we introduce a further mutant trait distorter and iterate our equations again. We elucidate the equilibrium trait distorter strength (k^*) by successively introducing mutant trait distorters (D_2) until one fails to invade, at which point the equilibrium strength (k^*) has been reached.

We find that, in the presence of the suppressor allele (S_1), weakly distorting X chromosomes (low- k) can evade suppression and successfully distort sex ratio. These weak trait distorters will be displaced by slightly more distorting mutants (D_2). If the cost of suppression (c_{sup}) is sufficiently low, this displacement causes the frequency of the suppressor allele (S_1) to increase in response. This trend means

Supplementary Table 3: Further Selection coefficients, drive values, and genotype frequency

notation. For each male and female genotype, its proportion in the population at generation t , and its probability of maturing from a zygote to an adult (viability, V) is given. For each male genotype, the proportion of X chromosomes in its sperm store (drive), and its probability of successfully fertilising the female's egg cell after copulation (fertility, F), is given. Male fertility (F) depends on the number of mates each female has per generation (λ) according to:

$$F_{normal} = \sum_{i=0}^{\lambda-1} \sum_{j=0}^{\lambda-i-1} \frac{1}{1 + (1-\frac{k}{2})^{i+j+(\lambda-i-j-1)}(1-\frac{\hat{k}}{2})} l^i (1-l-n)^j n^{\lambda-i-j-1},$$

$$F_{drive} = \sum_{i=0}^{\lambda-1} \sum_{j=0}^{\lambda-i-1} \frac{1-\frac{k}{2}}{(1-\frac{k}{2})^{i+j+(\lambda-i-j-1)}(1-\frac{\hat{k}}{2})} l(t)^i (1-l-n)^j n^{\lambda-i-j-1},$$

$$F_{driveMut} = \sum_{i=0}^{\lambda-1} \sum_{j=0}^{\lambda-i-1} \frac{1-\frac{\hat{k}}{2}}{(1-\frac{k}{2})^{i+j+(\lambda-i-j)}(1-\frac{\hat{k}}{2})} l^i (1-l-n)^j n^{\lambda-i-j-1}.$$

l , n , and $1-l-n$, are, respectively, the proportions of males in the population with: an unsuppressed D_1 ; an unsuppressed D_2 ; neither of these (all other males). k and \hat{k} respectively give the proportion of a male's Y bearing sperm that are killed by an unsuppressed D_1 and D_2 trait distorter, and c_{sup} gives the viability cost of trait distorter suppression.

		Females		Males	
		D_0/D_2	D_1/D_2	D_2/D_2	D_2/Y
S_1 / S_1	Proportion	p_{f10}	p_{f13}	p_{f16}	p_{m7}
	Fertility, F	/	/	/	F_{normal}
	Viability, V	1	1	1	$1-c_{sup}$
	Drive	/	/	/	0.5
S_1 / S_0	Proportion	p_{f11}	p_{f14}	p_{f17}	p_{m8}
	Fertility, F	/	/	/	F_{normal}
	Viability, V	1	1	1	$1-c_{sup}$
	Drive	/	/	/	0.5
S_0 / S_0	Proportion	p_{f12}	P_{f15}	p_{f18}	p_{m9}
	Fertility, F	/	/	/	$F_{driveMut}$
	Viability, V	1	1	1	1
	Drive	/	/	/	$(1+\hat{k})/2$

Supplementary Table 4: Recursions detailing the change in proportion of each genotype

across one generation (D_0 , D_1 and D_2 segregating at trait locus). Notation is defined in

Supplementary Table 1 & S3. T is the sum of the right sides of the system of equations such that

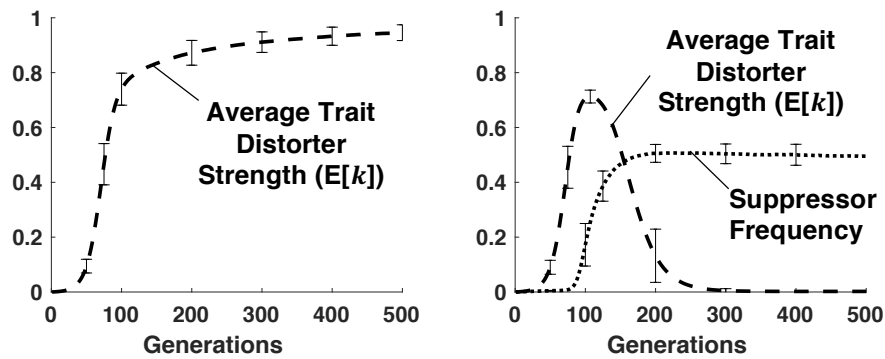
$\sum p=1$. It normalises the recursions to ensure that gene frequency changes reflect proportions.

$T p_{f1}' =$	$(p_{f1} + 0.5 p_{f10} + 0.25 p_{f11} + 0.5 p_{f12} + 0.5 p_{f14} + 0.25 p_{f15}) (0.5 p_{m1} + 0.25 p_{m2}) F_{normal}$
$T p_{f2}' =$	$((0.25 p_{f11} + 0.5 p_{f12} + 0.5 p_{f12} + p_{f3} + 0.25 p_{f5} + 0.5 p_{f6}) (0.5 p_{m1} + 0.25 p_{m2}) + (p_{f1} + 0.5 p_{f10} + 0.25 p_{f11} + 0.5 p_{f12} + 0.5 p_{f14} + 0.25 p_{f15}) (0.25 p_{m2} + 0.5 p_{m3})) F_{normal}$
$T p_{f3}' =$	$(0.25 p_{f11} + 0.5 p_{f12} + 0.5 p_{f12} + p_{f3} + 0.25 p_{f5} + 0.5 p_{f6}) (0.25 p_{m2} + 0.5 p_{m3}) F_{normal}$
$T p_{f4}' =$	$(0.5 p_{f13} + 0.25 p_{f14} + 0.5 p_{f14} + 0.25 p_{f5} + p_{f7} + 0.5 p_{f8}) (0.5 p_{m1} + 0.25 p_{m2}) F_{normal} + 1/4 (p_{f1} + 0.5 p_{f10} + 0.25 p_{f11} + 0.5 p_{f12} + 0.5 p_{f14} + 0.25 p_{f15}) (2 p_{m4} F_{drive} + p_{m5} F_{normal})$
$T p_{f5}' =$	$(0.25 p_{f14} + 0.5 p_{f15} + 0.25 p_{f5} + 0.5 p_{f6} + 0.5 p_{f8} + p_{f9}) (0.5 p_{m1} + 0.25 p_{m2}) F_{normal} + (0.5 p_{f13} + 0.25 p_{f14} + 0.5 p_{f14} + 0.25 p_{f5} + p_{f7} + 0.5 p_{f8}) (0.25 p_{m2} + 0.5 p_{m3}) F_{normal} + 1/4 (0.25 p_{f11} + 0.5 p_{f12} + 0.5 p_{f12} + p_{f3} + 0.25 p_{f5} + 0.5 p_{f6}) (2 p_{m4} F_{drive} + p_{m5} F_{drive}) + 1/4 (p_{f1} + 0.5 p_{f10} + 0.25 p_{f11} + 0.5 p_{f12} + 0.5 p_{f14} + 0.25 p_{f15}) (-2 (1 + k) F_{drive} p_{m6} + p_{m5} F_{normal})$
$T p_{f6}' =$	$(0.25 p_{f14} + 0.5 p_{f15} + 0.25 p_{f5} + 0.5 p_{f6} + 0.5 p_{f8} + p_{f9}) (0.25 p_{m2} + 0.5 p_{m3}) F_{normal} + 1/4 (0.25 p_{f11} + 0.5 p_{f12} + 0.5 p_{f12} + p_{f3} + 0.25 p_{f5} + 0.5 p_{f6}) (-2 (1 + k) F_{drive} p_{m6} + p_{m5} F_{normal})$
$T p_{f7}' =$	$1/4 (0.5 p_{f13} + 0.25 p_{f14} + 0.5 p_{f14} + 0.25 p_{f5} + p_{f7} + 0.5 p_{f8}) (2 p_{m4} F_{drive} + p_{m5} F_{normal})$
$T p_{f8}' =$	$1/4 ((0.25 p_{f14} + 0.5 p_{f15} + 0.25 p_{f5} + 0.5 p_{f6} + 0.5 p_{f8} + p_{f9}) (2 p_{m4} F_{drive} + p_{m5} F_{normal}) + (0.5 p_{f13} + 0.25 p_{f14} + 0.5 p_{f14} + 0.25 p_{f5} + p_{f7} + 0.5 p_{f8}) (-2 (1 + k) F_{drive} p_{m6} + p_{m5} F_{normal}))$
$T p_{f9}' =$	$1/4 (0.25 p_{f14} + 0.5 p_{f15} + 0.25 p_{f5} + 0.5 p_{f6} + 0.5 p_{f8} + p_{f9}) (-2 (1 + k) F_{drive} p_{m6} + p_{m5} F_{normal})$
$T p_{f10}' =$	$(0.5 p_{f10} + 0.25 p_{f11} + 0.5 p_{f13} + 0.25 p_{f14} + p_{f16} + 0.5 p_{f17}) (0.5 p_{m1} + 0.25 p_{m2}) F_{normal} - 1/4 (p_{f1} + 0.5 p_{f10} + 0.25 p_{f11} + 0.5 p_{f12} + 0.5 p_{f14} + 0.25 p_{f15}) (-2 p_{m7} - p_{m8}) F_{driveMut}$
$T p_{f11}' =$	$((0.25 p_{f11} + 0.5 p_{f12} + 0.25 p_{f14} + 0.5 p_{f15} + 0.5 p_{f17} + p_{f18}) (0.5 p_{m1} + 0.25 p_{m2}) + (0.5 p_{f10} + 0.25 p_{f11} + 0.5 p_{f13} + 0.25 p_{f14} + p_{f16} + 0.5 p_{f17}) (0.25 p_{m2} + 0.5 p_{m3})) F_{normal}$
$T p_{f12}' =$	$(0.25 p_{f11} + 0.5 p_{f12} + 0.25 p_{f14} + 0.5 p_{f15} + 0.5 p_{f17} + p_{f18}) (0.25 p_{m2} + 0.5 p_{m3}) F_{normal} + 1/4 (0.25 p_{f11} + 0.5 p_{f12} + 0.5 p_{f12} + p_{f3} + 0.25 p_{f5} + 0.5 p_{f6}) (-2 (1 + \hat{k}) F_{driveMut} p_{m9} + p_{m8} F_{driveMut})$
$T p_{f13}' =$	$1/4 ((0.5 p_{f10} + 0.25 p_{f11} + 0.5 p_{f13} + 0.25 p_{f14} + p_{f16} + 0.5 p_{f17}) (2 p_{m4} F_{drive} + p_{m5} F_{normal}) - (0.5 p_{f13} + 0.25 p_{f14} + 0.5 p_{f14} + 0.25 p_{f5} + p_{f7} + 0.5 p_{f8}) (-2 p_{m7} - p_{m8}) F_{driveMut})$
$T p_{f14}' =$	$1/4 ((0.25 p_{f11} + 0.5 p_{f12} + 0.25 p_{f14} + 0.5 p_{f15} + 0.5 p_{f17} + p_{f18}) (2 p_{m4} F_{drive} + p_{m5} F_{normal}) + (0.5 p_{f10} + 0.25 p_{f11} + 0.5 p_{f13} + 0.25 p_{f14} + p_{f16} + 0.5 p_{f17}) (-2 (1 + k) F_{drive} p_{m6} + p_{m5} F_{normal}) + (0.5 p_{f13} + 0.25 p_{f14} + 0.5 p_{f14} + 0.25 p_{f5} + p_{f7} + 0.5 p_{f8}) (-2 (1 + \hat{k}) F_{driveMut} p_{m9} + p_{m8} F_{driveMut}) - (0.25 p_{f14} + 0.5 p_{f15} + 0.25 p_{f5} + 0.5 p_{f6} + 0.5 p_{f8} + p_{f9}) (-2 p_{m7} - p_{m8}) F_{driveMut})$
$T p_{f15}' =$	$1/4 ((0.25 p_{f11} + 0.5 p_{f12} + 0.25 p_{f14} + 0.5 p_{f15} + 0.5 p_{f17} + p_{f18}) (-2 (1 + k) F_{drive} p_{m6} + p_{m5} F_{normal}) + (0.25 p_{f14} + 0.5 p_{f15} + 0.25 p_{f5} + 0.5 p_{f6} + 0.5 p_{f8} + p_{f9}) (-2 (1 + \hat{k}) F_{driveMut} p_{m9} + p_{m8} F_{driveMut}))$
$T p_{f16}' =$	$1/4 (0.5 p_{f10} + 0.25 p_{f11} + 0.5 p_{f13} + 0.25 p_{f14} + p_{f16} + 0.5 p_{f17}) (-2 p_{m7} - p_{m8}) F_{driveMut}$
$T p_{f17}' =$	$1/4 ((0.5 p_{f10} + 0.25 p_{f11} + 0.5 p_{f13} + 0.25 p_{f14} + p_{f16} + 0.5 p_{f17}) (-2 (1 + \hat{k}) F_{driveMut} p_{m9} + p_{m8} F_{driveMut}) - (0.25 p_{f11} + 0.5 p_{f12} + 0.25 p_{f14} + 0.5 p_{f15} + 0.5 p_{f17} + p_{f18}) (-2 p_{m7} - p_{m8}) F_{driveMut})$
$T p_{f18}' =$	$1/4 (0.25 p_{f11} + 0.5 p_{f12} + 0.25 p_{f14} + 0.5 p_{f15} + 0.5 p_{f17} + p_{f18}) (-2 (1 + \hat{k}) F_{driveMut} p_{m9} + p_{m8} F_{driveMut})$
$T p_{m1}' =$	$1/4 (p_{f1} + 0.5 p_{f10} + 0.25 p_{f11} + 0.5 p_{f12} + 0.5 p_{f14} + 0.25 p_{f15}) ((2 p_{m1} + p_{m2}) F_{normal} - 2 F_{drive} p_{m4} + p_{m5} F_{normal} + 2 p_{m7} F_{driveMut} + p_{m8} F_{driveMut})$
$T p_{m2}' =$	$1/4 ((p_{f1} + 0.5 p_{f10} + 0.25 p_{f11} + 0.5 p_{f12} + 0.5 p_{f14} + 0.25 p_{f15}) ((p_{m2} + 2 p_{m3}) F_{normal} + 2 (-1 + k) F_{drive} p_{m6} + 2 (-1 + \hat{k}) F_{driveMut} p_{m9} + p_{m5} F_{normal} + p_{m8} F_{driveMut}) + (0.25 p_{f11} +$

	$0.5 p_{f12} + 0.5 p_{f2} + p_{f3} + 0.25 p_{f5} + 0.5 p_{f6} ((2 p_{m1} + p_{m2}) F_{normal} - 2 F_{drive} p_{m4} + p_{m5} F_{normal} + 2 p_{m7} F_{driveMut} + p_{m8} F_{driveMut}))$
$T p_{m3}' =$	$1/4 (0.25 p_{f11} + 0.5 p_{f12} + 0.5 p_{f2} + p_{f3} + 0.25 p_{f5} + 0.5 p_{f6} ((p_{m2} + 2 p_{m3}) F_{normal} + 2 (-1 + k) F_{drive} p_{m6} + 2 (-1 + \hat{k}) F_{driveMut} p_{m9} + p_{m5} F_{normal} + p_{m8} F_{driveMut}))$
$T p_{m4}' =$	$1/4 V_{suppression} (0.5 p_{f13} + 0.25 p_{f14} + 0.5 p_{f4} + 0.25 p_{f5} + p_{f7} + 0.5 p_{f8} ((2 p_{m1} + p_{m2}) F_{normal} - 2 F_{drive} p_{m4} + p_{m5} F_{normal} + 2 p_{m7} F_{driveMut} + p_{m8} F_{driveMut}))$
$T p_{m5}' =$	$1/4 V_{suppression} ((0.5 p_{f13} + 0.25 p_{f14} + 0.5 p_{f4} + 0.25 p_{f5} + p_{f7} + 0.5 p_{f8} ((p_{m2} + 2 p_{m3}) F_{normal} + 2 (-1 + k) F_{drive} p_{m6} + 2 (-1 + \hat{k}) F_{driveMut} p_{m9} + p_{m5} F_{normal} + p_{m8} F_{driveMut})) + (0.25 p_{f14} + 0.5 p_{f15} + 0.25 p_{f5} + 0.5 p_{f6} + 0.5 p_{f8} + p_{f9} ((2 p_{m1} + p_{m2}) F_{normal} - 2 F_{drive} p_{m4} + p_{m5} F_{normal} + 2 p_{m7} F_{driveMut} + p_{m8} F_{driveMut}))$
$T p_{m6}' =$	$1/4 (0.25 p_{f14} + 0.5 p_{f15} + 0.25 p_{f5} + 0.5 p_{f6} + 0.5 p_{f8} + p_{f9} ((p_{m2} + 2 p_{m3}) F_{normal} + 2 (-1 + k) F_{drive} p_{m6} + 2 (-1 + \hat{k}) F_{driveMut} p_{m9} + p_{m5} F_{normal} + p_{m8} F_{driveMut}))$
$T p_{m7}' =$	$1/4 V_{suppression} (0.5 p_{f10} + 0.25 p_{f11} + 0.5 p_{f13} + 0.25 p_{f14} + p_{f16} + 0.5 p_{f17} ((2 p_{m1} + p_{m2}) F_{normal} - 2 F_{drive} p_{m4} + p_{m5} F_{normal} + 2 p_{m7} F_{driveMut} + p_{m8} F_{driveMut}))$
$T p_{m8}' =$	$1/4 V_{suppression} ((0.5 p_{f10} + 0.25 p_{f11} + 0.5 p_{f13} + 0.25 p_{f14} + p_{f16} + 0.5 p_{f17} ((p_{m2} + 2 p_{m3}) F_{normal} + 2 (-1 + k) F_{drive} p_{m6} + 2 (-1 + \hat{k}) F_{driveMut} p_{m9} + p_{m5} F_{normal} + p_{m8} F_{driveMut})) + (0.25 p_{f11} + 0.5 p_{f12} + 0.25 p_{f14} + 0.5 p_{f15} + 0.5 p_{f17} + p_{f18} ((2 p_{m1} + p_{m2}) F_{normal} - 2 F_{drive} p_{m4} + p_{m5} F_{normal} + 2 p_{m7} F_{driveMut} + p_{m8} F_{driveMut}))$
$T p_{m9}' =$	$1/4 (0.25 p_{f11} + 0.5 p_{f12} + 0.25 p_{f14} + 0.5 p_{f15} + 0.5 p_{f17} + p_{f18} ((p_{m2} + 2 p_{m3}) F_{normal} + 2 (-1 + k) F_{drive} p_{m6} + 2 (-1 + \hat{k}) F_{driveMut} p_{m9} + p_{m5} F_{normal} + p_{m8} F_{driveMut}))$

that, given sequential mutations on the X chromosome to increase sex ratio distortion, suppression will ultimately evolve, completely ($\lambda > 1$) or partially ($\lambda = 1$) restoring an equal sex ratio (Figure 3d & Supplementary Figure 6c & Supplementary Figure 6d). Consequently, we conclude that, with reasonable assumptions about the cost of suppression (c_{sup}), trait distorter suppression is the ultimate outcome of trait distorter evolution. For the sex ratio to be appreciably distorted (>60% females), suppression cost needs to exceed around $c_{sup} = 0.15$ ($\lambda = 1$) or $c_{sup} = 0.35$ ($\lambda > 1$) (Supplementary Figure 8).

By setting the frequencies of all genotypes bearing the suppressor (S_i) to zero, we can use our recursions in Supplementary Table 4 to find the equilibrium strength of trait distortion in the absence of suppression. We exactly recover the equilibrium derived in Supplementary Equation 5, which gives the ESS strength of X

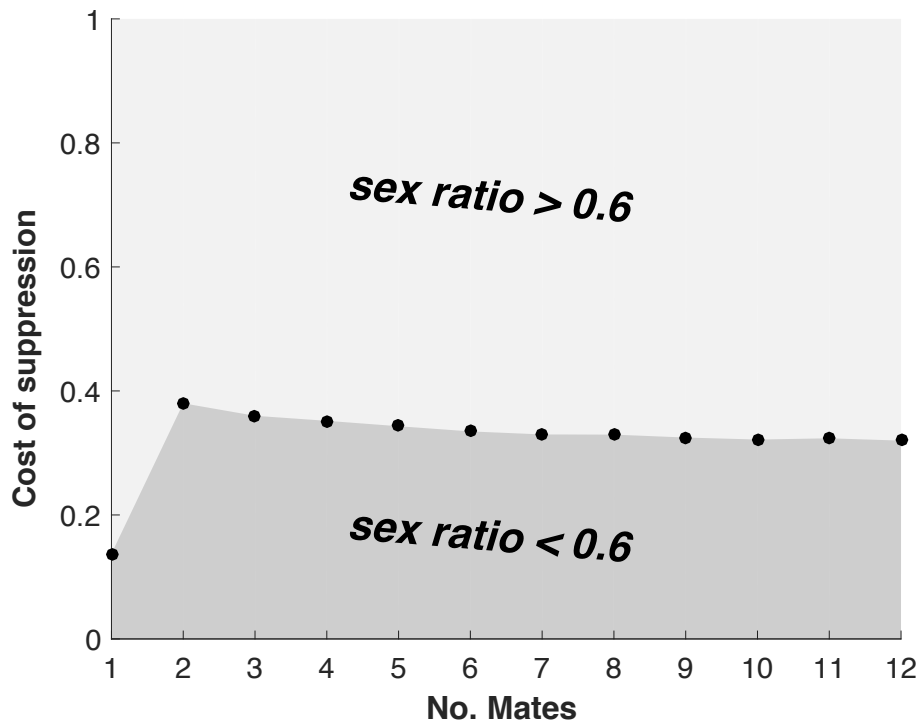


Supplementary Figure 7: Evolution of sex ratio distortion (dynamics). This figure plots the results of the agent-based simulation model, in which X chromosome drive can mutate and take any value within 0-1. In (a), there is no suppressor allele (S_1), and the population average level of X chromosome drive ($E[k]$) tends towards maximum strength ($k^*=1$). In (b), a suppressor of distortion (S_1) is introduced from rarity. The population average strength of X chromosome distortion ($E[k]$) increases alongside the suppressor (S_1) frequency. Eventually, a threshold is passed, after which, distorting X chromosomes ($k_a, k_b > 0$) are lost from the population. Sex ratio is restored to 0.5 at equilibrium. Double female mating ($\lambda=2$) and high suppression cost ($c_{sup}=0.3$) were assumed in these simulations. The plots show average values over 100 runs, for $N=100,000$ individuals, with error bars plotting one standard deviation in each direction of the mean.

chromosome trait distortion (k^*) in the absence of suppression (Supplementary Figure 6a).

Agent-based simulation

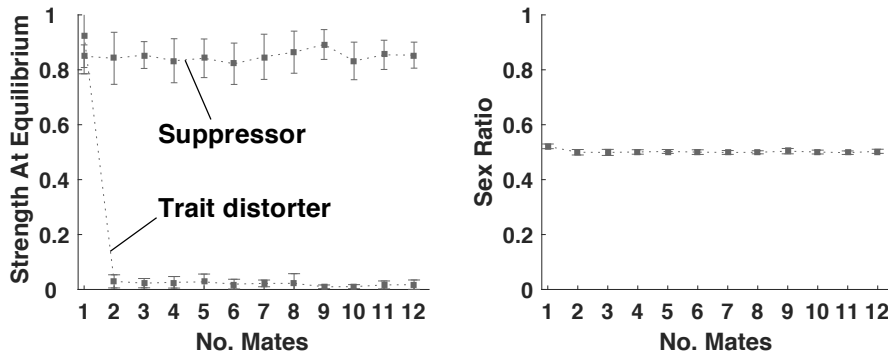
We construct an agent-based simulation to ask what level of sex ratio distortion evolves under strong selection, and when continuous variation is permitted at trait distorter and suppressor loci. We model a population of $N=10,000$ individuals and track evolution at an X chromosome trait distorter locus and an autosomal suppressor locus. Individuals either have two alleles at the X chromosome



Supplementary Figure 8. Cost of suppression required for appreciable sex ratio distortion. The equilibrium strength of X chromosome distortion was obtained for different costs of suppression (c_{sup}). The sex ratio at this equilibrium level of X chromosome was recorded. The cost of suppression required for sex ratio distortion to be appreciably distorted (>60% females produced) is plotted (solid circles). Sex ratio is significantly distorted for the region above this curve.

locus, with strengths denoted by k_a and k_b (females), or one allele at the X chromosome locus, with strength denoted by k_a (males), and one Y chromosome.

Each allele at the trait distorter locus can take any continuous value between zero and one. Individuals have two alleles at the suppressor locus, with strengths denoted by m_a and m_b (diploid). At the suppressor locus, we consider both the case of: (i) discrete variation, in which suppressor strengths are either zero or one, and (ii) continuous variation, in which suppressor strengths can take any continuous value between zero and one. We assume that the strongest (highest value)



Supplementary Figure 9. Equilibrium trait distorter strength and sex ratio under continuous

suppressor variation. (a) Equilibrium suppressor strength ($E[m^*]$) evolves to be high (~ 0.83).

Maximally distorting (but largely suppressed) X chromosomes ($E[k^*]=1$) evolve when females are singly mated ($\lambda=1$); otherwise, non-distorting X chromosomes ($E[k^*]=0$) evolve. (b) Owing to the evolution of strong suppression, sex ratio is completely ($\lambda>1$) or partially ($\lambda=1$) recovered at equilibrium. Suppression cost in these simulations was $c_{sup}=0.03$. Error bars show one standard deviation from the mean over 10 trials.

suppressor allele within an individual is dominant. The ejaculate size of a given male

is $1 - \frac{(1 - \max(m_a, m_b))k_a}{2}$, and his fertility is this value divided by the total sperm stored

in the females it mates with. The total sperm store is the sum of ejaculates of this

male and λ other males drawn at random from the population with replacement. The

viability of males with an active trait distorter ($k_a>0$) is given by $1 - \max(m_a, m_b)c_{sup}$.

The viability of all other individuals is 1.

Each generation, there are N breeding pairs. Females are drawn at random with

replacement to fill each female position in each breeding pair. Males are drawn from

the population, with replacement, with probabilities given by their fertility. Breeding

pairs then reproduce to produce one offspring, before dying (non-overlapping

generations). Alleles at the suppressor locus are inherited in Mendelian fashion.

Alleles at the trait distorter locus in males may drive, meaning the X chromosome is inherited, rather than the Y chromosome, with the probability $(1+k_a(1-\max(m_a, m_b)))/2$. Offspring then compete for spots in the adult population, of which there are N . To fill each spot, offspring are drawn with replacement with likelihood that is proportional to their viability. Each generation, alleles at the trait distorter locus, and for the case of continuous variation at the suppressor locus, alleles at the suppressor locus have a 0.0005 chance of mutating to a new value, which is drawn from a normal distribution centred around the pre-mutation value, with variance 0.5, and truncated between 0 and 1 (strong selection). For the case of discrete variation at the suppressor locus, alleles at the suppressor locus have a 0.001 chance of mutating each generation between suppressor and non-suppressor states.

We iterate this lifecycle over 5,000 generations. We see that, when discrete variation is permitted at the suppressor locus, the simulation quantitatively recovers the equilibrium level of distortion, and corresponding sex ratio, given by the game theoretic and population genetic models (Supplementary Figure 6). When continuous variation is permitted at the suppressor locus, qualitatively equivalent results are obtained: suppressor strength ($E[m]$) evolves to be high enough that sex ratio is fully ($\lambda \geq 2$) or partially ($\lambda = 1$) recovered at equilibrium (Supplementary Figure 9).

Supplementary Note 4

Genomic Imprinting and Altruism

Genomic imprinting occurs at a minority of genes in mammals and flowering plants. An imprinted allele has different epigenetic marks, and corresponding expression levels, when maternally and paternally inherited⁵⁰. We examined the evolution of an altruistic helping behaviour in a population capable of genomic imprinting. A behaviour is altruistic if it incurs a cost (c) to perform, by the actor, and provides a benefit (b) to another individual, the recipient. Altruism is favoured if the genetic relatedness (R) between the actor and recipient is sufficiently high, such that $Rb > c$ ⁵¹.

An individual may be more closely related to their social partners via their maternal or paternal genes⁵²⁻⁵⁴. For example, if a female mates two males, then on average her offspring would be related by $R_m = 1/2$ at maternal genes and $R_p = 1/4$ at paternal genes. If genes can 'gain information' about where they came from, by imprinting, then they could be selected to adjust traits accordingly. Assume that relatedness to social partners is R_p and R_m at paternal and maternal genes respectively. In this case, altruistic helping would be favoured at: maternally imprinted genes when $R_m b > c$; paternally imprinted genes when $R_p b > c$; and unimprinted genes when $((R_p + R_m)/2)b > c$ ⁵⁴⁻⁵⁶. Consequently, if $R_m b > c > ((R_p + R_m)/2)b$, then altruistic helping is favoured at maternally imprinted genes, when it is disfavoured at unimprinted genes (selfish trait distortion).

We consider an autosomal, maternally expressed selfish genetic element that may gain a propagation advantage by upregulating individual altruistic investment^{55,57-61}.

The genes that do not gain a propagation advantage from altruism upregulation comprise both paternally expressed and unimprinted genes. The conflict between maternally and paternally expressed genes, which can result in arms races and a ‘tug of war’ over organism phenotype, has been considered in previous theoretical work⁶²⁻⁶⁵. However, we focus on unimprinted suppressors, for simplicity, and because unimprinted genes comprise the larger group of genes, constituting the majority within the parliament of genes^{50,66,67}. We focus our analyses on when a maternally expressed trait distorter and an unimprinted suppressor can spread. We first describe our modelling assumptions, then successively analyse the cases of unimprinted, and imprinted, altruism. The purpose of this model is to illustrate how selection will act on selfish imprinted genes and their suppressors.

Modelling Assumptions

We track a large population of diploid individuals. We consider a gene that induces an altruistic investment of some amount ($k > 0$), at a fitness cost to the individual ($c(k)$), which is a monotonically increasing function of altruistic investment ($\frac{\partial c}{\partial k} \geq 0$), and a benefit to the social partner ($b(k) > c(k)$), which is a function that is monotonically increasing with the level of altruistic investment ($\frac{\partial b}{\partial k} \geq 0$) yet diminishing with the *cost* of altruistic investment ($\frac{d^2 b}{dc^2} \leq 0$)⁶⁸.

Each generation, male gametes (e.g. sperm) fuse at random with female gametes (e.g. eggs) to generate individuals (random mating). Individuals then pair up with other individuals who have matching maternally (egg-) inherited alleles at all loci;

pairs are random with respect to identity at the paternally (sperm-) inherited allele at all loci ($R_m=1$; $R_p=0$). Individuals may then invest in altruism directed towards their partner, before producing gametes in proportion to their fitness (fertility), and dying (non-overlapping generations).

In nature, relatedness asymmetries within a generation may be generated by sex biased migration patterns⁵⁸, or as a consequence of greater variance in reproductive success in males^{52,69}. They may alternatively be generated if kin recognition alleles are imprinted, which has been implicated in humans⁷⁰ and mice⁷¹⁻⁷³.

Unimprinted Altruism

We consider an unimprinted altruism gene, denoted by y_A , that, when homozygous, induces an altruistic investment of k_A ($k_A > 0$), and when heterozygous, induces an altruistic investment of h^*k_A , where h denotes the dominance. If we take g and g' as the population frequency of the altruism gene in two consecutive generations, then the population frequency of the altruism gene in the latter generation is:

$$\bar{w}g' = g^2 \left(\frac{b(k_A)g(1-g)h}{g^2+g(1-g)} + \frac{b(k_A)g^2}{g^2+g(1-g)} - c(k_A) + 1 \right) + \frac{1}{2}g(1-g) \left(\frac{g(1-g)b(k_A)h}{g^2+g(1-g)} + \frac{b(k_A)g^2}{g^2+g(1-g)} - c(k_A)h + 1 \right) + \frac{1}{2}(1-g)g \left(\frac{b(k_A)(1-g)gh}{(1-g)^2+(1-g)g} - c(k_A)h + 1 \right), \quad (6)$$

where the mean fitness of individuals is given by: $\bar{w} = g(1-g) \left(\frac{g(1-g)b(k_A)h}{g^2+g(1-g)} + \frac{b(k_A)g^2}{g^2+g(1-g)} - c(k_A)h + 1 \right) + g^2 \left(\frac{g(1-g)b(k_A)h}{g^2+g(1-g)} + \frac{b(k_A)g^2}{g^2+g(1-g)} - c(k_A) + 1 \right) + (1-g)g \left(\frac{(1-g)gb(k_A)h}{(1-g)^2+(1-g)g} - c(k_A)h + 1 \right) + (1-g)^2 \left(\frac{(1-g)gb(k_A)h}{(1-g)^2+(1-g)g} + 1 \right)$.

Each term relates to a different class of individual. For illustration, we derive the term relating to heterozygous individuals with a maternally derived y_A : $\frac{1}{2}g(1 -$

$g) \left(\frac{g(1-g)b(k_A)h}{g^2+g(1-g)} + \frac{b(k_A)g^2}{g^2+g(1-g)} - c(k_A)h + 1 \right)$. (1) Find the frequency of individuals with

this genotype: $g^*(1-g)$. (2) Multiply this by absolute fitness, which is 1 at baseline, with additively applied benefits weighted by the probability that the individual pairs with altruists, and additively applied costs applied if the individual is an altruist:

$\frac{g(1-g)b(k_A)h}{g^2+g(1-g)} + \frac{b(k_A)g^2}{g^2+g(1-g)} - c(k_A)h + 1$. (3) Weight this by the proportion of y_A -bearing

gametes produced by individuals: $\frac{1}{2}$.

The altruism gene decreases in population frequency when $g' < g$, which requires the condition: $\frac{1}{2}b(k_A) < c(k_A)$. Given that genetic relatedness is $(R_f + R_m)/2 = (0+1)/2 = \frac{1}{2}$, this condition corresponds to Hamilton's Rule^{51,74,75}. This $(\frac{1}{2}b(k_A) < c(k_A))$ is also the condition for the invasion of a weaker altruism gene (lower k_A) against a stronger one, owing to diminishing returns on altruistic investment $\left(\frac{d^2b}{dc^2} < 0\right)$. Taken together, this implies that, when $\frac{1}{2}b(k_A) < c(k_A)$, the optimal altruism investment for unimprinted genes is zero, and increased altruistic investment is increasingly suboptimal.

Trait Distorter Spread

We consider an imprinted altruism gene that is only expressed when maternally inherited, denoted by D_1 , and induces an altruistic investment of k ($k > 0$). If we take p and p' as the population frequency of the altruism gene in two consecutive

generations, then the population frequency of the altruism gene in the latter generation is:

$$\bar{w}p' = p(1-p)(b(k) - c(k) + 1)/2 + (1-p)p/2 + p^2(b(k) - c(k) + 1), \quad (7)$$

where the mean fitness of individuals is given by: $\bar{w} = 1 - p + (b(k) - c(k) + 1)p$.

Each term relates to a different class of individual. For illustration, we derive the term relating to heterozygous individuals with a paternally derived D_1 : $(1-p)p/2$. (1) Find the population frequency of individuals with this genotype: $(1-p)p$. (2) Weight by (absolute) individual fitness: 1. (3) Weight by the proportion of imprinted altruism gene-bearing gametes (D_1) produced: $1/2$.

We ask when a rare imprinted altruism gene (D_1) can invade a population fixed for the non-trait distorter (D_0). We take Supplementary Equation 7, set $p'=p=p^*$, and solve to find two possible equilibria: $p^*=0$ (non-trait distorter fixation) and $p^*=1$ (imprinted gene fixation). The imprinted gene (D_1) can invade from rarity when the $p^*=0$ equilibrium is unstable, which occurs when the differential of p' with respect to p , at $p^*=0$, is greater than one. The imprinted altruism gene invasion criterion is therefore $b(k) > c(k)$.

We now ask what frequency the imprinted altruism gene (D_1) will reach after invasion. The gene (D_1) can spread to fixation if the $p^*=1$ equilibrium is stable, which requires that the differential of p' with respect to p , at $p^*=1$, is less than one. This

requirement always holds true, demonstrating that there is no negative frequency dependence on the imprinted gene, and that it will always spread to fixation after its initial invasion.

Given that genetic relatedness is $R_m=1$, our condition for the spread of the imprinted altruism gene ($b(k)>c(k)$) corresponds to Hamilton's Rule^{51,74,75}. Combining with the result of the "Unimprinted altruism" model, altruistic investment (of $k=k_A=k$) is simultaneously favoured at maternally expressed genes and disfavoured at unimprinted genes, rendering the imprinted altruism gene a selfish trait distorter, when $\frac{1}{2}b(k_i)<c(k_i)<b(k_i)$.

Spread of an autosomal suppressor

We ask when an unimprinted suppressor (S_1), competing against a non-suppressor (S_0), will invade from rarity. We can write recursions detailing the generational change in the frequencies of the four possible gametes, D_0/S_0 , D_0/S_1 , D_1/S_0 , D_1/S_1 , with the respective frequencies in the current generation denoted by x_{00} , x_{01} , x_{10} and x_{11} , and the frequencies in the subsequent generation denoted by an appended dash ('):

$$\bar{w}x_{00}' = x_{00}x_{00} + \frac{x_{00}x_{01}}{2} + \frac{x_{00}x_{10}}{2} + \frac{x_{00}x_{11}}{4} + \frac{x_{01}x_{00}}{2} + \frac{x_{01}x_{10}}{4} + \frac{1}{4}x_{11}x_{00}(1 - c_{sup}) + \frac{1}{4}x_{10}x_{01}(1 - c_{sup})(1 + b(k)(x_{00} + x_{10})) + \frac{1}{2}x_{10}x_{00}(1 - c(k) + b(k)(x_{00} + x_{10})) \quad (8)$$

$$\bar{w}x_{01}' = \frac{x_{00}}{4} + \frac{x_{00}x_{01}}{2} + \frac{x_{00}x_{11}}{4} + \frac{x_{01}x_{00}}{2} + x_{01}x_{01} + \frac{x_{01}x_{10}}{4} + \frac{x_{01}x_{11}}{2} + \frac{1}{4}x_{11}x_{00}(1 - c_{sup}) + \frac{1}{2}x_{11}x_{01}(1 - c_{sup}) + \frac{1}{4}x_{10}x_{01}(1 - c_{sup})(1 + b(k)(x_{00} + x_{10})) \quad (9)$$

$$\begin{aligned} \bar{w}x_{10}' = & \frac{x_{00}x_{10}}{2} + \frac{x_{00}x_{11}}{4} + \frac{x_{01}x_{10}}{4} + \frac{1}{4}x_{11}x_{00}(1 - c_{sup}) + \frac{1}{2}x_{11}x_{10}(1 - c_{sup}) + \frac{1}{4}x_{10}x_{01}(1 - c_{sup}) \\ & (1 + b(k)(x_{00} + x_{10})) + \frac{1}{2}x_{10}x_{11}(1 - c_{sup})(1 + b(k)(x_{00} + x_{10})) + \frac{1}{2}x_{10}x_{00}(1 - \\ & c(k) + b(k)(x_{00} + x_{10})) + x_{10}x_{10}(1 - c(k) + b(k)(x_{00} + x_{10})) \end{aligned} \quad (10)$$

$$\begin{aligned} \bar{w}x_{11}' = & \frac{x_{00}x_{11}}{4} + \frac{x_{01}x_{10}}{4} + \frac{x_{01}x_{11}}{2} + \frac{1}{4}x_{11}x_{00}(1 - c_{sup}) + \frac{1}{2}x_{11}x_{01}(1 - c_{sup}) + \frac{1}{2}x_{11}x_{10}(1 - c_{sup}) \\ & + x_{11}x_{11}(1 - c_{sup}) + \frac{1}{4}x_{10}x_{01}(1 - c_{sup})(1 + b(k)(x_{00} + x_{10})) + \frac{1}{2}x_{10}x_{11}(1 - c_{sup})(1 \\ & + b(k)(x_{00} + x_{10})) \end{aligned} \quad (11)$$

\bar{w} is the average fitness of individuals in the current generation, and equals the sum of the equations' right-hand sides. Each term in each equation relates to a different class of individual. For illustration, we derive the term corresponding to the contribution of D_0/S_0 gametes to the next generation, by individuals with a maternally inherited D_1/S_0 gamete and a paternally inherited D_0/S_0 gamete; this is the $\frac{1}{2}x_{10}x_{00}(1 - c(k) + b(k)(x_{00} + x_{10}))$ term in the $\bar{w}x_{00}'$ recursion. (1) Find the population frequency of individuals with this genotype: $x_{10}x_{00}$. (2) Weight by (absolute) individual fitness: $1 - c(k) + b(k)(x_{00} + x_{10})$. (3) Weight by the proportion of trait distorter-bearing gametes (D_1) produced: $\frac{1}{2}$.

We derive the Jacobian stability matrix for the equilibrium in which the trait distorter (D_1) and non-suppressor (S_0) are at fixation ($x_{00}^*=0$, $x_{01}^*=0$, $x_{10}^*=1$, $x_{11}^*=0$). The suppressor can invade when the equilibrium is unstable, which occurs when the leading eigenvalue is greater than one. The leading eigenvalue is $\frac{(b(k)+2)(1-c_{sup})}{2(b(k)-c(k)+1)}$, meaning the suppressor invasion criterion is given by:

$$c_{sup}(1+b(k)/2) < c(k) - b(k)/2. \quad (12)$$

Therefore, the suppressor invades from rarity above a threshold level of distortion, k , when, from the perspective of an unimprinted locus, the number of relatives that die as a result of trait distortion $(c(k)-b(k)/2)$, exceeds the number of relatives that die as a result of trait distorter suppression $(c_{sup}(1+b(k)/2))$.

Consequences of suppressor spread for organism phenotype

We ask what frequency the trait distorter (D_1) and suppressor (S_1) will reach after initial suppressor (S_1) invasion. We assume that the suppressor is introduced from rarity when the trait distorter has reached the population frequency given by $f(x_{00} \rightarrow f, x_{10} \rightarrow 1-f, \{x_{01}, x_{11}\} \rightarrow 0)$. We numerically iterate Supplementary Equations 8-11, over successive generations, until equilibrium has been reached. At equilibrium, for all parameter combinations $(f, t, c_{sup}, c_{trait})$, the suppressor reaches an internal equilibrium and the trait distorter is lost from the population ($x_{00}^* + x_{01}^* = 1, x_{10}^* = 0, x_{11}^* = 0$). This equilibrium arises because trait distorter-presence gives the suppressor (S_1) a selective advantage, leading to high suppressor frequency, which in turn reverses the selective advantage of the trait distorter (D_1), leading to trait distorter loss and suppressor equilibration (Figure 3b).

Invasion of a mutant trait distorter

We ask when a mutant trait distorter (D_2) of strength \hat{k} will invade against a resident trait distorter (D_1) that is unsuppressed and at fixation ($\hat{k} \neq k$). We write recursions detailing the generational frequency changes in the six possible gametes, $D_0/S_0, D_0/S_1, D_1/S_0, D_1/S_1, D_2/S_0, D_2/S_1$, with current generation frequencies denoted

respectively by x_{00} , x_{01} , x_{10} , x_{11} , x_{20} , x_{21} , and next generation frequencies denoted with an appended dash ('):

$$\begin{aligned}\bar{w}x_{00}' &= x_{00}x_{00} + \frac{x_{00}x_{01}}{2} + \frac{x_{00}x_{10}}{2} + \frac{x_{00}x_{11}}{4} + \frac{x_{00}x_{20}}{2} + \frac{x_{00}x_{21}}{4} + \frac{x_{01}x_{00}}{2} + \frac{x_{01}x_{10}}{4} + \frac{x_{01}x_{20}}{4} \\ &+ \frac{1}{4}x_{11}x_{00}(1 - c_{sup}) + \frac{1}{4}x_{21}x_{00}(1 - c_{sup}) + \frac{1}{4}x_{10}x_{01}(1 - c_{sup})(1 + b(k)(x_{00} + x_{10} \\ &+ x_{20})) + \frac{1}{2}x_{10}x_{00}(1 - c(k) + b(k)(x_{00} + x_{10} + x_{20})) + \frac{1}{4}x_{20}x_{01}(1 - c_{sup})(1 + \\ &b(\hat{k})(x_{00} + x_{10} + x_{20})) + \frac{1}{2}x_{20}x_{00}(1 - c(\hat{k}) + b(\hat{k})(x_{00} + x_{10} + x_{20}))\end{aligned}\quad (13)$$

$$\begin{aligned}\bar{w}x_{01}' &= \frac{x_{00}x_{01}}{2} + \frac{x_{00}x_{11}}{4} + \frac{x_{00}x_{21}}{4} + \frac{x_{01}x_{00}}{2} + x_{01}x_{01} + \frac{x_{01}x_{10}}{4} + \frac{x_{01}x_{11}}{2} + \frac{x_{01}x_{20}}{4} + \frac{x_{01}x_{21}}{2} \\ &+ \frac{1}{4}x_{11}x_{00}(1 - c_{sup}) + \frac{1}{2}x_{11}x_{01}(1 - c_{sup}) + \frac{1}{4}x_{21}x_{00}(1 - c_{sup}) + \frac{1}{2}x_{21}x_{01}(1 - c_{sup}) + \\ &\frac{1}{4}x_{10}x_{01}(1 - c_{sup})(1 + b(k)(x_{00} + x_{10} + x_{20})) + \frac{1}{4}x_{20}x_{01}(1 - c_{sup})(1 + b(\hat{k})(x_{00} + \\ &x_{10} + x_{20}))\end{aligned}\quad (14)$$

$$\begin{aligned}\bar{w}x_{10}' &= \frac{x_{00}x_{10}}{2} + \frac{x_{00}x_{11}}{4} + \frac{x_{01}x_{10}}{4} + \frac{1}{4}x_{11}x_{00}(1 - c_{sup}) + \frac{1}{2}x_{11}x_{10}(1 - c_{sup}) + \frac{1}{4}x_{11}x_{20}(1 - c_{sup}) \\ &+ \frac{1}{4}x_{21}x_{10}(1 - c_{sup}) + \frac{1}{4}x_{10}x_{01}(1 - c_{sup})(1 + b(k)(x_{00} + x_{10} + x_{20})) + \frac{1}{2}x_{10}x_{11}(1 - \\ &c_{sup})(1 + b(k)(x_{00} + x_{10} + x_{20})) + \frac{1}{4}x_{10}x_{21}(1 - c_{sup})(1 + b(k)(x_{00} + x_{10} + x_{20})) + \\ &\frac{1}{2}x_{10}x_{00}(1 - c(k) + b(k)(x_{00} + x_{10} + x_{20})) + x_{10}x_{10}(1 - c(k) + b(k)(x_{00} + x_{10} + \\ &x_{20})) + \frac{1}{2}x_{10}x_{20}(1 - c(k) + b(k)(x_{00} + x_{10} + x_{20})) + \frac{1}{4}x_{20}x_{11}(1 - c_{sup})(1 + \\ &b(\hat{k})(x_{00} + x_{10} + x_{20})) + \frac{1}{2}x_{20}x_{10}(1 - c(\hat{k}) + b(\hat{k})(x_{00} + x_{10} + x_{20}))\end{aligned}\quad (15)$$

$$\begin{aligned}\bar{w}x_{11}' &= \frac{x_{00}x_{11}}{4} + \frac{x_{01}x_{10}}{4} + \frac{x_{01}x_{11}}{2} + \frac{1}{4}x_{11}x_{00}(1 - c_{sup}) + \frac{1}{2}x_{11}x_{01}(1 - c_{sup}) + \frac{1}{2}x_{11}x_{10}(1 - c_{sup}) \\ &+ x_{11}x_{11}(1 - c_{sup}) + \frac{1}{4}x_{11}x_{20}(1 - c_{sup}) + \frac{1}{2}x_{11}x_{21}(1 - c_{sup}) + \frac{1}{4}x_{21}x_{10}(1 - c_{sup}) \\ &+ \frac{1}{2}x_{21}x_{11}(1 - c_{sup}) + \frac{1}{4}x_{10}x_{01}(1 - c_{sup})(1 + b(k)(x_{00} + x_{10} + x_{20})) + \frac{1}{2}x_{10}x_{11}(1 - \\ &c_{sup})(1 + b(k)(x_{00} + x_{10} + x_{20})) + \frac{1}{4}x_{10}x_{21}(1 - c_{sup})(1 + b(k)(x_{00} + x_{10} + x_{20})) + \\ &\frac{1}{4}x_{20}x_{11}(1 - c_{sup})(1 + b(\hat{k})(x_{00} + x_{10} + x_{20}))\end{aligned}\quad (16)$$

$$\bar{w}x_{20}' = \frac{x_{00}x_{20}}{2} + \frac{x_{00}x_{21}}{4} + \frac{x_{01}x_{20}}{4} + \frac{1}{4}x_{11}x_{20}(1 - c_{sup}) + \frac{1}{4}x_{21}x_{00}(1 - c_{sup}) + \frac{1}{4}x_{21}x_{10}(1 - c_{sup}) +$$

$$\begin{aligned}
& \frac{1}{2}x_{21}x_{20}(1 - c_{sup}) + \frac{1}{4}x_{10}x_{21}(1 - c_{sup})(1 + b(k)(x_{00} + x_{10} + x_{20})) + \frac{1}{2}x_{10}x_{20}(1 - \\
& c(k) + b(k)(x_{00} + x_{10} + x_{20})) + \frac{1}{4}x_{20}x_{01}(1 - c_{sup})(1 + b(\hat{k})(x_{00} + x_{10} + x_{20})) + \\
& \frac{1}{4}x_{20}x_{11}(1 - c_{sup})(1 + b(\hat{k})(x_{00} + x_{10} + x_{20})) + \frac{1}{2}x_5x_6(1 - c_{sup})(1 + b(\hat{k})(x_{00} + \\
& x_{10} + x_{20})) + \frac{1}{2}x_{20}x_{00}(1 - c(\hat{k}) + b(\hat{k})(x_{00} + x_{10} + x_{20})) + \frac{1}{2}x_{20}x_{10}(1 - c(\hat{k}) + \\
& b(\hat{k})(x_{00} + x_{10} + x_{20})) + x_{20}x_{20}(1 - c(\hat{k}) + b(\hat{k})(x_{00} + x_{10} + x_{20})) \tag{17}
\end{aligned}$$

$$\begin{aligned}
\bar{w}x_{21}' = & \frac{x_{00}x_{21}}{4} + \frac{x_{01}x_{20}}{4} + \frac{x_{01}x_{21}}{2} + \frac{1}{4}x_{11}x_{20}(1 - c_{sup}) + \frac{1}{2}x_{11}x_{21}(1 - c_{sup}) + \frac{1}{4}x_{21}x_{00}(1 - c_{sup}) \\
& + \frac{1}{2}x_{21}x_{01}(1 - c_{sup}) + \frac{1}{4}x_{21}x_{10}(1 - c_{sup}) + \frac{1}{2}x_{21}x_{11}(1 - c_{sup}) + \frac{1}{2}x_{21}x_{20}(1 - c_{sup}) \\
& + x_{21}x_{21}(1 - c_{sup}) + \frac{1}{4}x_{10}x_{21}(1 - c_{sup})(1 + b(k)(x_{00} + x_{10} + x_{20})) + \frac{1}{4}x_{20}x_{01}(1 - \\
& c_{sup})(1 + b(\hat{k})(x_{00} + x_{10} + x_{20})) + \frac{1}{4}x_{20}x_{11}(1 - c_{sup})(1 + b(\hat{k})(x_{00} + x_{10} + x_{20})) + \\
& \frac{1}{2}x_{20}x_{21}(1 - c_{sup})(1 + b(\hat{k})(x_{00} + x_{10} + x_{20})). \tag{18}
\end{aligned}$$

\bar{w} is the average fitness of individuals in the current generation, and equals the sum of the right-hand side of the system of equations. The mutant trait distorter can invade when the equilibrium given by $x_{00}^*=0$, $x_{01}^*=0$, $x_{10}^*=1$, $x_{11}^*=0$, $x_{20}^*=0$, $x_{21}^*=0$ is unstable, which occurs when the leading eigenvalue of the Jacobian stability matrix for this equilibrium is greater than one. Testing for stability in this way, we find that the mutant trait distorter invades from rarity when $\Delta b > \Delta c$, where $\Delta b = b(\hat{k}) - b(k)$, $\Delta c = c(\hat{k}) - c(k)$.

The implication is that mutant trait distorters will invade if they approach a 'target' strength (k_{target}), corresponding to the level of trait distortion that would maximise the fitness of the gene⁵³, at which:

$$\frac{\partial b}{\partial k} = \frac{\partial c}{\partial k}. \quad (19)$$

In the absence of suppression, this target is the equilibrium level of distortion

$$(k^* = k_{target}).$$

Equilibrium trait distorter and suppressor frequencies (long term evolution)

We ask what equilibrium state will arise after the invasion of a mutant trait distorter.

We assume that the mutant trait distorter (D_2) is introduced from rarity when the resident trait distorter (D_1) has reached the population frequency given by q . We numerically iterate Supplementary Equations 13-18, over successive generations, until equilibrium has been reached. At equilibrium, for all parameter combinations $(q, t(k), t(\hat{k}), c_{sup}, c(k), c(\hat{k}))$, the resident trait distorter (D_1) is lost from the population ($x_{10}, x_{11} = 0$), with either the mutant trait distorter (D_2) and non-suppressor (S_0) at fixation ($x_{20}^* = 1$), or the non-trait distorter at fixation alongside the suppressor at an internal equilibrium ($x_{00}^* + x_{01}^* = 1$). The latter scenario arises if the mutant trait distorter triggers suppressor invasion ($c_{sup}(1 + b(\hat{k})/2) < c(\hat{k}) - b(\hat{k})/2$). This equilibrium arises because mutant trait distorter-presence gives the suppressor (S_1) a selective advantage, leading to high suppressor frequency, which in turn reverses the selective advantage of distortion, leading to trait distorter (D_1, D_2) loss and suppressor equilibration.

Given that mutant trait distorters will invade if they approach a 'target' strength (k_{target}), if the individual level cost associated with this target level of distortion ($c(k_{target})$) is sufficiently high relative to the cost of suppression (c_{sup}), so that the

following condition is satisfied, the equilibrium level of distortion will be $k^*=0$:

$c_{sup}(1+b(k_{target})/2) < c(k_{target}) - b(k_{target})/2$. If this condition is not satisfied the equilibrium level of distortion will be $k^*=k_{target}$ (Figure 3e).

Discussion

Although there have been no direct tests, our predictions are consistent with data on imprinted genes. There is no evidence that traits influenced by imprinted genes deviate significantly from individual level optima under normal development⁵².

Significant deviation is only observed when imprinted genes are deleted, implying that imprinted trait distorters are either suppressed, or counterbalanced by oppositely imprinted genes pulling the trait in the opposite direction^{63,76}. Furthermore, although many different parties (coreplicons) have vested interests in genomic imprinting, our analysis suggests why the unimprinted majority could win control⁷⁷. This could help explain both why imprinting appears to be relatively rare within the genome^{50,54,66}, and why imprints are removed and re-added every generation in mice, handing control of genomic patterns of imprinting to unimprinted genes^{54,77,78}.

Supplementary Note 5

Horizontal Gene Transfer and Public Goods

Bacteria produce and excrete many extracellular factors that provide a benefit to the local population of cells and so can be thought of as public goods⁷⁹. We modelled the evolution of investment in a public good in a large, clonally reproducing population. We assume a public good that costs c to produce, and provides a benefit b to the group. We assume a well-mixed population, meaning genetic relatedness at vertically inherited genes is zero ($R_{vertical}=0$), and so indirect fitness benefits cannot favour public good production at the individual level ($R_{vertical}b=0<c$)^{51,74,75,80,81}. There are also direct fitness benefits of public good production, which arise because producers of public goods receive a fraction of the benefit (b) they confer on the group, but we assume that the population is sufficiently large and well mixed that direct fitness benefits cannot favour public good production at the individual level. This means that public good production is disfavoured at the individual level.

We consider a selfish genetic element that resides on a mobile locus (horizontal & vertical transmission) and may gain a propagation advantage by upregulating individual public goods investment⁸²⁻⁸⁶. The genes that do not gain a propagation advantage from increased public goods production comprise the non-mobile loci (vertical transmission). Non-mobile loci comprise most of the genome, and so constitute the majority within the parliament of genes. We focus our analyses on when a mobile trait distorter and a non-mobile suppressor can spread. The purpose of this model is to illustrate how selection will act on selfish mobile genes and their suppressors.

Model assumptions

We consider a public goods gene (D_1) that competes against a non-trait distorter (D_0) at a mobile locus. The trait distorter (D_1) increases public goods investment by some amount (k), at a fitness cost to the individual ($c(k)$) and benefit shared within the group ($b(k) > c(k)$) that are both monotonically increasing functions of investment ($\frac{\partial\{b,c\}}{\partial k} \geq 0$).

We assume the following lifecycle. Individuals in a large, effectively infinite, population randomly aggregate into smaller social groups (*patches*). Individuals then randomly pair up within their patch, and horizontal gene transfer occurs, with certainty, within pairs that are genetically dissimilar at the mobile locus^{87,88}.

Alternative assumptions about the probability of horizontal gene transfer do not change our qualitative results (Scott, unpublished). Only one allele at the mobile locus is transferrable in each patch, and each allele at the mobile locus is transferrable in an equal proportion of patches. We denote those patches in which the non-trait distorter (D_0) is transferred as “type 1” patches, and those patches in which the trait distorter (D_1) is transferred as “type 2” patches. Individuals may then produce public goods, which are shared within patches, before the population re-merges, and individuals reproduce in proportion to their fitness before dying (non-overlapping generations), with progeny inheriting all alleles from their parent (perfect inheritance).

Trait Distorter Spread

We respectively take \bar{p} and \bar{p}'' as the population frequency of the trait distorter (D_1) at the start of two consecutive generations, and p'_j as the average frequency of the trait distorter (D_1) in patches of type j after horizontal gene transfer ($j \in \{1, 2\}$), with $p'_{j=1} = \bar{p} + \bar{p}(1-\bar{p})$ and $p'_{j=2} = \bar{p} - \bar{p}(1-\bar{p})$. The population frequency of the trait distorter in the latter generation (\bar{p}'') is:

$$\bar{p}'' = \frac{p'_{j=1}(1+p'_{j=1}b(k)-c(k))+p'_{j=2}(1+p'_{j=2}b(k)-c(k))}{2+(b(k)-c(k))(p'_{j=1}+p'_{j=2})}, \quad (20)$$

where the denominator denotes average individual fitness. Stable equilibria occur for

$$\bar{p}=\bar{p}''=p^* \text{ and } \left. \frac{\partial \bar{p}''}{\partial \bar{p}} \right|_{\bar{p}=p^*} < 1, \text{ which occurs when } p^* = \left\{ 0, \left(1 + \sqrt{1 - \frac{4c(k)}{b(k)}} \right) / 2 \right\}.$$

Unstable equilibria occur for $\bar{p}=\bar{p}''=\bar{p}^*$ and $\left. \frac{\partial \bar{p}''}{\partial \bar{p}} \right|_{\bar{p}=p^*} > 1$, which occurs when $p^* =$

$$\left\{ \left(1 - \sqrt{1 - \frac{4c(k)}{b(k)}} \right) / 2, 1 \right\}. \text{ Therefore, the trait distorter } (D_1) \text{ exhibits positive and}$$

negative frequency dependence, meaning it can only invade if introduced at high

enough frequency $\left(\bar{p} > \left(1 - \sqrt{1 - \frac{4c(k)}{b(k)}} \right) / 2 \right)$, reaching a polymorphism below

fixation $\left(p^* = \left(1 + \sqrt{1 - \frac{4c(k)}{b(k)}} \right) / 2 \right)$ (Supplementary Figure 10). Frequency

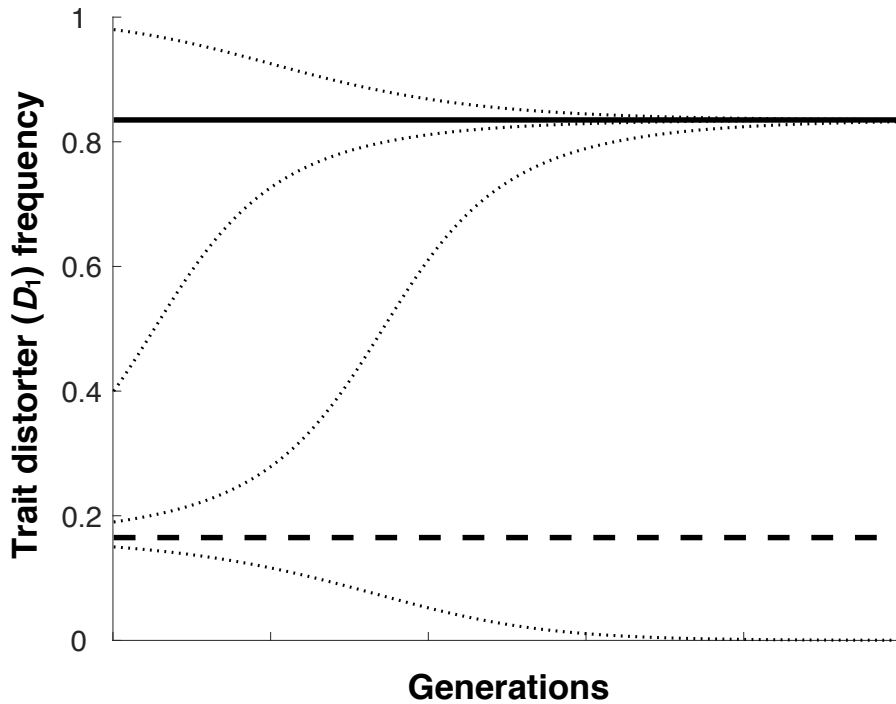
dependence arises because, when there is low genetic diversity at the mobile locus

($p \rightarrow 0/1$), there is less generational horizontal gene transfer, and correspondingly

lower patch relatedness, which dissipates the trait distorter's selective advantage⁸².

A trait distorter (D_1) is more likely to invade and reach a high population frequency if

it produces a public good associated with a large benefit to cost ratio (b/c)^{82-84,89-91}.



Supplementary Figure 10. Spread of the trait distorter (D_1) in the absence of suppression. The trait distorter (D_1) is introduced at some frequency, and equilibrates over successive generations (dotted lines indicate trajectories corresponding to different initial trait distorter frequencies). The solid

and dashed lines are, respectively, the stable $\left(p^* = \left(1 + \sqrt{1 - \frac{4c(k)}{b(k)}}\right)/2\right)$ and unstable

$\left(p^* = \left(1 - \sqrt{1 - \frac{4c(k)}{b(k)}}\right)/2\right)$ equilibria. If introduced at a frequency greater than the unstable internal

equilibrium, trait distorter (D_1) frequency reaches the stable internal equilibrium, without going to fixation (negative frequency dependence). If introduced at a frequency lower than the unstable internal equilibrium, trait distorter (D_1) frequency goes to zero (positive frequency dependence).

Spread of a suppressor and consequences for the organism

We consider a suppressor allele (S_1) that competes against a non-suppressor (S_0) at a non-mobile locus. Suppressors of mobile elements are widespread and may silence elements before they are translated, through gene methylation and RNAi⁹².

We respectively take \bar{x}_i and \bar{x}_i'' as the population genotype frequencies at the start of two consecutive generations, with the subscript $i \in \{00, 01, 10, 11\}$ denoting the

respective genotypes: $\{D_0/S_0, D_0/S_1, D_1/S_0, D_1/S_1\}$. We take x_{ij}' as the average frequency of genotype i in patches of type j after horizontal gene transfer ($j \in \{1, 2\}$), with $x_{001}' = \bar{x}_{i=00} + (\bar{x}_{i=00} + \bar{x}_{i=01})\bar{x}_{i=10}$, $x_{011}' = \bar{x}_{i=01} + (\bar{x}_{i=00} + \bar{x}_{i=01})\bar{x}_{i=11}$, $x_{101}' = \bar{x}_{i=10} - (\bar{x}_{i=00} + \bar{x}_{i=01})\bar{x}_{i=10}$, $x_{111}' = \bar{x}_{i=11} - (\bar{x}_{i=00} + \bar{x}_{i=01})\bar{x}_{i=11}$, $x_{002}' = \bar{x}_{i=00} - (\bar{x}_{i=10} + \bar{x}_{i=11})\bar{x}_{i=01}$, $x_{012}' = \bar{x}_{i=01} - (\bar{x}_{i=10} + \bar{x}_{i=11})\bar{x}_{i=01}$, $x_{102}' = \bar{x}_{i=10} + (\bar{x}_{i=10} + \bar{x}_{i=11})\bar{x}_{i=00}$, $x_{112}' = \bar{x}_{i=11} + (\bar{x}_{i=10} + \bar{x}_{i=11})\bar{x}_{i=01}$. The population genotype frequencies in the latter generation (\bar{x}_i'') are:

$$W\bar{x}_{00}'' = \sum_{j=1}^{j=2}(x_{00j}'(1 + x_{10j}'b(k))), \quad (21)$$

$$W\bar{x}_{01}'' = \sum_{j=1}^{j=2}(x_{01j}'(1 + x_{10j}'b(k))), \quad (22)$$

$$W\bar{x}_{10}'' = \sum_{j=1}^{j=2}(x_{10j}'(1 + x_{10j}'b(k) - c(k))), \quad (23)$$

$$W\bar{x}_{11}'' = \sum_{j=1}^{j=2}(x_{11j}'(1 + x_{10j}'b(k))(1 - c_{sup})), \quad (24)$$

where W is average individual fitness, equal to the sum of the right-hand sides of the system of equations.

We numerically iterated these recursions, for a range of parameter values (b, c, c_{sup}), and for different initial frequencies of the trait distorter (D_1) to find the trait distorter (D_1) and suppressor (S_1) frequencies at equilibrium, and the resulting average trait distortion ($x_{10} k$). We found that, when distortion is weak (low k), suppressors are not favoured, but the trait distorter has relatively little impact at the individual level. For example, when the cost of suppression is $c_{sup}=0.05$, and the cost and benefit of public goods production are $c_{HGT}=k$ (linear cost) and $b_{HGT}=8k^{0.9}$ (relatively large,

decelerating benefit), unsuppressed trait distorters cannot upregulate public goods by more than $k=c_{HGT}=0.08$ (Figure 3c).

We found that the suppressor invades from rarity, in response to a trait distorter at equilibrium $\left(x_{10}^* = \left(1 + \sqrt{1 - \frac{4c(k)}{b(k)}}\right)/2\right)$, above a threshold level of distortion. If the

suppressor invades, it increases in frequency until the trait distorter's (D_1) selective advantage is reversed and the trait distorter is lost from the population; the

suppressor (S_1) then equilibrates (Figure 3c). A trait distorter (D_1) is more likely to

evade suppression if it produces a public good associated with a large benefit to cost ratio ($b(k)/c(k)$) and if there is a high cost of suppression (c_{sup})⁹³.

Evolution of trait distortion

We ask when a mutant trait distorter (D_2) of strength (\hat{k}) will invade against a resident trait distorter (D_1) that is unsuppressed and at equilibrium ($\hat{k} \neq k$). We denote those patches in which the mutant trait distorter (D_2) is transferred as “type 3”

patches. We use the subscript $i \in \{00, 01, 10, 11, 20, 21\}$ to denote the respective

genotypes $\{D_0/S_0, D_0/S_1, D_1/S_0, D_1/S_1, D_2/S_0, D_2/S_1\}$, and $j \in \{1, 2, 3\}$ to denote patch type.

Average genotype frequencies in each patch type after horizontal gene transfer (x_{ij})

are given by: $x_{001}' = \bar{x}_{i=00} + (\bar{x}_{i=00} + \bar{x}_{i=01})(\bar{x}_{i=10} + \bar{x}_{i=20})$, $x_{011}' = \bar{x}_{i=01} + (\bar{x}_{i=00} + \bar{x}_{i=01})(\bar{x}_{i=11} + \bar{x}_{i=21})$, $x_{101}' = \bar{x}_{i=10} - (\bar{x}_{i=00} + \bar{x}_{i=01})\bar{x}_{i=10}$, $x_{111}' = \bar{x}_{i=11} - (\bar{x}_{i=00} + \bar{x}_{i=01})\bar{x}_{i=11}$, $x_{201}' = \bar{x}_{i=20} - (\bar{x}_{i=00} + \bar{x}_{i=01})\bar{x}_{i=20}$, $x_{211}' = \bar{x}_{i=21} - (\bar{x}_{i=00} + \bar{x}_{i=01})\bar{x}_{i=21}$, $x_{002}' = \bar{x}_{i=00} - (\bar{x}_{i=10} + \bar{x}_{i=11})\bar{x}_{i=00}$, $x_{012}' = \bar{x}_{i=01} - (\bar{x}_{i=10} + \bar{x}_{i=11})\bar{x}_{i=01}$, $x_{102}' = \bar{x}_{i=10} + (\bar{x}_{i=10} + \bar{x}_{i=11})(\bar{x}_{i=00} + \bar{x}_{i=20})$, $x_{112}' = \bar{x}_{i=11} + (\bar{x}_{i=10} + \bar{x}_{i=11})(\bar{x}_{i=01} + \bar{x}_{i=21})$, $x_{202}' = \bar{x}_{i=20} - (\bar{x}_{i=10} + \bar{x}_{i=11})\bar{x}_{i=20}$, $x_{212}' = \bar{x}_{i=21} - (\bar{x}_{i=10} + \bar{x}_{i=11})\bar{x}_{i=21}$, $x_{003}' =$

$\bar{x}_{i=00} - (\bar{x}_{i=20} + \bar{x}_{i=21})\bar{x}_{i=00}$, $x_{013}' = \bar{x}_{i=01} - (\bar{x}_{i=20} + \bar{x}_{i=21})\bar{x}_{i=01}$, $x_{103}' = \bar{x}_{i=10} -$
 $(\bar{x}_{i=20} + \bar{x}_{i=21})\bar{x}_{i=10}$, $x_{113}' = \bar{x}_{i=11} - (\bar{x}_{i=20} + \bar{x}_{i=21})\bar{x}_{i=11}$, $x_{201}' = \bar{x}_{i=20} + (\bar{x}_{i=20} +$
 $\bar{x}_{i=21})(\bar{x}_{i=00} + \bar{x}_{i=10})$, $x_{211}' = \bar{x}_{i=21} + (\bar{x}_{i=20} + \bar{x}_{i=21})(\bar{x}_{i=01} + \bar{x}_{i=11})$. We write
 recursions detailing the generational genotype frequency changes:

$$W\bar{x}_{00}'' = \sum_{j=1}^{j=3}(x_{00j}'(1 + x_{10j}'b(k) + x_{20j}'b(\hat{k}))), \quad (25)$$

$$W\bar{x}_{01}'' = \sum_{j=1}^{j=3}(x_{01j}'(1 + x_{10j}'b(k) + x_{20j}'b(\hat{k}))), \quad (26)$$

$$W\bar{x}_{10}'' = \sum_{j=1}^{j=3}(x_{10j}'(1 + x_{10j}'b(k) + x_{20j}'b(\hat{k}) - c(k))), \quad (27)$$

$$W\bar{x}_{11}'' = \sum_{j=1}^{j=3}(x_{11j}'(1 + x_{10j}'b(k) + x_{20j}'b(\hat{k}))(1 - c_{sup}))), \quad (28)$$

$$W\bar{x}_{20}'' = \sum_{j=1}^{j=3}(x_{20j}'(1 + x_{10j}'b(k) + x_{20j}'b(\hat{k}) - c(\hat{k}))), \quad (29)$$

$$W\bar{x}_{21}'' = \sum_{j=1}^{j=3}(x_{21j}'(1 + x_{10j}'b(k) + x_{20j}'b(\hat{k}))(1 - c_{sup}))), \quad (30)$$

where W is average individual fitness, equal to the sum of the right-hand sides of the system of equations.

We assume that trait distorter strength (k) is initially low, and introduce successive mutant trait distorters (D_2), each deviating only slightly from the trait distorters from which they are derived, until one fails to displace the resident trait distorter. The strength of the non-invadable allele gives the equilibrium level of distortion under δ -weak selection⁴⁹. We find that, if the rate of decrease in marginal cooperative benefits $\left(-\frac{d^2b}{dk^2}\right)$ is high relative to the rate of increase in marginal cooperative costs $\left(\frac{d^2c}{dk^2}\right)$, distortion (k^*) evolves to be low, and the suppressor (S_1) may not invade.

Otherwise, stronger trait distorters (D_2) successively invade, bringing trait distorter

strength above the threshold level at which the suppressor (S_1) spreads, with the end result that trait distorters are suppressed and lost from the population, with no trait distortion at equilibrium ($k^*=0$) (Figure 3f).

Discussion

We lack empirical data that would allow us to test our model of mobile public goods genes. Genes associated with extracellular traits, which could represent cooperative public goods, appear to be overrepresented on mobile elements⁹¹. However, this may be nothing to do with cooperation *per se* – genes involved with adaptation to new environments might be more likely to be horizontally acquired, and extracellular traits might be especially important in adaptation to new environments^{84-87,94}.

Supplementary Note 6

Suppressor Conditionality

We assumed in our Equilibrium and Dynamics models (Main Text) that suppressors are only expressed in the presence of their target trait distorters (facultative). We generalise our Equilibrium models (Main Text) by defining the parameter ψ as the “conditionality” of the suppressor ($0 \leq \psi \leq 1$). For full conditionality ($\psi=1$), the suppressor is facultative. For zero conditionality ($\psi=0$), the suppressor is obligate, meaning it is fully expressed when the trait distorter is absent. For intermediate conditionality ($0 < \psi < 1$), the suppressor is partially expressed when the trait distorter is absent. As a result, the suppressor incurs a cost of c_{sup} on the individual when the trait distorter is present, and a cost of $(1-\psi) * c_{sup}$ when the trait distorter is absent.

In the facultative suppressor case ($\psi=1$), considered in the main text, the fitness of D_0/S_0 D_0/S_1 and D_1/S_1 D_0/S_1 individuals, which have a suppressor but not a trait distorter, is 1. Now, in the generalised scenario, the fitness of these individuals is:

$$1 - (1 - \psi) * c_{sup}. \quad (31)$$

Amending Equations 2-5 & 8-13 (main text) according to this small change, and repeating the analysis described in the Methods section (main text), reveals that the suppressor invasion condition ($c_{sup} < c_{trait}(k)$) and the stronger-trait distorter invasion condition ($\Delta t(1 - c_{trait}(\hat{k})) > \Delta c_{trait}$) are unchanged. The suppressor invasion condition is unchanged because an invading suppressor can only gain a selective advantage if it

finds itself in the same individual as a trait distorter, and in such a scenario, it will confer the full cost of c_{sup} regardless of its conditionality (ψ). The stronger-trait distorter invasion condition is unchanged because it is derived for an equilibrium in which the suppressor is absent.

However, suppressor conditionality affects trait distorter-suppressor dynamics in a subtle way. In the facultative suppressor case ($\psi=1$), considered in the main text, the spread of the suppressor from rarity causes the trait distorter to lose its selective advantage and be eliminated from the population, leading to an absence of distortion at the individual level. This occurs because, under suppression, the trait distorter (D_1) gains no transmission advantage over the non-trait distorter (D_0), but is associated with a cost of c_{sup} arising from facultative suppressor expression. The non-trait distorter (D_0) does not pay this cost, so gains a selective advantage under suppression, and spreads at the expense of the trait distorter (D_1).

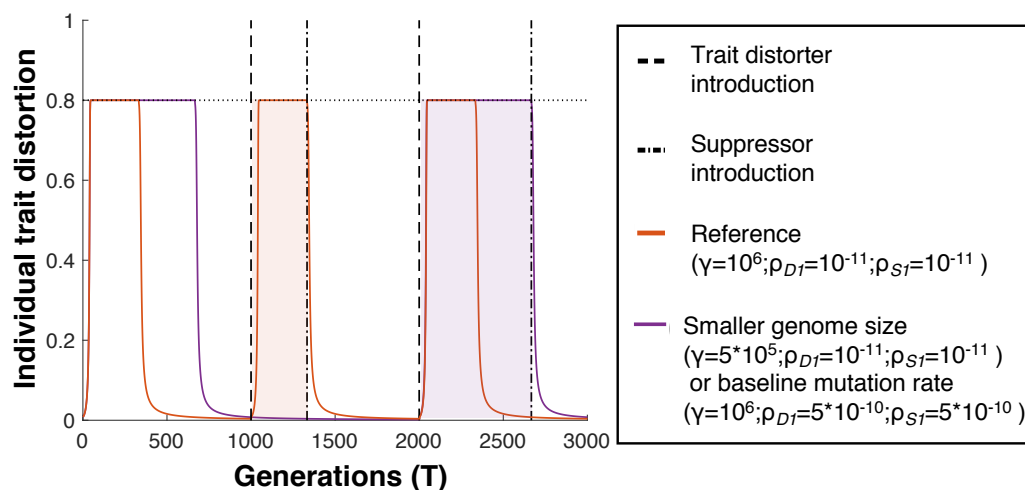
For suppressors with intermediate conditionality ($0<\psi<1$), suppressor spread means that the trait distorter (D_1) pays a cost of c_{sup} and the non-trait distorter (D_0) pays a smaller cost of $((1-\psi)*c_{sup})$. As a result, as the suppressor spreads to high population frequency, the non-trait distorter (D_0) gains a selective advantage, and spreads at the expense of the trait distorter (D_1). However, the selective advantage of the non-trait distorter (D_0), over the trait distorter (D_1), under suppression, is weaker when the suppressor has intermediate conditionality ($0<\psi<1$), compared to when it is fully facultative ($\psi=1$). As a result, the time taken for the trait distorter (D_1) to fall to

negligible population frequency is increased if suppressors are not fully facultative ($0 < \psi < 1$).

For obligate suppressors ($\psi=0$), the trait distorter (D_1) has equal fitness to the non-trait distorter (D_0) under suppression, as both face the full cost of c_{sup} , owing to obligate suppressor expression. This means that trait distorters are not purged after suppression, and though the trait is fully restored to optimality as a result of suppressor fixation, individuals continue to pay the cost of c_{sup} at equilibrium. The residual cost (c_{sup}) is an artefact of the conflict, and will remain, to the detriment of population (absolute) mean fitness, until a conditional ($\psi > 0$) suppressor arises by mutation and selectively displaces the obligate one.

Supplementary Note 7

Dynamics Models (Additional Figures)



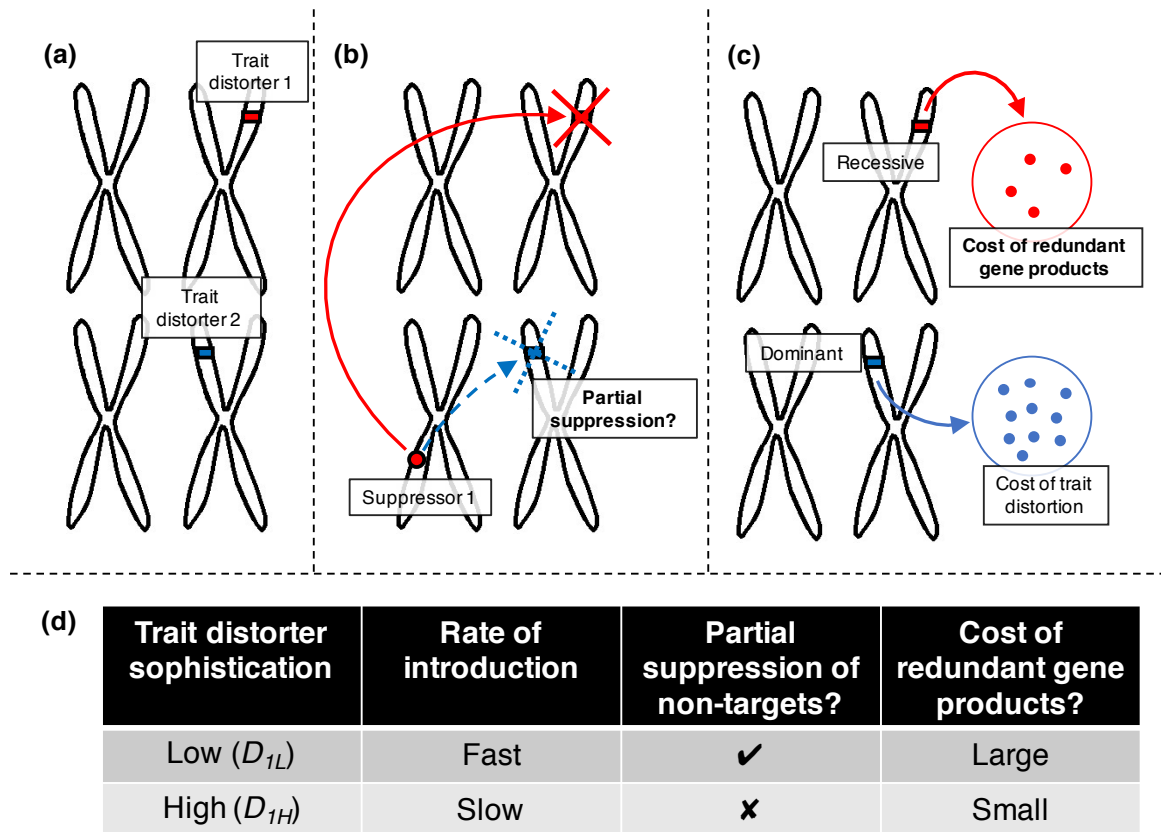
Supplementary Figure 11. The effect of genome size (γ) / mutation rate (ρ) when trait distorters do not co-segregate. Individual trait distortion is plotted over evolutionary time. We introduced trait distorters (D_i) deterministically at new loci every $1/(\theta\gamma\rho_{D_i})$ generations, and their dedicated suppressors after a lag of $1/((1-\theta)\gamma\rho_{sup})$ generations. Individual trait distortion increases and decreases cyclically over evolutionary time, between peaks of k and troughs of 0.

Individual trait distortion is plotted for three different parameter regimes. The first parameter regime is plotted as a reference, and represented by the red line ($\gamma=10^6$; $\rho_{D_i}=10^{-11}$, $\rho_{S_i}=10^{-11}$). The second parameter regime has a half-sized genome size relative to the reference, with an unchanged baseline mutation rate ($\gamma=5*10^5$; $\rho_{D_i}=10^{-11}$, $\rho_{S_i}=10^{-11}$). The third parameter regime has a half-sized baseline mutation rate relative to reference, with an unchanged genome size ($\gamma=10^6$; $\rho_{D_i}=5*10^{-10}$, $\rho_{S_i}=5*10^{-10}$). Proportional changes in genome size (γ) have identical effects to proportional changes in baseline mutation rate (ρ), and therefore, the second and third parameter regimes lead to the same outcome, which is represented by the purple line.

Owing to rapid gene frequency equilibration after trait distorter / suppressor introduction, the periodic functions (red and purple lines) can be approximated as rectangles. A decrease in genome size (γ) or baseline mutation rate (ρ) leads to an increase in the width, and therefore area (shaded regions), of the rectangles, but a corresponding decrease in the density of the rectangles. Therefore,

the average trait distortion across evolutionary time, which is given by the area under the curve, and approximated by $\frac{k\theta\rho_{D_1}}{(1-\theta)\rho_{S_1}}$ (Equation 6), is unaffected by genome size (γ) and baseline mutation rate (ρ).

These numerical solutions assume the following parameter values: $c_{sup}=0.1$; $t=k$, $c_{trait}=k/2$, $k=0.8$.

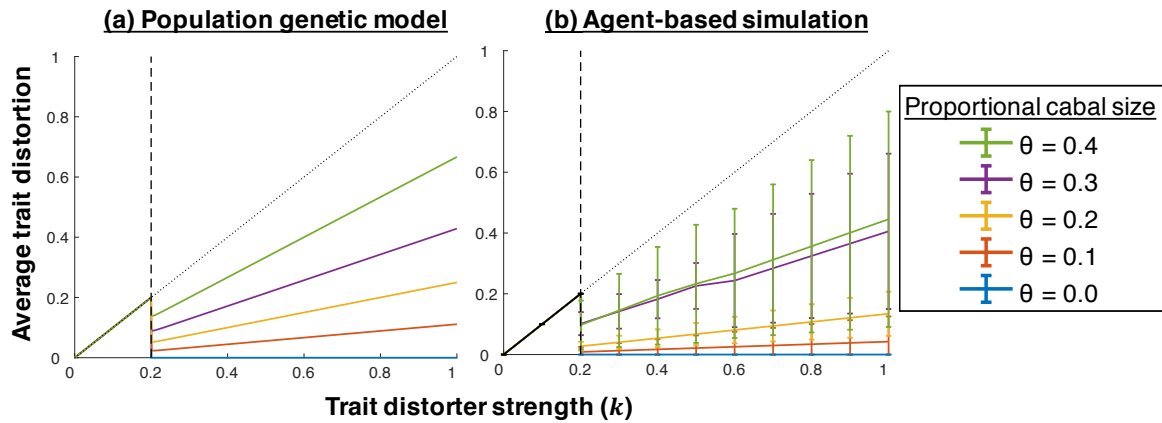


Supplementary Figure 12. Types of trait distorter interaction. (a) Trait distorters may arise at different loci within the cabal (e.g. red and blue rectangle markers). Low-sophistication trait distorters are more mutationally accessible than high-sophistication trait distorters ($\rho_{D1L}=2*\rho_{D1H}$), so arise more frequently.

(b) A dedicated suppressor of a trait distorter at a specific locus (e.g. red rectangle marker) may arise at some locus within the commonwealth (e.g. red circle marker). The dedicated suppressor suppresses its target trait distorter with full strength (red solid arrow). If trait distorters are low-sophistication, dedicated suppressors also suppress non-target trait distorters (e.g. blue rectangle marker) with partial ($z=0.5$) strength (blue dashed arrow). High-sophistication trait distorters are invulnerable to non-target suppression ($z=0$).

(c) Of all trait distorters across a genome, the one that is most trait-distorting after suppression exhibits inter-locus dominance, and distorts the individual trait. Expression of the inter-locus recessive trait distorters results in an individual-level cost (c_{rec}), which is greater for low-sophistication trait distorters ($c_{trait}(Dist) = \frac{c_{rec}(Waste)}{|I_{distorter}|-1} \geq 0$) than high-sophistication ones

$$\left(c_{trait}(Dist) = \frac{5(c_{rec}(Waste))}{3(|I_{distorter}|-1)} \geq 0 \right).$$

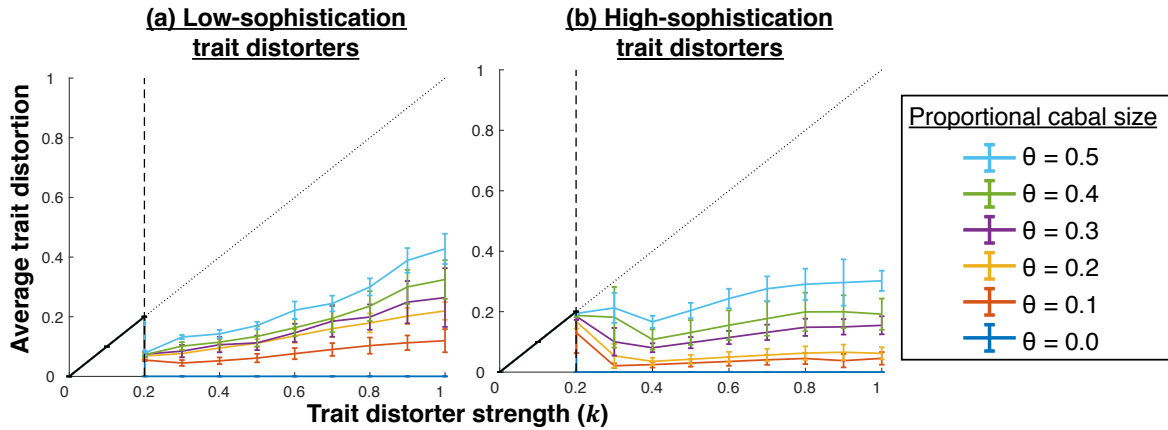


Supplementary Figure 13. Comparison of population genetic and agent-based simulation

results when trait distorters do not co-segregate. Trait distorters (D_i) and their dedicated suppressors (S_i) are continuously introduced, from rarity, at different loci within the cabal and the commonwealth respectively. The resulting average individual trait distortion, taken over evolutionary time, is plotted, for different proportional cabal sizes (θ), against the extent to which the trait distorters cause trait values to deviate from the individual optimum (k). The following parameter values are assumed: $c_{sup}=0.1$; $t=k$, $c_{trait}=k/2$, $\gamma=10^6$, $\rho_{S1}=10^{-11}$, $\rho_{D1}=10^{-11}$. On these assumptions, trait distorters favouring suppression ($c_{sup} < c_{trait}(k)$), which lie to the right of the dashed lines, scarcely co-segregate.

Part (a) plots numerical solutions for the population genetic model described in *Methods: Long term trait distortion (exact numerical solution)*. Part (b) plots the average results from 4 runs of the agent-based simulation model described in *Methods: Agent-based simulation (multiple loci; discrete)*, where each simulation is run for $T_{end}=10^6$ generations, and where error bars represent one standard deviation in each direction. For the simulations, we arbitrarily assume that trait distorters are low-sophistication (D_{1L}) as opposed to high-sophistication (D_{1H}). However, this choice is inconsequential given that the characteristics of trait distorter interaction do not affect average trait distortion when trait distorters scarcely co-segregate.

Given the exceedingly low probabilities of trait distorter / suppressor introduction in these parameterisations (very high stochasticity), the simulation results underestimate average trait distortion, and are highly variable (large error bars). The simulation results are underestimates because, as individual simulation runs are finite (T_{end}), they may end before rare trait distorters have been completely purged. Nevertheless, the results of the two models are consistent, and increasingly converge as simulation run times (T_{end}) are increased.



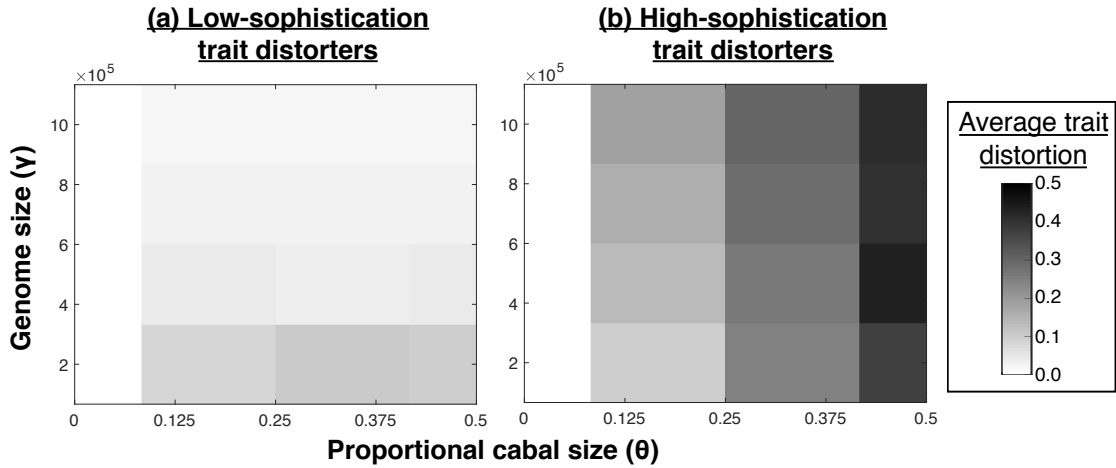
Supplementary Figure 14. The effect of trait distorter strength (k) when trait distorters may co-segregate. Trait distorters (D_i) and their dedicated suppressors (S_i) are continuously introduced, from rarity, at different loci within the cabal and the commonwealth respectively (*Methods: Agent-based simulation (multiple loci; discrete)*). The resulting average individual trait distortion, taken over $T_{end}=30,000$ generations, is plotted as an average over 4 simulation runs, with error bars representing one standard deviation in each direction. Average trait distortion is plotted for different levels of trait distorter sophistication (low; high), and different proportional cabal sizes (θ), against the extent to which the trait distorters cause trait values to deviate from the individual optimum (k).

Weaker trait distorters (k) are suppressed and purged more slowly than stronger ones, and are therefore more likely to co-segregate. As a result, when trait distorters are (a) low-sophistication (D_{1L}), average trait distortion is pulled below linearity (Equation 6; Main Text) when trait distorters are weak, leading to an accelerating relationship between average trait distortion and trait distorter strength (k). When trait distorters are (b) high-sophistication (D_{1H}), average trait distortion is pulled above linearity (Equation 6; Main Text) when trait distorters are weak, meaning the difference in average trait distortion caused by trait distorters of different strengths (k) is reduced (flatter relationship).

For reduced proportional cabal size (θ), average trait distortion is reduced, and the relationship between trait distorter strength (k) and average trait distortion tends towards linearity (Equation 6; Main Text).

The following parameter values were taken: $c_{sup}=0.1$, $t=k$, $c_{trait}=Dist/2$, $\gamma=10^6$, $\rho_{S1}=4*10^{-9}$, $\rho_{D1L}=4*10^{-9}$, $\rho_{D1H}=2*10^{-9}$. For these parameters, genome size (γ) and baseline mutation rate

$(\rho_{S1}/\rho_{D1L}/\rho_{D1H})$ are not exceedingly high. As a result, the increased trait distortion achieved by high-sophistication trait distorters as a result of productive interaction whilst co-segregating is roughly offset by the increased trait distortion achieved by low-sophistication trait distorters as a result of higher mutational accessibility. This is why trait distorter sophistication has a relatively small effect on average trait distortion, as can be seen by comparing (a) and (b).



Supplementary Figure 15. Evolution of trait distortion and suppression across the genome.

The figures show how the average level of trait distortion depends upon the cabal (θ) and genome (γ) size, for both (a) low- and (b) high-sophistication trait distorters. The results are from our agent-based simulation (*Methods: Agent-based simulation (multiple loci; continuous)*), where trait distorters (D_i) and their dedicated suppressors (S_i) are continuously introduced. Trait distorters and suppressors vary continuously in strength, and are free to evolve. Each block is the average of 30 simulation runs, each over $T_{end}=30,000$ generations. Average trait distortion increases with cabal size (θ). Low-sophistication trait distorters interact counter-productively whilst co-segregating, and so average trait distortion decreases with genome size (γ). High-sophistication trait distorters interact productively whilst co-segregating, and so average trait distortion increases with genome size ($c_{sup}=0.01$, $t=k$, $c_{trait}=Dist/2$, $\rho_{S1}=4*10^{-9}$, $\rho_{D1L}=4*10^{-9}$, $\rho_{D1H}=2*10^{-9}$).

Supplementary Note 8

Real-World Estimates of Proportional Cabal Size

Y chromosome cabal in *Drosophila melanogaster*

- A review of *Drosophila* Y chromosome evolution pulls together indirect evidence to suggest that, although only 12 genes are currently known, there is an upper bound of 20 genes on the *D.melanogaster* Y chromosome¹⁰⁵. We use this upper bound for our calculation of proportional cabal size, meaning proportional cabal size is likely to be an overestimate.
- The total number of genes in the *D.melanogaster* genome is 17,684,¹⁰⁶ meaning the proportional cabal size can be calculated as $\theta = 20/17684 = 0.001$ (1sf).

Cytoplasmic element & X chromosome cabal in humans

- In humans, the only cytoplasmic elements that carry transcribed genes are the mitochondria. Human mitochondria bear 37 genes¹⁰⁷.
- The number of genes on the human X chromosome (protein coding genes plus non-coding RNA genes) is 1515 (Ensembl release 97 - July 2019)¹⁰⁸.
- The total number of genes in the human genome is 42611¹⁰⁹.
- In human females, the proportional size of the cabal favouring female sex ratio distortion can be calculated as $\theta = (37 + 1515) / 42611 = 0.04$ (1sf).

Plasmid cabal in *Escherichia coli*

- Different *E.coli* individuals will carry different numbers and types of plasmids. We therefore draw a random sample of 139 *E.coli* strains from the 875 *E.coli*

strains for which complete genome sequences are publicly available

(Genbank Refseq; <ftp://ftp.ncbi.nih.gov/genomes>).

- For each strain in our sample, we calculate proportional cabal size by counting the number of genes that are on plasmids, and dividing this by the total number of genes in the individual.
- Averaging over strains, we calculate proportional cabal size as $\theta=0.036$ (2 *sf*).

Supplementary Discussion 1

Relation to Gardner & Úbeda (2017) and Grafen's Formal Darwinism

Gardner & Welch (2011)⁹⁵ provided formal justification for the view that genes evolve so as to maximise their own fitness (*selfish gene theory*). Specifically, they showed that, over evolutionary time, the allele-variants that come to occupy positions at loci are those variants that maximise the inclusive fitness of the gene. Following this formalism, Gardner & Úbeda (2017)⁵³ defined intra-genomic conflict as instances where the evolutionary interests of different genes, as determined by what maximises their respective inclusive fitnesses, do not coincide. The approach of Gardner & Úbeda (2017)⁵³ is useful because it clarifies the evolutionary “battleground” over which intra-genomic conflict can play out. If a “battleground” model establishes the *causes* of intra-genomic conflict, then a “resolution” model addresses the *consequences*. Our models are resolution models, and complementary to the battleground models described in Gardner & Úbeda (2017)⁵³.

The theoretical justification for organismal fitness maximisation is found the optimisation models of Grafen's Formal Darwinism project⁹⁶. This formalism assumes that organism phenotype is controlled solely by genes in a single co-replicon, in which genes are unimprinted, autosomal and inherited in Mendelian fashion⁶. Our models follow Burt & Trivers (2006)⁵² in taking organismal fitness maximisation as a starting point, and then addressing how robust this formalism is once nascent selfish genetic elements, residing in minority co-replicons, can gain

some control over organism phenotype. Our models are therefore complementary to those of Grafen's Formal Darwinism project.

Individual fitness maximisation: emergence versus maintenance

Specifically, our models show that, if individuals are maximising their fitness with respect to a given trait under potential conflict, then attempts to distort the trait from individual fitness maximisation, driven by selfish genetic elements arising in coreplicons representing minority-interests in the genome (cabal), will by and large be futile, unless the cabal is relatively large in size (approaching half of the genome). Therefore, our models provide justification for the idea that, once an organism has obtained fitness maximisation, it cannot, in general, be appreciably distorted by the subsequent invasion of trait-distorting elements.

However, there is a bias in our methodology. We assumed that the organism is already maximising its fitness, and then showed that subsequent distortions from this maximand will often be negligible. This bias is evident in the strategy set afforded to different alleles across the genome: the minority-interest within the genome (cabal) can only exert influence over the trait via 'trait distorters' (they distort the trait– they cause a shift away from the norm), whereas the majority-interest within the genome (commonwealth) can only exert influence over the trait via 'suppressors' (they restore – they cause a shift back towards the norm). Therefore, what our models really show is that, for traits under intragenomic conflict, individual fitness maximisation can be maintained. They do not show that individual fitness maximisation is obtained in the first place (emerges). The question of whether

individual fitness maximisation emerges when traits are underpinned by conflicting coreplicons is a direction for future research.

Supplementary Discussion 2

Simple Selfish Genetic Elements vs Trait Distorters

We draw a distinction between two types of selfish genetic element. “Simple” selfish genetic elements (SGEs), such as transposons and simple meiotic drivers, do not need to manipulate organism traits in order to give themselves a selfish propagation advantage. The spread of a simple SGE may have detrimental consequences for the organism, for example due to the disruptive act of driving itself⁹⁷. However, these costs are disfavoured across the whole genome, including by the simple SGE in question, and there will be unanimous selection across the genome to attenuate these costs. The spread of simple SGEs therefore does not generate intra-genomic conflict over organism form. Their existence therefore does not compromise organismal design (individual fitness maximisation).

Simple meiotic drivers may incur fertility costs from the act of driving, which will be alleviated once the driver reaches fixation and stops driving⁹⁷. Meiotic drivers may bring costly linked genes to fixation, but these costs will not be recoverable via the suppression of the meiotic driver itself, which is uncostly at fixation. Therefore, meiotic drivers will not generate selection for suppression once they have reached fixation⁹⁸. Furthermore, the evolution of meiotic drivers, to increase or decrease their drive strength, will not affect the strength of selection for their suppression, so long as the new mutant drivers reach fixation before a relevant suppressor arises. Finally, because meiotic drivers at fixation do not have a predictable effect on organism phenotype, hybrid crosses, revealing selfish genetic elements, cannot tell apart meiotic drivers that are at fixation from meiotic drivers that are under suppression⁹⁹.

The second type of selfish genetic element, which are considered in this study, are “trait distorters”. For these SGEs to selfishly propagate themselves, they need to manipulate an organism trait in a specific direction. Our illustrative models focused on a hypothetical trait distorter that manipulated some undefined organism trait, which allowed it to gain a propagation advantage by driving at meiosis. Trait distortion was necessary to facilitate drive, which is what distinguishes our hypothetical trait distorter from a simple meiotic driver. Trait distorters have lasting, predictable effects on the organism after they have reached fixation in the population, compromising individual fitness maximisation. Suppression is therefore favoured even after the trait distorter has reached fixation, and is increasingly favoured as the trait distorter evolves to be more trait-distorting. Finally, hybrid crosses can reveal trait distorters under suppression. If the hybrids express trait distortion whereas the parents do not, the trait distorters were under suppression in the parents¹⁰⁰.

Relation to Eshel (1984) and Eshel (1985)

Eshel¹⁰¹ highlighted the conflict between individual fitness maximisation and selfish genetic elements. Eshel⁹⁸ also pointed out that suppressors of simple meiotic drivers will spread as long as (i) the driver is below fixation, and (ii) the suppressor is unlinked. In doing so, he pointed out that fair meiosis can be stabilised if there is free-recombination between genes. There are a few key differences between our models and the models of meiotic drive suppression developed by Eshel⁹⁸ and others^{48,102,103}.

Eshel⁹⁸ modelled a simple meiotic driver and an unlinked suppressor. The driver may exert an individual cost in heterozygous and/or homozygous form. The suppressor completely suppresses drive, at no cost to the individual. However, upon suppression, the individual cost of drive is not recovered. On these assumptions, the suppressor is favoured via individuals that are heterozygous for the driver. For driver-heterozygotes, those individuals that bear the suppressor will have a lower proportion of offspring that inherit the costly driver. As a result, driver-heterozygotes that bear the suppressor will have more grandchildren than driver-heterozygotes that lack the suppressor^{48,98}. An implication of this is that, once the driver has gone to fixation and driver-heterozygotes have consequently diminished, there is no further selection on the suppressor. A second implication is that, because the suppressor is cost-free, it will always spread whilst the driver is present and below fixation, regardless of the individual cost or the transmission advantage associated with the driver.

In our illustrative model, and in contrast to Eshel's⁹⁸, the individual cost associated with the driver is recovered when the driver is suppressed. As a result of this, the suppressor can be favoured as a direct consequence of recovering the costs associated with the driver. Suppressor selection need not rely on increasing the number of grandchildren of driver-heterozygotes. In our model, suppressors are therefore still favoured even after the driver has gone to fixation. Furthermore, in our model, and in contrast to Eshel's, the act of suppression incurs an individual cost. As

a result of this, suppressors are not universally favoured, but rather, they are only favoured when drivers are sufficiently costly.

Ours and Eshel's⁹⁸ model address different biological scenarios. Non-recoverable costs of drive, as assumed by Eshel, are likely to stem from linked deleterious genes, and not from any systematic distortion of individual traits. This scenario applies, for example, to *Segregation Distorter (SD)* in *Drosophila melanogaster*, which drives without systematically biasing organism traits. In accordance with Eshel's model, empirical observation of *Segregation Distorter (SD)* in natural populations demonstrates that unlinked suppressors spread easily, but only if the driver is below fixation¹⁰⁴.

In contrast, if the individual costs stem directly from the expression of the driver, rather than any unlinked genes, recoverable costs of drive are appropriate⁹⁷. In this case, suppression of the driver also removes the individual level cost. This scenario applies to cases where the meiotic driver is not just a meiotic driver *per se*, but rather, a "trait distorter" that gains the ability to drive at meiosis as a consequence of distorting an organism trait. Our models demonstrate that trait distorters, unlike the simple meiotic drivers considered by Eshel⁹⁸, can promote suppressor spread even after they have gone to fixation. As a result, costly trait distorters ($c_{trait}(k) > c_{sup}$) will be suppressed in the evolutionary long term, even if they can reach fixation in the evolutionary short term.

Our models also expand on Eshel⁹⁸ in addressing *how likely* it is that a suppressor will spread. Eshel⁹⁸ demonstrated that a cost-free unlinked suppressor can spread in response to a costly meiotic driver. Our models account for a cost of suppression, to show that the likelihood that a trait distorter is suppressed correlates with the costliness of the driver to the individual, which serves to limit deviation from individual fitness maximisation.

Supplementary Discussion 3

Relation to Cosmides & Tooby (1981): Coreplicons, Cabal & Commonwealth

An anonymous referee suggested that, were we to extend our models to permit trait distorter introduction at any locus in the genome, rather than at a subsection of loci that are chosen *a priori* (cabal), the resulting trait distortion may be greater. In this section, we explicitly clarify why cabals are defined *a priori* by showing how they follow from the ‘coreplicon’ concept introduced by Cosmides & Tooby (1981)⁶. We then undertake this suggested modelling extension, showing that the scenario it depicts: (i) leads to the same results as our models, but (ii) is biologically implausible.

The coreplicon concept

Cosmides & Tooby (1981)⁶ pointed out that we can divide a genome up into ‘coreplicons’. A coreplicon comprises a collection of loci within the genome that are inherited in the same way, and so share the same maximand. Autosomal loci and X chromosome loci do not form part of the same coreplicon, because the former are transmitted equally through males and females and the latter are transmitted predominantly through females. Coreplicons are assigned, *a priori*, based on inheritance patterns – not on the basis of trait-affecting alleles that have been observed empirically or within the context of a theoretical model. The coreplicon concept has been employed regularly in the study of intragenomic conflict and evolutionary adaptation^{52,110,111}.

Coreplicons have the potential to be in conflict over organism traits. If, for a given trait, loci within coreplicon X are propagated best when the organism trait value is x , but loci within coreplicon Y are propagated best when the organism trait value is y , then the coreplicons have the potential to be in conflict if the current organism trait value is between x and y . This evolutionary battleground ('potential conflict') is derived *a priori* based on a purely theoretical, first principles optimisation approach, as detailed in Gardner & Úbeda (2017)⁵³. The evolutionary battleground for conflict is independent of whether any conflicting, trait-affecting alleles actually exist at any of the loci ('actual conflict')¹¹².

Sometimes, different coreplicons may form alliances, because they both benefit from a particular kind of trait distortion. For instance, if coreplicon Z is propagated best when the organism trait is z , where z lies in between x and y but is closer to x , coreplicon Z may ally with coreplicon X if the current organism trait value lies at y . Though the coreplicons may ally here, they may disagree over the form of other traits. This is where the concepts of 'cabal' and 'commonwealth' are useful. For example, in humans, cytoplasmic elements are inherited exclusively through females, and X chromosomes are inherited predominantly though not exclusively through females, meaning they represent different coreplicons. However, the coreplicons form a cabal with respect to sex ratio, favouring a female bias.

The cabal / commonwealth concept

The cabal comprises all coreplicons that favour the distortion of a particular trait, along a particular axis, in a particular direction, away from individual fitness maximisation. The commonwealth comprises the remaining coreplicons. Cabals and commonwealths are therefore trait-specific. It is useful, when analysing a specific trait, to partition the genome along these lines, because it is the resolution of this conflict – between the cabal and commonwealth – that gives the evolved deviation of a trait from individual fitness maximisation. Cabals and commonwealths are defined *a priori*, by partitioning and summing up the coreplicons that respectively disfavour and favour the trait distortion under study.

Our models address whether selfish genetic elements can distort organism traits away from individual fitness maximisation, where the ‘individual’ here really means the majority interest within the parliament of genes¹¹¹. This is why we only considered cabal sizes of up to a half. If the cabal was greater than half of the genome, it would reflect the majority interest within the parliament, so would cease to be a cabal. Our models therefore consider the full range of scenarios depicting potential distortion of organism traits from individual fitness maximisation.

Modelling extension

Having justified our approach, which defines the cabal and commonwealth *a priori*, we now undertake the theoretical exercise suggested by the anonymous reviewer, and allow trait distorters to arise at any locus in the genome, and not just at an *a priori* subsection (cabal).

We first note that this scenario is biologically implausible. In reality, most sites in the genomes of biological organisms *cannot* become trait distorters. Most loci in a genome are unimprinted, vertically inherited and autosomal. Therefore, for an organism approximating individual fitness maximisation, no conceivable distortion of an organism trait could possibly give these loci a propagation advantage. Meiotic drivers or transposons could arise at any of these loci, and the resulting selfish genetic elements could spread through the population as a result. However, *trait distorters* could not arise at these loci – the transmission of alleles at these loci is maximised when the organism trait values are those which lead to individual fitness maximisation^{110,111}. The key difference here is between meiotic drive (could be favoured at any locus; selfish benefit does not arise via distorting a trait) and selfish genetic elements that gain a benefit by distorting a trait (such as the specific examples that we consider and model in this paper)^{52,53}.

Nevertheless, we will imagine a hypothetical organism where any site in its genome could give rise to a trait distorter. The question then becomes: what type of trait distortion is favoured at each locus? It could firstly be the case that each locus gains its selfish propagation advantage by distorting a unique trait, or by distorting a common trait but along a unique dimension (axis) and direction. If this is the case, each locus in the genome would effectively form its own cabal, with a proportional size within the genome approximating zero ($\theta \rightarrow 0$). It could alternatively be the case that groups of loci favour the same type of trait distortion (same trait, dimension and direction), meaning proportional cabal sizes can be larger ($\theta > 0$). However, given that the size of any one cabal cannot exceed a half (else that group of loci would cease

to be a cabal), it must logically be the case that (at least two) different types of trait distortion are favoured across the genome.

We now assume that the rate of trait distorter introduction, per generation, per locus, in some genome within the population, is ρ_{D1} . We take the number of loci within a genome to be γ , which means that new trait distorters are introduced into the population every $1/(\rho_{D1}\gamma)$ generations. This is a faster rate than previously considered in our Dynamics models, which was dependent on proportional cabal size ($1/(\theta\rho_{D1}\gamma)$). As was the case in our Dynamics models, the suppressor of a given trait distorter will be expected to arise after a lag of $(1/(1-\theta)\rho_{S1}\gamma)$ generations, where ρ_{S1} is the rate of suppressor introduction, per generation, per locus, for any locus situated outside of the target trait distorter's cabal.

So in this new theoretical scenario, compared to our previous Dynamics models, trait distorters are arising at a faster rate, but they are suppressed at the same rate as before. This would apparently suggest that average trait distortion should be more appreciable in this new scenario. However, this is not the case. The rate that trait distorters *that distort a given trait* are introduced is the same as our Dynamics models ($1/(\theta\rho_{D1}\gamma)$). This new formulation appears to favour increased deviation of organisms from individual fitness maximisation, but this is not the case, as the new scenario is implicitly considering the distortion of multiple traits simultaneously.

The distortion of any *given trait* from individual fitness maximisation in this new theoretical scenario is still accurately given by our Dynamics models. Specifically, in

this new theoretical scenario, if trait distorters belonging to the same cabal arise at new loci in the genome very slowly compared to the rate at which gene frequencies equilibrate after trait distorter / suppressor introduction (separation of timescales), the trait that the cabal is attempting to distort assumes an average value, in individuals over evolutionary time, given by Equation 6 in the main text. If trait distorters belonging to the same cabal arise more quickly than this, such that they may co-segregate, the trait that the cabal is attempting to distort assumes an average value that is given by the simulation results of our Dynamics models. This holds regardless of the overall rate of trait distorter introduction across the whole genome.

Therefore, the scenario in which trait distorters may arise at any locus in the genome implicitly refers to a scenario where multiple traits are being distorted and restored simultaneously, in the context of a single model. However, there is no reason why the evolution of distortion and suppression at one trait should be affected by the evolution of distortion and suppression at any other trait. Consequently, the results of the new theoretical scenario converge on our Dynamics models once we consider a single type of trait distortion in isolation. Our models cover the full range of scenarios depicting potential distortion of an organism trait from individual fitness maximisation. The modelling extension, as well as being biologically implausible, provides no additional insight.

Supplementary References

1. Fisher, R. A. *The Genetical Theory of Natural Selection*. (Oxford University Press, 1930).
2. Charnov, E. L. *The Theory of Sex Allocation*. (Princeton University Press, 1982).
3. West, S. *Sex Allocation*. (Princeton University Press, 2009).
doi:10.1515/9781400832019
4. Bull, J. J. *Evolution of sex determining mechanisms*. (Benjamin-Cummings Publishing Company, 1983).
5. Jaenike, J. Sex Chromosome Meiotic Drive. *Annual Review of Ecology and Systematics* **32**, 25–49 (2001).
6. Cosmides, L. M. & Tooby, J. Cytoplasmic inheritance and intragenomic conflict. *Journal of Theoretical Biology* **89**, 83–129 (1981).
7. Helleu, Q., Gérard, P. R. & Montchamp-Moreau, C. Sex Chromosome Drive. *Cold Spring Harbor Perspectives in Biology* **7**, a017616–16 (2015).
8. Hickey, W. A. & Craig, G. B. Genetic Distortion of Sex Ratio in a Mosquito *Aedes Aegypti*. *Genetics* **53**, 1177–& (1966).
9. Policansky, D. & Ellison, J. Sex-Ratio in *Drosophila-Pseudoobscura* - Spermio-genic Failure. *Science* **169**, 888–+ (1970).
10. Hauschteckjungen, E. & Maurer, B. Sperm Dysfunction in Sex-Ratio Males of *Drosophila-Subobscura*. *Genetica* **46**, 459–477 (1976).
11. Wu, C. I. Virility Deficiency and the Sex-Ratio Trait in *Drosophila-Pseudoobscura*. I. Sperm Displacement and Sexual Selection. *Genetics* **105**, 651–662 (1983).

12. Montchamp-Moreau, C. & Joly, D. Abnormal spermiogenesis is associated with the X-linked sex-ratio trait in *Drosophila simulans*. *Heredity* **79**, 24–30 (1997).
13. Presgraves, D. C., Severance, E. & Wilkinson, G. S. Sex chromosome meiotic drive in stalk-eyed flies. *Genetics* **147**, 1169–1180 (1997).
14. Wu, C. I. Virility Deficiency and the Sex-Ratio Trait in *Drosophila-Pseudoobscura*. II. Multiple Mating and Overall Virility Selection. *Genetics* **105**, 663–679 (1983).
15. Taylor, D. R., Saur, M. J. & Adams, E. Pollen performance and sex-ratio evolution in a dioecious plant. *Evolution* **53**, 1028–1036 (1999).
16. Wilkinson, G. S. & Sanchez, M. I. Sperm development, age and sex chromosome meiotic drive in the stalk-eyed fly, *Cyrtodiopsis whitei*. *Heredity* **87**, 17–24 (2001).
17. Wilkinson, G. S., Johns, P. M., Kelleher, E. S., Muscedere, M. L. & Lorsch, A. Fitness effects of X chromosome drive in the stalk-eyed fly, *Cyrtodiopsis dalmanni*. *Journal of Evolutionary Biology* **19**, 1851–1860 (2006).
18. Angelard, C., Montchamp-Moreau, C. & Joly, D. Female-driven mechanisms, ejaculate size and quality contribute to the lower fertility of sex-ratio distorter males in *Drosophila simulans*. *BMC Evol Biol* **8**, 326–12 (2008).
19. Price, T. A. R. *et al.* Sex ratio distorter reduces sperm competitive ability in an insect. *Evolution* **62**, 1644–1652 (2008).
20. Thomson, G. J. & Feldman, M. W. Population genetics of modifiers of meiotic drive: IV. On the evolution of sex-ratio distortion. *Theoretical Population Biology* **8**, 202–211 (1975).

21. Hamilton, W. D. Extraordinary sex ratios. A sex-ratio theory for sex linkage and inbreeding has new implications in cytogenetics and entomology. *Science* **156**, 477–488 (1967).
22. Jaenike, J. Sex-Ratio Meiotic Drive in the *Drosophila quinaria* Group. *The American Naturalist* **148**, 237–254 (1996).
23. Atlan, A., Joly, D., Capillon, C. & Montchamp-Moreau, C. Sex-ratio distorter of *Drosophila simulans* reduces male productivity and sperm competition ability. *Journal of Evolutionary Biology* **17**, 744–751 (2004).
24. Beckenbach, A. Multiple Mating and the ‘Sex-Ratio’ Trait in *Drosophila pseudoobscura*. *Evolution* **35**, 275–281 (1981).
25. Taylor, J. E. & Jaenike, J. Sperm competition and the dynamics of X chromosome drive: stability and extinction. *Genetics* **160**, 1721–1731 (2002).
26. Edwards, A. W. F. The population genetics of ‘sex-ratio’ in *Drosophila pseudoobscura*. *Heredity* **16**, 291–304 (1961).
27. Curtsinger, J. W. & Feldman, M. W. Experimental and Theoretical Analysis of the ‘Sex-Ratio’ Polymorphism in *Drosophila Pseudoobscura*. *Genetics* **94**, 445–466 (1980).
28. Stalker, H. D. The Genetic Systems Modifying Meiotic Drive in *Drosophila Paramelanica*. *Genetics* **46**, 177–202 (1961).
29. Klaczko, L. B. & Carvalho, A. B. Autosomal Suppressors of Sex-Ratio in *Drosophila-Mediopunctata*. *Heredity* **71**, 546–551 (1993).
30. Jaenike, J. Suppression of sex-ratio meiotic drive and the maintenance of Y-chromosome polymorphism in *Drosophila*. *Evolution* **53**, 164 (1999).
31. Cazemajor, M., Landre, C. & Montchamp-Moreau, C. The Sex-Ratio Trait in

- Drosophila simulans*: Genetic Analysis of Distortion and Suppression. *Genetics* **147**, 635–642 (1997).
32. Tao, Y., Hartl, D. L. & Laurie, C. C. Sex-ratio segregation distortion associated with reproductive isolation in *Drosophila*. *Proc Natl Acad Sci USA* **98**, 1–6 (2001).
 33. Tao, Y., Masly, J. P., Araripe, L., Ke, Y. & Hartl, D. L. A sex-ratio meiotic drive system in *Drosophila simulans*. I: an autosomal suppressor. *PLoS Biol* **5**, e292 (2007).
 34. Ferree, P. M. & Barbash, D. A. Distorted Sex Ratios: A Window into RNAi-Mediated Silencing. *PLoS Biol* **5**, e303–5 (2007).
 35. Hornett, E. A. *et al.* The Evolution of Sex Ratio Distorter Suppression Affects a 25 cM Genomic Region in the Butterfly *Hypolimnas bolina*. *PLoS Genet* **10**, e1004822–12 (2014).
 36. Vaz, S. C. & Carvalho, A. B. Evolution of autosomal suppression of the sex-ratio trait in *Drosophila*. *Genetics* **166**, 265–277 (2004).
 37. Carvalho, Vaz, S. C. & Klaczko, L. B. Polymorphism for Y-linked suppressors of sex-ratio in two natural populations of *Drosophila mediopunctata*. *Genetics* **146**, 891–902 (1997).
 38. Capillon, C. & Atlan, A. Evolution of Driving X Chromosomes and Resistance Factors in Experimental Populations of *Drosophila simulans*. *Evolution* **53**, 506 (1999).
 39. Wu, C. I. The fate of autosomal modifiers of the sex-ratio trait in *Drosophila* and other sex-linked meiotic drive systems. *Theoretical Population Biology* **24**, 107–120 (1983).

40. Atlan, A., Capillon, C., Derome, N., Couvet, D. & Montchamp-Moreau, C. The Evolution of Autosomal Suppressors of Sex-Ratio Drive in *Drosophila Simulans*. *Genetica* **117**, 47–58 (2003).
41. Rood, E. S. & Freedberg, S. Intragenomic conflict produces sex ratio dynamics that favor maternal sex ratio distorters. *Ecol Evol* **6**, 8085–8093 (2016).
42. Bastide, H., Gérard, P. R., Ogereau, D., Cazemajor, M. & Montchamp-Moreau, C. Local dynamics of a fast-evolving sex-ratiosystem in *Drosophila simulans*. *Molecular Ecology* **22**, 5352–5367 (2013).
43. Werren, J. H. The coevolution of autosomal and cytoplasmic sex ratio factors. *Journal of Theoretical Biology* **124**, 317–334 (1987).
44. Maynard Smith, J. & Price, G. R. The Logic of Animal Conflict. *Nature* **246**, 15–18 (1973).
45. Tao, Y. *et al.* A sex-ratio meiotic drive system in *Drosophila simulans*. II: an X-linked distorter. *PLoS Biol* **5**, e293 (2007).
46. Wu, C. I. & Beckenbach, A. T. Evidence for Extensive Genetic Differentiation Between the Sex-Ratio and the Standard Arrangement of *Drosophila-Pseudoobscura* and *Drosophila-Persimilis* and Identification of Hybrid Sterility Factors. *Genetics* **105**, 71–86 (1983).
47. Dyer, K. A., Charlesworth, B. & Jaenike, J. Chromosome-wide linkage disequilibrium as a consequence of meiotic drive. *Proc Natl Acad Sci USA* **104**, 1587–1592 (2007).
48. Crow, J. F. Why is mendelian segregation so exact? *Bioessays* **13**, 305–312 (1991).

49. Wild, G. & Traulsen, A. The different limits of weak selection and the evolutionary dynamics of finite populations. *Journal of Theoretical Biology* **247**, 382–390 (2007).
50. Peters, J. The role of genomic imprinting in biology and disease: an expanding view. *Nature Publishing Group* **15**, 517–530 (2014).
51. Hamilton, W. D. The genetical evolution of social behaviour. I. *Journal of Theoretical Biology* **7**, 1–16 (1964).
52. Burt, A. & Trivers, R. *Genes in Conflict*. (Harvard University Press, 2006).
doi:10.4159/9780674029118
53. Gardner, A. & Úbeda, F. The meaning of intragenomic conflict. *Nature Ecology & Evolution* **1**, 1–9 (2017).
54. Haig, D. *Genomic Imprinting and Kinship*. (Rutgers University Press, 2002).
55. Úbeda, F. & Gardner, A. A model for genomic imprinting in the social brain: juveniles. *Evolution* **64**, 2587–2600 (2010).
56. Queller, D. C. Theory of genomic imprinting conflict in social insects. *BMC Evol Biol* **3**, 15 (2003).
57. Haig, D. Parental antagonism, relatedness asymmetries, and genomic imprinting. *Proceedings of the Royal Society B: Biological Sciences* **264**, 1–6 (1997).
58. Haig, D. Genomic imprinting, sex-biased dispersal, and social behavior. *Ann. N. Y. Acad. Sci.* **907**, 149–163 (2000).
59. Úbeda, F. & Gardner, A. A model for genomic imprinting in the social brain: adults. *Evolution* **65**, 462–475 (2011).
60. Úbeda, F. & Gardner, A. A model for genomic imprinting in the social brain:

- elders. *Evolution* **66**, 1567–1581 (2012).
61. Haig, D. Coadaptation and conflict, misconception and muddle, in the evolution of genomic imprinting. *Heredity* **113**, 96–103 (2013).
 62. Kondoh, M. & Higashi, M. Reproductive Isolation Mechanism Resulting from Resolution of Intragenomic Conflict. *The American Naturalist* **156**, 511–518 (2000).
 63. Wilkins, J. F. & Haig, D. Genomic imprinting of two antagonistic loci. *Proceedings of the Royal Society B: Biological Sciences* **268**, 1861–1867 (2001).
 64. Wilkins, J. F. Genomic imprinting and conflict-induced decanalization. *Evolution* **65**, 537–553 (2011).
 65. Wilkins, J. F. Antagonistic coevolution of two imprinted loci with pleiotropic effects. *Evolution* **64**, 142–151 (2010).
 66. Gregg, C. *et al.* High-Resolution Analysis of Parent-of-Origin Allelic Expression in the Mouse Brain. *Science* **329**, 643–648 (2010).
 67. Galbraith, D. A. *et al.* Testing the kinship theory of intragenomic conflict in honey bees (*Apis mellifera*). *Proc Natl Acad Sci USA* **113**, 1020–1025 (2016).
 68. Foster, K. R. Diminishing returns in social evolution: the not-so-tragic commons. *Journal of Evolutionary Biology* **17**, 1058–1072 (2004).
 69. Haig, D. The Kinship Theory of Genomic Imprinting. *Annual Review of Ecology and Systematics* **31**, 9–32 (2000).
 70. Jacob, S., McClintock, M. K., Zelano, B. & Ober, C. Paternally inherited HLA alleles are associated with women's choice of male odor. *Nat Genet* **30**, 175–

- 179 (2002).
71. Isles, A. R. *et al.* A possible role for imprinted genes in inbreeding avoidance and dispersal from the natal area in mice. *Proceedings of the Royal Society B: Biological Sciences* **269**, 665–670 (2002).
 72. Isles, A. R., Baum, M. J., Ma, D., Keverne, E. B. & Allen, N. D. Urinary odour preferences in mice. *Nature* **409**, 783–784 (2001).
 73. Hager, R. & Johnstone, R. A. The genetic basis of family conflict resolution in mice. *Nature* **421**, 533–535 (2003).
 74. Gardner, A., West, S. A. & Wild, G. The genetical theory of kin selection. *Journal of Evolutionary Biology* **24**, 1020–1043 (2011).
 75. Rousset, F. *Genetic Structure and Selection in Subdivided Populations (MPB-40)*. (Princeton University Press, 2013).
 76. Gardner, A. & Ross, L. Mating ecology explains patterns of genome elimination. *Ecol Lett* **17**, 1602–1612 (2014).
 77. Burt, A. & Trivers, R. Genetic conflicts in genomic imprinting. *Proceedings of the Royal Society B: Biological Sciences* **265**, 2393–2397 (1998).
 78. Kafri, T., Gao, X. H. & Razin, A. Mechanistic Aspects of Genome-Wide Demethylation in the Preimplantation Mouse Embryo. *Proc Natl Acad Sci USA* **90**, 10558–10562 (1993).
 79. West, S. A., Diggle, S. P., Buckling, A., Gardner, A. & Griffin, A. S. The Social Lives of Microbes. *Annu. Rev. Ecol. Evol. Syst.* **38**, 53–77 (2007).
 80. Frank, S. A. A general model of the public goods dilemma. *Journal of Evolutionary Biology* **23**, 1245–1250 (2010).
 81. West, S. A. & Buckling, A. Cooperation, virulence and siderophore production

- in bacterial parasites. *Proceedings of the Royal Society B: Biological Sciences* **270**, 37–44 (2003).
82. Mc Ginty, S. E., Lehmann, L., Brown, S. P. & Rankin, D. J. The interplay between relatedness and horizontal gene transfer drives the evolution of plasmid-carried public goods. *Proceedings of the Royal Society B: Biological Sciences* **280**, 20130400–20130400 (2013).
83. Mc Ginty, S. É., Rankin, D. J. & Brown, S. P. Horizontal gene transfer and the evolution of bacterial cooperation. *Evolution* **65**, 21–32 (2011).
84. Dimitriu, T. *et al.* Genetic information transfer promotes cooperation in bacteria. *Proc Natl Acad Sci USA* **111**, 11103–11108 (2014).
85. Dimitriu, T. *et al.* Indirect Fitness Benefits Enable the Spread of Host Genes Promoting Costly Transfer of Beneficial Plasmids. *PLoS Biol* **14**, e1002478–28 (2016).
86. Dimitriu, T. *et al.* Selection of horizontal gene transfer through public good production. *bioRxiv* **106**, 1
87. Niehus, R., Mitri, S., Fletcher, A. G. & Foster, K. R. Migration and horizontal gene transfer divide microbial genomes into multiple niches. *Nat. Commun.* **6**, 1–9 (2015).
88. Ghaly, T. M. & Gillings, M. R. Mobile DNAs as Ecologically and Evolutionarily Independent Units of Life. *Trends in Microbiology* **26**, 904–912 (2018).
89. Smith, J. The social evolution of bacterial pathogenesis. *Proceedings of the Royal Society B: Biological Sciences* **268**, 61–69 (2001).
90. Rankin, D. J., Rocha, E. P. C. & Brown, S. P. What traits are carried on mobile genetic elements, and why? *Heredity* **106**, 1–10 (2011).

91. Nogueira, T. *et al.* Horizontal Gene Transfer of the Secretome Drives the Evolution of Bacterial Cooperation and Virulence. *Current Biology* **19**, 1683–1691 (2009).
92. Johnson, L. J. The Genome Strikes Back: The Evolutionary Importance of Defence Against Mobile Elements. *Evolutionary Biology* **34**, 121–129 (2007).
93. Mc Ginty, S. É. & Rankin, D. J. The evolution of conflict resolution between plasmids and their bacterial hosts. *Evolution* **66**, 1662–1670 (2012).
94. Ghoul, M. *et al.* Pyoverdinin cheats fail to invade bacterial populations in stationary phase. *Journal of Evolutionary Biology* **29**, 1–9 (2016).
95. Gardner, A. & Welch, J. J. A formal theory of the selfish gene. *Journal of Evolutionary Biology* **24**, 1801–1813 (2011).
96. Grafen, A. The formal darwinism project in outline. *Biol Philos* **29**, 155–174 (2014).
97. Zanders, S. E. & Unckless, R. L. Fertility Costs of Meiotic Drivers. *Current Biology* **29**, R512–R520 (2019).
98. Eshel, I. Evolutionary Genetic Stability of Mendelian Segregation and the Role of Free Recombination in the Chromosomal System. *The American Naturalist* **125**, 412–420 (1985).
99. Corbett-Detig, R., Medina, P., Frérot, H., Blassiau, C. & Castric, V. Bulk pollen sequencing reveals rapid evolution of segregation distortion in the male germline of *Arabidopsis* hybrids. *Evolution Letters* **129**, 1393–11 (2019).
100. Blows, M. W., Berrigan, D. & Gilchrist, G. W. Rapid evolution towards equal sex ratios in a system with heterogamety. *Evol Ecol Res* 277–283 (1999).
101. Eshel, I. Are intragametic conflicts common in nature? Do they represent an

- important factor in evolution? *Journal of Theoretical Biology* **108**, 159–162 (1984).
102. Feldman, M. W. & Otto, S. P. A Comparative Approach to the Population-Genetics Theory of Segregation Distortion. *The American Naturalist* **137**, 443–456 (1991).
103. Prout, T., Bundgaard, J. & Bryant, S. Population genetics of modifiers of meiotic drive I. The solution of a special case and some general implications. *Theoretical Population Biology* **4**, 446–465 (1973).
104. Larracuenta, A. M. & Presgraves, D. C. The selfish Segregation Distorter gene complex of *Drosophila melanogaster*. *Genetics* **192**, 33–53 (2012).
105. Bernardo Carvalho, A., Koerich, L. B. & Clark, A. G. Origin and evolution of Y chromosomes: *Drosophila* tales. *Trends in Genetics* **25**, 270–277 (2009).
106. Matthews, B. B. *et al.* Gene Model Annotations for *Drosophila melanogaster*: Impact of High-Throughput Data. *G3 (Bethesda)* **5**, 1721–1736 (2015).
107. Taanman, J. W. The mitochondrial genome: structure, transcription, translation and replication. *Biochim. Biophys. Acta* **1410**, 103–123 (1999).
108. Zerbino, D. R. *et al.* Ensembl 2018. *Nucleic Acids Research* **46**, D754–D761 (2017).
109. Perteira, M. *et al.* CHES: a new human gene catalog curated from thousands of large-scale RNA sequencing experiments reveals extensive transcriptional noise. *Genome Biol.* **19**, 1–14 (2018).
110. Haig, D. Genetic dissent and individual compromise. *Biol Philos* **29**, 233–239 (2014).
111. Grafen, A. Optimization of inclusive fitness. *Journal of Theoretical Biology*

238, 541–563 (2006).

112. Queller, D. C. & Strassmann, J. E. Evolutionary Conflict. *Annu. Rev. Ecol. Evol. Syst.* **49**, annurev-ecolsys-110617-062527-21 (2018).

4

The evolution of genetic kin discrimination

The evolution of genetic kin discrimination

Abstract

Kin discrimination has the potential to promote the evolution of altruism via kin selection. However, genetic kin discrimination, in which individuals recognise kin based on a shared genetically-encoded signal ('tag'), is thought to be inherently unstable, because individuals with common tags will find social partners at a faster rate than individuals with rare tags (common-tag advantage), meaning rare tags are purged from the population ('Crozier's paradox'; Crozier 1986). However, we show that the common-tag advantage is often reversed because 'cheaters' (non-altruists) build up within groups of individuals using common tags (Grafen 1990). We show that, for genetic kin discrimination to evolve: (1) it must be *favoured*, over indiscriminate cooperation and indiscriminate defection, at the individual level (it must maximise individual fitness). (2) It must be *stabilised*. Stabilisation occurs when there are large number of different tags segregating in the population, or when rare tags do not lead to a reduced social interaction rate for the individual (no common-tag advantage). Overall, our results suggest that, when we look in the right parameter space, genetic kin discrimination is relatively easy to evolve.

Introduction

A behaviour is altruistic if it incurs a cost (c) to perform, by the actor, and provides a benefit (b) to another individual, the recipient. Altruism is favoured if the genetic relatedness (R) between the actor and recipient is sufficiently high, such that $Rb > c$

(Hamilton 1964). Positive genetic relatedness ($R > 0$) may come about if limited dispersal (population structure) leads to individuals interacting predominantly with their relatives (population viscosity; Hamilton 1964; Taylor 1992a). Positive genetic relatedness ($R > 0$) may alternatively, or additionally, come about if individuals are able to recognise and preferentially interact with their relatives within social groups (kin discrimination; Grafen 1990; Hamilton 1964).

It is well documented empirically that both population viscosity and kin discrimination are important factors promoting positive relatedness and therefore cooperation in nature (West et al. 2007). Population viscosity is a particularly salient selective force in microbial populations (though microbes are also capable of kin discrimination; Strassmann et al. 2011; West et al. 2006). Kin discrimination is a particularly salient selective force in animal societies, particularly when animals allow non-kin into their social groups (Cornwallis et al. 2009; Penn and Frommen 2010). We have a good understanding of how, and under what conditions, population viscosity leads to cooperation – there is an abundance of theoretical work on this topic, which explains large swathes of natural diversity in cooperative behaviour (Lehmann and Rousset 2010; West et al. 2007).

In contrast, there is a major disconnect between theory and empirical work on kin discrimination, particularly *genetic* kin discrimination, which uses the matching of alleles ('tags') at a given locus ('tag locus') to indicate kinship (Penn and Frommen 2010). It is generally thought, based on arguments originally made by Crozier (1986), that genetic kin discrimination ('tag-based cooperation') is inherently unstable

(Gardner and West 2007; Penn and Frommen 2010; Rousset and Roze 2007).

Counter claims – that Crozier’s argument is misguided and genetic kin discrimination is not in fact unstable – have been largely rejected on the basis of formal mathematical models (Rousset and Roze 2007). How can we square this with the vast empirical observation of kin discrimination, much of it apparently tag-based, across animal, microbial and plant populations (Grosberg and Quinn 1986; Karban et al. 2013; Manning et al. 1992; Strassmann 2016; Strassmann et al. 2011)?

The main objective of this chapter is to examine arguments – both for and against the inherent instability of genetic kin discrimination – and to subsequently clarify when (under what conditions) we should expect genetic kin discrimination (tag-based cooperation) to evolve. The structure of the paper is as follows. First, we clarify what we mean by genetic kin discrimination (tag-based cooperation), before verbally explicating Crozier’s argument for the instability of genetic kin discrimination, and Grafen’s counter-argument (Crozier 1986; Grafen 1990). We then use Hamilton’s rule to clarify the reach of Crozier’s and Grafen’s arguments – under what parameter space do they hold (Axelrod et al. 2004)? We focus primarily on when individual-level selection: (i) favours, and (ii) stabilises genetic kin discrimination. We then briefly consider how a different kind of selection, for selfish genetic elements, can lead to a limited form of genetic kin discrimination. After this, we explain why previous mathematical treatments of kin discrimination – finding it to be inherently unstable – are limited. We finally construct our own mathematical (population genetic) model of kin discrimination to show these verbal arguments in action. Our key message is that genetic kin discrimination is evolutionarily stable under more

permissive conditions than previously recognised (Gardner and West 2007; Penn and Frommen 2010).

1) Genetic kin discrimination (tag-based cooperation)

We are interested in whether genetic kin discrimination can evolve and be stably maintained in a population. By ‘genetic kin discrimination’ – we specifically mean ‘discrimination of kin based on a genetic marker (tag), followed by kin-directed cooperation’. Therefore, following Hamilton (1964), Crozier (1986) and Grafen (1990), we are concerned with the evolution of *kin-directed cooperation* via tag-matching. Of course, genetic kin discrimination may evolve for other reasons, such as mate choice / inbreeding avoidance (Penn and Frommen 2010). However, here we are concerned with when kin discrimination in a more narrow sense can evolve as a means of directing cooperation towards kin?

For narrow-sense genetic kin discrimination (kin-directed cooperation via tag-matching), there needs to be two loci involved. Firstly, there is a ‘tag locus’.

Individuals only engage in social interactions with others who have a matching allele (tag) at this locus. Secondly, there is a ‘trait locus’, which determines the behaviour of individuals during social interactions (social interactions are always with same-tag conspecifics). If an individual has a ‘conditional cooperator’ allele at the trait locus, it cooperates during social interactions, incurring a cost, c , to give their social partner a benefit, b ($b > c > 0$). If an individual has a ‘defector’ allele at the trait locus, it withholds cooperation during social interactions.

Genetic kin discrimination is stable if, at equilibrium: (i) each tag segregating in the population is at equal frequency, and (ii) the conditional cooperator allele is approximately at fixation (Grafen 1990). Satisfaction of (i) means that, at equilibrium, each individual in the population is equally good at discriminating kin (no one tag provides more reliable information about kinship than any other, given that tags are at equal frequency). Satisfaction of (ii) means that each individual directs cooperation towards kin (no individuals withhold cooperation). We are interested in whether this equilibrium can be reached.

2) Crozier's scenario

Crozier (1986) argued that genetic kin discrimination is generally unstable. He modelled a population where all individuals are conditional cooperators, meaning individuals exhibit cooperation towards other individuals with their tag, but do not interact socially with individuals lacking their tag. In Crozier's model, no individuals are 'defectors' (there is no 'defection' allele available at the 'trait locus'), which means that social interactions between pairs of individuals are *always* beneficial for both individuals involved.

Individuals with common tags will find social partners (same-tag conspecifics) at a faster rate than individuals with rare tags. Therefore, the more common an individual's tag, the faster it can engage in social interactions, and given that social interactions are always beneficial, the greater its fitness. This leads to positive frequency dependence at the tag locus (selection of common tags) and a corresponding erosion of tag diversity. This process, by which kin-directed altruism

destabilises the mechanism through which individuals can recognise kin, is known as ‘Crozier’s paradox’ (Crozier 1986; Figure 1a).

3) Grafen’s scenario

Grafen (1990) argued that Crozier’s scenario is misleading and that the supposed ‘paradox’ is in fact a fallacy. In Crozier’s scenario, individuals are constrained to be conditional cooperators. Grafen generalised this scenario by assuming that, in social interactions (pairings between same-tag conspecifics), individuals may defect and withhold cooperation (‘cheat’). ‘Conditional cooperation’ and ‘defection’ represent two different strategies available to individuals, and they are encoded by two different, competing alleles at the ‘trait locus’. The ‘trait locus’ is different from the ‘tag locus’, and in Grafen’s scenario, there is full recombination between these two loci (no physical linkage).

Grafen noted that, for individuals using a given tag, defection will be increasingly favoured, relative to conditional cooperation, as the population frequency of the tag increases. The reason is that, if a tag is *rare*, then there is a high probability that same-tag conspecifics are genuine genealogical kin. This means that tag-matching causes kin to associate with each other, and interactions between kin favour cooperation (‘kin selection’; Hamilton 1964; Maynard Smith 1964). In contrast, if a tag is *common*, such that most individuals in the population have the tag regardless of kinship, then there is a relatively low probability that same-tag conspecifics are genuine genealogical kin. This means that tag-matching does not cause kin to associate with each other, and interactions between non-kin favour defection.

Therefore, if a defector (cheater) arises by mutation, it will fail to invade if it arises in an individual bearing a rare tag, but it will successfully invade, and begin to spread, if it arises in an individual bearing a common tag. Defectors (cheaters) will be continually purged from individuals bearing rare tags, but selected amongst individuals bearing common tags. This process means that defectors (cheaters) will become disproportionately found amongst individuals bearing common tags relative to individuals bearing rare tags. In other words, for a given 'defector' allele segregating in the population, it will be statistically likely to be sharing a genotype with a common, as opposed to a rare, tag. This statistical association between alleles is called linkage disequilibrium (LD). Over time, as defectors continue to spread amongst individuals bearing common tags, LD will increase.

As a result of LD build-up, social interactions (interactions between same-tag conspecifics) will have different payoffs for different individuals depending on the rarity of their tag. For individuals with rare tags, there is a high chance that a given social partner will be a cooperator, meaning social interactions will be net beneficial. For individuals with common tags, there is a high chance that a given social partner will be a defector, meaning social interactions will be net neutral or detrimental.

Crozier's insight about social interaction rate – that individuals with more common tags will engage in social interactions (interactions between same-tag conspecifics) at a faster rate than individuals with less common tags – is undisputed by Grafen. However, in Grafen's scenario, faster social interaction rate does not lead to

increased fitness, because, owing to the build-up of linkage disequilibrium (association) between common tags and defection (cheating), individuals that interact at faster rate also have a reduced net payoff per interaction. This leads to negative frequency dependence at the tag locus (selection of rare tags).

At equilibrium, each tag has equal frequency, and because, at this equilibrium, interactions are largely restricted to genealogical kin, each individual is a conditional cooperator (defection is disfavoured). Genetic kin discrimination is therefore stable (Figure 1b).

4) Routes to stable genetic kin discrimination

Genes may be selected if they improve the reproductive success (fitness) of the organism. Genes may alternatively be selected if they gain a sufficient, selfish transmission advantage at the possible expense of the reproductive success (fitness) of the individual (Burt and Trivers 2006). The latter are called 'selfish genetic elements'. Here, we do not consider selfish genetic elements (we return to them in Section 7), and ask whether stable kin discrimination can evolve via standard natural selection to improve organism fitness (individual-level selection).

For genetic kin discrimination to evolve by individual-level selection, two things must be true. Firstly, genetic kin discrimination must provide a higher fitness return than alternative possible strategies. The alternative possible strategies are indiscriminate cooperation and indiscriminate defection (Axelrod et al. 2004). Indiscriminate cooperation would correspond to a strategy of socially interacting with all possible

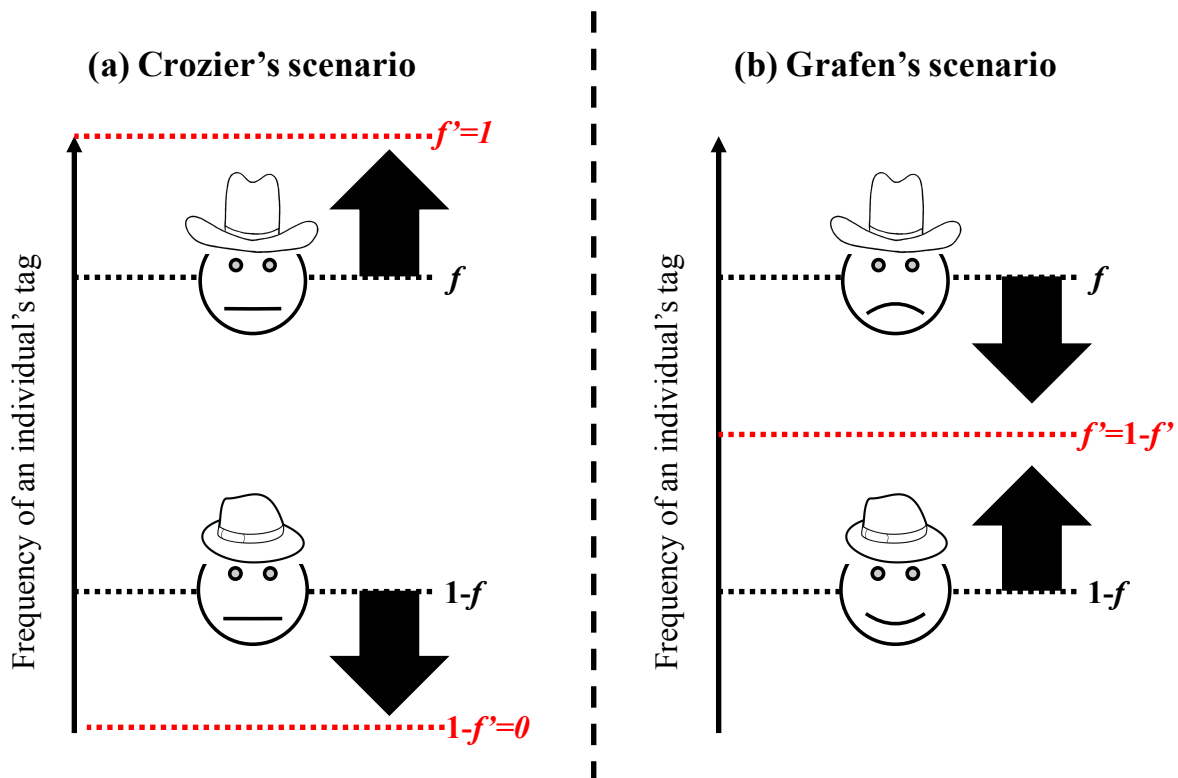


Figure 1. Is genetic kin discrimination stable? Part (a) depicts Crozier's argument for the inherent instability of genetic kin discrimination. Part (b) depicts Grafen's argument for genetic kin discrimination stability. We assume there are two tags segregating in a population. Each tag expresses a unique observable signal. Individuals can only socially interact with others with whom they share a tag. Individuals with the 'cowboy hat' tag are initially at population frequency f , and individuals with the 'trilby hat' tag are initially at population frequency $1-f$ ($f > 1/2$; black dotted lines). The population frequencies of each tag at equilibrium are denoted using an appended dash (f' ; $1-f'$; red dotted lines).

In part (a), Crozier's scenario, the payoff received per social interaction is similar for 'cowboy hat' and 'trilby hat' pairs (all faces have 'neutral', horizontal smiles). However, the 'cowboy hat' tag is more common ($f > 1/2$), meaning individuals with the 'cowboy hat' tag socially interact at a faster rate, and therefore have higher fitness. This means the more common tag ('cowboy hat') goes to fixation ($f'=1$), resulting in a loss of tag diversity at equilibrium.

In part (b), Grafen's scenario, individuals with the 'cowboy hat' tag are predominantly defectors (faces are 'sad'), because the tag is too common to reliably identify kin. Conversely, individuals with the 'trilby hat' tag are predominantly cooperators (faces are 'happy'), because the tag is rare enough to reliably identify kin. As a result, the payoff received per social interaction is net positive for 'trilby hat' pairs, but net negative for 'cowboy hat' pairs, meaning individuals with the rarer 'trilby hat' tag have higher fitness. This leads to equilibration of tag frequencies at equilibrium ($f'=1/2$).

If defectors spread through common tag groups slowly (slow *linkage disequilibrium* build-up), relative to the extent that having a common tag leads to an increased social interaction rate, Crozier's scenario will be more appropriate (instability of genetic kin discrimination); else, Grafen's scenario will be more appropriate (stability of genetic kin discrimination).

individuals (not restricting interactions based on tag-matching), and being cooperative to all social partners. Indiscriminate defection would correspond to a strategy of socially interacting with all possible individuals (not restricting interactions based on tag-matching), and defecting (cheating) against all social partners.

Secondly, there must be an evolutionary trajectory that leads to all individuals adopting the strategy that provides the highest fitness return (individual fitness maximisation). That is, there must be a route, along an evolutionary fitness landscape, through which a population can come to rest at a 'peak' where all individuals are adopting the favoured strategy (maximising their fitness).

Given both that (i) genetic kin recognition provides a higher fitness return than alternative strategies (maximises individual fitness), and (ii) there is an evolutionary trajectory leading to individual fitness maximisation, genetic kin discrimination will evolve and be stably maintained at equilibrium. We evaluate Crozier's and Grafen's arguments in light of this.

5) When is genetic kin discrimination favoured at the individual level?

Grafen's argument is that kin discrimination can, in principle, evolve and be stably maintained – not that kin discrimination will *always* arise whenever individuals have an opportunity to cooperate.

For kin discrimination to arise by individual level selection, (i) kin discrimination must maximise individual fitness, and (ii) there must be an evolutionary trajectory leading to individual fitness maximisation. Here, we focus first on condition (i), and consider, specifically, when kin discrimination (tag-based cooperation) provides a greater payoff to the actor than either indiscriminate defection or indiscriminate helping (we return to condition ii in Section 6).

To elucidate when kin discrimination maximises individual fitness, we first define some coefficients of relatedness between various members of the population. We then use ‘Hamilton’s Rule’ to see when different social strategies (indiscriminate / conditional; cooperation / defection) will be favoured at the individual level.

If we imagine a population where individuals are aggregated into social groups, and within these social groups, some individuals are kin (e.g. siblings) and some are non-kin, and similarly, some individuals have matching tags, and some individuals have dissimilar tags, we can define three different coefficients of relatedness (R_{group} , R_{kin} , R_{tag}). Relatedness is the correlation between trait-values (as caused by allelic identity at the trait locus) among social partners (Grafen 1985; Hamilton 1964; Queller 1992).

- R_{group} is the relatedness between individuals drawn randomly from within a social group (tags need not match).
- R_{kin} is the relatedness between *kin* drawn randomly from within a social group.

- For example, in a large panmictic population, if the 'kin' within a social group are siblings, then $R_{kin}=1/2$. Conversely, if the 'kin' within a social group are cousins, then $R_{kin}=1/8$.
- R_{tag} is the relatedness between *same-tag* individuals drawn randomly from within a social group.
 - For limitingly rare tags, a social partner (same-tag conspecific within the social group) is virtually guaranteed to have inherited it by common descent, meaning the social interactants are genuine genealogical kin (and therefore will be correlated at the trait locus), meaning $R_{tag}=R_{kin}$.
 - For limitingly common tags (tags near fixation), everyone in the population has the same tag, meaning a social partner (same-tag conspecific within the social group) is unlikely to have inherited it by common descent, meaning tag-matching provides no information about kinship (partners will be largely uncorrelated at the trait locus), meaning $R_{tag}=R_{group}$.
 - It is therefore always the case, regardless of population tag frequency, that $R_{kin} \geq R_{tag} \geq R_{group}$.

Indiscriminate defection favoured if $R_{kin} < c/b$

If $R_{kin} < c/b$, there is no evolutionary incentive to direct cooperation towards kin. That is – the payoff to helping (b) is too low relative to the cost (c) to favour cooperation directed towards the kin that are present in the social group. Rather, for the individual, it is better to withhold cooperation towards kin (as well as non-kin). We should therefore expect indiscriminate defection to maximise individual fitness when

$R_{kin} < c/b$. There is no reason to expect kin discrimination (kin-directed cooperation) in this parameter space ($R_{kin} < c/b$), regardless of the mechanism of discriminating kin (tag-matching; phenotypic-matching; familiarity; etc.).

Result 1: If Hamilton's Rule is unsatisfied for social interactions between kin ($R_{kin} < c/b$), there is no evolutionary incentive for individuals to recognise kin – indiscriminate defection will be favoured by individual-level selection.

Indiscriminate defection favoured if $R_{tag}(f=1/L) < c/b < R_{kin}$

Conversely, if $R_{kin} > c/b$, cooperation directed exclusively towards kin would be favoured by individual level selection over indiscriminate defection. However, individuals using tag-matching to identify kin will necessarily interact with some non-kin as well as kin, unless they are using a vanishingly rare tag ($R_{tag} \leq R_{kin}$). Whether or not kin discrimination is ultimately favoured will depend on how reliably tags indicate kinship at a hypothetical equilibrium at which genetic kin discrimination (tag-based cooperation) is stable.

For kin discrimination to be stable, there needs to be negative frequency dependence at the tag locus, which prevents common tags from running away to fixation (Crozier's paradox). If there is negative frequency dependence at the tag locus, all tags segregating in a population will equilibrate in frequency. For instance, if there are L tags segregating at a locus in a population, negative frequency dependence will bring each tag to the population frequency $f=1/L$ at a hypothetical equilibrium where genetic kin discrimination (tag-based cooperation) is stable.

If, at this hypothetical equilibrium where genetic kin discrimination (tag-based cooperation) is stable, the relatedness between social partners (same-tag conspecifics within the social group), which we define as $R_{tag}(f=1/L)$, is low enough that $R(f=1/L) < c/b$, kin discrimination is disfavoured at the individual level, and indiscriminate defection is favoured. In this scenario, kin discrimination is disfavoured at the individual level – not because restricting cooperation to kin is disfavoured ($R_{kin} > c/b$), but because tag-matching cannot reliably identify those kin at equilibrium ($R(f=1/L) < c/b$).

Tag-matching cannot lead to reliable kin discrimination when $R(f=1/L) < c/b$. However, a more reliable means of discriminating kin could still be favoured so long as $R_{kin} > c/b$. If such a means of discriminating kin exists – for instance, through familiarity or phenotype-matching rather than tag-matching – this might be favoured rather than indiscriminate defection in this parameter space ($R_{tag}(f=1/L) < c/b < R_{kin}$). However, we remain focused on genetic kin discrimination (tag-matching) in the present study, and do not explore this possibility.

Result 2: If tag-based discrimination cannot lead (at equilibrium) to sufficiently strong associations between genealogical kin, such that $R_{tag}(f=1/L) < c/b$, there is no evolutionary incentive for individuals to use tag-based discrimination – indiscriminate defection will be favoured over tag-based discrimination by individual-level selection.

Other means of kin discrimination (familiarity; phenotype-matching) might be favoured over indiscriminate defection in this parameter space, given that there is

still an evolutionary incentive to direct cooperation solely to genealogical kin as long as $c/b < R_{kin}$.

High tag availability (L) increases the likelihood that genetic kin discrimination (tag-based cooperation) is favoured over indiscriminate defection ($R_{tag}(f=1/L) > c/b$).

As the number of tags segregating at the tag locus (L) increases, the frequency of each tag at the hypothetical equilibrium where genetic kin discrimination (tag-based cooperation) is stable ($f=1/L$) decreases. As a result, the equilibrium probability that same-tag individuals in a social group are genuine genealogical kin increases, meaning the relatedness of same-tag individuals in a social group at equilibrium ($R_{tag}(f=1/L)$) increases, meaning genetic kin discrimination (tag-based cooperation) is more likely to be favoured over indiscriminate defection ($R_{tag}(f=1/L) > c/b$).

As the number of tags segregating in a population approaches infinity ($L \rightarrow \infty$), the equilibrium probability that same-tag individuals in a social group are genuine genealogical kin approaches unity, meaning the relatedness of same-tag individuals in a social group at equilibrium approaches the relatedness between kin in a social group: $R_{tag}(f=1/L) = R_{kin}$. Therefore, for an infinite number of tags at the tag locus ($L \rightarrow \infty$), genetic kin discrimination (tag-based cooperation) will be favoured, over indiscriminate defection, at the individual level, whenever there is an evolutionary incentive to direct cooperation solely towards genealogical kin: $R_{kin} > c/b$.

Result 3: As the number of tags segregating at the tag locus (L) increases, the more likely it is that genetic kin discrimination is favoured over indiscriminate defection by individual-level selection.

Indiscriminate cooperation favoured if $R_{group} > c/b$

However, just because genetic kin discrimination (tag-based cooperation) is favoured over indiscriminate defection ($R_{tag}(f=1/L) > c/b$), does not mean it is also favoured over indiscriminate cooperation (cooperation directed towards non-kin as well as kin). We therefore address when indiscriminate cooperation is selectively favoured (at the individual level) over genetic kin discrimination (tag-based cooperation).

Given that tags have the potential to indicate kinship, the relatedness between same-tag individuals in a social group (R_{tag}) is always greater than or equal to the relatedness between individuals drawn at random (tags need not match) within a social group ($R_{tag} \geq R_{group}$). Therefore, if $R_{group} > c/b$ holds, it must be the case that $R_{tag} > c/b$ holds for any tag frequency ($0 < f \leq 1$), meaning cooperation is always favoured between same-tag individuals. Therefore, when this condition holds ($R_{group} > c/b$), defectors will be disfavoured and lost from the population at equilibrium, leaving only (conditional) cooperators.

When there are only conditional cooperators, all pairwise social interactions are beneficial for both interactants (no chance of being cheated), meaning kin discrimination is disfavoured by individual-level selection, because it limits the rate of

social interaction. Conversely, indiscriminate cooperation will be favoured by individual-level selection, because it maximises the rate of social interaction.

Result 4: If Hamilton's Rule is satisfied for random individuals drawn from a social group ($R_{group} > c/b$), there is no evolutionary incentive for individuals to recognise kin – indiscriminate cooperation will be favoured by individual-level selection.

It is notable that most models of genetic kin discrimination only allow two alleles at the trait locus – ‘indiscriminate defection’ and ‘conditional (tag-based) cooperation’ (Antal et al. 2009; Axelrod et al. 2004; Crozier 1986; Hammond and Axelrod 2006; Jansen and van Baalen 2006; Rousset and Roze 2007; Taylor and Grafen 2010; Traulsen and Nowak 2007). They do not allow a third allele – ‘indiscriminate cooperation’ – which would correspond to a strategy of socially interacting with, and being cooperative towards, all members of the social group, regardless of tag-identity.

An ‘indiscriminate cooperation’ allele would be selected when $R_{group} > c/b$. However, in lieu of such an allele, individuals can approximate the indiscriminate cooperation phenotype if they have the ‘conditional cooperation’ allele alongside a high-frequency tag. This is why tag diversity is lost in models, like Crozier's (1986), that implicitly consider social interactions that satisfy $R_{group} > c/b$.

Tag diversity is *not* lost in these models because genetic kin discrimination is inherently unstable, as is often claimed (Crozier 1986; Rousset and Roze 2007).

Rather, tag diversity is lost when $R_{group} > c/b$ because it allows individuals to approximate a strategy – indiscriminate cooperation – that confers greater fitness than genetic kin discrimination.

Genetic kin discrimination favoured if $R_{tag}(f=1/L) > c/b > R_{group}$

If $R_{tag}(f=1/L) > c/b > R_{group}$, then genetic kin discrimination (tag-based cooperation) is favoured by individual level selection, over both indiscriminate defection and indiscriminate helping.

6) If genetic kin discrimination is favoured (optimal) at the individual level, when is this optimum reachable?

The analysis so far has addressed when genetic kin discrimination (tag-based cooperation) maximises the fitness of individuals. It has therefore only addressed when kin discrimination (tag-based cooperation) is *favoured*. We now take this result for granted – we assume that genetic kin discrimination maximises individual fitness, which requires $R_{tag}(f=1/L) > c/b > R_{group}$. We address when there exists an evolutionary trajectory through which individual fitness maximisation is obtainable. In other words, we now address when kin discrimination is *stable*.

High-frequency tags have an initial selective advantage because they allow their bearers to engage in social interactions at a faster rate than rare tags do. As pointed out by Crozier (1986), this is the destabilising force for genetic kin discrimination (Section 2). For genetic kin discrimination to be stable, the *rate* at which defectors spread through groups of individuals using high-frequency tags (rate of *LD* build-up)

must be fast enough to reverse the initial selective advantage of high-frequency tags, before a tag runs away to fixation. As pointed out by Grafen (1990), LD build-up is the stabilising force for genetic kin discrimination (Section 3).

For genetic kin discrimination to be stable, the rate of LD build-up must be fast relative to the initial fitness benefit (faster social interaction rate) of having a high-frequency tag. The rate of LD build-up will depend on: (i) the fitness consequences of cooperation, and (ii) the number of tags (L) segregating in the population. The initial fitness benefit of having a high-frequency tag will depend on (iii) the extent that Crozier's claim – that having a high-frequency tag increases an individual's rate of social interaction – is *actually* true, biologically speaking.

High fitness consequences of cooperation stabilise genetic kin discrimination

If the fitness consequences of cooperation (the magnitudes of b and c) are increased, the strength of selection for defection will correspondingly increase amongst individuals using high-frequency tags. This increases the rate of LD build-up, promoting the stable acquisition of genetic kin discrimination. This effect has been demonstrated in a previous theoretical treatment of genetic kin discrimination (Rousset and Roze 2007). We also find this effect in our own theoretical model detailed in Sections 9-14.

Result 5: Genetic kin discrimination is more likely to be stabilised if the fitness consequences of cooperation (b,c) are large.

High tag availability stabilises genetic kin discrimination

If the number of tags segregating in a population (L) are increased, the average frequency of a given tag is lower ($1/L$), meaning tags more reliably indicate kinship, and so the relatedness between same-tag conspecifics (R_{tag}) is higher on average. This means that conditional cooperators are selected more strongly on average, and correspondingly, that defectors are selected more strongly (relatively) in high-frequency tag groups. This effect has been demonstrated in previous theoretical treatments of genetic kin discrimination (Jansen and van Baalen 2006; Rousset and Roze 2007). We also find this effect in our own theoretical model detailed in Sections 9-14.

Result 6: Genetic kin discrimination is more likely to be stabilised if the number of tags segregating in the population (L) is large.

If having a common tag does not result in an increased social interaction rate, there is no Crozier's paradox

Crozier's insight – that individuals with more common tags will be able to engage in social interactions at a faster rate than individuals with less common tags – is the reason why common tags may be positively selected, leading to positive frequency dependence at the tag locus, destabilising genetic kin discrimination. However, this insight might not hold in general. For some social behaviours, having a common tag might *not* result in an increased rate of social interaction (interaction with same-tag conspecifics).

If an individual, with a specific tag, forms consecutive, random, pairwise associations within its social group, the average number of associations required before a same-tag conspecific is found, will decrease with the frequency of the focal individual's tag. If individuals have the *potential* to interact socially with every individual they consecutively associate with (no extrinsic upper limit to social interaction rate), then yes – having a rare tag will result in fewer *actual* social interactions.

However, it may be the case that individuals do *not* have the potential to interact socially with every individual they associate with. In the extreme, for certain social behaviours, individuals might only interact once per generation (extrinsic upper limit to social interaction rate). This includes many major life history decisions, such as the decision faced by cooperatively breeding birds, made once per lifetime, regarding which nest to help at.

In this extreme scenario (one social interaction per lifetime), an individual bearing a rarer tag might have to 'try out' more random associations within the social group before it finds a same-tag conspecific, resulting in a greater time lag before social interaction. However, given that individuals only socially interact (join a nest to help at) once per lifetime, this time lag may not result in a reduced *rate* of social interaction. As a result, in this extreme scenario, the selective advantage of having a common tag – increased rate of social interaction – does not exist, meaning there is no Crozier's paradox, and kin discrimination will evolve whenever it is favoured by individual-level selection ($R_{tag}(f=1/L) > c/b > R_{group}$).

Result 7: If individuals only socially interact a limited number of times in their lifetime (extrinsic upper limit), then having a rare tag might not result in a reduced rate of social interaction. In these cases, there is no common tag-advantage (no Crozier's paradox), and genetic kin discrimination will evolve whenever it is favoured at the individual level.

Many natural social behaviours will fall in between the extremes of (i) no extrinsic upper limit to social interaction rate, and (ii) an extreme upper limit to social interaction rate, such that individuals only socially interact once per lifetime. That is - an individual's rate of social interaction is likely be *somewhat* limited by extrinsic factors, like the rate with which it can gather enough resources (food / energy / territory etc.) to be cooperative (there will be *some* extrinsic upper limit to social interaction rate). However, an individual's rate of social interaction is also likely be *somewhat* limited by the rate with which it can find social partners (same-tag conspecifics), which will be affected by tag frequency.

Previous theoretical treatments of genetic kin discrimination have not considered how the relationship between tag frequency and social interaction rate affects kin discrimination stability (Jansen and van Baalen 2006; Rousset and Roze 2007). We consider this factor in our own mathematical treatment, detailed in a Sections 9-14 .

7) An alternative route to genetic kin discrimination (selfish genetic elements)

We have explained how individual level selection can favour genetic kin discrimination (*Results 1-4*), and when this individual-level optimum can be stabilised (*Results 5-7*). We have shown how individual-level selection can lead to ‘true’ genetic kin discrimination, where every tag segregating in the population is at equal frequency at equilibrium ($f=1/L$), meaning no one tag confers more reliable information about kinship than any other, and where the conditional cooperation allele is approximately at fixation in the population (meaning all individuals in the population have equal kin-discrimination ability).

We briefly consider whether genetic kin discrimination could evolve via another route – via selfish genetic elements – in which natural selection ‘chooses’ genes or gene-coalitions based on the extent to which they propagate themselves selfishly (at the possible expense of the fitness of the individual; Burt and Trivers 2006). For selfish genetic elements to lead to genetic kin discrimination (tag-based cooperation), it would need to be the case that: (i) tag-based cooperation maximises the fitness a gene or gene-coalition (maximises gene *propagation*); and (ii) there is an evolutionary trajectory through which gene fitness maximisation is obtainable (Gardner and Úbeda 2017; Gardner and Welch 2011).

Tag-trait gene coalitions could gain a selfish advantage if there is physical linkage between the tag and trait loci. Physical linkage between loci is a phenomenon absent from Grafen’s (1990) description of how kin discrimination could be stabilised – he, after Hamilton (1964), was focused on how individual-level selection can lead to genetic kin discrimination.

Given that there is a net benefit to cooperation ($b > c$), any tag that is physically linked to the conditional cooperation allele will transiently spread in the population. This works because, owing to linkage, cooperators are transiently protected from cheaters until a rare recombination event occurs, generating a cheater of that tag. If physical linkage is sufficiently tight, a rare tag linked to conditional cooperation can invade a population even when indiscriminate defection is favoured at the individual level ($R_{tag}(f=1/L) < c/b$).

Physical linkage therefore promotes conditional cooperation, not because it associates identical-by-descent cooperators (kin), but because it associates non-identical-by-descent cooperators (non-kin). Under physical linkage, tag frequencies are governed (to some extent) by the timings of recombination events (a non-adaptive factor). The population will therefore fail to rest at the equilibrium where all tag frequencies are equal ($f=1/L$).

Conversely, tag frequencies may differ from each other ($f \neq 1/L$), and may cycle temporally, according to the rate of recombination events. If physical linkage between tag and trait is 'loose' (intermediate recombination), tag frequencies may cycle stably, such that no one tag runs all the way to fixation. Loose linkage means that recombination events are rare enough that cooperators can gain a transient advantage, but common enough that no one tag goes all the way to fixation.

Therefore, selection for selfish genetic elements can allow *some* degree of tag diversity to persist alongside *some* degree of conditional cooperation. This has been captured in previous theoretical models (Jansen and van Baalen 2006; Rousset and Roze 2007). However, this is not true genetic kin discrimination, which requires that all tags are maintained at equal frequency at equilibrium and that all individuals are conditional cooperators.

Furthermore, linkage-mediated tag-based cooperation may compromise individual fitness maximisation (Jansen and van Baalen 2006). It is driven by two-locus coalitions of selfish genetic elements, recognising themselves in other individuals (not necessarily genealogical kin) and directing cooperation towards themselves, spreading initially through the population before being thwarted by rare recombination events generating 'cheating' coalitions.

Result 8: 'Loose' physical linkage between tag and trait loci can allow some degree of tag diversity and some degree of conditional cooperation to be stably maintained in a population over time. However, this is not true genetic kin discrimination, because tag frequencies are not equal at equilibrium.

8) Previous theoretical treatments of genetic kin discrimination (tag-based cooperation)

An influential mathematical population genetic model found that genetic kin discrimination (tag-based cooperation) only evolves under highly restrictive conditions: loose physical linkage between tag and trait loci, coupled with high

fitness consequences of cooperation (large b and c) and high tag availability (L ; Rousset and Roze 2007). However, this model is limited in important respects.

Firstly, it assumes that there is no extrinsic upper limit to social interaction rate – individuals socially interact with *every* same-tag conspecific they come across. This means that an individual's rate of social interaction increases sharply (maximally) with the population frequency of its tag. If, conversely, individuals only had the potential to interact a finite number of times per lifetime (extrinsic upper limit to social interaction rate), tag frequency would have had reduced influence on social interaction rate. This would have meant genetic kin discrimination is more easily stabilised.

Secondly, and more importantly, the results presented in the main figures of Rousset and Roze (2007) (their central case) correspond to regions of parameter space where kin discrimination is disfavoured at the individual level ($R_{tag}(f=1/L) < c/b$ or $c/b < R_{group}$). Therefore, in their treatment, individual-level selection cannot lead to kin discrimination – only selection for selfish genetic elements can. This is likely to be why they found physical linkage between tag and trait loci to be a necessary requirement for genetic kin discrimination. We argue that, in fact, physical linkage is not a necessary requirement for genetic kin discrimination – it is only a necessary requirement for the more limited version of tag-based cooperation driven by selfish genetic elements (*Result 8*).

9) New model to address previous shortcomings

To properly assess whether genetic kin discrimination is evolvable in general, a mathematical treatment that explicitly focuses on the parameter space where kin discrimination is favoured at the individual level ($R_{tag}(f=1/L) > c/b > R_{group}$) is required. We construct a model, inspired by Rousset and Roze (2007), to address this. Our model retains the key features of Rousset and Roze (2007) but simplifies it in some respects with regards to features that are not expected to have causal significance for the evolution of kin discrimination (for instance, social group formation is simpler in our model). Our aim is to give Grafen's (1990) argument a fair chance, by focusing on the region of parameter space where it is predicted to work

$$(R_{tag}(f=1/L) > c/b > R_{group}).$$

As explained in Sections 3 & 6, the evolution of genetic kin discrimination does not merely require that it is favoured at the individual level (maximises individual fitness). It requires *also* that genetic kin discrimination is stable, and this depends in part on linkage disequilibrium (LD). To track genetic details like LD , we need to construct a fully explicit population genetic model. A phenotypic optimisation model would be inappropriate given that fitness maximisation is only a necessary but not a sufficient condition for the evolution of genetic kin discrimination. We need to track all genetic details.

Having said that, we will combine our population genetic analysis with a Hamilton's rule analysis, to identify the regions of parameter space where genetic kin discrimination is favoured at the individual level (when it maximises the fitness of

individuals). This shows us where to focus our attention in our subsequent population genetic analysis of the evolutionary *stability* of genetic kin discrimination.

10) Lifecycle Assumptions

We assume a large population of haploid individuals that each have one of L possible alleles at the 'tag locus'. Each allele at the tag locus is denoted by a number, i , within the set $i \in \{1, 2, \dots, L\}$. Each individual also has one of two possible alleles at the 'trait locus', which are denoted by 1 for conditional cooperator or 0 for indiscriminate defector.

Each generation, each individual enters a social group, where social groups are large. For a given individual, a proportion R of the other individuals in the social group are genotypically identical (clones) to the focal individual as a result of proximity. The remaining proportion $1-R$ of the other individuals in the social group are drawn from the population at random. An individual is related to a clone of itself by 1, and to a random individual by 0 (Grafen 1985). This means that the coefficient of relatedness between 'kin' within social groups, for our model, is $R_{kin}=1$.

Furthermore, it means that an individual is related to a random member of its social group (tags need not match) by $R_{group}=R*1+(1-R)*0=R$.

After individuals have entered social groups, they engage in an arbitrary (but large) number of potential social interactions. A proportion of these *potential* social interactions are *actual* social interactions. For each potential social interaction, individuals randomly aggregate themselves into pairs within their social groups. An

individual with a same-tag partner engages in social interaction (*actual* social interaction). An individual with a different-tag partner repeatedly re-aggregates itself randomly into new pairs, until either: (i) with probability α , it eventually associates with a same-tag partner, with whom it socially interacts with (*actual* social interaction), or (ii) with probability $1-\alpha$, time runs out, and the opportunity to engage in social interaction has passed, meaning it receives zero payoff (*potential* social interaction doesn't translate into an *actual* social interaction; Figure 2).

Therefore, α mediates the extent to which an individual's rate of social interaction is affected by the population frequency of its tag. When $\alpha=1$, population tag frequency has no bearing on social interaction rate – for each *potential* social interaction, individuals are free to cycle through their whole social group, if need-be, to find a same-tag conspecific to interact with. When $\alpha=0$, population tag frequency has maximal bearing on social interaction rate – for each round of *potential* social interaction, any individual who doesn't randomly aggregate with a same-tag conspecific on its first try does not get to engage in an *actual* social interaction.

For each *actual* social interaction, which comprises pairs of individuals sharing the same tag, individuals engage in a cooperation game. Individuals are cooperative if they have the conditional cooperation allele (1), suffering a fitness cost of c ($c>0$) to give a benefit of b to their social partner. Individuals withhold cooperation if they have the indiscriminate defector allele (0). There is a net benefit to cooperation ($b>c$).

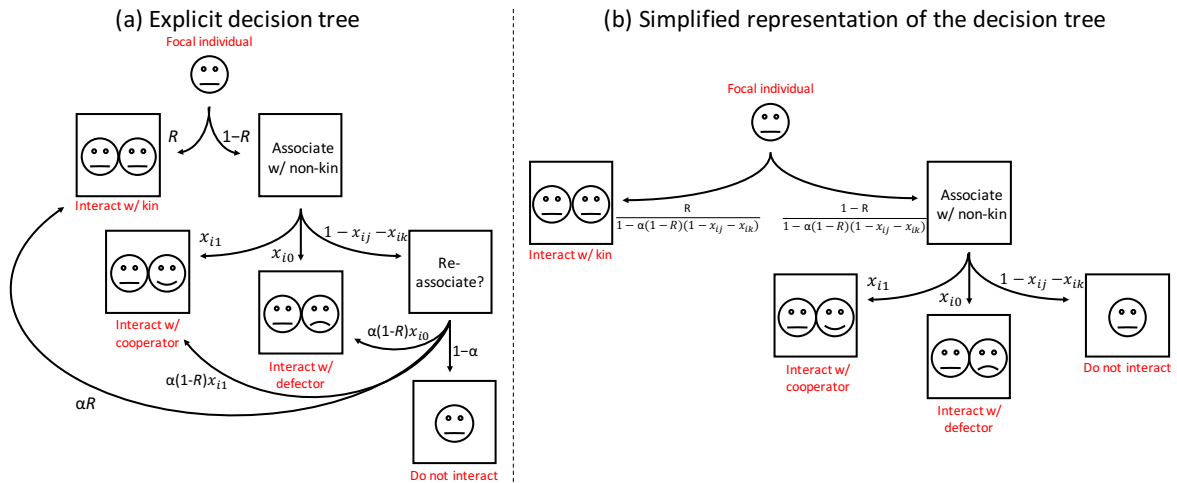


Figure 2: Decision tree. Parts (a) and (b) represent two equivalent descriptions of the ‘decision tree’ that each individual navigates each for each *potential* social interaction: in part (a), the sequence of ‘decisions’ is made explicit; part (b) is a simplified representation of the same process. The focal individual, and its kin (clones), are represented by neutral faces. The focal individual and its clones have the same allele at the trait locus: they are all either cooperators or defectors. Smiling faces represent individuals that are definitely (conditional) cooperators. Frowning faces represent individuals that are definitely defectors. As described in part (a), the focal individual associates, and socially interacts, with kin (clones), with probability R . With probability $(1-R)$, the focal individual associates with non-kin. With probability x_{i1} , the non-kin partner is a same-tag cooperator, and the individuals socially interact. With probability x_{i0} , the non-kin partner is a same-tag defector, and the individuals socially interact. With probability $1-x_{i0}-x_{i1}$, the non-kin partner has a different tag, meaning the individuals cannot socially interact. With probability $1-\alpha$, the focal individual (associated with a non-kin, different-tag partner) does not engage in social interaction. With probability α , the focal individual (associated with a non-kin, different-tag partner) re-associates and socially interacts with a same-tag individual. In part (b), the decision tree is simplified.

After the succession of potential social interactions, each haploid individual reproduces asexually in proportion to its fitness, which is the average of the accumulated payoffs over all *potential* social interactions, added to a baseline fitness of 1. After asexual reproduction, individuals die (non-overlapping generations). Haploid offspring then fuse with a random member of the population (random mating throughout the population) to form diploid individuals. Recombination between the tag and trait loci within diploids occurs with a probability denoted by r ($0 \leq r \leq 0.5$), before the diploids dissociate into haploids again, and the lifecycle is complete.

11) Recursions

At the start of a given generation, the frequency of genotype ij , comprising individuals with tag allele i ($i \in \{1, 2, \dots, L\}$) and trait allele j ($j \in \{0, 1\}$), is denoted by x_{ij} . The frequency of genotype ij after asexual reproduction but before mating and recombination is denoted by x_{ij}' . The frequency of genotype ij at the start of the next generation is denoted by x_{ij}'' . For the conditional cooperators ($i1$; Equation 1a) and indiscriminate defectors ($i0$; Equation 1b) in tag-group i , we first write a general-form recursion describing the change in population genotype frequency from the beginning of a generation (x_{ij}) until after asexual reproduction (x_{ij}'). We will subsequently (Equation 2) write a general-form recursion describing the change in the population frequency of a given genotype (ij) from before mating and recombination (x_{ij}') to the start of the next generation (x_{ij}''),

Genotype frequency change from the start of the generation (x_{ij}) to after fitness-dependent asexual reproduction (x_{ij}')

W is average individual fitness, and is equal to the sum of the right-hand sides of all the specific genotype recursions generated from the general form given below:

$$Wx_{i1}' = \frac{x_{i1} \left(R(1 + b - c) + (1 - R) \left((1 - \alpha)(1 - x_{i1} - x_{i0}) + x_{i1}(1 + b - c) + x_{i0}(1 - c) \right) \right)}{1 - \alpha(1 - R)(1 - x_{i1} - x_{i0})} \quad (1a)$$

$$Wx_{i0}' = \frac{x_{i0} \left(R + (1 - R) \left((1 - \alpha)(1 - x_{i1} - x_{i0}) + x_{i0} + x_{i1}(1 + b) \right) \right)}{1 - \alpha(1 - R)(1 - x_{i1} - x_{i0})}. \quad (1b)$$

We can explain the logic of these general-form recursions as follows. The frequency of a given genotype after fitness-dependent reproduction (x_{ij}') is equal to the

frequency of the genotype at the start of the generation (x_{ij} ; first term, outside of the brackets on the numerator of Equations 1a & 1b), weighted by the average fitness of individuals with that genotype after social interactions, which we can break down term-by-term as follows.

For individuals of a given genotype (ij), of all potential social interactions, a fraction $\frac{R}{1-\alpha(1-R)(1-x_{i1}-x_{i0})}$ of these will be actual social interactions with genealogical kin (clone mates). These interactions result in an average fitness of $1+b-c$ for conditional cooperators ($i1$), leading to the first bracketed term in Equation 1a $\left(\frac{R(1+b-c)}{1-\alpha(1-R)(1-x_{i1}-x_{i0})}\right)$, and an average fitness of 1 for indiscriminate defectors ($i0$), leading to the first bracketed term in Equation 1b $\left(\frac{R}{1-\alpha(1-R)(1-x_{i1}-x_{i0})}\right)$.

Similarly, of all potential social interactions, a fraction $\frac{(1-R)(1-\alpha)(1-x_{i1}-x_{i0})}{1-\alpha(1-R)(1-x_{i1}-x_{i0})}$ of these will not be actualised, owing to failure in finding a fellow tag group member within the social group. These wasted potential interactions are neutral with respect to fitness, which leads to the second bracketed term within Equations 1a & 1b $\left(\frac{(1-R)(1-\alpha)(1-x_{i1}-x_{i0})}{1-\alpha(1-R)(1-x_{i1}-x_{i0})}\right)$.

Next, of all potential social interactions, a fraction $\frac{(1-R)x_{ij}}{1-\alpha(1-R)(1-x_{i1}-x_{i0})}$ of these will be actual social interactions between individuals that are not genealogical kin (clones) but who nevertheless happen to share the same allele at the trait locus as well as the same tag. These interactions result in an average fitness of $1+b-c$ for conditional cooperators ($i1$), leading to the third bracketed term in Equation 1a $\left(\frac{(1-R)x_{i1}(1+b-c)}{1-\alpha(1-R)(1-x_{i1}-x_{i0})}\right)$,

and an average fitness of 1 for indiscriminate defectors ($i0$), leading to the third bracketed term in Equation 1b $\left(\frac{(1-R)x_{i0}}{1-\alpha(1-R)(1-x_{i1}-x_{i0})}\right)$.

Finally, of all potential social interactions, a fraction $\frac{(1-R)x_{ik}}{1-\alpha(1-R)(1-x_{i1}-x_{i0})}$ of these, where k here denotes the alternative allele to j at the trait locus (i.e. $k=1$ if $j=0$; $k=0$ if $j=1$), will be actual social interactions between individuals that are not genealogical kin (clones) and who happen to have different alleles at the trait locus despite sharing the same tag. These interactions result in an average fitness of $1-c$ for conditional cooperators ($i1$), leading to the fourth bracketed term in Equation 1a $\left(\frac{(1-R)x_{i0}(1-c)}{1-\alpha(1-R)(1-x_{i1}-x_{i0})}\right)$, and an average fitness of $1+b$ for indiscriminate defectors ($i0$), leading to the fourth bracketed term in Equation 1b $\left(\frac{(1-R)x_{i1}(1+b)}{1-\alpha(1-R)(1-x_{i1}-x_{i0})}\right)$.

Genotype frequency change from before mating & recombination (x_{ij}') to the start of the next generation (x_{ij}'')

This recursion details how genotype frequencies change as a result of random association of haploids into diploids, followed by recombination, then dissociation of the diploid into haploid offspring. k denotes the alternative allele to j at the trait locus (i.e. $k=1$ if $j=0$; $k=0$ if $j=1$):

$$x_{ij}'' = x_{ij}' \left(x_{ij}' + x_{ik}' + (1-r) \sum_{l \neq i}^L x_{lk}' + \sum_{l \neq i}^L x_{lj}' \right) + r x_{ik}' \sum_{l \neq i}^L x_{lj}' \quad (2)$$

We can explain the logic of this general form recursion as follows. Under random association, the proportion of diploid associations comprised of two ij haploids is given by $x_{ij}'^2$, and these diploid associations exclusively give rise to ij haploid

offspring, meaning the first term in Equation 2 is $x_{ij}^{\prime 2}$. The proportion of diploid associations comprised of an ij haploid and an ik haploid (same tag; different trait allele) is $2x_{ij}^{\prime}x_{ik}^{\prime}$, and half of the haploid progeny from these associations have the ij genotype, regardless of recombination, meaning the second term is $x_{ij}^{\prime}x_{ik}^{\prime}$. The proportion of diploid associations comprised of an ij haploid, and a haploid with a different tag and trait allele, is $2x_{ij}^{\prime}\sum_{l\neq i}^L x_{lk}^{\prime}$. From these associations, the proportion $(1-r)/2$ of haploid progeny have the ij genotype, meaning the third term is $(1-r)x_{ij}^{\prime}\sum_{l\neq i}^L x_{lk}^{\prime}$. The proportion of diploid associations comprised of an ij haploid, and a haploid with a different trait allele but the same tag, is $2x_{ij}^{\prime}\sum_{l\neq i}^L x_{lj}^{\prime}$. From these associations, half of the haploid progeny have the ij genotype, regardless of recombination, meaning the fourth term is $x_{ij}^{\prime}\sum_{l\neq i}^L x_{lj}^{\prime}$. Finally, the proportion of diploid associations comprised of an ik haploid, and a haploid with a different tag and different trait allele, is $2x_{ik}^{\prime}\sum_{l\neq i}^L x_{lj}^{\prime}$. From these associations, the proportion $r/2$ of haploid progeny have the ij genotype, meaning the fifth term is $rx_{ik}^{\prime}\sum_{l\neq i}^L x_{lj}^{\prime}$.

For much of our population genetic analysis, we convert our general-form recursions (Equations 1a, 1b & 2) into specific recursions tied to a population with a given number of segregating tags (L). The more tags there are (L), the more genotypes there are to keep track of. For instance, if there are four tags present in the population, we need to write recursions for each of the 8 possible genotypes ($10, 11, 20, 21, 30, 31, 40, 41$), where each recursion is easily written based on the general forms (Equations 1a, 1b & 2). Taken together, the series of specific genotype recursions arising from Equations 1a, 1b and 2 describe all population genotype frequency changes across a single generation.

12) Specific model: when is genetic kin discrimination favoured at the individual level?

We first elucidate the parameter space where kin discrimination is favoured at the individual level (individual fitness is maximised by genetic kin discrimination). This requires that: (i) indiscriminate helping is disfavoured, and (ii) indiscriminate defection is disfavoured.

To elucidate this parameter space, we take our general-form results for when indiscriminate behaviours will be favoured by individual-level selection (Results 1, 2 & 4), then write the general-form coefficients of relatedness (R_{kin} , R_{kin} , R_{kin}) in terms of our specific model parameters ('closing the model'; Cooper et al. 2018). We then verify this parameter space with an invasion analysis based on our population genetics recursions (Equations 1a, 1b & 2).

Our aim here is to verify our first-principles Hamilton's rule analysis, which should hold in general (outside the confines of our specific population genetic model), with an invasion analysis on our specific model.

When is indiscriminate defection favoured by individual-level selection in our specific model?

We first note that, for our specific model, kin within social groups are related to each other by $R_{kin}=1$ (clones). Given that there is a net advantage to cooperation ($b>c$),

there is always an evolutionary incentive, in our model, to direct cooperation towards kin ($R_{kin} > c/b$; *Result 1*).

However, if tag-matching cannot reliably associate kin at equilibrium, indiscriminate defection will still be favoured over genetic kin discrimination in our specific model.

This will occur when $R_{tag}(f=1/L) < c/b$, where $R_{tag}(f=1/L)$ is the coefficient of relatedness between same-tag conspecifics, drawn at random from a common social group, when tag frequencies are equal ($f=1/L$), as would be the case at a stable equilibrium at which individuals are discriminating kin (*Result 2*).

For our specific lifecycle assumptions, the coefficient of relatedness between same-tag conspecifics (R_{tag}) varies with tag frequency (f) according to: $R_{tag} = \frac{R}{(1-R)f+R}$.

When tag frequencies are equal ($f=1/L$), this coefficient of relatedness can be written in terms of model parameters as $R_{tag}(f = 1/L) = \frac{R}{(1-R)(1/L)+R}$. Substituting this explicit form ($R_{tag}(f=1/L)$) into our condition for the individual-level selection of indiscriminate defection ($R_{tag}(f=1/L) < c/b$), and rearranging, gives the following model-specific condition: $R < \frac{c}{c+L(b-c)}$.

To verify our condition for the individual-level advantage of indiscriminate defection ($R < \frac{c}{c+L(b-c)}$), we undertake an invasion analysis for our specific model. We ask when indiscriminate defection is uninvadable by genetic kin discrimination. We therefore consider an equilibrium where the indiscriminate defection allele (0) is at fixation, and tags are at equal population frequency ($1/L$; which permits potential kin discrimination). For a given tag availability (L), we evaluate the Jacobian stability

matrix at this equilibrium. The Jacobian stability matrix is a matrix of each genotype frequency (x_{10}'' , x_{11}'' , x_{20}'' , x_{21}'' , ..., x_{L0}'' , x_{L1}'') differentiated by each genotype frequency in the prior generation (x_{10}' , x_{11}' , x_{20}' , x_{21}' , ..., x_{L0}' , x_{L1}'). Indiscriminate helping is linearly stable when the eigenvalues of this Jacobian matrix are all less than one. We find that the stability condition varies (predictably) according to tag availability (L), and so we derive the stability condition as a function of tag availability (L). We find that indiscriminate defection cannot be invaded, indicating evolutionary stability, when $R < \frac{c}{c+L(b-c)}$. This recovers the Hamilton's Rule-derived result.

Therefore, our specific invasion analysis verifies our Hamilton's rule-derived result (*Result 2*),

When is indiscriminate cooperation favoured in our specific model?

Indiscriminate helping is favoured at the individual level when $R_{group} > c/b$, where R_{group} is the coefficient of relatedness between individuals drawn at random (tags need not match) from a common social group (*Result 4*). This coefficient of relatedness (R_{group}) can be written in terms of our specific model parameters as $R_{group} = R$, where R gives the proportion of each social group that are genetically identical (clones) due to proximity. Therefore, for our model, indiscriminate helping will be favoured at the individual level when social groups are made up of a sufficiently high proportion of kin (clones) ($R > c/b$).

To verify our condition for the individual-level advantage of indiscriminate helping ($R > c/b$), we undertake an invasion analysis for our specific model. We ask when the

‘indiscriminate helping equilibrium’ – where the conditional cooperation allele (1) is at fixation alongside one tag ($i \in \{1, 2, \dots, L\}$) – is invulnerable to the invasion of mutant genotypes (meaning the equilibrium is evolutionarily stable). For a given tag availability (L), we evaluate the Jacobian stability matrix at the equilibrium position given by $x_{it}^* = 1$ ($i \in \{1, 2, \dots, L\}$). Indiscriminate helping is linearly stable when the eigenvalues of this Jacobian matrix are all less than one. We find that, regardless of tag availability (L), this occurs when: $R > c/b$.

Therefore, our specific invasion analysis verifies our Hamilton’s rule-derived (*Result 4*).

13) Specific model: if genetic kin discrimination is favoured (optimal) at the individual level, when is this optimum reachable?

We focus on the parameter space where genetic kin discrimination is favoured at the individual level ($\frac{c}{b} > R > \frac{c}{c+L(b-c)}$), and ask when this individual-level optimum – genetic kin discrimination – is obtainable. This will depend on whether linkage disequilibrium between rare tags and conditional cooperation can build up quickly enough to stop common tags running away to fixation as a result of increased social interaction rate.

For different tag availabilities (L), we numerically iterate our recursions (Equations 1a, 1b and 2) over many generations, to find the equilibrium frequency of the

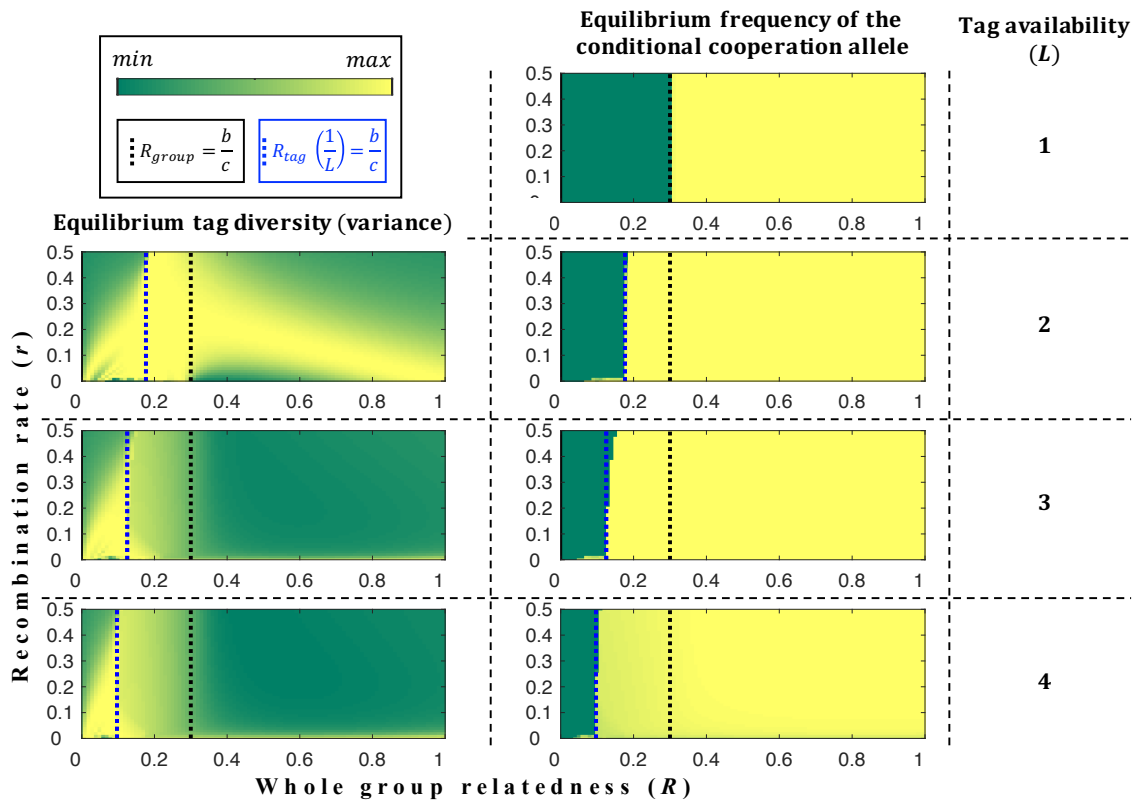


Figure 3. Evolution of genetic kin discrimination when social interaction rate is unaffected by tag frequency ($\alpha=1$). For different: tag availabilities (L ; varies along rows), recombination rates (r ; y-axes), and whole group relatedness ($R_{group}=R$; x-axes), we numerically iterated our recursions describing generational genotype frequency changes (Equations 1a, 1b & 2) over many (1000) generations, and recorded the resulting tag variance (left-hand column) and frequency of the conditional altruism allele (right-hand column), both of which range from being minimal in the green shaded areas and maximal in the yellow shaded areas. Stable genetic kin discrimination (tag-based cooperation) is found in regions of high tag diversity alongside high conditional altruism, which corresponds to the positions on the graphs (values of R and r) where both left- and right-hand columns concurrently display yellow shaded areas of parameter space. We find that, when search efficacy is maximal ($\alpha=1$), stable genetic kin discrimination (tag-based cooperation) evolves whenever it is favoured at the individual level, which is when indiscriminate cooperation is disfavoured (to the left of the black dotted lines), and indiscriminate defection is similarly disfavoured (to the right of the blue dotted lines). These numerical results assumed $b=1$ and $c=0.3$.

conditional cooperation allele and the equilibrium tag diversity. Tag diversity is calculated as a variance, such that $\text{Var}(X) = E[X^2] - E[X]^2$, where the random variable X gives the equilibrium frequencies of different tags, and where $E[X]=1/L$.

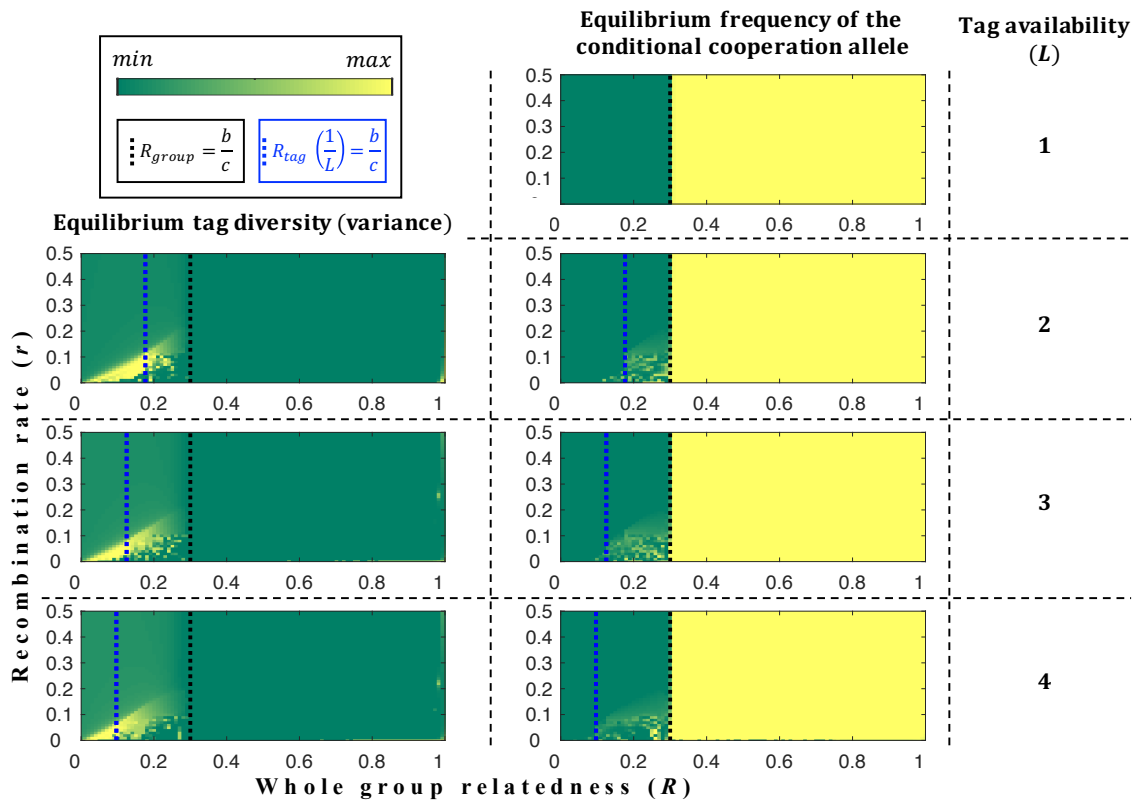


Figure 4. Evolution of genetic kin discrimination when social interaction rate is affected by tag frequency ($\alpha=0$). For different: tag availabilities (L ; varies along rows), recombination rates (r ; y -axes), and whole group relatedness ($R_{group}=R$; x -axes), we numerically iterated our recursions describing generational genotype frequency changes (Equations 1a, 1b & 2) over many (1000) generations, and recorded the resulting tag variance (*left-hand column*) and frequency of the conditional altruism allele (*right-hand column*), both of which range from being minimal in the green shaded areas and maximal in the yellow shaded areas. Stable genetic kin discrimination (tag-based cooperation) is found in regions of high tag diversity alongside high conditional altruism, which corresponds to the positions on the graphs (values of R and r) where both left- and right-hand columns concurrently display yellow shaded areas of parameter space. We find that, when search efficacy is minimal ($\alpha=0$), strict genetic kin discrimination (concurrent yellow regions) does not evolve for low tag availabilities ($1 \leq L \leq 4$), though limited genetic kin discrimination (*some* tag diversity alongside *some* conditional altruism) is promoted by intermediate recombination (r) and tag availability (L). For genetic kin discrimination (tag-based cooperation) to be favoured at the individual level, indiscriminate helping must be disfavoured (which occurs to the left of the black dotted lines), and indiscriminate defection must be similarly disfavoured (which occurs to the right of the blue dotted lines). Tag-based cooperation evolves primarily in the region where it is favoured at the individual level (region within blue and black dotted lines), but may also evolve in the region where indiscriminate defection is favoured at the individual level (region to the left of the blue line) under very low recombination (r), owing to selfish genetic element ('greenbeard') selection. These numerical results assumed $b=1$ and $c=0.3$.

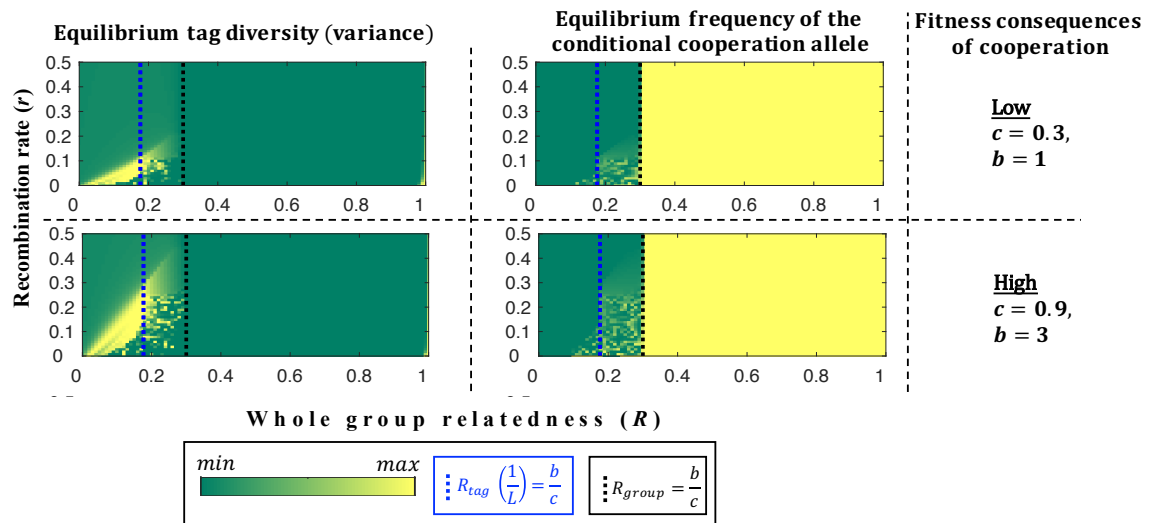


Figure 5. High fitness consequences of cooperation promote the stabilisation of genetic kin discrimination. For different: fitness consequences of cooperation (b, c ; varies along rows), recombination rates (r ; y -axes), and whole group relatedness ($R_{group}=R$; x -axes), we numerically iterated our recursions describing generational genotype frequency changes (Equations 1a, 1b & 2) over many (1000) generations, and recorded the resulting tag variance (*left-hand column*) and frequency of the conditional altruism allele (*right-hand column*), both of which range from being minimal in the green shaded areas and maximal in the yellow shaded areas. Stable genetic kin discrimination (tag-based cooperation) is found in regions of high tag diversity alongside high conditional altruism, which corresponds to the positions on the graphs (values of R and r) where both left- and right-hand columns concurrently display yellow shaded areas of parameter space. We find that, for greater fitness consequences of cooperation, some degree of genetic kin discrimination (*some* tag diversity alongside *some* conditional altruism) is found across a greater region of parameter space where genetic kin discrimination is favoured at the individual level (in between blue and black dotted lines). These numerical results assumed $b=1, c=0.3, L=2$ and $\alpha=0$.

We find that higher fitness consequences of cooperation (greater b, c) lead to an increased probability that genetic kin discrimination is stabilised when favoured (*Result 5*; Figure 5). Increased tag availability (L) leads to an increased probability that genetic kin discrimination is stabilised when favoured (*Result 6*; Figure 4).

When the rate of social interaction is unaffected by tag frequency ($\alpha=1$), kin discrimination is *always* stabilised when it is favoured at the individual level,

irrespective of tag availability ($L > 1$) or physical linkage between tag and trait (*Result 7*; Figure 3). By contrast, when the rate of social interaction is maximally affected by tag frequency ($\alpha = 0$), kin discrimination is less likely to be stabilised when favoured, especially when there aren't many segregating tags (L ; Figure 4).

14) Specific model: an alternative route to genetic kin discrimination (selfish genetic element selection)

We find that 'loose' physical linkage between tag and trait can lead to a limited form of tag-based cooperation – one that is associated with *some* (below-maximal) degree of tag diversity and *some* (below-maximal) degree of conditional cooperation. We find that, under sufficiently tight linkage, limited tag-based cooperation can arise even when indiscriminate defection is favoured at the individual level (*Result 8*; Figure 4).

15) Discussion

Individual-level selection

Our analysis clarifies that genetic kin discrimination can evolve via individual selection, if: (i) genetic kin discrimination maximises the fitness of individuals (genetic kin discrimination is *favoured*), and (ii) there is an evolutionary trajectory leading to individual fitness maximisation (genetic kin discrimination is *stabilised*). For genetic kin discrimination to maximise the fitness of individuals, genealogical relatedness within a social group (R_{group}) cannot be too high, else indiscriminate cooperation is favoured ($R_{group} < c/b$; *Result 4*). Additionally, genealogical relatedness

between same-tag individuals (R_{tag}) cannot be too low at equilibrium, else indiscriminate defection is favoured ($R_{tag}(1/L) > c/b$; *Result 2*). Genetic kin discrimination is therefore most likely to be *favoured* in natural populations where social group relatedness (R_{group}) is not too high, and there is a large diversity of tags (L) available for individuals to identify kin with (*Result 3*). Genetic kin recognition is more likely to be *stabilised*, after being favoured, if: (i) the fitness consequences of cooperation (b, c) are high (*Result 5*); (ii) there is a large diversity of tags (L) available for kin-identification (*Result 6*); and, (iii) the rate that individuals can engage in social interactions is relatively unaffected by the population frequency of their tag (α ; *Result 7*).

Our findings are consistent with empirical findings. A meta-analysis found that kin discrimination is most likely to be found where social group relatedness (R_{group}) is not too high, and where the fitness consequences of cooperation are high (Cornwallis et al. 2009). Two of the best characterised cases of genetic kin discrimination (tag-based cooperation) are: somatic fusion amongst planktonic larvae in a colonial ascidian, and cooperative breeding behaviour in a mouse (Grafen 1990; Grosberg and Quinn 1986; Manning et al. 1992). In both examples, the 'tag locus' is involved in histocompatibility (immunity), and houses vast allelic diversity (high L) as a result. Furthermore, in both examples, the social behaviours (somatic fusion / nest choice) are taken only once per lifetime, meaning an individual's social interaction rate might not be appreciably restricted by having a rare tag ($\alpha \rightarrow 1$). Finally, in both examples, the social behaviours in question are major life history decisions, and therefore will be associated with high fitness consequences (b, c). Therefore, both of these

systems appear to exhibit qualities (high L ; high a ; high b, c) that, as we have shown, promote the *stabilisation* of genetic kin discrimination.

Selfish genetic elements / the 'greenbeard' concept

Our analysis also clarifies that a limited form of tag-based cooperation can arise, via selfish genetic elements, if there is 'loose' physical linkage between tag and trait loci. Physical linkage can generate selfish tag-trait coalitions (*Result 9*). Linkage-mediated tag-based cooperation is limited because it does not lead to equilibria where all tags are at equal population frequency ($f=1/L$). Instead, tag frequencies might differ from each other ($f\neq 1/L$) and may cycle temporally, meaning individuals vary in their discriminatory ability, both within populations and over time (individual fitness maximisation is compromised).

We may be tempted to refer to this kind of tag-based cooperation as 'beard chromodynamics', after the 'greenbeard' concept, as Jansen & van Baalen (2006) did. In support of this semantic choice, the tag-trait coalitions held together by linkage are indeed selfish, running against the interest of the individual as a whole to direct cooperation towards copies of themselves in other individuals – they gain their advantage in exactly the same way theoretical greenbeard genes do (Madgwick et al. 2019).

In opposition to this semantic choice, a 'greenbeard' is historically defined as a tag-trait coalition that has *zero* recombination, whereas the selfish coalitions driving non-adaptive tag-based cooperation require *some* recombination within them (loose

linkage; Dawkins 1976; Gardner and West 2010; Hamilton 1964). We may choose, in the spirit of the original expositions by Hamilton (1964) and Dawkins (1976), to reserve the greenbeard concept to tag-trait coalitions that have zero recombination. This narrow-sense greenbeard concept has a particular usefulness – it was originally devised, and successfully so, for the purpose of explaining that the ‘relatedness’ relevant for inclusive fitness is defined specifically at the trait locus (Hamilton 1964).

Future directions

High tag availability ($L \rightarrow \infty$)

There are a number of obvious directions for future research on the topic of genetic kin discrimination. One direction is a more rigorous analysis of exceedingly large tag availabilities ($L \rightarrow \infty$). We showed that, for an increase in tag availability (L), genetic kin discrimination is more likely to be *favoured* (over indiscriminate defection; *Result 3*) as well as *stabilised* (*Result 6*). We also showed that, as tag availability tends to infinity ($L \rightarrow \infty$), genetic kin discrimination is favoured over indiscriminate defection whenever there is an evolutionary incentive to direct cooperation solely towards kin ($R_{kin} > c/b$).

However, an open question concerns whether, for very high tag availability ($L \rightarrow \infty$), kin discrimination is *always* stabilised when it is favoured at the individual level ($R_{kin} > c/b$), irrespective of the relationship between social interaction rate and tag frequency (α), and of physical linkage between tag and trait. We are yet to run the necessary numerical simulations of our recursions (Equations 1a, 1b & 2 for high L) to test this. If this is found to be the case, it would imply that genetic kin recognition

should evolve very easily. Specifically, it would imply that stable genetic kin discrimination should only require two things.

Firstly, there would need to be an individual-level benefit of tag-based cooperation ($R_{kin} > c/b > R_{group}$). Secondly, there would need to be a locus within the population that comprises a vast diversity of alleles, where the different alleles encode subtly different proteins, and where the proteins have observable and differentiable phenotypes (e.g. odours) that can be sensed by potential social partners. Such a locus could then be utilised for tag-based cooperation, allowing very fine-tuned kin discrimination ($L \rightarrow \infty$). Such loci should not be particularly rare – most eukaryotes have histocompatibility (immunity) loci, and these are likely to fit the bill (Grafen 1990; Grosberg and Quinn 1986; Manning et al. 1992).

Fitness cost of having a rare tag

Another direction for future research is a closer assessment of Crozier's argument that the rate with which an individual can engage in social interactions will depend on the population frequency of its tag. We explained that, in some scenarios, having a rare tag might result in a longer time lag between finding same-tag social partners, but this might not necessarily result in a reduced rate of social interaction, especially if individuals don't engage in social interactions very often.

However, even if the time lag associated with having a rare tag doesn't result in a reduced rate of social interaction, it might be associated with other costs. For instance, having a rare tag might delay important life history behaviours, such as

entering a nest to engage in cooperative breeding, with fitness consequences. Similarly, having to cycle through a larger number of individuals before finding a same-tag conspecific might expose individuals bearing rare tags to increased predation risk or resource depletion. These issues could be addressed by incorporating a rare tag cost into our model. We expect that a rare tag cost would make genetic kin discrimination harder to stabilise.

Other mechanisms of kin discrimination

We elucidated a region of parameter space where strict kin-directed cooperation is favoured, but genetic kin discrimination (tag-based cooperation) is disfavoured, because tags are insufficiently reliable indicators of kinship at equilibrium ($R_{tag}(f=1/L) < c/b < R_{kin}$; *Result 2*). This raised the possibility that other forms of kin discrimination, such as those based on phenotype-matching or familiarity, could be selected in this region.

In addition to this, these other forms of kin discrimination might be selected at the expense of genetic kin discrimination (tag-based cooperation) even outside of this parameter space ($R_{tag}(f=1/L) < c/b < R_{kin}$). This will depend on the relative costs and benefits of different mechanisms of discriminating kin – what mechanism is more reliable; does increased reliability come at a fitness cost (e.g. an energetic cost); etc.? A direction for future work could be to address this, to gain an understanding of when different mechanisms of differentiating kin are favoured.

Re-analysis of Rousset and Roze (2007)

Our model is a simplified version of Rousset & Roze (2007). We found stable kin discrimination under a fairly broad parameter space, but Rousset & Roze (2007) failed to find kin discrimination. The model of Rousset & Roze (2007) is arguably more realistic than ours. It is based on the infinite island model, and allows explicit migration between demes every generation. Our model, by contrast, assumed that demes are formed each generation via clonal reproduction, followed by an association of this group of clones (R) with individuals drawn randomly from the rest of the population ($1-R$).

Therefore, an important direction for future research is to check that our findings still hold in Rousset & Roze's (2007) lifecycle. To do this, we would have to write explicit recursions detailing how different coefficients of relatedness (R_{group} , $R_{tag}(f)$) change over each generation as a result of migration between demes. We could then solve these recursions to find the coefficients of relatedness at equilibrium (Cooper et al. 2018; Taylor 1992b). We could then perform a Hamilton's Rule analysis on their model, as we have done on our model, to elucidate the parameter space where kin discrimination is favoured at the individual level ($R_{tag}(f=1/L) > c/b > R_{group}$).

We predict that, in Rousset & Roze's (2007) lifecycle, stable genetic kin discrimination will evolve when: (i) it is favoured at the individual level ($R_{tag}(f=1/L) > c/b > R_{group}$), and (ii) there are a high number of segregating tags (L), meaning genetic kin discrimination can be stabilised. It is not clear that Rousset & Roze's (2007) explicitly considered this region of parameter space in their original analysis, which may be why they failed to find stable genetic kin discrimination. We

predict that, in Rousset & Roze's (2007) lifecycle, stable genetic kin discrimination will alternatively evolve if: (i) it is favoured at the individual level ($R_{tag}(f=1/L) > c/b > R_{group}$), and (ii) we introduce our parameter α , and focus on scenarios where having a rare tag does not reduce social interaction rate (high α), meaning genetic kin discrimination can be stabilised.

References

- Antal, T., Ohtsuki, H., Wakeley, J., Taylor, P. D., & Nowak, M. A. (2009). Evolution of cooperation by phenotypic similarity. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(21), 8597–8600.
doi:10.1073/pnas.0902528106
- Axelrod, R., Hammond, R. A., & Grafen, A. (2004). Altruism via kin-selection strategies that rely on arbitrary tags with which they coevolve. *Evolution*, *58*(8), 1833–1838.
- Burt, A., & Trivers, R. (2006). *Genes in Conflict*. Cambridge, MA and London, England: Harvard University Press. doi:10.4159/9780674029118
- Cooper, G. A., Levin, S. R., Wild, G., & West, S. A. (2018). Modeling relatedness and demography in social evolution. *Evolution Letters*, *2*(4), 260–271.
doi:10.1002/evl3.69
- Cornwallis, C. K., West, S. A., & Griffin, A. S. (2009). Routes to indirect fitness in cooperatively breeding vertebrates: kin discrimination and limited dispersal. *Journal of Evolutionary Biology*, *22*(12), 2445–2457. doi:10.1111/j.1420-9101.2009.01853.x
- Crozier, R. H. (1986). Genetic clonal recognition abilities in marine invertebrates

- must be maintained by selection for something else. *Evolution*, 40(5), 1100–1101. doi:10.1111/j.1558-5646.1986.tb00578.x
- Dawkins, R. (1976). *The Selfish Gene*. Oxford University Press.
- Gardner, A., & Úbeda, F. (2017). The meaning of intragenomic conflict. *Nature Ecology & Evolution*, 1(12), 1–9. doi:10.1038/s41559-017-0354-9
- Gardner, A., & Welch, J. J. (2011). A formal theory of the selfish gene. *Journal of Evolutionary Biology*, 24(8), 1801–1813. doi:10.1111/j.1420-9101.2011.02310.x
- Gardner, A., & West, S. A. (2007). Social Evolution: The Decline and Fall of Genetic Kin Recognition. *Current Biology*, 17(18), R810–R812.
doi:10.1016/j.cub.2007.07.030
- Gardner, A., & West, S. A. (2010). Greenbeards. *Evolution*, 64(1), 25–38.
doi:10.1111/j.1558-5646.2009.00842.x
- Grafen, A. (1985). A geometric view of relatedness. *Oxford surveys in evolutionary biology*, 262(3), 391–397. doi:10.1016/j.jtbi.2009.10.012
- Grafen, A. (1990). Do animals really recognize kin? *Animal Behaviour*, 39(1), 42–54.
- Grosberg, R. K., & Quinn, J. F. (1986). The Genetic-Control and Consequences of Kin Recognition by the Larvae of a Colonial Marine Invertebrate. *Nature*, 322(6078), 456–459. doi:10.1038/322456a0
- Hamilton, W. D. (1964). The genetical evolution of social behaviour. I. *Journal of Theoretical Biology*, 7(1), 1–16. doi:10.1016/0022-5193(64)90038-4
- Hammond, R. A., & Axelrod, R. (2006). Evolution of contingent altruism when cooperation is expensive. *Theoretical Population Biology*, 69(3), 333–338.
doi:10.1016/j.tpb.2005.12.002
- Jansen, V. A. A., & van Baalen, M. (2006). Altruism through beard chromodynamics.

Nature, 440(7084), 663–666. doi:10.1038/nature04387

Karban, R., Shiojiri, K., Ishizaki, S., Wetzell, W. C., & Evans, R. Y. (2013). Kin recognition affects plant communication and defence. *Proceedings of the Royal Society B: Biological Sciences*, 280(1756), 20123062–5.

doi:10.1098/rspb.2012.3062

Lehmann, L., & Rousset, F. (2010). How life history and demography promote or inhibit the evolution of helping behaviours. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1553), 2599–2617.

doi:10.1098/rstb.2010.0138

Madgwick, P. G., Belcher, L. J., & Wolf, J. B. (2019). Greenbeard Genes: Theory and Reality. *Trends in Ecology & Evolution*, 1–12. doi:10.1016/j.tree.2019.08.001

Manning, C. J., Wakeland, E. K., & Potts, W. K. (1992). Communal Nesting Patterns in Mice Implicate Mhc Genes in Kin Recognition. *Nature*, 360(6404), 581–583.

doi:10.1038/360581a0

Maynard Smith, J. (1964). Group selection and kin selection. *Nature*.

doi:10.4324/9780203790427-8

Penn, D. J., & Frommen, J. G. (2010). Kin recognition: An overview of conceptual issues, mechanisms and evolutionary theory. In P. Kappeler (Ed.), *Animal Behaviour: Evolution and Mechanisms* (Vol. 9, pp. 55–85). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-02624-9_3

Springer Berlin Heidelberg. doi:10.1007/978-3-642-02624-9_3

Queller, D. C. (1992). A General Model for Kin Selection. *Evolution*, 46(2), 376–380.

doi:10.2307/2409858?refreqid=search-

gateway:4b0b4e00ea4a577ac44ebb74cd224a17

Rousset, F., & Roze, D. (2007). Constraints on the origin and maintenance of

- genetic kin recognition. *Evolution*, 61(10), 2320–2330. doi:10.1111/j.1558-5646.2007.00191.x
- Strassmann, J. E. (2016). Kin Discrimination in Dictyostelium Social Amoebae. *The Journal of eukaryotic microbiology*, 63(3), 378–383. doi:10.1111/jeu.12307
- Strassmann, J. E., Gilbert, O. M., & Queller, D. C. (2011). Kin Discrimination and Cooperation in Microbes. *Annual Review of Microbiology*, 65(1), 349–367. doi:10.1146/annurev.micro.112408.134109
- Taylor, P. D. (1992a). Altruism in viscous populations — an inclusive fitness model. *Evolutionary Ecology*, 6(4), 352–356. doi:10.1007/BF02270971
- Taylor, P. D. (1992b). Inclusive Fitness in a Homogeneous Environment. *Proceedings: Biological Sciences*, 249(1326), 299–302. doi:10.2307/49757?refreqid=search-gateway:c790938b8208a9f97893fe3134051aed
- Taylor, P. D., & Grafen, A. (2010). Relatedness with different interaction configurations. *Journal of Theoretical Biology*, 262(3), 391–397. doi:10.1016/j.jtbi.2009.10.012
- Traulsen, A., & Nowak, M. A. (2007). Chromodynamics of cooperation in finite populations. *PLoS ONE*, 2(3), e270. doi:10.1371/journal.pone.0000270
- West, S. A., Griffin, A. S., & Gardner, A. (2007). Evolutionary explanations for cooperation. *Current Biology*, 17(16), R661–72. doi:10.1016/j.cub.2007.06.004
- West, S. A., Griffin, A. S., Gardner, A., & Diggle, S. P. (2006). Social evolution theory for microorganisms. *Nature Reviews Microbiology*, 4(8), 597–607. doi:10.1038/nrmicro1461

Acknowledgements

We thank Sam Levin and Alan Grafen for very helpful discussion and feedback.

Discussion

Chapters 2, 3 and 4 each contain their own discussion. In this chapter, I will attempt to place the work presented in Chapters 2, 3 and 4 in its broader context within the field of behavioural ecology.

Behavioural Ecology

The field of behavioural ecology attempts to rationalise the behaviour of organisms. As pointed out by Tinbergen (1963), we can rationalise a behaviour in any of four ways, and a complete account of a behaviour must address each: (1) how did it develop (ontogeny)?; (2) how does it work (mechanism)?; (3) via what path did it evolve? (phylogeny)?; (4) what is its function (adaptation)? Although behavioural ecologists are ostensibly interested in each of the four questions, the field has come to be particularly associated with the fourth question, concerning adaptation (Davies, Krebs, & West, 2012). Behavioural ecologists are primarily concerned with rationalising organismal behaviours in terms of the selective forces that have led to their evolution (Grafen, 1984).

The field has been phenomenally successful. We have a very good understanding of the selective forces that have led to: foraging behaviour, strategies of resource competition, mate choice, strategies of parental care, sex allocation, strategies of signalling and cooperation (Davies et al., 2012). By 'very good understanding', I mean that there is a close fit between theory and data. We have an overarching theoretical framework, built by Fisher (1930), Hamilton (1964), Price (1970),

Maynard Smith (1982) and others, from which trait- and organism-specific theory follows deductively. Our specific theory has accounted – sometimes quantitatively, in the case of sex allocation – for natural diversity in behaviour (West, 2009). This has the dual effect of validating our overarching theoretical framework, as well as our explanations for specific behaviours.

The first organisms to be analysed with the tools of behavioural ecology were, as you would expect, common, not-unusual ones, such as mammals, insects and birds ('paradigm organisms'). The organisms are 'unitary', meaning they are built from a single cell, meaning there is high cooperation amongst the constituent genes and cells comprising it. In recent years, the reach of behavioural ecology has expanded to weirder organisms, such as fungi, bacteria and even viruses ('non-paradigm organisms'; Domingo-Calap, Segredo-Otero, Durán-Moreno, & Sanjuán, 2019; Kiers et al., 2011; West, Griffin, Gardner, & Diggle, 2006). These organisms are often 'modular', meaning their constituent cells retain reproductive capability, reducing cooperation within the organism. These organisms are also often subject to heightened internal conflict. For example, many genes within bacteria – those on plasmids, for instance – readily hop between bacterial cells, and therefore will enter into conflict with the less-mobile genes, the latter of which will 'care' more about the reproductive success of the organism (bacterium).

As the field begins to grapple with the weirder branches of the tree of life, we should pause and ask: is our usual practice still justified? Are the methods that we employ as behavioural ecologists, which have been so successful in rationalising the

behaviours of paradigm-organisms, still appropriate for the study of non-paradigm organisms? To address this issue, I will first tie down what exactly 'usual practice' within behavioural ecology is, and how this practice is justified for the study of paradigm organisms. I will then turn to the question of whether this practice is still justified for the study of non-paradigm organisms.

The phenotypic gambit

As clarified by Grafen (1984), behavioural ecologists typically take the 'phenotypic gambit' when studying the selective forces responsible for a trait. Specifically, when studying a particular behaviour, a behavioural ecologist will typically assume that each individual in a natural population adopts the behavioural form (strategy) that maximises its fitness. This assumption – that individuals will always maximise their fitness with respect to a trait, regardless of the genetics underpinning the trait – allows us to focus on phenotypes, not genetics.

Behavioural ecologists know that this assumption is trivially false – there are many theoretical and empirical instances where fitness maximisation is not fully achieved, and is contingent of the genetic architecture of the trait (Grafen, 1984). A classic example is the empirical case of sickle-cell anaemia and malaria resistance in humans. The optimal phenotype is a simultaneous resistance to malaria as well as normal development (no sickle cell anaemia). However, only a fraction of individuals in natural populations obtain this optimal phenotype, because it is encoded by the heterozygous genotype at a specific locus (heterozygote advantage). This genetic architecture means that the two suboptimal homozygous genotypes, one conferring

sickle-cell anaemia and the other conferring malaria susceptibility, arise constantly as a result of Mendelian segregation (Allison, 1954). However, if the phenotypic gambit is usually approximately true, across traits and organisms, it provides a window for studying the selective forces responsible for a trait.

The phenotypic gambit, which frees our analysis from the tyranny of genetic detail, is extremely powerful. A typical way the gambit is used is as follows (Frank, 1998).

Empirical workers might observe that different behavioural strategies are prominent in different populations of a given, or closely related, species. They might also observe that the differences in strategy across populations correlate with differences in ecology or environment. For instance, they might observe that strategy X is correlated with some aspect of the selective environment Y. Employing the gambit, they can then say that the behavioural form X has evolved as a result of the selection pressure Y.

This empirical work may be verified, or contested, by theoretical workers, also employing the gambit. They might make a specific phenotypic fitness maximisation model, which is consistent with and based upon overarching, foundational theory (Taylor & Frank, 1996). The specific model can be used to elucidate selection pressures that (logically) govern what behavioural strategy maximises fitness. The selection pressures identified by the model can be tested against the selection pressures identified by the empirical work. Explanatory power lies in the fit between theory and data.

Justifying the gambit

One justification for the gambit is that it works. A closeness of fit between empirical and theoretical predictions validates both our explanation of the specific behaviour, as well as the soundness of the gambit itself. However, this type of validation is no substitute for a rigorous, formal, defence of the approach itself, based on the fundamental 'laws' of population genetics. A formal justification for the phenotypic gambit is required if, as behavioural ecologists, we truly want to defend our practices. Grafen (1984) pointed out the need for this kind of justification, and went on to provide it in subsequent decades (to a large extent, at least; Grafen, 2014a).

The justification can be explained in simple terms as follows. If a behaviour is controlled exclusively by a single, haploid, un-imprinted, vertically inherited locus, and each variant phenotype is encoded by its own allele at this locus, then, over many generations of allele frequency changes, natural selection will lead to a population of individuals that maximise their fitness with respect to the behaviour (Grafen, 1984). Therefore, the gambit is literally true for traits that are controlled by single, haploid, un-imprinted, vertically inherited loci. However, in reality, very few if any traits of interest will be exclusively determined by such a locus. Can we justify the phenotypic gambit when traits are encoded by multiple loci, with arbitrary ploidy, epistasis and dominance? Yes, we can, broadly, owing to the foundational work by Grafen (2014a). Grafen showed that the dynamical process of gene frequency change *still* generally results in individual fitness maximisation, even if traits are not underpinned by single haploid loci.

Gaps in Grafen's justification

However, Grafen only justified the gambit when all loci influencing a trait are inherited in the same way (they all belong to the same coreplicon). For example, a trait that is affected by loci on the X chromosome as well as autosomes does not fall within the remit – there is no formal justification that such a trait will be optimised, with respect to individual fitness, by natural selection. The assumption that a trait is only influenced by loci belonging to a common coreplicon is more likely to be (approximately) true for genomes that are not particularly 'fragmented' into multiple coreplicons. In paradigm organisms, genome fragmentation appears to be relatively low – the vast majority of genes are autosomal, un-imprinted and vertically inherited. Most genes will therefore belong to a common coreplicon, meaning the vast majority of, if not all, loci contributing to a given trait will belong to the same coreplicon, meaning the phenotypic gambit is justified. However, Grafen's justification of the gambit does not extend well to organisms with highly fragmented genomes (high internal conflict), such as bacteria, because traits are unlikely to be underpinned by genes that lie exclusively in one coreplicon.

Similarly, the gambit is only justified for organisms whose cells are either genetically identical or unable to reproduce selfishly (repression of competition), as is the case for unitary organisms (Gardner & Grafen, 2009). By contrast, there is currently no formal justification for the idea that modular organisms – which have genetically dissimilar, competing cells – will evolve by natural selection to maximise their fitness. This then leads to two questions. Firstly, if there is no formal justification for employing the phenotypic gambit when studying 'fragmented' organisms like

bacteria, or modular organisms like filamentous fungi, should we proceed as normal, and employ it anyway, or should we look to different approaches? Secondly, could we, in principle, provide formal justification for using the phenotypic gambit on these fragmented or modular organisms?

Pragmatic justification for extending the reach of the gambit

A *pragmatic* rather than a *formal* justification for using the gambit on modular and fragmented organisms is that, where it has so far been employed, it has worked, so far at least. That is, studies of modular and fragmented organisms that have employed the gambit have found a close fit between theory and data. For instance, strategies of cooperation and signalling in bacteria are being well rationalised by researchers who assume that individual bacterial cells, which might be subject to high genome fragmentation and internal conflict, are nevertheless fitness maximisers (Davies et al., 2012). Similarly, the ‘bargaining’ strategy of arbuscular mycorrhizal fungi engaging in mutualistic interactions with plants is being well rationalised on the assumption that the fungal networks, which are modular, are nevertheless fitness maximisers (Kiers et al., 2011; Kiers, Gardner, West, & Wyatt, 2019; Scott, Kiers, Cooper, Santos, & West, 2019).

Success stories like these are reason-enough to continue employing the phenotypic gambit in the study of non-paradigm organisms, at least in the short and medium term. However, if we want to continue using this approach in the long term, we should hope to obtain formal, first principles, justification for treating non-paradigm

organisms as if they are fitness maximisers, regardless of genetics. I will attempt to sketch out a way in which the practice might be justified for organisms with fragmented (potentially conflicting) genomes. I will not address the issue of how we might go about justifying the gambit for modular organisms (Gardner & Grafen, 2009).

Formal justification for extending the reach of the gambit is difficult to conceive

Firstly, it should be noted that, although Grafen's justification of the phenotypic gambit does not extend to traits that are underpinned by loci spanning multiple coreplicons, there is no reason to expect that, in itself, the fragmentation of genomes into coreplicons, will be a barrier to individual fitness maximisation. We would only expect genome fragmentation to prevent individual fitness maximisation if the coreplicons are actually in conflict, which occurs when different coreplicons are propagated best at different values of a given trait (Ågren, Davies, & Foster, 2019; Cosmides & Tooby, 1981; Queller & Strassmann, 2018).

Therefore, even though Grafen wasn't able to justify individual fitness maximisation for traits underpinned by multiple coreplicons, it is reasonable to assume that an extension of the *same kind of approach* should, in principle, be able to justify individual fitness maximisation for traits that are underpinned by many, but unconflicting, coreplicons. In other words, there is no conceptual reason to think that a trait underpinned by a single coreplicon should be any more optimised than a trait

underpinned by many coreplicons that are in agreement with each other over the value of the trait (Bourke, 2014; Haig, 2014).

What about traits that are underpinned by multiple, conflicting coreplicons? This includes, for example, bacterial social traits (e.g. public good production) that are affected by plasmids as well as chromosomes (Mc Ginty & Rankin, 2012; Nogueira et al., 2009). What might a justification for applying the phenotypic gambit to these traits look like? We could construct a dynamical model of allele frequency change, where trait-affecting alleles can arise at loci belonging to different, conflicting coreplicons. We could track the construction of a complex trait due to successive invasions of trait-affecting alleles. The emerging trait, as it increases in complexity, would be pulled back and forth according to the divergent preferences of the conflicting coreplicons underpinning it (Bourke, 2014; Grafen, 2014b; Haig, 2014).

We would seek to provide formal justification that this dynamical process generates an organismal trait that, despite being a 'compromise' between divergent interests, nevertheless maximises the reproductive success (offspring number) of the individual. However, it is not clear that individual fitness maximisation would be achieved in such a scenario (Grafen, 2014b; Haig, 2014). It is not even clear how such a scenario would be appropriately modelled – Grafen (2014b) suggested that it would require a new science of "teleonomy", called for by Williams (1966) but never developed.

Adaptation and the parliament of genes

However, the problem seems less intractable in the light of results collated in Chapter 3. We might be able to simplify our hypothetical model of the evolution of a complex trait underpinned by conflicting coreplicons. Firstly, I will recap exactly what was shown in Chapter 3, and what wasn't shown.

In Chapter 3 I showed that, if individuals are maximising their fitness with respect to a given trait under potential conflict, then attempts to distort the trait from individual fitness maximisation, driven by selfish genetic elements arising in coreplicons representing minority-interests in the genome (cabal), will by and large be futile, unless the cabal is relatively large in size (approaching half of the genome; Chapter 3, Figure 6). Therefore, in Chapter 3, I provide formal justification for the idea that, once an organism has obtained fitness maximisation, it cannot, in general, be appreciably distorted by the subsequent invasion of trait-distorting elements.

However, there is an implicit bias in the methodology I adopted in Chapter 3. I assumed that the organism is already maximising its fitness, and then showed that subsequent distortions from this maximand will often be negligible. This bias is evident in the strategy set afforded to different alleles across the genome: the minority-interest within the genome (cabal) can only exert influence over the trait via 'trait distorters' (they distort the trait– they cause a shift away from the norm), whereas the majority-interest within the genome (commonwealth) can only exert influence over the trait via 'suppressors' (they restore – they cause a shift back towards the norm). Therefore, what I really showed in this chapter is that, for traits

under intragenomic conflict, individual fitness maximisation can be *maintained*. I do not show that individual fitness maximisation is obtained in the first place.

Towards a justification for extending the reach of the gambit

The justification, in Chapter 3, for individual fitness maximisation *maintenance*, hints at a possible justification for individual fitness maximisation *emergence* in traits underpinned by conflicting coreplicons, such as those wielded by ‘fragmented’ organisms like bacteria. A sketch of such a possible justification can be described as follows.

We would have to amend the strategy sets of alleles in our Chapter 3 scenario. Previously, I assumed that coreplicons representing a minority interest within the genome only have access to ‘trait distorters’, which are alleles that cause a shift in trait value (constructive). I assumed that coreplicons representing a majority interest within the genome only have access to ‘suppressors’, which are alleles that specifically counteract the effects of existing ‘trait distorters’ (destructive). A proper justification for fitness maximisation *emergence* would require that all coreplicons have access to ‘trait-affecting’ alleles (constructive) as well as suppressors (destructive). We would need to track how a complex trait is built, through the successive invasions of trait-affecting alleles as well as suppressors, contributed by all coreplicons, with layers of trait complexity being added as new trait-affecting alleles invade.

A key result from Chapter 3 is that suppressors spread rapidly, quickly removing the influence of their target allele over the trait. An implication of this is that, if suppressors are allowed to arise within different conflicting coreplicons, the focal trait will not evolve in a purely constructive manner – the alleles underpinning it would shift over time, as old trait-affecting alleles are suppressed, and new ones arise. Notably, the coreplicons representing the majority interest within the genome would (by nature of being larger in size) be able to generate suppressors at a faster rate than coreplicons representing minority perspectives. This means that trait-affecting alleles contributed by minority-interest coreplicons will lose their influence over the organism trait at a faster rate than trait-affecting alleles contributed by majority-interest coreplicons.

Furthermore, as layers of complexity are added to a complex trait, trait-affecting alleles responsible for the *foundations* of a complex trait may not be suppressible without the whole trait collapsing. Conversely, trait-affecting alleles responsible for *tweaking* the trait, after most of the foundational work has been done, may be suppressed without such destructive effects on the organism. Therefore, as a complex trait is built, the *opportunity* for minority-influence coreplicons to gain some influence over it, by suppressing trait-affecting alleles contributed by the majority-interest coreplicons, is reduced. As a result, as trait complexity is built, the control of majority-interest coreplicons over traits is entrenched, because minority-interest coreplicons lose the option of suppressing most of the trait-affecting genes, contributed by the majority-interest coreplicons, underpinning the trait.

This might lead to the scenario where, even if multiple conflicting coreplicons have the *potential* to contribute to the building of a trait, the vast majority of trait-affecting responsible for a *complex* trait are likely to have been derived from majority-interest coreplicons. Therefore, the ‘conflict’ might be largely illusory, with trait evolution proceeding according to the majority-interest within the genome, leading to individual fitness maximisation.

This is only a sketch of how the phenotypic gambit might be justified for the analysis of traits, such as those employed by ‘fragmented’ organisms like bacteria, that have the potential to be influenced by multiple conflicting coreplicons. It provides a worthwhile direction for future, more rigorous, research.

References

- Allison, A. C. (1954). Protection afforded by sickle-cell trait against subtertian malarial infection. *British Medical Journal*, *1*(4857), 290–294.
<http://doi.org/10.1136/bmj.1.4857.290>
- Ågren, J. A., Davies, N. G., & Foster, K. R. (2019). Enforcement is central to the evolution of cooperation. *Nature Ecology & Evolution*, *3*(7), 1–12.
<http://doi.org/10.1038/s41559-019-0907-1>
- Bourke, A. F. G. (2014). The gene’s-eye view, major transitions and the formal darwinism project. *Biology & Philosophy*, *29*(2), 241–248.
<http://doi.org/10.1007/s10539-013-9422-y>
- Cosmides, L. M., & Tooby, J. (1981). Cytoplasmic inheritance and intragenomic conflict. *Journal of Theoretical Biology*, *89*(1), 83–129.

- [http://doi.org/10.1016/0022-5193\(81\)90181-8](http://doi.org/10.1016/0022-5193(81)90181-8)
- Davies, N. B., Krebs, J. R., & West, S. A. (2012). *An Introduction to Behavioural Ecology*. John Wiley & Sons.
- Domingo-Calap, P., Segredo-Otero, E., Durán-Moreno, M., & Sanjuán, R. (2019). Social evolution of innate immunity evasion in a virus. *Nature Microbiology*, *4*(6), 1–11. <http://doi.org/10.1038/s41564-019-0379-8>
- Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*. Oxford University Press.
- Frank, S. A. (1998). Foundations of Social Evolution. *Heredity* (Vol. 82, pp. 268–344). Princeton University Press. <http://doi.org/10.1038/sj.hdy.6885351>
- Gardner, A., & Grafen, A. (2009). Capturing the superorganism: a formal theory of group adaptation. *Journal of Evolutionary Biology*, *22*(4), 659–671. <http://doi.org/10.1111/j.1420-9101.2008.01681.x>
- Grafen, A. (1984). Natural selection, kin selection and group selection. *Behavioural Ecology: an Evolutionary Approach*. <http://doi.org/10.1515/9781400820108.92>
- Grafen, A. (2014a). The formal darwinism project in outline. *Biology & Philosophy*, *29*(2), 155–174. <http://doi.org/10.1007/s10539-013-9414-y>
- Grafen, A. (2014b). The formal darwinism project in outline: response to commentaries. *Biology & Philosophy*, *29*(2), 281–292. <http://doi.org/10.1007/s10539-013-9413-z>
- Haig, D. (2014). Genetic dissent and individual compromise. *Biology & Philosophy*, *29*(2), 233–239. <http://doi.org/10.1007/s10539-013-9418-7>
- Hamilton, W. D. (1964). The genetical evolution of social behaviour. I. *Journal of Theoretical Biology*, *7*(1), 1–16. [http://doi.org/10.1016/0022-5193\(64\)90038-4](http://doi.org/10.1016/0022-5193(64)90038-4)

- Kiers, E. T., Gardner, A., West, S. A., & Wyatt, G. A. K. (2019). Restricting mutualistic partners to enforce trade reliance. *Nature Communications*, 7(1), 1–7. <http://doi.org/10.1038/ncomms10322>
- Kiers, E. T., Kiers, E. T., Duhamel, M., Duhamel, M., Beesetty, Y., Beesetty, Y., et al. (2011). Reciprocal Rewards Stabilize Cooperation in the Mycorrhizal Symbiosis. *Science*, 333(6044), 880–882. <http://doi.org/10.1126/science.1208473>
- Maynard Smith, J. (1982). *Evolution and the Theory of Games*. Cambridge: Cambridge University Press. <http://doi.org/10.1017/CBO9780511806292>
- Mc Ginty, S. É., & Rankin, D. J. (2012). The evolution of conflict resolution between plasmids and their bacterial hosts. *Evolution*, 66(5), 1662–1670. <http://doi.org/10.1111/j.1558-5646.2011.01549.x>
- Nogueira, T., Rankin, D. J., Touchon, M., Taddei, F., Brown, S. P., & Rocha, E. P. C. (2009). Horizontal Gene Transfer of the Secretome Drives the Evolution of Bacterial Cooperation and Virulence. *Current Biology*, 19(20), 1683–1691. <http://doi.org/10.1016/j.cub.2009.08.056>
- Price, G. R. (1970). Selection and covariance. *Nature*, 227(5257), 520–521. <http://doi.org/10.1038/227520a0>
- Queller, D. C., & Strassmann, J. E. (2018). Evolutionary Conflict. *Annual Review of Ecology, Evolution, and Systematics*, 49(1), annurev–ecolsys–110617–062527–21. <http://doi.org/10.1146/annurev-ecolsys-110617-062527>
- Scott, T. W., Kiers, E. T., Cooper, G. A., Santos, dos, M., & West, S. A. (2019). Evolutionary maintenance of genomic diversity within arbuscular mycorrhizal fungi. *Ecology and Evolution*, 9(5), 2425–2435. <http://doi.org/10.1002/ece3.4834>
- Taylor, P. D., & Frank, S. A. (1996). How to make a kin selection model. *Journal of*

Theoretical Biology, 180(1), 27–37. <http://doi.org/10.1006/jtbi.1996.0075>

Tinbergen, N. (1963). On aims and methods of Ethology. *Ethology*, 20(4), 410–433.

<http://doi.org/10.1111/j.1439-0310.1963.tb01161.x>

West, S. (2009). *Sex Allocation*. Princeton: Princeton University Press.

<http://doi.org/10.1515/9781400832019>

West, S. A., Griffin, A. S., Gardner, A., & Diggle, S. P. (2006). Social evolution theory for microorganisms. *Nature Reviews Microbiology*, 4(8), 597–607.

<http://doi.org/10.1038/nrmicro1461>

Williams, G. C. (1966). *Adaptation and Natural Selection*. Princeton University Press.

<http://doi.org/10.2307/j.ctv39x5jt>

6

Appendix

Darwin's aliens

Review Article

Cite this article: Levin SR, Scott TW, Cooper HS, West SA (2019). Darwin's aliens. *International Journal of Astrobiology* **18**, 1–9. <https://doi.org/10.1017/S1473550417000362>

Received: 30 May 2017

Accepted: 8 September 2017

First published online: 1 November 2017

Key words:

aliens; astrobiology; evolution; extraterrestrial life; individuality; major transitions

Author for correspondence:

Samuel R. Levin, E-mail: samuel.levin@zoo.ox.ac.uk

Darwin's aliens

Samuel R. Levin¹, Thomas W. Scott¹, Helen S. Cooper² and Stuart A. West¹

¹Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK and ²37 Beech Croft Road, Oxford OX2 7AY, UK

Abstract

Making predictions about aliens is not an easy task. Most previous work has focused on extrapolating from empirical observations and mechanistic understanding of physics, chemistry and biology. Another approach is to utilize theory to make predictions that are not tied to details of Earth. Here we show how evolutionary theory can be used to make predictions about aliens. We argue that aliens will undergo natural selection – something that should not be taken for granted but that rests on firm theoretical grounds. Given aliens undergo natural selection we can say something about their evolution. In particular, we can say something about how complexity will arise in space. Complexity has increased on the Earth as a result of a handful of events, known as the major transitions in individuality. Major transitions occur when groups of individuals come together to form a new higher level of the individual, such as when single-celled organisms evolved into multicellular organisms. Both theory and empirical data suggest that extreme conditions are required for major transitions to occur. We suggest that major transitions are likely to be the route to complexity on other planets, and that we should expect them to have been favoured by similarly restrictive conditions. Thus, we can make specific predictions about the biological makeup of complex aliens.

Introduction

There are at least 100 billion planets in our Galaxy alone (Cassan *et al.* 2012), and at least 20% of them are likely to fall in the habitable zone (Petigura *et al.* 2013), the region of space capable of producing a biosphere. Even if 0.001% of those planets evolved life, that would mean 200 000 life-harboring planets in our Galaxy; and it would only take *one* alien life form for our conception of the Universe to change dramatically. It is no wonder, then, that hundreds of millions of dollars have recently been invested in astrobiology research (Schneider 2016), the USA and Europe have rapidly growing astrobiology initiatives (Des Marais *et al.* 2008; Horneck *et al.* 2016), and myriad new work has been done to try and predict what aliens will be like (Benner 2003; Davies *et al.* 2009; Rothschild 2009; Rothschild 2010; Shostak 2015). The challenge, however, is that when trying to predict the nature of aliens, we have only one sample – Earth – from which to extrapolate. As a result, making these predictions is hard.

So far, the main approach to making predictions about extra-terrestrial life has been relatively mechanistic (Domagal-Goldman *et al.* 2016). We have used observations about how things have happened on the Earth to make statistical statements about how likely they are to have happened elsewhere. For example, certain traits have evolved many times on the Earth, and so we posit that extraterrestrial life forms will converge on the same earthly mechanisms. Because eye-like organs have evolved at least 40 times (von Salvini-Plawen & Mayr 1977), and are relatively ubiquitous, we predict that they would evolve on other planets, too (Conway Morris 2003; Flores Martinez 2014). Similarly, we have used a mechanistic understanding of chemistry and physics to make predictions about what is most probable on other planets. For example, carbon is abundant in the Universe, chemically versatile, and found in the interstellar medium, so alien life forms are likely to be carbon-based (Cohen & Stewart 2001). These kinds of predictions come from a mixture of mechanistic understanding and extrapolating from what has happened on the Earth. There is no theoretical reason why aliens could not be silicon-based and eyeless.

An alternative approach is to use theory. When making predictions about life on other planets, a natural theory to use would be evolutionary theory. Evolutionary theory has been used to explain a wide range of features of life on the Earth, from behaviour to morphology. For example, it has allowed us to predict when some organisms, especially insects, should manipulate the sex of their offspring, to produce an excess of sons or daughters, how some birds should forage for food, and why males tend to be larger than females (Darwin 1871; Clutton-Brock & Harvey 1977; Davies & Houston 1981; West 2009; Davies *et al.* 2012). If life arises on other planets, then the evolutionary theory should be able to make similar predictions about it. Neither approach – theoretical or mechanistic – is more or less valid than the other. But each has different advantages and can be used to make different sorts of predictions.

Here, we examine how theoretical and mechanistic approaches can be combined to better understand what to expect from alien life. We consider whether aliens will undergo natural selection, and what implications would follow if they do. That aliens undergo natural selection is something often taken for granted, but which needs justification on firm theoretical grounds. We then turn our attention to a specific subset of aliens: complex ones. We examine how complexity has arisen on the Earth, and make predictions about how complexity would arise elsewhere in the Universe. Finally, we describe some biological features we would expect to find in complex extraterrestrial life.

Natural selection

On Earth

Darwin (1859) showed that just a few simple features of life on Earth lead to evolutionary change via natural selection. Individual organisms differ in how they look and act – there is natural *variation*. These differences are *heritable* – offspring tend to look and act like their parents. These heritable differences are linked to *differential success* – some individuals, as a result of how they are made or behave, leave more offspring than others. These three features, with heritable variation leading to differential success, result in natural selection (Darwin 1859; Fisher 1930). Any traits or behaviours linked to the greater production of offspring (higher fitness or success) will build up in the population over time. As the environment changes, different traits lead to higher success. This leads to changes in the population or evolutionary change.

Thus, the ingredients required for natural selection are incredibly simple. Given a collection of entities (a population) that has:

(1) heredity; (2) variation; and (3) differential success linked to variation, then natural selection will follow. The entities that are more successful will become more prevalent in the population, as a result of being ‘selected’. Natural selection does not depend on a specific genetic system (Darwin knew nothing of modern genetics) or a specific genetic material, elemental makeup or planet-type. Given that 1, 2 and 3 exist, natural selection occurs (Fig. 1).

Natural selection not only explains evolutionary change, it also explains adaptation. When we look around at the natural world, we cannot help but see what looks like design: a giraffe’s neck is for reaching high up leaves, a stick insect’s body for camouflage, a tree’s leaf for photosynthesizing. Organisms look designed or ‘adapted’ for the world in which they live. Through the gradual selection of small improvements, traits associated with success in the environment accrue in the population. Consequently, over time, natural selection will lead to organisms that appear *as if* they were designed for success in the environment. The clause ‘*as if*’ is key here – natural selection leads to the appearance of design (adaptation), without a designer (Grafen 2003; Gardner 2009).

In fact, natural selection is the *only* explanation we have for the appearance of design without a designer (Gardner 2009). Other processes can cause evolutionary change. For example, a mutation can cause a change from one generation to the next. But, without natural selection, random mutation is incredibly unlikely to produce the complex traits that we see around us, like limbs or eyes. Things that appear purposeful, such as limbs, organs and cells, require the gradual selection of improvements.

Another way to say this is that natural selection is unique because it is a *directional force*. The entities that increase in

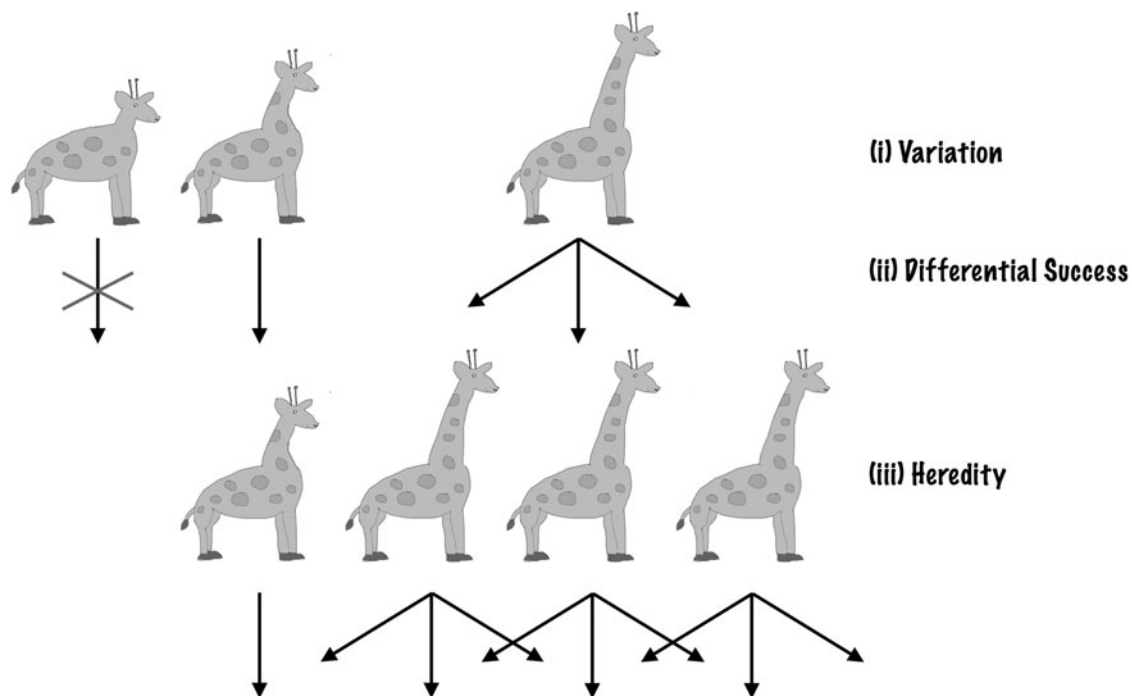


Fig. 1. Natural selection. Natural Selection operates if three conditions are satisfied: variation, differential success linked to variation and heredity. Here, we illustrate with an example: the evolution of long necks in giraffes. (i) Initially, there are natural variations in giraffes’ neck lengths. (ii) Longer-necked giraffes have access to more food, high up in the trees and so live longer to have more offspring. (iii) Giraffes’ offspring resemble their parents. As a result of (i), (ii) and (iii), the population gradually shifts to be dominated by long-necked giraffes.

representation in the population are a *specific subset* of the population – those that are better at replicating. Natural selection increases fitness (Fisher 1930). As a result of these ‘successful’ entities accruing in the population, over time entities become adapted for the *apparent* purpose of success. They look like ‘well-designed’ machines, with the ‘purpose’ of their ‘design’ being successful replication.

In space

Natural selection is the only way we know to get the kinds of life forms we are familiar with, from viruses to trees. By familiar, we are not restricting ourselves to life forms that look earthly. Instead, they are familiarly life-like in the sense that they stand out from the background of rocks and gases because they appear to be busy trying to replicate themselves. A simple replicator could arise on another planet. But without natural selection, it won’t acquire apparently purposeful traits like metabolism, movement or senses. It won’t be able to adapt to its environment, and in the process, become a more complex, noticeable and interesting thing.

We can ask, then, will aliens undergo natural selection? Evolutionary theory tells us that, for all but the most transient and simple molecules, the answer is yes. Without a designer, the only way to get something with the apparent purpose of replicating itself (something like a cell or a virus), is through natural selection. Consequently, if we are able to notice it as life, then it will have undergone natural selection (or have been designed by something that itself underwent natural selection).

It is easy to quibble about the definition of life, and as some authors have pointed out, trying to do so can reveal more about human language than about the external world (Cleland & Chyba 2002). Our goal here is not to thoroughly define life. We adopt a functional stance – what separates life from non-life is

its apparent purposiveness, leading to tasks such as replication and metabolism (Maynard Smith & Szathmáry 1995). Further, without natural selection, entities cannot adapt to their environment, and are therefore transient and will not be discovered. If we identified an extra-terrestrial entity that we deemed to be a foreign life form, but that had no degree of adaptedness, this prediction would not hold.

Picture an alien (Fig. 2). If what you are picturing is a simple replicating molecule, then this ‘alien’ *might* not undergo natural selection (Fig. 2a). For example, it could replicate itself perfectly every time, and thus there would be no variation, and it would never improve. Or it might have such a high error rate in replication that it quickly deteriorates. If we count things like that as life, then there could be aliens that do not undergo natural selection. But if you are picturing anything more complex or *purposeful* than a simple molecule, then the alien you are picturing has undergone natural selection (Fig. 2b). This is the kind of prediction that theory can make. Given heredity, variation and differential success, aliens will undergo natural selection. Or, more interestingly, without those three things, aliens could not be more complicated than a replicating molecule. Given an adapted alien, one with an appearance of design or purpose, it *will have undergone natural selection*.

Complexity

What is complexity?

We have established that aliens will undergo natural selection. It also seems reasonable that, given the sliding scale from replicating molecules to large creatures with many ‘body parts’, and beyond, some alien discoveries would be more interesting than others. In particular, the more complex the aliens we find, the more interesting and exciting they will be, irrespective of whether they appear anything like the life forms on the Earth. Something similar to a

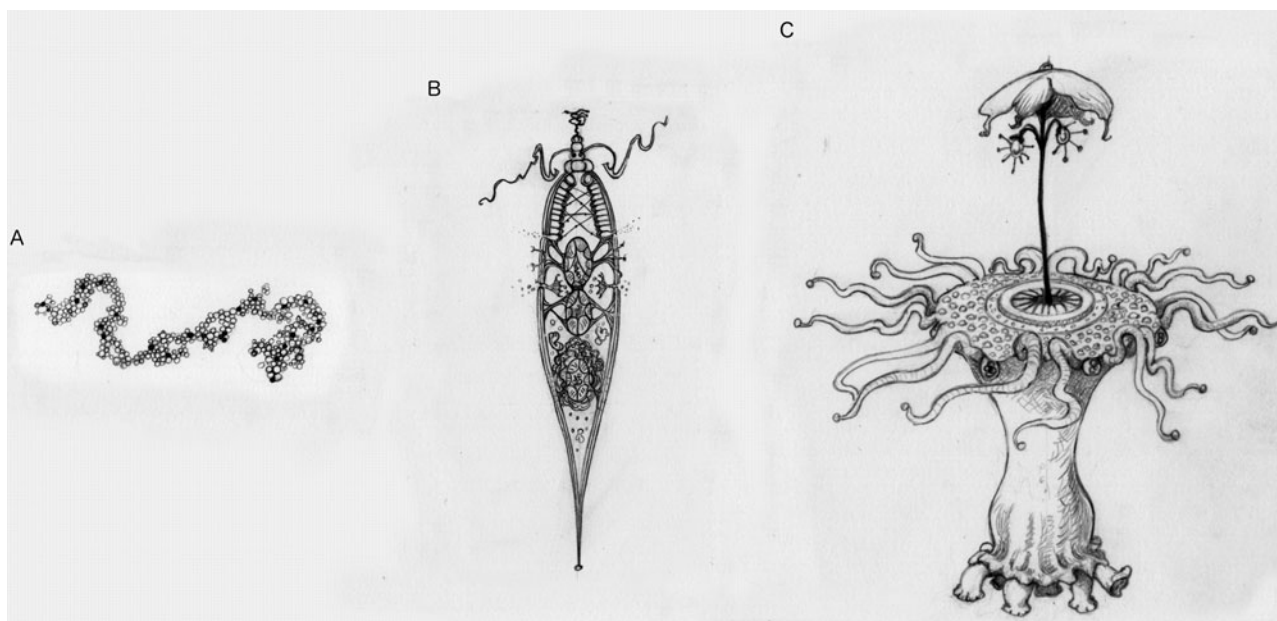


Fig. 2. Picture an alien. These illustrations represent different levels of adaptive complexity we might imagine when thinking about aliens. (a) A simple replicating molecule, with no apparent design. This may or may not undergo natural selection. (b) An incredibly simple, cell-like entity. Even something this simple has sufficient contrivance of parts that it must undergo natural selection. (c) An alien with many intricate parts working together is likely to have undergone major transitions.

colony of Ewoks from Star Wars or the Octomite in Fig. 4 would likely be more interesting than a simple chemical replicator.

Complexity is difficult to define, and there is certainly no hard and fast rule about what is and is not complex. In biology, it is common to define complexity in terms of functional parts. Things with more parts taking on more tasks and containing more functional interactions are more complex (Maynard Smith & Szathmáry 1995; Corning & Szathmáry 2015). A tree is more complex than a virus, and a beehive is more complex than a protein. Importantly, with organisms as with machines, the parts need to be working towards a common purpose, such as assembling a car or surviving to reproduce. Again, our goal here is not to provide definitions. The challenge comes at the boundaries, for example between a virus and a cell, where the definitions become murky. In the following sections, we are not focusing on the boundaries, but things, like the vast majority of life on the Earth, which clearly have a multitude of parts working in concert. Astrobiology is a largely empirical field, and the kinds of things programs like SETI are searching for are undeniably complex.

Complexity on Earth

What do we know about how complexity arises on the Earth? The theory of natural selection itself is silent about *whether* complexity will arise. The theory is useful for making predictions about what kinds of conditions or environments will lead to what kinds of evolutionary adaptations – not for making long-term predictions about the form of specific traits or creatures. However, recent advances in the field of evolutionary biology have shed light on how complexity has arisen on the Earth, on what points on the tree of life this has happened, and on what theoretical conditions favour it (Maynard Smith & Szathmáry 1995; Queller 1997; Bourke 2011; West *et al.* 2015).

In particular, the evolution of complex life on the Earth appears to have depended upon a small number of what have been termed major evolutionary transitions in individuality. In each transition, a group of individuals that could previously replicate independently cooperate to form a new, more complex life form or higher level organism. For example, genes cooperated to form genomes, different single-celled organisms formed the eukaryotic cell, cells cooperated to form multicellular organisms, and multicellular organisms formed eusocial societies (Maynard Smith & Szathmáry 1995; Queller 1997; Bourke 2011; West *et al.* 2015).

Major transitions

Major transitions on the Earth

Major evolutionary transitions are defined by two features. First, entities that were capable of replication before the transition can replicate only as part of a larger unit after it (interdependence). For example, the cells in our bodies cannot evolve back into single-celled organisms. Second, there is a relative lack of conflict within the larger unit, such that it can be thought of as an organism (individual) in its own right (Queller & Strassmann 2009; West *et al.* 2015). For example, it is common to think of a single bird as an individual, and not as a huge community of cells each doing their own thing.

Major transitions are important because the new higher-level organisms that they produce can lead to a great jump in

complexity. For example, the evolution of multicellularity involved a transition from an entity with one part (the single-celled organism) working for the success of itself, to an entity with many parts (the multicellular organism), working for the success of the whole group. The cells can now have very different functions (a division of labour), as each is just a component of a multicellular machine, sacrificing itself for the good of the group, to get a sperm or egg cell into the next generation. As a result, diverse specialized forms such as eyes, kidneys, and brains were able to develop. The rise in complexity on Earth has been mediated by a handful of such jumps, when units with different goals (genes, single cells, individual insects) became intricately linked collectives with a single common goal (genomes, multicellular organisms, eusocial societies). Increases in complexity can also occur through mutations, gene duplications, or even whole genome duplications, but these are not major transitions. These other changes tend to be reversible and gradual, while major transitions are irreversible and cause large leaps in complexity.

The identification of major evolutionary transitions was an empirical observation about how complexity has increased on earth (Maynard Smith & Szathmáry 1995). The next step was to use evolutionary theory to provide insight about when (or under what conditions) we can expect major transitions to occur (Maynard Smith & Szathmáry 1995; Queller 1997; Gardner & Grafen 2009; Bourke 2011; West *et al.* 2015). Major transitions involve the original entities completely subjugating their own interests for the interests of the new collective. This represents an incredibly extreme form of cooperation. Think of the skin or liver cells in your body sacrificing for your sperm or eggs, or the worker ants in a eusocial colony sacrificing for the queen. Evolutionary theory tells us what conditions lead to such extraordinary cooperation.

What conditions drive major transitions?

Consider a multicellular organism, such as yourself. Why don't your hand and heart cells try to reproduce themselves, as opposed to helping your sperm or egg cells? The answer involves genetic similarity or 'relatedness' (Hamilton 1964). Your hand cells contain the same genes as your sperm cells because they are clonal copies. A hand cell could in principle get the same fraction of its genes into the next generation (all of them) by either copying itself, or by helping copy the sperm cells. A similar phenomenon occurs in eusocial insects, such as some ants, bees, wasps and termites. A worker termite can pass on half her genes to her offspring. But a random sibling in the colony (her brother or sister) also contains, on average, half her genes. Thus, a worker can get the same fraction of gene copies into the next generation by reproducing or by helping her mother, the queen, to reproduce (Hamilton 1964; Boomsma 2009). Helping their mother is likely to be more efficient than reproducing on their own, and so our termite can better get their genes into the next generation by helping rather than reproducing (Hamilton 1964; Queller & Strassmann 1998; Bourke 2011).

These are two examples of *alignment of interests*. The 'interests' are evolutionary interests in getting genes into future generations. The hand and the sperm cells both act as if they 'want' to get copies of their genes into the next generation, because as we discussed above, natural selection will have led to them being adapted in this way (Grafen 2003; Gardner 2009). The interests between them are aligned because they share the same genes. When individuals share genes, we say that they are genetically

related. Relatedness is a statistical measure of the extent to which individuals share genes (Grafen 1985).

In the case of eusocial ant colonies and human bodies, the interests are aligned through genetic relatedness. But there are other ways for evolutionary interests to be aligned. Consider, for example, a mutualism between two species. Some aphids carry bacteria in their gut (Moran 2007). The aphids provide the bacteria with sugars and other nutrients to survive and the bacteria provide the aphids with vital amino acids missing from their diet. The aphid and the bacteria do not share the same genes, but neither can reproduce without the other. To reproduce itself, the aphid has to help reproduce the bacteria and vice versa. Again, their evolutionary interests are *aligned*.

The very cells that make up our bodies – known as eukaryotic cells – evolved through a similar kind of alignment of interests (Margulis 1970; Thiergart *et al.* 2012; Archibald 2015). Early in the evolution of life, one bacterial species engulfed another. Over time, the two species took on different roles, with one specializing in replication and the other in energy production. The nucleus of our cells is the descendant of the former, and the mitochondria the latter. Neither can reproduce without the other. Their interests are aligned through reproductive dependence on each other.

All cooperation in nature requires alignment of interests (West *et al.* 2007). Consider, for example, flower pollination by bees. The bee benefits by receiving food from the flower, and the flower benefits by being pollinated. But major transitions are a particularly *extreme* form of cooperation. Compare the pollination scenario to the cells *within* the flower or the bee. Major transitions involve organisms cooperating so completely that they give up their status as individuals, becoming parts of a whole (Queller & Strassmann 2009). Unsurprisingly, then, major transitions require the extreme condition of *effectively* complete or perfect alignment of interests (Gardner & Grafen 2009; West *et al.* 2015).

It is also useful to consider the biology of organisms that do not have interests sufficiently aligned, and thus where conflict remains and major transitions have not occurred. For example, in single-celled organisms, we can compare non-clonal cooperative groups of things like slime moulds with clonal groups such as those that make up multicellular organisms such as humans and trees. These non-clonal groups have evolved only relatively limited division of labour, and never complex multicellular organisms (Fisher *et al.* 2013). Numerous experimental studies have shown that this is because in non-clonal groups non-cooperative ‘cheats’ can spread, limiting the extent of cooperation (Griffin *et al.* 2004; Diggle *et al.* 2007; Kuzdzal-Fick *et al.* 2011; Rumbaugh *et al.* 2012; Pollitt *et al.* 2014; Popat *et al.* 2015; Inglis *et al.* 2017).

Thus, there must be something in place to maintain the alignment of interests (Bourke 2011; West *et al.* 2015). Evolutionary theory can suggest what these somethings would have to be. In multicellular organisms, the something is the single-celled bottleneck (Buss 1987; Queller 2000). Multicellular organisms start each new generation as a single-celled zygote, such that all the cells in the resulting body are clonal (it could also be a spore giving rise to a haploid cell). Eusocial insect colonies evolved from colonies founded by a singly mated queen (Boomsma 2007, 2009, 2013; Hughes *et al.* 2008). If the queen had multiple mating partners, a worker would have half-sisters, and be less related to her siblings than her offspring, breaking down the alignment. The monogamous mating pair is the eusocial colony’s equivalent of a zygote or a bottlenecking event (Boomsma 2013). With unrelated units, like

mitochondria and the nucleus, the individual parts must be co-dependent for joint reproduction (Foster & Wenseleers 2006; West *et al.* 2015) – which can be thought of as a different form of bottleneck. The rarity of conditions like these – conditions under which alignment is so complete – explains the rarity of major transitions in individuality in the history of life.

Biology of organisms that have undergone major transitions

Do the conditions required for major transitions tell us anything about the biology of organisms that have undergone major transitions? Yes. Organisms are a nested hierarchy, where each nested level is the vestige of a former individual (Fig. 3). Eusocial ant colonies function as a single individual, but are made up of multicellular organisms. Those organisms themselves are made up of cells. In turn, those cells resulted from the fusion of two simple species early in evolution. Each of those organisms had a genome that evolved from the union of the individual, replicating molecules.

Further, at each level of the hierarchy, there must be something to *align the interests* of the parts. This usually happens through some form of population bottlenecking. When the parts are related, it is a relatedness bottleneck, such as the single-celled stage in multicellular organisms, or the singly mated female in the social insects (Boomsma 2009, 2013; West *et al.* 2015). When the parts are unrelated, it is usually another form of a bottleneck, such as enforced vertical transmission with joint reproduction (Foster & Wenseleers 2006; West *et al.* 2015). We use the term ‘bottleneck’ to refer to new generations being founded by a strict unit (the zygote, the mutualist pair, etc.), but another way to think of this is that the parts require each other for reproduction (e.g. the soma and the germ line, or the mitochondria and the nucleus). Other, further aligners may be required (e.g. in multicellular organisms, there may need to be a cap on somatic mutations), but these are more likely to be life-form specific.

To conclude so far, empirical observation tells us that complexity has increased on earth through major transitions. Evolutionary theory tells us that for major transitions to occur, the conflict must be eliminated. The theory also tells us what conditions lead to the elimination of conflict. The empirical data agree with the predictions of the theory, in that major transitions have only occurred in the extreme conditions that effectively remove conflict (Boomsma 2007; Hughes *et al.* 2008; Fisher *et al.* 2013; West *et al.* 2015; Fisher *et al.* 2017).

Complex aliens

Complexity and major transitions in space

We can now ask: what does the major evolutionary transition approach tell us about aliens? Will extraterrestrial life undergo major transitions? Not necessarily. Natural selection cannot predict a specific course of evolution. However, as we have said, we might be particularly interested in *complex* aliens. Complexity requires different parts or units working together towards a common goal or purpose. Under natural selection, units are selected to be selfish, striving to replicate themselves at the expense of others. Theory tells us that for units to unite under a common purpose, the evolutionary conflict between them must effectively eliminate (Gardner & Grafen 2009; West *et al.* 2015).

Once again, picture an alien (Fig. 2). If you are picturing something like unlinked replicating molecules or undifferentiated blobs

of slime, then your aliens might not have undergone major transitions. But if what you are picturing has different parts with specialized functions, then your alien is likely to have undergone major transitions (Fig. 2c). What matters is not that we call them ‘major transitions’, but rather that complexity requires multiple parts of an organism striving to the same purpose, and that theory predicts that this requires restrictive conditions (Gardner & Grafen 2009; West *et al.* 2015). Consequently, if we find complex organisms, we can make predictions about what they will be like.

Are there other ways to get complexity? To do so, natural selection would have to sculpt separate parts with unique functions out of a single replicator. Could, for example, the alien equivalent of a single copy of a gene, housed in one ‘cell’ generate the equivalent of limbs and organs? If so, it would disprove our prediction. However, both empirical (major transitions are how complexity has increased on Earth) and theoretical (functional parts requires the elimination of conflict) evidence support the argument that complex aliens will have undergone major transitions.

The biology of complex aliens

Given that complex aliens will have undergone major transitions, we can make a number of predictions about their biology (Fig. 4).

1. They will be entities that are made up of smaller entities – a nested hierarchy of individuality with as many levels as completed transitions. This could mean a collection of replicators, like the first genomes on the Earth, or some hideously complex nesting of groups on a planet where many more transitions have occurred than on our own. For example, you might imagine a ‘society of societies’, where many different social colonies collaborate, with each society specializing on different tasks, such that they are completely dependent on each other. Versions of the simpler entities are likely to be found free-living on the planet as well.
2. Whatever the number of transitions, there will be something that aligns interests, or eliminates conflict within the entities, at the level of each transition.

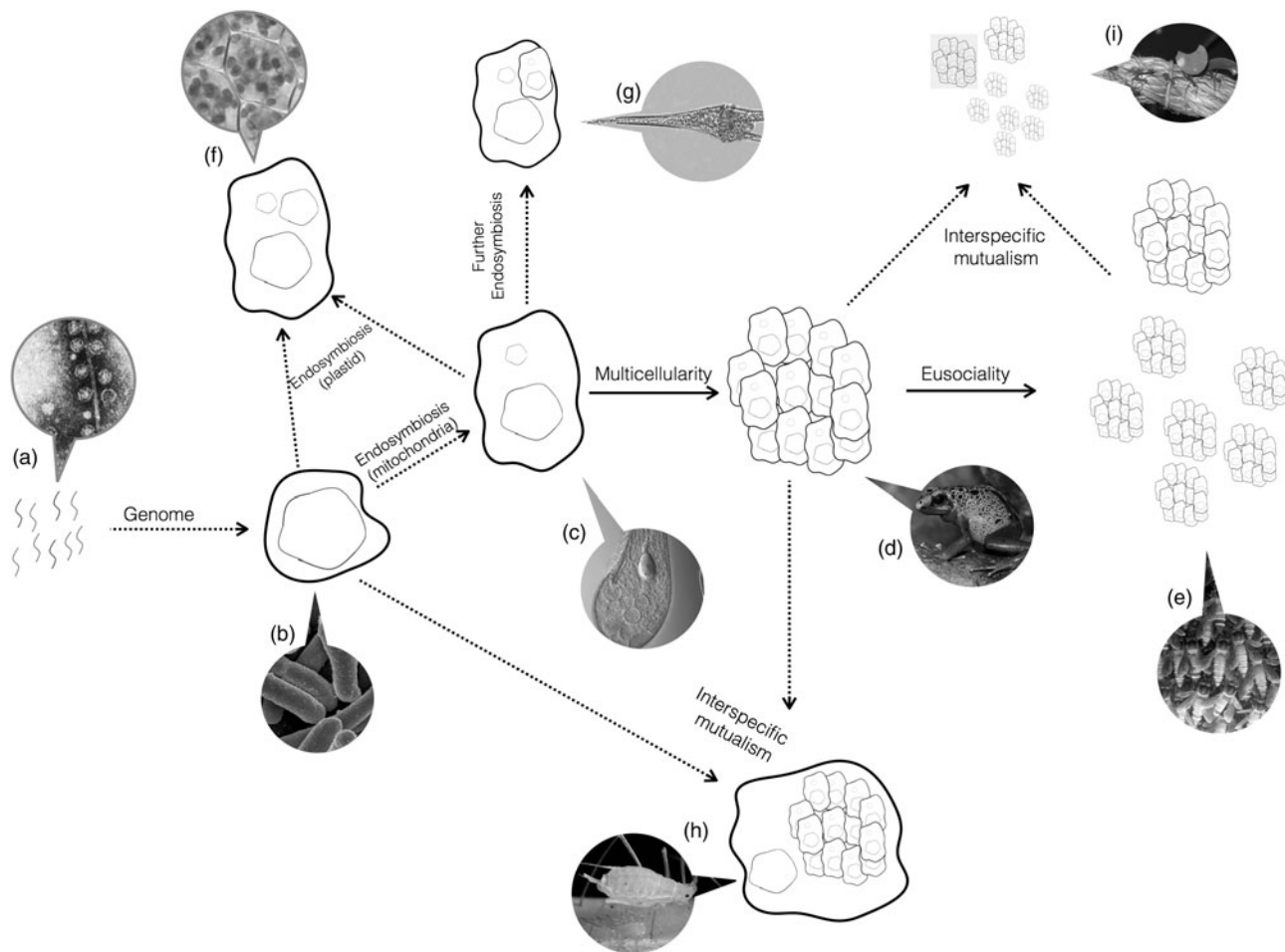


Fig. 3. Major Transitions. Life started with naked replicating molecules, and has since undergone a series of major transitions. Arrows show the occurrence of major transitions in individuality. Dotted arrows represent transitions between dislike things and solid lines represent transitions between like things. Callouts show examples of the present-day organisms that have undergone that transition but no further ones. (a) As we have not yet identified the earliest replicators, Spiegelman's monster, a simple replicating RNA molecule, is shown as an example candidate. (b) A single-celled bacteria, such as *Escherichia coli*. (c) A single-celled eukaryote, like *Blepharisma japonicum*. (d) A multicellular organism, like frogs. (e) An obligate eusocial colony, such as honeybees. (f) Secondary endosymbiosis events, such as the origin of the chloroplast. (g) Further endosymbiosis events, such as those leading to Dinoflagellates. (h) Obligate interspecific mutualisms, such as between aphids and Buchnera bacteria. (i) Obligate mutualisms between a multicellular organism and eusocial colony, such as between leaf-cutter ants and fungi. All images courtesy of Wikipedia.

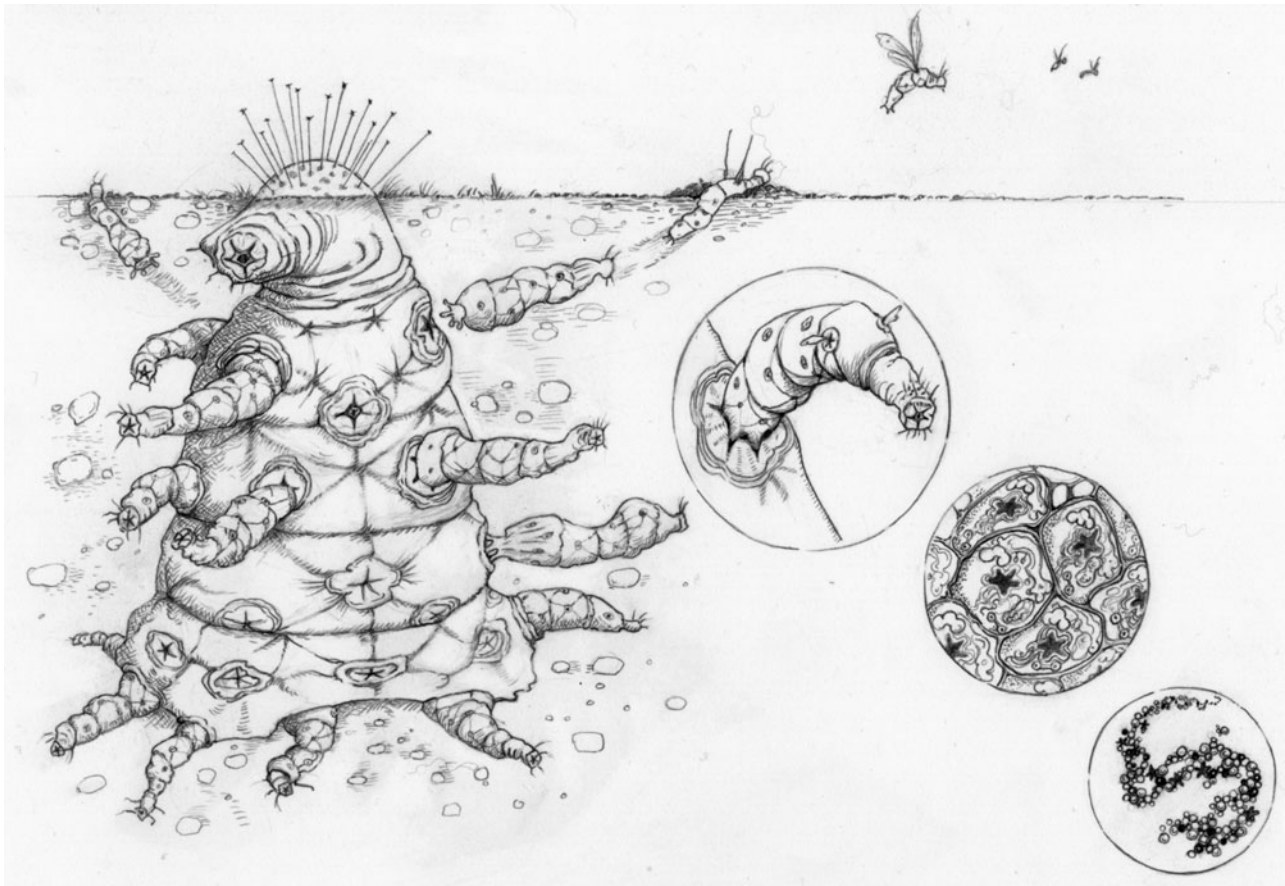


Fig. 4. Major transitions in space: 'The Octomite'. A complex alien that comprises a hierarchy of entities, where each lower-level collection of entities has aligned evolutionary interests such that conflict is effectively eliminated. These entities engage in a division of labour, with various parts specializing on various tasks, such that the parts are mutually dependent.

3. Theory suggests that some sort of population bottlenecking will be key to aligning interests. Bottlenecking is not necessarily the only way to eliminate conflict, but it is probably the easiest evolutionary route to take. In particular, it does not require additional mechanisms of enforcement, such as kin discrimination, policing or randomization. The specific kinds of bottlenecking will depend on whether like or dislike units are united.

- a. When like entities come together, interests can be aligned through a bottleneck similar to our single-celled bottleneck in multicellular organisms or the single mating pair in eusocial colonies, which maximizes relatedness between entities.
- b. If the organisms are made up different types of entities, we can expect something similar to the bottleneck that forces mitochondria and nuclei to pass to the next generation together, with joint reproduction. By trapping individuals together over evolutionary time, their interests become aligned.
- c. Some aliens, like us, may contain both types of conflict reduction, for having both like and dislike types joined within them.

Conclusion

When using evolutionary theory to make predictions about extraterrestrial life, it is important to avoid circularity. Our chain of

argument is: (1) Extraterrestrial life will have undergone natural selection. (2) Knowing that aliens undergo natural selection, we can make further predictions about their biology, based on the theory of natural selection. In particular, we can say something about complex aliens – that they will likely have undergone major transitions. (3) Theory tells us that restrictive conditions, which eliminate conflict, are required for major transitions. (4) Consequently, complex aliens will be composed of a nested hierarchy of entities, with the conditions required to eliminate conflict at each of those levels.

When making predictions about aliens, we must take advantage of our entire scientific toolkit. Mechanistic understanding is a good way to extrapolate from what we see on Earth. The theory is a good way to make predictions that are independent of the details of the Earth. Combining both approaches is the best way to make predictions about the many hundreds, thousands or millions of hypothetical aliens. Now we just need to find them.

Acknowledgements. We thank The Clarendon Fund, Hertford College, and the Natural Environment Research Council for funding; and Magdalen College for emergency housing.

Author disclosure statement. No competing financial interests exist.

References

Archibald JM (2015). Endosymbiosis and Eukaryotic Cell Evolution. *Current biology*: CB 25(19), R911–921.

- Benner SA** (2003). Synthetic biology: Act natural. *Nature* **421**(6919), 118.
- Boomsma JJ** (2007). Kin selection versus sexual selection: why the ends do not meet. *Current biology* : *CB* **17**(16), R673–R683.
- Boomsma JJ** (2009). Lifetime monogamy and the evolution of eusociality. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **364**(1533), 3191–3207.
- Boomsma JJ** (2013). *Nature's True Self. Science* (New York, N.Y.) **340**(6129), 145–146.
- Bourke AFG** (2011). *Principles of Social Evolution*. Oxford University Press.
- Buss LW** (1987) *The Evolution of Individuality*. Princeton University Press.
- Cassan A, Kubas D, Beaulieu J-P, Dominik M, Horne K, Greenhill J, Wambsgans J, Menzies J, Williams A, Jørgensen UG, Udalski A, Bennett DP, Albrow MD, Batista V, Brillant S, Caldwell JAR, Cole A, Coutures C, Cook KH, Dieters S, Prester DD, Donatowicz J, Fouqué P, Hill K, Kains N, Kane S, Marquette J-B, Martin R, Pollard KR, Sahu KC, Vinter C, Warren D, Watson B, Zub M, Sumi T, Szymanski MK, Kubiak M, Poleski R, Soszynski I, Ulaczyk K, Pietrzynski G and Wyrzykowski L** (2012). One or more bound planets per Milky Way star from microlensing observations. *Nature* **481**(7380), 167–169.
- Cleland CE and Chyba CF** (2002) Defining 'life'. *Origins of Life and Evolution of the Biosphere* **32**(4), 387–393.
- Clutton-Brock TH, Harvey PH and Rudder B** (1977). Sexual dimorphism, socio-economic sex ratio and body weight in primates. *Nature* **269**(5631), 797–800.
- Cohen J and Stewart I** (2001). Where are the dolphins? *Nature* **409**(6823), 1119–1122.
- Corning PA and Szathmáry E** (2015). "Synergistic selection": a Darwinian frame for the evolution of complexity. *Journal of theoretical biology* **371**: 45–58.
- Darwin C** (1859). *On the origins of species by means of natural selection*. London: Murray 247.
- Darwin C** (1871). *The descent of man, and selection in relation to sex*. By Charles Darwin. New York, D. Appleton and company.
- Davies NB and Houston AI** (1981). Owners and Satellites: The Economics of Territory Defence in the Pied Wagtail, *Motacilla alba*. *The Journal of Animal Ecology* **50**(1), 157.
- Davies NB, Krebs JR and West SA** (2012) *An Introduction to Behavioural Ecology*. John Wiley & Sons.
- Davies PCW, Benner SA, Cleland CE, Lineweaver CH, McKay CP and Wolfe-Simon F** (2009). Signatures of a shadow biosphere. *Astrobiology* **9** (2), 241–249.
- Des Marais DJ, Nuth JA, Allamandola LJ, Boss AP, Farmer JD, Hoehler TM, Jakosky BM, Meadows VS, Pohorille A, Runnegar B and Spormann AM** (2008). The NASA Astrobiology Roadmap. *Astrobiology* **8** (4), 715–730.
- Diggle SP, Griffin AS, Campbell GS and West SA** (2007). Cooperation and conflict in quorum-sensing bacterial populations. *Nature* **450**(7168), 411–414.
- Domagal-Goldman SD, Wright KE, Adamala K, Arina de la Rubia L, Bond J, Dartnell LR, Goldman AD, Lynch K, Naud M-E, Paulino-Lima IG, Singer K Walter-Antonio M, Abrevaya XC, Anderson R, Arney G, Atri D, Azúa-Bustos A, Bowman JS, Brazelton WJ, Brenneka GA, Carns R, Chopra A, Colangelo-Lillis J, Crockett CJ, DeMarines J, Frank EA, Frantz C, de la Fuente E, Galante D, Glass J, Gleeson D, Glein CR, Goldblatt C, Horak R, Horodyskyj L, Kaçar B, Kereszturi A, Knowles E, Mayeur P, McGlynn S, Miguel Y, Montgomery M, Neish C, Noack L, Rugheimer S, Stüeken EE, Tamez-Hidalgo P, Imari Walker S and Wong T** (2016). The Astrobiology Primer v2.0. *Astrobiology* **16**(8), 561–653.
- Fisher RA** (1930). *The genetical theory of natural selection: a complete variorum edition*. Oxford University Press.
- Fisher RM, Cornwallis CK and West SA** (2013) Group formation, relatedness, and the evolution of multicellularity. *Current Biology*, Cell Press, **23** (12), 1120–1125.
- Fisher RM, Henry LM, Cornwallis CK, Kiers ET and West SA** (2017) The evolution of host-symbiont dependence. *Nature Communications*, Nature Publishing Group, **8**, 15973.
- Flores Martinez CL** (2014). SETI in the light of cosmic convergent evolution. *Acta Astronautica* **104**(1), 341–349.
- Foster KR and Wenseleers T** (2006). A general model for the evolution of mutualisms. *Journal of Evolutionary Biology* **19**(4), 1283–1293.
- Gardner A** (2009). Adaptation as organism design. *Biology Letters* **5**(6), 861–864.
- Gardner A and Grafen A** (2009). Capturing the superorganism: a formal theory of group adaptation. *Journal of Evolutionary Biology* **22**(4), 659–671.
- Grafen A** (1985). A geometric view of relatedness. *Oxford surveys in evolutionary biology* **2**, 28–89.
- Grafen A** (2003). Fisher the evolutionary biologist. *Journal of the Royal Statistical Society: Series D (The Statistician)* **52**(3), 319–329.
- Griffin AS, West SA and Buckling A** (2004). Cooperation and competition in pathogenic bacteria. *Nature* **430**(7003), 1024.
- Hamilton WD** (1964). The genetical evolution of social behaviour I and II. *Journal of theoretical biology* **7**(1), 1–52.
- Horneck G, Walter N, Westall F, Grenfell JL, Martin WF, Gomez F, Leuko S, Lee N, Onofri S, Tsiganis K, Saladino R, Pilat-Lohinger E, Palomba E, Harrison J, Rull F, Muller C, Strazzulla G, Brucato JR, Rettberg P and Capria MT** (2016). AstRoMap European Astrobiology Roadmap. *Astrobiology* **16**(3), 201–243.
- Hughes WOH, Oldroyd BP, Beekman M and Ratnieks FLW** (2008). Ancestral monogamy shows kin selection is key to the evolution of eusociality. *Science* (New York, N.Y.) **320**(5880), 1213–1216.
- Inglis RF, Ryu E, Asikhia O, Strassmann JE and Queller DC** (2017). Does high relatedness promote cheater free multicellularity in synthetic life-cycles?. *Journal of Evolutionary Biology* **30**(5), 985–993.
- Kuzdzal-Fick JJ, Fox SA, Strassmann JE and Queller DC** (2011). High relatedness is necessary and sufficient to maintain multicellularity in *Dictyostelium*. *Science* **334**(6062), 1548–1551.
- Margulis L** (1970). Recombination of non-chromosomal genes in *Chlamydomonas*: assortment of mitochondria and chloroplasts? *Journal of theoretical biology* **26**(2), 337–342.
- Moran NA** (2007). Symbiosis as an adaptive process and source of phenotypic complexity. *Proceedings of the National Academy of Sciences* **104** Suppl 1 (Supplement 1), 8627–8633.
- Morris SC** (2003). The navigation of biological hyperspace. *International Journal of Astrobiology* **2**(2), 149–152.
- Petigura EA, Howard AW and Marcy GW** (2013). Prevalence of Earth-size planets orbiting Sun-like stars. *Proceedings of the National Academy of Sciences of the United States of America* **110**(48), 19273–19278.
- Pollitt EJ, West SA, Cruz SA, Burton-Chellew MN and Diggle SP** (2014). Cooperation, quorum sensing, and evolution of virulence in *Staphylococcus aureus*. *Infection and immunity* **82**(3), 1045–1051.
- Popat R, Pollitt EJ, Harrison F, Naghra H, Hong KW, Chan KG, Griffin AS, Williams P, Brown SP, West SA and Diggle SP** (2015) Conflict of interest and signal interference lead to the breakdown of honest signaling. *Evolution. The Society for the Study of Evolution* **69**(9), 2371–2383.
- Queller DC** (1997). Cooperators Since Life Began The Major Transitions in Evolution. John Maynard Smith, Eors Szathmáry. *The Quarterly Review of Biology* **72**(2), 184–188.
- Queller DC** (2000). Relatedness and the fraternal major transitions. *Philosophical Transactions of the Royal Society B: Biological Sciences* **355** (1403), 1647–1655.
- Queller DC and Strassmann JE** (1998) Kin selection and social insects. *Bioscience*, Oxford University Press, **48**(3), 165–175.
- Queller DC and Strassmann JE** (2009). Beyond society: the evolution of organismality. *Philosophical Transactions of the Royal Society B: Biological Sciences* **364**(1533), 3143–3155.
- Rothschild LJ** (2009). Defining the envelope for the search for life in the Universe. *Proceedings of the International Astronomical Union* **5**(H15): 697–698.
- Rothschild LJ** (2010). A powerful toolkit for synthetic biology: Over 3.8 billion years of evolution. *BioEssays* **32**(4), 304–313.
- Rumbaugh KP, Trivedi U, Watters C, Burton-Chellew MN, Diggle SP and West SA** (2012). Kin selection, quorum sensing and virulence in pathogenic bacteria. *Proceedings of the Royal Society B: Biological Sciences* **279**(1742), 3584.

- Schneider D** (2016). \$100 million seti initiative starts listening for E.T. *IEEE Spectrum* **53**(1), 41–42.
- Shostak S** (2015). Searching for Clever Life. *Astrobiology* **15**(11), 949–950.
- Smith JM and Szathmáry E** (1995). The major evolutionary transitions. *Nature* **374**(6519), 227–232.
- von Salvini-Plawen L and Mayr E** (1977). *On the Evolution of Photoreceptors and Eyes*. Boston, MA, Springer US: 207–263.
- Thiergart T, Landan G, Schenk M, Dagan T and Martin WF** (2012). An evolutionary network of genes present in the eukaryote common ancestor polls genomes on eukaryotic and mitochondrial origin. *Genome Biology and Evolution* **4**(4), 466–485.
- West S** (2009). *Sex Allocation*. Princeton, Princeton University Press.
- West SA, Fisher RM, Gardner A, and Kiers ET** (2015). Major evolutionary transitions in individuality. *Proceedings of the National Academy of Sciences* **112**(33), 10112–10119.
- West SA, Griffin AS and Gardner A** (2007). Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. *Journal of evolutionary biology* **20**(2), 415–432.