



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

The selfishness of law-abiding genes

Citation for published version:

Bird, A 2019, 'The selfishness of law-abiding genes', *Trends in Genetics*.
<https://doi.org/10.1016/j.tig.2019.10.002>

Digital Object Identifier (DOI):

[10.1016/j.tig.2019.10.002](https://doi.org/10.1016/j.tig.2019.10.002)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Trends in Genetics

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Opinion

The Selfishness of Law-Abiding Genes

Adrian Bird^{1,*}

Selfish genes were once controversial, but it is now accepted that the genome contains parasitic elements in addition to a complement of conventional genes. This opinion article argues that 'law-abiding' genes also indulge in game playing to ensure their propagation, so that initially nonessential processes secure a genetic heritage. A gene-centered view of this kind can help to explain otherwise puzzling aspects of biology, including the complexity and stability of living systems.

Modes of Gene Self-Interest

It is generally believed that the genome comprises a collection of genes working as a team to generate a reproductively competent organism. Each gene, so the argument goes, is there because it enhances this goal. If it did not do so, it would be lost due to rigorous scrutiny by natural selection. This consensus has been challenged in the past, firstly by neutral theory, which established that much of the variability among genes is due to stochastic events that become fixed by chance, and secondly by the selfish gene hypothesis [1], which argued that while most genes indulge their selfishness by bolstering the reproductive success of the host, some exploit the genomic landscape to accomplish their own spread, despite contributing little or nothing to organismal fitness. The idea that the hitherto 'safe space' of our DNA might be a playground for self-interested genetic elements was initially anathema to many biologists, but has now been assimilated into conventional wisdom. We accept that transposons and certain other types of DNA, such as the t-complex in mice or B chromosomes in plants, are either of negligible benefit to the host or even somewhat detrimental [2]. They nevertheless persist due to their ability to drive their own selection, for example, by out-proliferating the host genome or by insuring that cells failing to contain them are disadvantaged [3]. Members of this class can be viewed as 'outlaws', in contrast to the majority of genes, which are 'law-abiding'.

The argument here is that the importance of self-interest to explain the presence of many apparently law-abiding genes is underestimated. While one way of securing a genomic niche in perpetuity is to help insure that the host organism remains a thriving reproductive entity, there are strategies that involve 'gaming the system', whereby nonessential or even deleterious processes become assimilated into the genome. This distinction may seem prosaic, but it is highly relevant to the quest for mechanistic understanding of living processes at the molecular level. The underlying concepts aired here are not new, but they are often overlooked at the molecular extreme of biology, particularly in relation to multicellular eukaryotes, which are the main focus here. One aim of this article is to address this omission. Specifically, it discusses how genes lacking special mechanisms to drive their own proliferation may, in the fullness of time, acquire properties that allow them to establish a foothold in the genome, despite minimal or no intrinsic benefit to life processes of the host organism. According to this scenario, genes acquire properties that contribute to two functions that may or may not overlap: organismal fitness and genome membership in perpetuity. A second aim of this piece is to argue that recognition of this duality of purpose opens up a broader view of biological function, with the potential to shed light on several perplexing features, including robustness and biological complexity.

Law-Abiding Genes with Ulterior Motives

An illustration of the prevailing bias in interpretation of gene function is provided by the sup-35/pha-1 gene pair in *Caenorhabditis elegans*. The pha-1 gene was originally classified based on functional studies as a developmental gene essential for normal pharyngeal differentiation [4]. Analysis of natural *C. elegans* strains revealed, however, that pha-1 is always associated with the presence of sup-35, which turns out to be a lethal toxin whose action is suppressed by pha-1 [5]. Removal of both gene products restores normal development, demonstrating that this toxin-antitoxin pair

Highlights

The presence of selfish elements in the genome is accepted, but the remainder of genes are usually considered to be 'law-abiding'.

'Law-abiding' genes can also indulge in game playing to ensure their propagation, allowing nonessential processes to secure a genetic heritage.

Appreciation of this aspect of gene 'function' can shed new light on the origin of molecular mechanisms that underlie biological processes.

More generally, a bottom-up 'gene-centric' view of genome evolution may help to explain the complexity and robustness of biological entities.

Molecular complexity of living organisms may therefore be driven from below by the opportunism of genes.

¹Wellcome Centre for Cell Biology, University of Edinburgh, Michael Swann Building, The King's Buildings, Max Born Crescent, Edinburgh, EH9 3BF, UK

*Correspondence: a.bird@ed.ac.uk

constitutes a selfish element. The authors speculate that other essential genes may turn out to be antidotes to as yet unknown toxins. While toxin–antitoxin gene pairs are abundant in prokaryotes [6], they are less obvious in eukaryotes, although unsuspected examples have recently emerged in budding yeast [7]. Neutralization of a toxin can of course be considered unremarkable, as it enhances the fitness of the organism in the conventional sense. What makes the case noteworthy is that its root driver is propagation of the two genes concerned, rather than enhancement of organismal development as originally assumed. A better-known example of a biological phenomenon that only made sense when viewed from a gene-centric perspective is genomic imprinting, whereby a gene donated to the zygote by one of the parents arrives in a silenced state that often endures throughout development [8]. The role played by that gene must consequently be provided by an active copy from the other parent. The natural tendency to seek a benefit to the organism in such parental asymmetry was short-circuited by the striking proposal that this phenomenon was the result of a competitive game played between parental genomes to benefit their genetic heritage [9]. The opportunity arises in organisms where the embryo develops within the mother’s body, leading to competition for limited resources between male and female genomes. Here again, most imprinting does not appear to enhance development and is therefore, in conventional biological terms, pointless.

For some genes, selfish¹ and beneficial aspects are less easy to disentangle. For example, the RNA editases known as ADARs convert certain adenines in double-stranded RNA to inosine, which mimics guanine. Codon alteration via this post-transcriptional base change converts an errant glutamine codon to arginine at a critical site within the glutamate receptor 2 mRNA. This change is obligatory if the gene is to produce a fully functional neurotransmitter receptor [10]. In other words, a severely deleterious mutation in *GIA2* is rescued by *ADAR2*, allowing both genes to persist in the population. Mutual dependence ensures that the otherwise dispensable activity of the *ADAR2* protein becomes essential. In agreement with this notion, mice lacking *ADAR2* that also have the missense mutation in *GIA2* corrected have no obvious phenotypic defects [11]. However, many members of the ADAR family can also be viewed as law-abiding genes whose presence fits neatly into the orthodox biology of the organism due to their role in innate immunity and genome defense [12]. This duality highlights a difficulty in identifying examples of gene-centered rather than organism-centered evolution: namely, the tendency of genes to eventually acquire new functions that can be viewed as (and often are) advantageous to the organism. Transitions of this kind are well documented for transposons, which are evidently purely selfish to start with, but, given time, can be assimilated into the genomic mainstream; for example, by being coopted as novel promoters or derivative proteins that enhance organismal fitness [13,14]. For many conventional genes, the tendency to assume that such acquired functions provide their sole biological rationale may mask an underlying history of self-interest.

In what ways could apparently law-abiding genes indulge in a degree of self-interest to insure their persistence? A general mechanism would be for a dispensable gene to render the host organism addicted to its presence, just as cancer cells can become addicted to oncogenes [15]. The genes concerned would essentially hold the host organism to ransom, but this does not distinguish them from those whose role is to conventionally enhance organism fitness. A possible way of detecting an ulterior motive is to assume that genes that have ‘inveigled’ their way into the genome opportunistically will betray their presence by functional volatility during evolution. Most of us prefer to think of the genes we study as agents of biological progress rather than insidious exploiters of the genome, but it is a useful exercise to entertain alternative scenarios. The DNA methyltransferases, for example, are effectively absent in some animals (the fruit fly *Drosophila* and nematode worm *C. elegans*) and therefore, to a degree, dispensable for metazoan life. Interestingly, their biological significance, which at the molecular level involves adding a methyl group to DNA cytosine, has fluctuated widely during evolution. In bacteria their proposed role in defense of the genome against transposable

¹Needless to say, the present discussion in no way intends to bestow human foibles on DNA, as genes are self-evidently not sentient and cannot therefore be selfish, law-abiding, or act strategically. As discussed by others (see [1]), it is regrettably almost impossible to describe such systems without using words that bear the load of their everyday meaning.

elements and bacteriophages is well established. Less appreciated is the notion that they are commonly part of a selfish toxin–antitoxin gene pair comprising an endonuclease able to destroy the host genome and its obligatory partner, the protective DNA methyltransferase [16,17]. Cytosine methylation is implicated in multiple functions in eukaryotes. In some it represses mobile genetic elements [18], in certain mosses its role is to phase nucleosomes [19], in the fungus *Neurospora* it facilitates rapid DNA sequence destruction via active mutagenesis [20], and in vertebrates it can provoke chromatin modification [21] and reinforce gene silencing [22]. This array of roles might suggest that this enzyme activity is intrinsically biologically valuable. Alternatively, the persistence of DNA methyltransferase genes may reflect a high degree of evolutionary opportunism [23], which has allowed them to make their activity indispensable in diverse ways.

DNA methylation readers could conceivably aid and abet this quest for a biological future. One such protein, MeCP2, can fit with this scenario, as its absence subtly affects the expression of large numbers of genes [24,25]. Loss of MeCP2 via mutation in humans leads to serious defects in the brain [26], which might suggest that MeCP2 is a fundamental regulator of nervous system development. Evolutionary considerations question this view, however, as most animals have nervous systems, but only vertebrates, which account for a small proportion of the animal kingdom, have MeCP2. This protein therefore appears to be a late arrival in evolutionary terms, rather than being a core ancestral component of brain assembly. A conventional view of MeCP2 function is that by exerting global transcriptional restraint it tunes gene expression in neurons to optimize their identity, but it is also possible to devise a scenario based on self-interest. Initially, the argument goes, MeCP2 was present at low levels, as it is in non-neuronal tissues, and therefore played little or no role in creating an optimal nervous system. Because DNA methylation is sparse in the great majority of the genome, sporadic mutations that led to mildly increased MeCP2 expression would have had a minimal dampening effect on transcription that may initially have been selectively neutral. If not eliminated by drift, further chance increases might have followed, with neuronal development incrementally adjusting to each minor hike in MeCP2-mediated repression through compensatory mutations in other genes. Mechanisms that lead to ‘constructive neutral evolution’ of this kind have been proposed [27–29]. Gradually, brain development would accommodate the encroachment of MeCP2 until it became an essential feature. So, in response to the question ‘why do brains need MeCP2?’, the answer under this speculative scenario would be: ‘they do not; MeCP2 has made itself indispensable by stealth’.

The gene-centric view adopted above springs from a long tradition in genetics [30,31], but is little used in molecular cell biology, as preferred scenarios are invariably top-down. Consider genes encoding transcription factors, which are cast as dominant agents (e.g., ‘master-regulators’) whose product proteins select their servile target genes from above in order to impose rational control. The opposite state of affairs, whereby genes exploit chance DNA sequence changes to opt-in to the regulatory opportunity afforded by an abundant factor is rarely articulated. For example, HOX transcription factors are archetypal master regulators of development, but the A/T-rich DNA sequences recognized by HOX proteins are vague and consequently do not confer much intrinsic specificity (discussed in [32]). The high abundance of HOX proteins in certain tissues during development, together with their low specificity, represents a standing invitation to potential target genes to buy into, or drop out of, their ability to affect transcription. Thus, it is mutations in the target genes that define the biological effects of the transcription factor from below. This bottom-up perspective can help to explain why the transcription factors that drive genes change extraordinarily rapidly during evolution, varying between species, between individuals of the same species, and even between alleles in the same individual [33]. The curiously relaxed target sequence specificity of many eukaryotic transcription factors may be driven in part by the need to maximize the likelihood of being ‘chosen’ by evolving control regions of downstream genes. A gene-centric perspective also sheds light on why some factors can be either activators or repressors of gene expression. The factors simply provide discrete surfaces that client target genes exploit in any way that is expedient. Rather than asserting their anointed role as an activator or a repressor, the factor is the creature of its clients, to be used in any manner that furthers their or the organism’s interests (or both), whether as a stimulant or restrainer of transcription or in some other role.

The Genome as an Ecosystem Can Help to Explain Biological Complexity

If we consider the genome to be an exquisitely regulated information repository, finely honed over millions of years under the ruthless scrutiny of natural selection to generate the optimal organisms that we see today, then how could so much game playing be tolerated? A system with so little apparent regard for the 'common good' surely risks collapse under pressure of mass self-interested exploitation. In fact, the genome of multicellular eukaryotes appears to be remarkably tolerant of surplus DNA that contains nascent and obsolete genetic elements. In unicellular organisms, whose proliferation rate is an important determinant of their survival, there is more pressure to off-load surplus DNA than in multicellular plants and animals, whose reproductive success is unrelated to cell division rate and therefore relatively unconstrained by genome size. The evolutionary 'slack' that accompanies multicellularity therefore creates the space as well as (crucially) the time needed to indulge in long-term genomic game playing. Put another way, the ability of many functionally neutral or even slightly deleterious genes to persist for long periods without elimination means that genes that by chance acquire a selfish agenda may inveigle their way into the genetic logic of the organism without the need for a driver. Their incorporation will be inevitably slower than that of transposons, but arguably feasible over evolutionary time.

Might an increased appreciation of the selfish strategies adopted by genes help us to understand biology? The flamboyant complexity of organisms at the molecular level, for example, poses an on-going challenge. Alluding to the difficulty in understanding how genome sequences can be interpreted, Sydney Brenner famously suggested that 'we are drowning in a sea of data, but starving for knowledge' (Nobel Lecture, 2002). A gene-centric view may help here. According to the scenarios outlined above, complexity arises because gene variants are constantly probing for new genomic strategy that will guarantee their place for posterity. From this point of view, the genome resembles an ecosystem in which genes explore every available niche that could permit their survival into the next generation and beyond. A free-for-all conducted in geological time. An ecosystem analogy for the genome has been invoked previously in relation to the colonization and spread of transposable elements [34], but is here extended to all genes.

The conception of genomes as ecosystems in which genes compete to participate carries with it the implication that capacity is limited, otherwise more and more genes would join the party leading to a rampant expansion of gene number. In fact, while genome size is highly variable in animals and plants, the number of protein-coding genes appears to be constrained. Most vertebrates have ~20 000–25 000 protein-coding genes, which is similar to the number in many invertebrates, including small-genomed nematode worms (~20 000) and fruit flies (~15 000), as well as the angiosperm plant *Arabidopsis* (~25 000). Even the microscopic unicellular budding yeast has ~6000 genes. The origin of an apparent limitation on gene number is poorly understood, but of considerable interest. One candidate enforcer is the mutation rate, which relentlessly introduces genetic changes that are in most cases destructive. Perhaps on-going degradation due to mutation limits the amount of genetic information that can be reliably transmitted, leading to an approximate ceiling on the permissible number of genes. Regardless of the underlying cause, the availability of abundant genomic DNA, only a small proportion of which encodes proteins, creates the conditions whereby uncommitted genomic segments can compete to join the cohort of essential genes.

Genetics Meets Sociology?

A conventional view of complexity is that it is adaptive and therefore somehow a response to selection pressure. Other scenarios have also been proposed, whereby complexity arises due to evolutionary accumulation of neutral changes, including gene duplication or the activity of mobile genetic elements [27,35]. Under models of this kind, such features build up nonadaptively in organisms whose effective population size is too small to purge them via purifying selection. The self-interest of genes, often dependent on neutral intermediates, would insure occupation of interlocking niches, inevitably leading to increased complexity. Stability is an emergent feature of many complex systems involving biology, but just how complexity and stability are connected is a matter of debate, for example, in relation to the apparent self-regulation of the biosphere [36]. It is worth considering

the possibility that the unending quest for niche occupation by genes will lead to stability as a by-product. The inexorable accumulation of new molecular strategies naturally increases complexity, but at the same time improves robustness, assuming that potential gene niches often correspond to phenotypic weaknesses. By plugging such potential gaps, the competition among genes for membership of the genome will shore up the entire enterprise. Looked at this way, both biological complexity and robustness are a necessary by-product of gene self-interest.

Biologists often invoke parallels between molecular processes of life and computer logic, but a gene-centered approach suggests that economics or social science may be a more appropriate model [14]. While aware of the seductive and often misleading nature of such anthropomorphic analogies, it is informative to invoke another: human bureaucracy. Here too the large-scale framework of a higher purpose creates the freedom for constituent members to indulge their self-seeking aspirations. Human bureaucratization can be seen as driven from below by the relentless urges of human self-interest, just as the reproductive imperative of biological entities may be driven by the self-interest of genes. Bureaucracies conjure up a vision of bloated inefficiency and needless complexity, but this overlooks two key attributes: robustness and stability. These human conglomerations are marvels of survivability, often despite a markedly suboptimal performance of their overarching role. They are sustained, it is suggested here, by the mixed, predominantly selfish, motivations of their constituent members. In the same way, a biological entity in which genes struggle to persist, but which requires the activity of those same genes to thrive reproductively, might also acquire robustness. A corollary of this speculation is that a mathematical treatment of human bureaucracy in the terms outlined above, rather than in terms of collectively shared high-level goals, could be usefully applied to the evolution of genomes (and vice versa). The convergence of biology and social science, which this conjecture implies, is a daunting prospect, but may be inevitable. It may even benefit both disciplines.

Concluding Remarks

Conventionally, gene function is considered almost exclusively from the standpoint of organism-centred evolution. Accepting that evolution does not occur exclusively at the level of organismal fitness, I argue here that a gene-centred perspective improves the explanatory power of molecular genetics. Although the underpinning ideas have been around for many years, there are relatively few examples where the self-interest of genes has been invoked to explain biological phenomena, especially in multicellular eukaryotes. This could of course reflect the rarity of this mode of gene-centred evolution, but more likely it is a by-product of a utilitarian cultural tradition that sees organisms exclusively as purpose-driven machines. The approach raises numerous unanswered questions, some of which are listed (see [Outstanding Questions](#)). Most interesting is the possibility that unexplained aspects of biology at the molecular level, notably robustness and complexity, are emergent properties generated as a passive consequence of genome ecology.

Acknowledgments

I thank Eugene Koonin, Maria Luisa Cochella, and Matt Lyst for commenting on the manuscript. A.B. is a Wellcome Investigator (107930), holder of a European Research Council Advanced Grant (EC 694295), and a member of the Simons Initiative for the Developing Brain.

References

1. Dawkins, R. (1976) *The Selfish Gene* (Oxford University Press)
2. Burt, A. and Trivers, R. (2008) *Genes in Conflict: The Biology of Selfish Genetic Elements* (Belknap Press)
3. Lindholm, A.K. et al. (2016) The ecology and evolutionary dynamics of meiotic drive. *Trends Ecol. Evol.* 31, 315–326
4. Schnabel, H. and Schnabel, R. (1990) An organ-specific differentiation gene, pha-1, from *Caenorhabditis elegans*. *Science* 250, 686–688
5. Ben-David, E. et al. (2017) A maternal-effect selfish genetic element in *Caenorhabditis elegans*. *Science* 356, 1051–1055
6. Van Melderen, L. and Saavedra De Bast, M. (2009) Bacterial toxin-antitoxin systems: more than selfish entities? *PLoS Genet* 5, e1000437
7. Nuckolls, N.L. et al. (2017) *wtf* genes are prolific dual poison-antidote meiotic drivers. *eLife* 6, e26033
8. Barlow, D.P. and Bartolomei, M.S. (2014) Genomic imprinting in mammals. *Cold Spring Harb. Perspect. Biol.* 6, a018382

Outstanding Questions

Does a re-evaluation of attributed gene functions expose potential examples of self-interest that may have been overlooked?

Can further examples of toxin-antitoxin gene pairs be identified (e.g., in vertebrates)?

Is the total number of protein-coding genes constrained; if so, by what?

Can multicomponent entities that depend on the self-interested strategies of their constituent members be described mathematically?

9. Moore, T. and Haig, D. (1991) Genomic imprinting in mammalian development: a parental tug-of-war. *Trends Genet* 7, 45–49
10. Sommer, B. et al. (1991) RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. *Cell* 67, 11–19
11. Higuchi, M. et al. (2000) Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2. *Nature* 406, 78–81
12. Krishnan, A. et al. (2018) Diversification of AID/APOBEC-like deaminases in metazoa: multiplicity of clades and widespread roles in immunity. *Proc. Natl. Acad. Sci. U. S. A.* 115, E3201–E3210
13. Jordan, I.K. et al. (2003) Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends Genet* 19, 68–72
14. Haig, D. (2016) Transposable elements: self-seekers of the germline, team-players of the soma. *Bioessays* 38, 1158–1166
15. Weinstein, I.B. (2002) Cancer. Addiction to oncogenes—the Achilles heel of cancer. *Science* 297, 63–64
16. Naito, T. et al. (1995) Selfish behavior of restriction-modification systems. *Science* 267, 897–899
17. Mruk, I. and Kobayashi, I. (2014) To be or not to be: regulation of restriction-modification systems and other toxin-antitoxin systems. *Nucleic Acids Res* 42, 70–86
18. Kim, M.Y. and Zilberman, D. (2014) DNA methylation as a system of plant genomic immunity. *Trends Plant Sci* 19, 320–326
19. Huff, J.T. and Zilberman, D. (2014) Dnmt1-independent CG methylation contributes to nucleosome positioning in diverse eukaryotes. *Cell* 156, 1286–1297
20. Selker, E.U. et al. (1993) Dense nonsymmetrical DNA methylation resulting from repeat-induced point mutation in *Neurospora*. *Science* 262, 1724–1728
21. Nan, X. et al. (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* 393, 386–389
22. Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev* 16, 6–21
23. Colot, V. and Rossignol, J.L. (1999) Eukaryotic DNA methylation as an evolutionary device. *Bioessays* 21, 402–411
24. Kinde, B. et al. (2016) DNA methylation in the gene body influences MeCP2-mediated gene repression. *Proc. Natl. Acad. Sci. U. S. A.* 113, 15114–15119
25. Lagger, S. et al. (2017) MeCP2 recognizes cytosine methylated tri-nucleotide and di-nucleotide sequences to tune transcription in the mammalian brain. *PLoS Genet* 13, e1006793
26. Guy, J. et al. (2011) The role of MeCP2 in the brain. *Annu. Rev. Cell Dev. Biol.* 27, 631–652
27. Lynch, M. and Conery, J.S. (2003) The origins of genome complexity. *Science* 302, 1401–1404
28. Gray, M.W. et al. (2010) Cell biology. Irremediable complexity? *Science* 330, 920–921
29. Stoltzfus, A. (1999) On the possibility of constructive neutral evolution. *J. Mol. Evol.* 49, 169–181
30. Haig, D. (1997) *The Social Gene* (Blackwell Scientific)
31. Smith, J.M. (1998) *Evolutionary Genetics*, 2nd edn (Oxford University Press)
32. Slattery, M. et al. (2011) Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell* 147, 1270–1282
33. Dowell, R.D. (2010) Transcription factor binding variation in the evolution of gene regulation. *Trends Genet* 26, 468–475
34. Brookfield, J.F. (2005) The ecology of the genome - mobile DNA elements and their hosts. *Nat. Rev. Genet.* 6, 128–136
35. Wolf, Y.I. et al. (2018) Physical foundations of biological complexity. *Proc. Natl. Acad. Sci. U. S. A.* 115, E8678–E8687
36. Doolittle, W.F. (2017) Darwinizing Gaia. *J. Theor. Biol.* 434, 11–19