

Genetic Architecture Promotes the Evolution and Maintenance of Cooperation

Antoine Frénoy*, François Taddei, Dusan Misevic

INSERM U1001, Université Paris Descartes, Sorbonne Paris Cité, Faculté de Médecine Paris Descartes, Paris, France

Abstract

When cooperation has a direct cost and an indirect benefit, a selfish behavior is more likely to be selected for than an altruistic one. Kin and group selection do provide evolutionary explanations for the stability of cooperation in nature, but we still lack the full understanding of the genomic mechanisms that can prevent cheater invasion. In our study we used Aevol, an agent-based, *in silico* genomic platform to evolve populations of digital organisms that compete, reproduce, and cooperate by secreting a public good for tens of thousands of generations. We found that cooperating individuals may share a phenotype, defined as the amount of public good produced, but have very different abilities to resist cheater invasion. To understand the underlying genetic differences between cooperator types, we performed bio-inspired genomics analyses of our digital organisms by recording and comparing the locations of metabolic and secretion genes, as well as the relevant promoters and terminators. Association between metabolic and secretion genes (promoter sharing, overlap via frame shift or sense-antisense encoding) was characteristic for populations with robust cooperation and was more likely to evolve when secretion was costly. In mutational analysis experiments, we demonstrated the potential evolutionary consequences of the genetic association by performing a large number of mutations and measuring their phenotypic and fitness effects. The non-cooperating mutants arising from the individuals with genetic association were more likely to have metabolic deleterious mutations that eventually lead to selection eliminating such mutants from the population due to the accompanying fitness decrease. Effectively, cooperation evolved to be protected and robust to mutations through entangled genetic architecture. Our results confirm the importance of second-order selection on evolutionary outcomes, uncover an important genetic mechanism for the evolution and maintenance of cooperation, and suggest promising methods for preventing gene loss in synthetically engineered organisms.

Citation: Frénoy A, Taddei F, Misevic D (2013) Genetic Architecture Promotes the Evolution and Maintenance of Cooperation. PLoS Comput Biol 9(11): e1003339. doi:10.1371/journal.pcbi.1003339

Editor: Claus O. Wilke, University of Texas at Austin, United States of America

Received: May 16, 2013; **Accepted:** September 30, 2013; **Published:** November 21, 2013

Copyright: © 2013 Frénoy et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was funded by the French "Agence Nationale de la Recherche" (<http://www.agence-nationale-recherche.fr/>) via the project grant "COOPINFO" (ANR-10-BLAN-1724). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: frenoy@gmx.com

Introduction

The evolution of cooperation in microbial populations is a fascinating, rich and controversial evolutionary problem [1–6]. The theoretical understanding of cooperation has been gradually advancing for decades, and recently those insights have also been applied to practical, medical problems, such as the treatment of infections triggered by cooperating, pathogenic bacteria [7,8]. Most evolutionary explanations of cooperation rely on kin selection and group selection theories and are constantly being improved and refined by a host of mathematical tools [9,10]. Among them, the game theory and meta-population models have proved to be especially useful in the analysis of long term versus short term, as well as the individual versus population benefit of cooperation [11–14]. However, those methods tell us practically nothing about the evolutionary pressure on the structure of genomes that encode the cooperative traits. They typically do not distinguish between genotypes and phenotypes and consider only a finite set of possible behaviors (often only two: cooperate or not) with a constant extrinsic probability of switching between them. Although some recent papers do go further than evolving classical binary behavior by considering more complex stochastic strategies that take into account past interactions [15], they also remain "one

locus = one parameter" models, unable to consider genetic architecture of cooperation genes. Several experimental studies have shown the need to go beyond these limitations to understand cooperation in microbial systems. Specifically, Foster *et al.* demonstrated that the pleiotropic effect of a *Dictyostelium discoideum* gene involved in a cooperative behavior (differentiation into prestalk cells) causes the mutations inducing cheating behavior to be associated with a direct fitness cost to the individual [16]. Similarly, cheating mutations induce a cost in *Pseudomonas aeruginosa* because of co-regulation of public and "private" goods via the same quorum-sensing mechanism [17].

We postulate that genomic architecture of metabolic and secretion genes – achieved by sense-antisense coding or frameshifts – can provide a mechanism for the evolution and maintenance of cooperation that is similar but more basic than ones relying on genetic pleiotropy or co-regulation. Here we investigate how two specific types of genomic architecture of cooperation genes may affect the evolutionary fate of cooperation itself. The first type relies on the concept of operons, already well described and investigated in the context of co-regulation or co-transfer of genes in the same operon [18,19]. We specifically consider metabolic and secretion genes that have the same promoter and terminator sequence, thus sharing an operon. The second architecture type is

Author Summary

Cooperation is a much studied and debated phenomena in the microbial world marked by a key question: Given the survival of the fittest evolutionary paradigm, why do individuals act in seemingly altruistic ways, paying a cost to help others? Kin selection and group selection, together with mathematical tools from areas such as economics and game theory, have provided some answers. However, they largely ignored the underlying genetic and genomic mechanisms that drive the evolution of cooperation. In this study, we show that the architecture of the genomes has a major role in shaping the fate of cooperating populations. Specifically, we use an *in silico* evolution platform and discover that genes for cooperative traits are “hiding” behind metabolic ones by overlapping their sequences or sharing operons. In conditions where cheaters may outcompete the cooperators, this entangled architecture evolves spontaneously and effectively protects cooperation from invasion by cheater mutants. We describe a novel genetic mechanism for the evolution and maintenance of cooperation and, by taking into account the second order selection pressures on the genomes, highlight the need for going beyond simple game theory models in its study.

the overlap, base-pair sharing between metabolic and secretion genes due to being in different reading frames or on different DNA strands. Although more rare in bacterial context, gene overlaps may be caused by the strong constraints on maximum genome size and have similar evolutionary explanations and properties as operons [20]. We describe and quantify the role of both these genetic architecture types and show that physical association of cooperation and metabolic genes, via operon and overlap, introduces an evolutionary constraint, pleiotropy in the broad sense, which prevents non-cooperating, cheater individuals from prospering and protects cooperation. Even though the same DNA coding for multiple proteins can, in a broad sense, be viewed as pleiotropy at the sequence level, as far as we know, its importance has never been described in the context of cooperation.

In all our experiments we use the Aevol platform [21,22], an *in silico* experimental evolution system. While similar to existing individual-based, genetic-algorithm simulations, Aevol embodies a number of features inspired by microbial genetics that make it especially well suited for our study. For example, the phenotype of an Aevol digital organism is a continuous function comprised of a potentially unlimited number of biological processes and their performance level, which in turn allows for a continuous cooperating phenotype instead of the classical binary one. When it evolves, the cooperation among individuals is based on a public good molecule that diffuses and degrades in the environment. Individuals live in a spatially structured world, suitable for the evolution of cooperation [23,24] and more similar to natural microbial populations than classical meta-population models. The public good is costly to secrete but may benefit any neighboring organisms. Both indirectly selected secretion genes and metabolic genes contributing to fitness directly are encoded in the double-stranded genomes strings of zeros and ones. A set of rules for transcription, translation and protein synthesis governs the complex genotype to phenotype to fitness mapping. Phenotypically similar or even identical individuals can have different genotypes, thus also having different evolvability, robustness, and evolutionary fate [25]. All these properties of Aevol set the stage for evolutionary experiments where genetic architecture con-

straints of cooperation can be both observed and described. We first demonstrate the existence of differences in the resistance to cheater invasion among several phenotypically equivalent populations. We then correlate the maintenance of cooperation genes with the abundance of promoter sharing or overlapping between metabolic and secretion genes. We hypothesize that such non-random encoding of the secretion is indirectly selected for in situations when cooperation is favored. Indeed, when evolving populations start from a naive, non-secreting ancestor, the cooperators employed this protective encoding, and more so when the cooperation cost was high. Mutational analysis confirmed that the constrained genetic architecture resulted in cooperation-destroying mutations also having a direct negative fitness effect. Overall, our results highlight the need for considering appropriately detailed and realistic computational systems and generally show the importance of second-order selection pressures and genetic architecture in the study and understanding of the evolution and maintenance of cooperation.

Results

Creating a bank of cooperators: General properties

In our modified version of Aevol dedicated to the study of cooperation, the phenotype is divided into two groups of traits: metabolism (biological processes allowing the individual to live and reproduce) and secretion (processes relating to the costly secretion of a diffusible public good molecule). Starting from an ancestor with a single, metabolic gene, we independently evolved 50 populations for 20,000 generations. We effectively put cooperation under direct selection by using a particular fitness calculation in which secretion genes were treated the same as the metabolic ones during evolution. At the end of this phase, we chose the fittest individual from each replicate and, by simply reassigning half of the phenotype from metabolism to secretion, obtained 50 cooperators with high secretion levels. Specifically, their average secretion was 96.2% of the maximal secretion in Aevol, and the standard deviation in secretion was 3.89% of the mean. These individuals had generally comparable metabolic and secretion part of their phenotype with on average 19.2 genes in each.

Cheater invasion dynamics differs between populations

Using the cooperators from previous experiments, we started with 50 clonal populations that we then let evolve for an additional 5,000 generations with a possibility of secreting at a moderately high cost ($c=0.4$, see Materials and Methods for the effect of public good cost and fitness calculation details). Each of these populations was replicated 50 times, for a total of 2500 experiments. In all cases the amount of secretion greatly decreased, but not by the same amount or at the same rate (Fig. 1). To quantify these differences, we performed a one-way ANOVA on the average secretion between generation 200 and generation 1,000, the visually chosen time interval during which cooperation is stabilizing to a new level after a quick and strong decay. We found a highly significant between groups effect ($F = 138.2$, $p < 10^{-37}$), each group consisting of the 50 populations that share a common ancestor, confirming that some cooperators are intrinsically more resistant to cheater invasion than others, even though they initially had very similar phenotypes. Moreover, there was no significant correlation between the ancestral cooperation level and the final one ($r = 0.11$, $p = 0.44$), eliminating the possibility of our results being driven by an initial difference in the population cooperation level.

When visually inspecting the phenotype of a randomly chosen cooperator and its descendants from the previous experiment, we

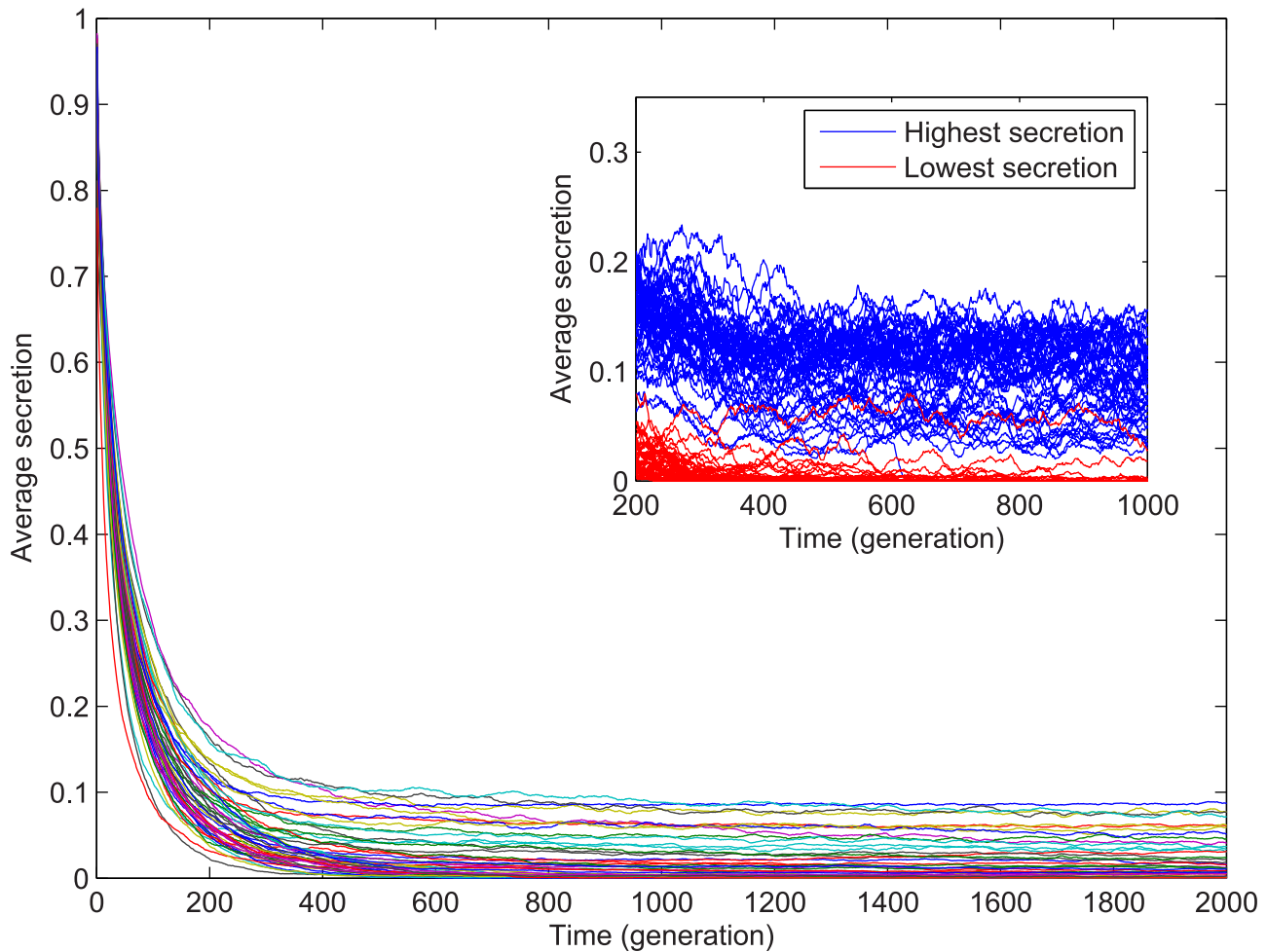


Figure 1. Cheater invasion dynamics in phenotypically similar starting populations. In the main figure each line represents the average secretion, over time, of 50 replicate populations started from the same ancestor. For clarity, we did not include the standard error of each group of replicate populations. The insert figure shows all 50 replicate populations for the two most extreme population groups, zoomed in on the time period of interest for our analyses. Specifically, of all the population groups sharing a common ancestor, the blue populations had the highest and red the lowest average secretion between generation 200 and generation 1,000. Note that average secretion in the inset represents the average secretion within each population, whilst in the main figure it is the average of the secretion of all the individuals from the 50 replicate populations that share the same ancestor.

doi:10.1371/journal.pcbi.1003339.g001

also noticed that it was the same secretion genes that survived cheater invasion between several independent replicates of evolution. One such example, where the phenomenon was especially striking, is presented in Fig. 2. While we did not perform any statistical analysis because of the computational difficulty of tracking every protein for several thousands of generations, this observation motivated further experiments: it supports the idea that our 50 populations are different (in their resistance to cheater invasion) because their secretion genes are somehow different.

Genetic architecture and resistance to cheater invasion

To quantify the genetic architecture of 50 ancestral organisms we measured the percentage of secretion genes that (1) share an operon with at least one metabolic gene, (2) overlap with at least one metabolic gene, (3) satisfy at least one of (1) and (2), or (4) satisfy both (1) and (2) (see Material and Methods for more details). We then compared the genetic architecture measures with the resistance to cheater invasion, expressed as the average remaining

secretion between generation 200 and generation 1,000, as before (Fig. 3). We found that all four genetic architecture properties strongly correlate with the remaining secretion amount ($r=0.60$ and $p<10^{-5}$ for operon sharing, $r=0.57$ and $p<10^{-4}$ for overlapping, $r=0.63$ and $p<10^{-5}$ for at least one of them, $r=0.56$ and $p<10^{-4}$ for both of them), supporting our hypothesis that physical linkage between secretion and metabolic genes confers resistance to cheater invasions.

Mutation effects and resistance to cheater invasion

In order to confirm the effect of genetic architecture on cheater resistance, rather than examining the exact locations and interactions between genes and using them to infer population's evolutionary fate, we directly quantified the effect of mutations on secretion and fitness. We constructed 10,000,000 mutants of each of the 50 ancestors, and calculated the mutational effect as the percentage of mutations that decrease the amount secreted without decreasing metabolic fitness, weighted by their negative effect on secretion. We found significant correlations between the

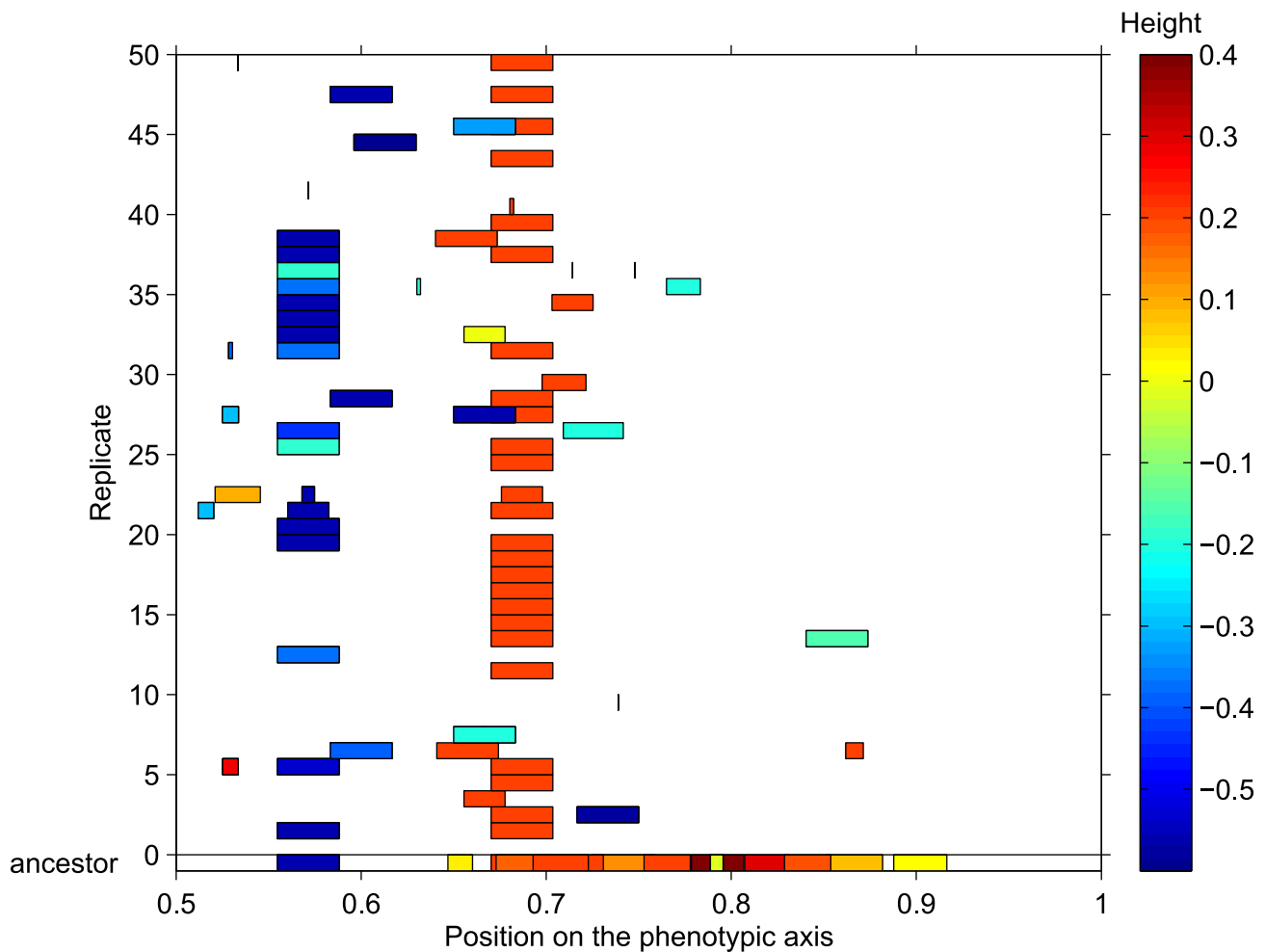


Figure 2. Example of the preferential maintenance of certain secretion genes of a single cooperator from our bank. The bottom row of the graph (ancestor) shows the location on the phenotypic axis of the genes coding for secretion proteins in one cooperator organism from our bank. Above, the secretion genes from the best individual after 1,000 generations of evolution in each of the 50 replicate populations descending from this ancestor are shown. Colors represent the height of the proteins encoded by the genes (see Materials and Methods for detailed explanation of protein properties in Aevol). doi:10.1371/journal.pcbi.1003339.g002

mutation effect and both the robustness to cheater invasion (calculated as before, $r = -0.367$, $p = 8.8 \times 10^{-3}$) and the genetic architecture (here defined as the percentage of secretion genes sharing an operon or overlapping with at least one metabolic gene, $r = -0.430$, $p = 1.8 \times 10^{-3}$). Simply put, the individuals with genetic architecture that groups together metabolism and secretion genes exhibit higher resistance to cheaters, because they are subject to fewer mutations that would convert cooperators into cheaters without any direct fitness loss.

We thus have two measures that predict well the resistance to cheater invasion: genetic architecture and accessibility of mutations. Since the generation range used to quantify cheater resistance was chosen *ad hoc*, we also examined the effect of different ranges on the correlations. Interestingly, we found that genetic architecture is better correlated with cheater resistance when it is measured between generations 1,000 and 2,000 ($r = 0.62$ and $p = 1.4 \times 10^{-6}$) than between generations 1 and 500 ($r = 0.50$ and $p = 2.1 \times 10^{-4}$). Conversely, mutational effects are better correlated with cheater resistance measured in the early (generations 1 to 500, $r = -0.56$ and $p = 2.1 \times 10^{-5}$) than late interval (generation 1,000 to 2,000, $r = -0.26$ and

$p = 6.6 \times 10^{-2}$). Overall, both genetic architecture and mutation effects are good predictors of how easily cooperators may be invaded by cheaters, but genetic architecture is better at predicting long-term effects, while mutational effects are more strongly correlated with short-term ones. Mutations may affect long-term maintenance of cooperation in many ways and genetic architecture captures but one of them. As we elaborate in the Discussion section below, these results indicate that while all mutational constraints play a role, it is the overlap and operon ones that have the strongest long-term evolutionary consequences.

Preferential maintenance of secretion genes based on genetic architecture

We tested the importance of gene overlap and operon sharing in maintenance of cooperation by examining the extent of genomic connections between secretion and metabolism before and after the increase in secretion cost and the accompanying decrease in cooperation. After 2,000 generations of evolution at a higher cost, the secretion genes still present are over 7 times more likely to overlap or share an operon with metabolic genes than the secretion genes at the start of the experiment (Fig. 4), with the

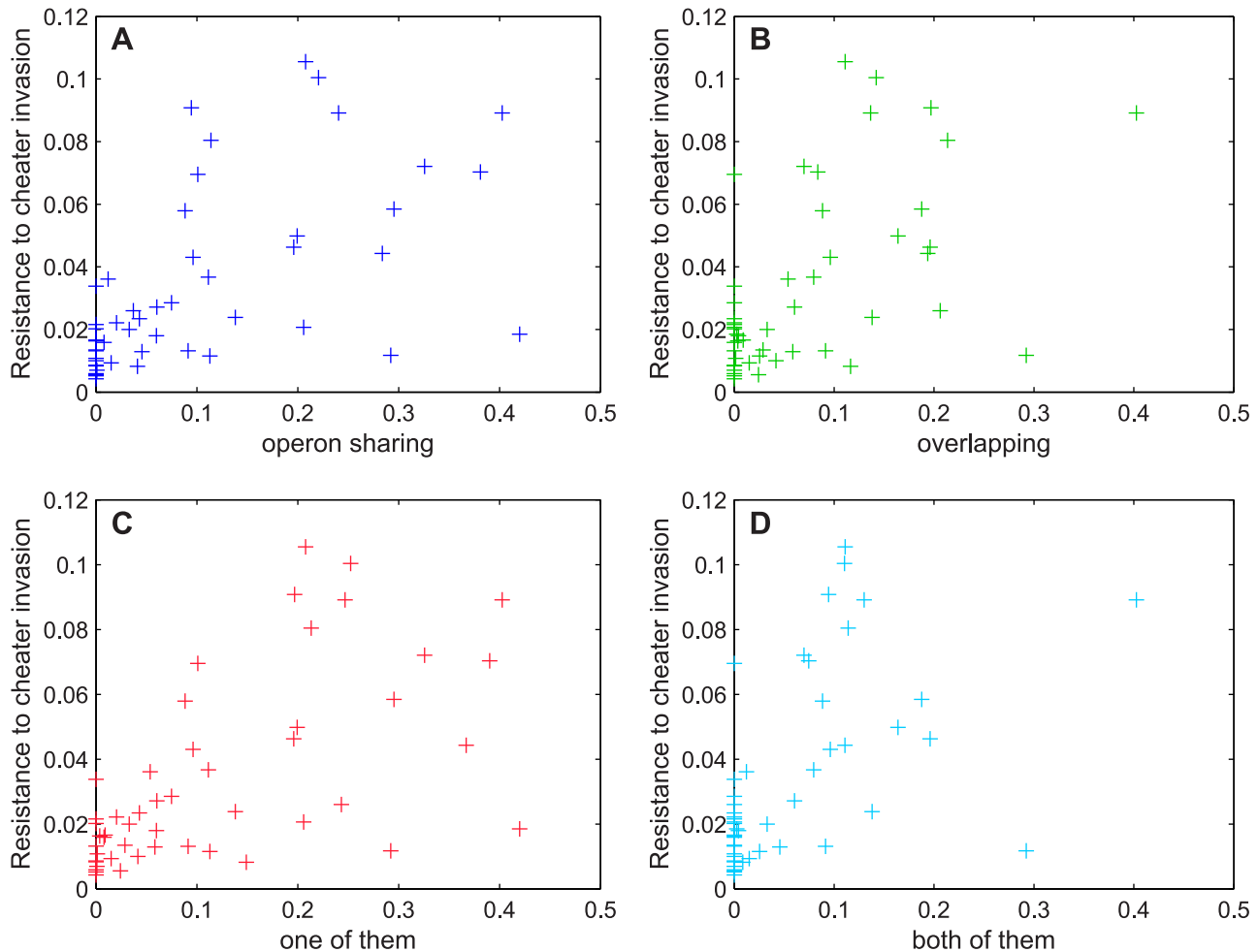


Figure 3. Correlation between the genetic architecture of cooperation genes and the resistance to cheater invasion. Resistance is measured by the amount of secretion surviving cheater invasion. The four types of associations between secretion and metabolic genes shown here are: sharing an operon (A, dark blue), overlapping (B, green), having at least one of the previous two properties (C, red), and having both of them (D, light blue).

doi:10.1371/journal.pcbi.1003339.g003

difference being highly significant (Welch's t test, $p < 10^{-4}$). The proportion of all four categories of association between metabolic and secretion genes (share an operon, overlap, do at least one of them, do both) has increased, and all increases were significant (Welch's t test, $p < 10^{-5}$ for operon sharing, $p = 1.9 \times 10^{-3}$ for overlapping, $p = 1.1 \times 10^{-5}$ for doing at least one of them, $p = 6.8 \times 10^{-4}$ for doing both). Note that these categories do not exactly correspond to the partitioning done on Fig. 4 (see Material and Methods for detailed explanation), but capture the same general properties of genetic architecture.

Evolution of genetic links between metabolic and cooperation genes

In the previous experiments we worked with already evolved cooperators, measured their resistance to cheater invasion and genetic architecture. We now turn to *de novo* evolution of cooperation, in order to show that gene overlap and operon sharing will evolve, via indirect selection pressures, in conditions moderately favorable for cooperation. Naive ancestors evolved for 20,000 generations with cooperation cost of 0.3. In the final populations, individuals on average had 19 metabolic genes and

1.8 secretion genes. While the number of secretion genes is low, by pooling data of all 1024 individuals from each population we obtained a large number of genes that we could analyze. We compare the shared operons and gene overlap for metabolic and secretion genes with the same measures applied only within metabolic genes, as a control. The genetic architecture links between metabolic and secretion genes are on average 3 times stronger than within metabolic genes alone, the difference being highly significant (Fig. 5a, comparing the sum of the three bottom categories – dark blue, light blue and green –, Welch's t test, $p = 1.0 \times 10^{-3}$). The proportion of all four genetic architecture categories differed between the two gene groupings, and all differences were significant (Welch's t test, $p = 2.6 \times 10^{-3}$ for operon sharing, $p = 3.0 \times 10^{-3}$ for overlapping, $p = 1.0 \times 10^{-3}$ for at least one of them, $p = 6.4 \times 10^{-3}$ for both). We repeated the analysis for 50 more populations that evolved under a lower secretion cost ($c = 0.01$) and we observed no difference in genetic association between metabolic and secretion genes compared to associations with metabolic genes alone (Fig. 5b, Welch's t test, $p = 0.50$). Comparison of the two sets of experiments performed at different secretion costs shows that the preferential association

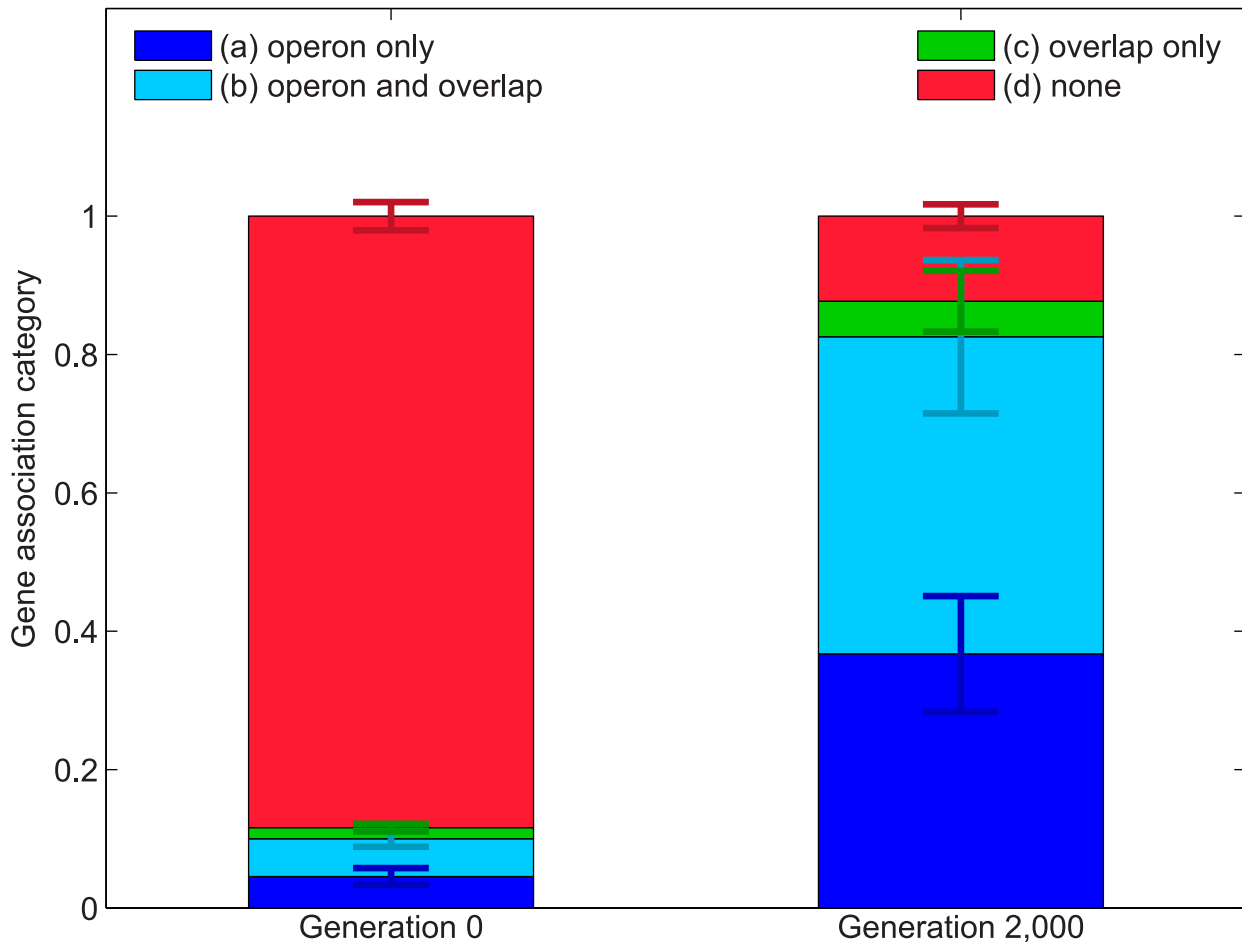


Figure 4. Secretion genes at generation 0 and at generation 2,000 partitioned between four categories: (a) sharing a promoter (*i.e.* being on the same operon) without overlapping with a metabolic gene (dark blue), (b) overlapping and sharing an operon with a metabolic gene (light blue), (c) overlapping without sharing an operon with a metabolic gene (green), (d) neither sharing an operon nor overlapping with a metabolic gene (red). Error bars represent one standard error of the mean (fifty original cooperators). The color of the error bars corresponds to the genetic architecture category which they relate to. doi:10.1371/journal.pcbi.1003339.g004

between secretion and metabolic genes evolves only when the cost of cooperation is relatively high.

Discussion

Our research was motivated by two simple observations: (1) phenotypically near-identical populations have different evolutionary fates (Fig. 1), and (2) for a given cooperator, the secretion genes that survive cheater invasion seem to be always the same between several replicates (Fig. 2); and by two straightforward questions: what were the differences between these populations, and what were the differences between these genes? In our initial experiments we found that when populations cooperating at near-maximal level suddenly faced a direct, high cost of public good secretion, the cheaters were quick to invade. However, the invasion dynamic was qualitatively different across populations, with some ancestors predictably evolving into populations without any cooperation, while others kept low but non-zero levels of secretion (Fig. 1).

We propose that these diverging evolutionary fates for otherwise phenotypically similar populations are due to differences in the ancestral genetic architecture of cooperative traits. In a previous

paper we showed that the accessibility of mutations impacting secretion may lead to different future secretion dynamics for phenotypically similar individuals [25]. Here we show that in the case of cooperation decay, beyond the simple effect triggered by the accessibility of mutations impacting secretion, the way secretion genes “share” the genome with metabolic genes also has an effect on the selection of these mutations. More precisely, we suggest that secretion genes that are physically connected with a metabolic gene, for example belonging to the same operon, or physically overlapping, are more robust to cheater invasion: a deleterious mutation in one of these genes is more likely to also deleteriously affect a metabolic gene, and thus is less likely to be selected for. Similar mechanisms, coupling cooperative traits with metabolic, individualistic traits have been described before [16,17], but instead of relying on gene coregulation or genes with multiple effect, we report a more basic genetic mechanism of entanglement for genes with singular effects. Additionally, these studies remain two isolated data-points and are thus not enough to show the existence of second order selection pressures leading to such architecture. Bio-informatics methods could provide much more information to support or deny this hypothesis, however there have been very few relevant studies of the genes involved in

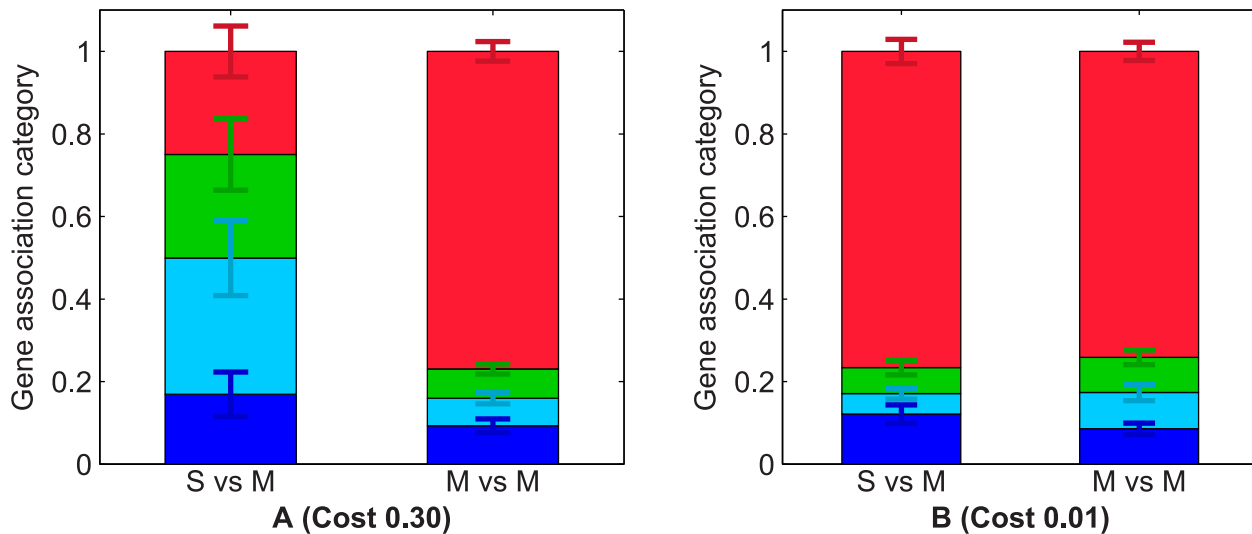


Figure 5. Genetic architecture associations after 20,000 generation of *de novo* evolution of cooperation, at (A) high, $c=0.3$ and (B) low, $c=0.01$ secretion cost. In both cases, we quantified the presence of operons or overlaps between secretion genes and metabolic genes (S vs M) as well as between metabolic genes and other metabolic genes (M vs M). Genes were partitioned in four categories, labeled as in the Fig. 4: sharing a promoter (an operon) without overlapping with a metabolic gene (dark blue), overlapping and sharing an operon with a metabolic gene (light blue), overlapping without sharing an operon with a metabolic gene (green), not sharing an operon nor overlapping with any metabolic gene (red). Error bars represent plus and minus one standard error of the mean for the fifty replicate populations and their color corresponds to the genetic architecture category they relate to.
doi:10.1371/journal.pcbi.1003339.g005

cooperative behavior, the primary reason being the difficulty of identifying such genes. There is one notable exception [26], where the authors use the prediction of cellular localization: outer membrane and excreted proteins are more likely to be related to cooperative traits than cytoplasmic, inner membrane and periplasmic proteins. Their main result is about the role of horizontal transfer in the evolution of cooperation, however they also show that genes coding for outer membrane and excreted proteins are more likely than others to be “genome neighbors” of additive systems (e.g. toxin-antitoxin). As the experimental data is suggestive but overall still insufficient, the use of an appropriate individual based model dedicated to the study of evolutionary processes and selection pressures acting on the genome, such as Aevol, is a way to fill the void. Of course, each model has biases and limitations, however, the strong point of Aevol is that we implement only simple, easy to understand, small-scale rules inspired by bacterial genomics, and all other properties and processes are emergent. For example, in Aevol there is no parameter like “probability that two neighboring genes overlap”. Thus, the outcomes we describe here are not directly driven by the model and are not something we necessarily expected to evolve.

Using Aevol we were able to directly test our hypothesis about the effects of genetic architecture on the evolution and maintenance of cooperation we generated and analyzed a total of 500,000,000 mutant organisms. About half of the mutants are phenotypically perfectly similar to the original individual (no mutations or only neutral mutations happened). The calculated probability of having no mutations for typical organism with genome length of 10,000 is 0.18. The calculated probability of having exactly one mutation is 0.31, and the one of having strictly more than one mutation is 0.51. We emphasize that a large part of the analyzed mutations are neutral but of course focus on the ones that change organisms fitness and secretion. We recorded the effect of all mutations, specifically searching among “cheating” mutations, the ones that would decrease the amount of secretion, for mutations that do not simultaneously decrease metabolic

fitness, and may thus be selected for, or at least not immediately purged by selection. We find that the proportion of these mutations, weighted by their negative effect on secretion, directly and significantly correlates with the population’s vulnerability to cheater invasion, which supports our hypothesis. Interestingly, when we measured the remaining cooperation later in time it correlated more strongly with amount of genetic architecture linkage between secretion and metabolism than the measured effect of the introduced mutations. The higher durability of genetic architecture constraints, compared to immediate mutation effects, may be due to the higher durability of the genetic architecture itself. As we saw from the comparison between different gene association categories at generation 0 and generation 2,000, it is exactly the secretion genes that do overlap or share an operon with metabolic genes that may be preserved (Fig. 4). On the other hand, our mutational analysis explored only a small, nearby portion of the immense fitness landscape. As organisms evolve, and move around that landscape, the particular mutants we constructed and analyzed may become less accessible via mutations and thus also less relevant for the evolutionary dynamics. Finally, the enrichment of genotypically connected metabolic and secretion genes among all secretion genes present in the individuals strongly suggests that this type or architecture may generally be created and maintained via indirect selection in cooperative systems.

We should note that in our first set of experiments we used individuals that evolved under an altered fitness calculation regime that enabled us to directly select for future cooperators of similar phenotype but different genotype. We could have as well designed these individuals by hand, directly writing the zeros and ones in their genomes. However, this would have likely generated fragile and generally poorly adapted individuals, as evolution is typically better in organism design than us humans. The change between the alternative fitness calculation and the regular one may appear somewhat artificial or arbitrary, but it can also be seen as a transition from producing a private, non-secreted good, to a public, secreted one. Still, while these results show the effect that

architecture may have on cooperation, they do not by themselves prove the existence of second order selection pressure that would be sufficient to create operon and overlap type constraints during the evolution and maintenance of cooperation. We thus turn to our second set of experiments, in which secretion evolves *de novo*, without ever being directly selected for.

The results of this second set of evolution experiments strongly support the hypothesis that when costly cooperation does evolve and persist, there is a selection pressure grouping secretion genes with metabolic genes to protect them from removal. Such selection pressure is necessarily indirect, since cooperation via public good secretion does not directly increase the fitness of the cooperating organisms. Additionally, it would not prevent any “cheaters” from appearing but would reduce their likelihood of having a greater fitness than cooperators and spreading, making overall conditions more favorable for cooperation by reducing the effective mutation rate for switching from being a cooperator to being a potentially invading cheater. Previous work has already established this “mutation rate” (in a binary, game theory vocabulary) is one of the very important parameters controlling the dynamics of cooperation in spatially structured populations [27,28]. Here we extend these results by showing a specific genetic mechanisms that would allow evolution to modulate the rate of switching between potentially invading cheaters and cooperators. The total rate of production of cheaters may not be different between populations, but because of the genetic entanglement of cooperation and metabolism, a large proportion of cheaters is unable to invade and thus such cheater mutants are evolutionary dead-ends.

The role of constraints introduced by second-order selection, such as the one we exemplified here, in assuring the best long-term outcomes has been proposed before in a more general and abstract context [29]. Specifically, our results may provide a set of new potential explanations for the evolution of operons and overlaps, important building blocks of life. While in the past operon and overlap existence has been linked to co-regulation and co-transfer of genes working together and belonging to the same sets of biological processes [19], here we highlight their role as an evolutionary constraint. Specifically, operons and overlaps may protect genes that are at risk of removal because of a short-term cost and in spite of the long-term benefit they may provide. Of course, the particular combination of short-term cost and long-term benefit is not unique to cooperation and it underlies other biological processes, most notably sex and recombination, which also continue to be intensely studied [30]. In terms of cooperation itself, genetic architecture constraints may be highly relevant in understanding the much studied siderophore-mediated cooperation in *P. aeruginosa*, where cooperative as well as essential metabolic traits are under the control of a quorum sensing mechanism [17]. However, our idea also has large implications outside of the field of cooperation: going beyond explaining evolutionary outcomes, the genetic architecture coupling mechanism we describe here could be actively used to prevent mutations from removing of genes introduced into bio-engineering organisms, one of the major problems in the field of synthetic biology [31].

Conclusion

The study of the evolution and maintenance of cooperation is rich in theories, majority of which rely on higher level properties of individuals, such as relatedness, fitness, or group structure. Our experiments investigate basic, genome-level properties and show that the presence of genetic associations between metabolism and secretion genes aids the maintenance of cooperation across thousands of generations. Operon sharing and gene overlap are

selected for when cooperation is costly and directly change populations’ evolutionary fate. Second order selection is known to play a major role in the rapid evolution of microbial populations [32] and here we contribute to understanding the specific and much studied case of cooperation via public good secretion. We used an *in silico* experimental platform, Aevol, which has enabled us to collect and analyze genetic architecture and evolutionary dynamics data in detail previously unattainable with either mathematical or experimental systems. The role of second-order selection and genetic constraints in evolution will undoubtedly continue to motivate experimental and theoretical research but in our case it also has the potential to inform bio-engineering and synthetic biology applications.

Materials and Methods

Aevol digital evolution system

In this study we use the Aevol platform, an individual-based model of evolution, especially well suited for the study of selection pressures on genomic architecture [21,22,25,33]. It is free and open-source software and is downloadable from <http://www.aevol.fr/download>. The specific version of the platform we used in this study, including analysis routines, parameter files, other minor modifications, is available on request. Aevol has already been used in several peer-reviewed publications including some that studied cooperation, so we invite the reader to refer to [34] for more information on how cooperation has been implemented and for characterization of the related parameters, and to [22,35] for more general details about the original version of Aevol that did not incorporate cooperation.

In Aevol, the individuals are living on a toroidal, two-dimensional square grid, with each location being occupied by exactly one individual. In our experiments the grid contains 32×32 positions, for a total of 1024 individuals. Selection and reproduction are performed locally in a synchronous way: at each generation, for each position in the grid, we compete the nine individuals in the neighborhood to determine which one’s descendant in going to occupy this position in the next generation.

The phenotype of an individual is represented by a two-dimensional curve describing the level of performance for each point of a continuous set of abstract biological processes. This is a very general way of encoding a phenotype without any restriction on the type of biological processes that can be represented. The genotype is a string of zeros and ones, which is transcribed and translated according to a bacterial genomics-inspired process: promoters and terminators are identified to allow transcription, then the transcribed sequences are searched for ribosomal binding site and start codon, followed by what will be a gene and then by an in-frame stop codon, to allow translation.

Our genetic code is an abstract mathematical function transforming the gene, i.e. the binary sequence between the start and stop codons, into three numbers, interpreted as a triangle on the axis of biological processes. These three numbers are M (mean position of the triangle on the phenotypic axis), W (half-width of the triangle), and H (height of the triangle). Base-pairs are read three by three, and our amino-acid space has eight symbols: START and STOP, M0 and M1 which are used to specify M , W0 and W1 which are used to specify W , H0 and H1 which are used to specify H (Fig. 6). Each of these eight amino-acids is assigned to exactly one of the eight (2^3) possible triplets. There is no redundancy in the codon–amino-acid mapping, however there is still a large redundancy in the gene–protein mapping because codons inside genes can be reordered without impacting phenotype. Specifically, what matters is the order of codons

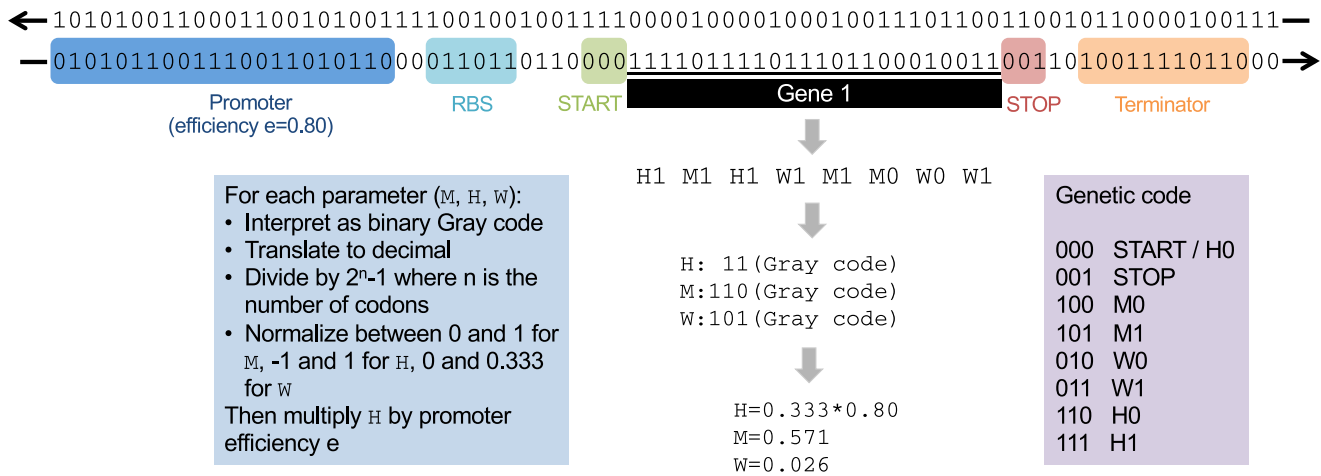


Figure 6. Aevol genetic code. Here we use an example of a functioning gene from an Aevol individual to explain the transcription and the translation processes. The gene is flanked by a promoter and terminator regions and preceded by a ribosome binding site (RBS). The codons for mean position, the width, and the height of the protein are identified, transformed into Gray code using the Genetic code table (box on the right), and finally scaled and normalized, as we summarize in the box on the left and describe in more detail in the Methods. Note that a gene with re-shuffled codons, for example H1 H1 M1 M1 M0 W1 W0 W1, would encode exactly the same protein. START codon may occasionally be found inside a gene, in which case it is interpreted as H0. The promoter differs from the consensus sequence by 1 base out of the maximal 4 differences allowed, giving it a 0.80 efficiency.

doi:10.1371/journal.pcbi.1003339.g006

specifying the same triangle property (M , W , or H), while the relative order of codons for different properties can be altered freely. Once a coding sequence has been detected using the rules explained above and transformed into an amino-acid sequence, we extract from there three binary words (for M , W and H) according to the following process: amino acid $X0$ adds a 0 to the binary word of X and $X1$ adds a 1 to the binary word of X , where X is any of M , W , or H . We obtain an integer value for each of the three binary words by interpreting them using Gray code. Gray code is an alternative binary encoding in which two successive integers are encoded by binary numbers differing in only one digit. The integer values are then normalized by $2^n - 1$ where n is the number of codons used, and scaled to a $[0, 1]$ interval for M , $[0, 0.033]$ for W , and $[-1, 1]$ for H . Finally the H value is multiplied by the transcription efficiency – a property of the promoter explained below. The mean position specifies the primary trait the protein affects, and as it is a real number, it allows for an infinite number of different traits. The height specifies protein's performance level for the primary trait, while the width determines all the traits a protein affects. Individual's phenotype is computed by summing up all the triangles encoded in its genome. There is no genetic regulation via transcription factors in this version of Aevol, however there are protein-protein interactions (two proteins contributing to the same biological processes) and transcription efficiency is regulated by the strength of the promoter (defined as the distance to a consensus sequence). As in natural systems such as bacteria or phages, this genomics allows two genes to cluster on the same mRNA (operon) or to physically use the same DNA basis in different reading frames or different senses (overlap). Examples of these different configurations are represented on Fig. 7.

The environment is represented by a two-dimensional curve indicating what is, for every possible biological process, the optimal level of performance in the given environment. The fitness of an individual is a decreasing function of the distance between the individual's phenotype and the optimal phenotype. Individuals are locally selected according to their rank in the neighborhood, with a probability of reproduction exponentially decreasing with the rank. The chosen individual will undergo reproduction with

mutations (insertion, deletion or substitution of a small number of basis and duplication, inversion, translocation or deletion of a larger portion of the genome). The rates of different mutation types are parameters of the model and have been set to 5×10^{-5} per basis for small mutations, and 5×10^{-6} for large mutations. Ancestral genome is 5,000 bases long and contains a single gene, while the typical genome length after several thousands of generations of evolution is around 10^4 basis. Aevol is a stochastic simulation, the variability coming from the randomness of mutations and the probabilistic selection. One of the parameter is the random seed used to initialize the random number generator. We can replicate an experiment by running it several times with the same exact parameters, but different random seeds.

In our experiments, we distinguish two categories of biological processes: the “metabolic” ones (all traits positioned before $M=0.5$ on the axis of the biological processes), that allow an individual performing them to live and reproduce, and the “secretion” ones (position after $M=0.5$ on the axis), that determine the level of the production of the public good. We note that under our setup, while genes are generally pleiotropic, simultaneously influencing multiple traits, it is not possible for a gene to affect both metabolic and secretion traits. The public good is costly to secrete, but diffuses in the environment and is beneficial to every individual that comes in contact with it. The cost for the production of one unit of public good varies in our experiments, but is always equal to the cost coefficient (parameter we set) multiplied by the amount of the public good produced. The fitness of an individual is given by this equation:

$$F_m \times (1 + 0.2 \times (Get_{PG} - cost \times Produce_{PG}))$$

Where F_m is metabolic fitness (calculated as explained before but only considering the left part of the axis), Get_{PG} is the amount of public good present in the environment at the location the individual inhabits, $cost$ is the per-unit cost of the public good production, and $Produce_{PG}$ is the amount of public good produced by the individual (computed similarly to metabolic fitness but

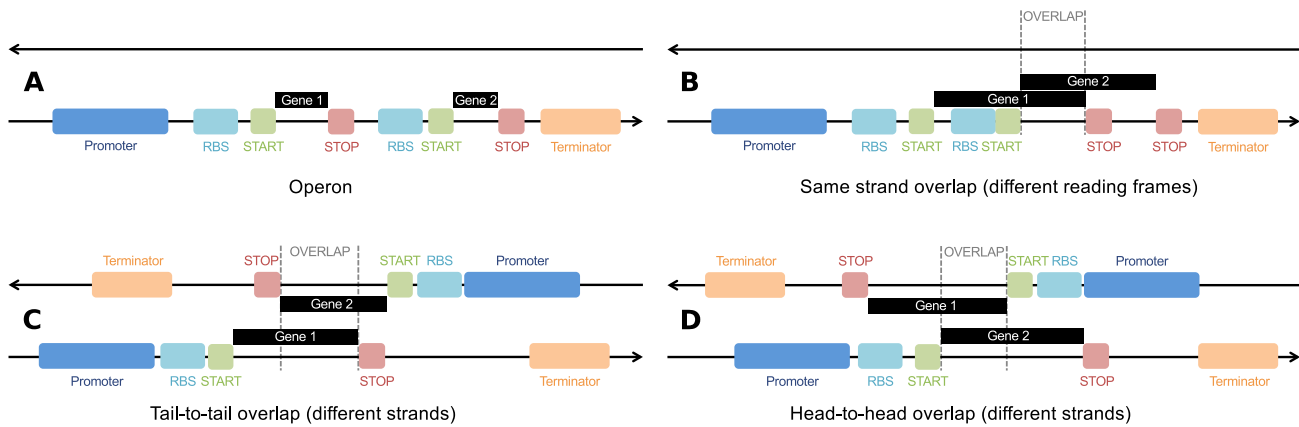


Figure 7. Examples of constrained genetic architecture in Aevol. (A) As is often the case in natural systems, here the two digital genes belong to the same operon. They share a promoter and a terminator sequence and are thus being expressed at the same level. These hypothetical genes would belong to the “operon only” category from Fig. 4 and Fig. 5. (B) These two genes also share the same operon but additionally their sequences overlap, putting them in the “operon and overlap” category. In this case, the genes are in different reading frames and do not share a STOP codon, although such configuration is also possible. As the black gene boxes indicate, the left STOP codon corresponds to the left START codon, while the right STOP codon corresponds to the right START codon. (C, D) Two examples of genes from the “overlap only” category which are encoded on different strands. This is not an exhaustive list of possible genetic constraints, as a gene may, for example, share an operon with a gene on the same strand while simultaneously overlapping with a gene on an opposite strand.
doi:10.1371/journal.pcbi.1003339.g007

considering the right part of the axis). 0.2 is a constant chosen based on previous experiments [25].

The diffusion parameter is 0.05 per generation, meaning that five percent of the public good present at one position will diffuse in each of the eight neighboring positions during one generation. The degradation rate is set to 0.10 per generation, meaning that ten percent of the public good at each location will degrade during one generation. This degradation can be thought of as replacing any explicit consumption of the public good, but also as specifying the public good durability. Overall, in all our experiments, 54% of the public good present at generation n at some position will remain at this position at generation $n + 1$. In the Supporting Text S1, we experimentally show that the secretion mechanism implemented in Aevol, as described above, leads to the usual cooperation dilemma.

Evolving a bank of cooperators

To evolve a large number of strong cooperators, we assigned biological processes that were usually in the secretion part of the phenotype to the metabolic part of the phenotype, allowing a strong direct selection on them. After 20,000 generations of evolution under these conditions, the whole phenotype of the individuals closely matches the target phenotype. Thus, when picking the best individual and re-assigning half of the trait axis back to secretion, we get a “near-perfect” cooperator, one that secretes close to the maximal possible amount of the public good. Evolving cooperators in this way makes secretion genes evolve in the same way as the metabolic ones, to a high level, increasing the potential signal in further experiments. We repeated this experiment 50 times, extracted the fittest individual from each population, and obtained a bank of 50 independently evolved cooperators.

Analysis of the genomic architecture

For each of the 50 cooperators we evolved in the first set of experiments, we analyzed the architecture of all its secretion genes and classified them in four different categories: (1) genes that share an operon with at least one metabolic gene without overlapping

with a metabolic gene, (2) genes that overlap with at least one metabolic gene without sharing an operon with a metabolic gene (this is possible because our digital DNA is double stranded and thus allows for two reading senses, in addition to three reading frames for each sense), (3) genes that overlap with at least one metabolic gene and share an operon with at least one metabolic gene (not necessarily the same one), and (4) genes that share neither operon nor overlap with a metabolic gene. There are multiple ways one could classify the different genes, for example, by distinguishing the number of metabolic genes a secretion gene overlaps or shares an operon with. The four categories we chose have the benefit of intuitive simplicity in addition to including all secretion genes in exactly one category, and we have used them in Fig. 4 and Fig. 5. The number of genes in each category is always shown as a percentage of all the secretion genes and standardized by the genes’ phenotypic area. Here, the phenotypic area refers to the area of the protein (triangle) the gene encodes for, and allows us to give more weight to the genes that have a strong impact on secretion as well as enable comparison between replicate experiments that may have different secretion levels.

However, when performing the statistical analyses to determine the correlation between the presence of overlap and the resistance to cheater invasion, it does not make sense to, for example, exclude the secretion genes that also share an operon (in addition to overlapping) with a metabolic gene. So we use slightly different, larger, categories for secretion genes: share an operon with at least one metabolic gene (which is exactly the addition of categories 1 and 3 of our previously explained partitioning), overlap with at least one metabolic gene (addition of categories 2 and 3), do at least one of them (addition of categories 1, 2 and 3), do both of them (same than category 3), do none of them (same than category 4). The difference is that these categories are no longer exclusive: one gene can be in more than one of the new categories at the same time. The genes are standardized by their phenotypic area, as before. These regrouped categories are used in the statistical analyses throughout the paper, and can easily be visually inferred from the categories of the bar graphs in Fig. 4 and Fig. 5.

De novo evolution of cooperation

In these experiments, each population starts from a randomly constructed organism with a 5,000 base pair genome. As random sequences of 0's and 1's are generated, they are screened for the presence of open reading frames with genes. Thousands of sequences are tested and the first one that has exactly one metabolic gene with a positive effect on fitness is selected. This genome is then cloned to fill the population grid and form the starting population. Reason for starting with a single, valid gene rather than an organism with effectively empty genome is that in both cases all the genes except the first one have a very high probability of evolving from duplication followed by divergence of one already existing gene. Indeed, promoters and ribosome binding sites are hard to evolve from scratch. Starting from purely random sequence would only greatly slow down the evolution process (genomes could evolve for thousands of generation before the first gene appears [22]) without qualitatively changing the understanding of the evolutionary process in our system. After 20,000 generations of *de novo* evolution, we pooled the proteins from all the individuals in each replicate to obtain a measure of average genetic architecture within a population. As before, rather than using just a protein count, we standardized the contribution of each protein by its phenotypic area.

Supporting Information

Figure S1 Individuals are tempted to stop cooperating. For each of the 50 populations, we plot the average fitness increase an individual would experience if it would individually stop cooperating, *i.e.* the temptation to defect, against the average amount secreted by an individual in the population. Except in the populations where no cooperation has evolved (red points), the temptation is always greater than zero. (PDF)

Figure S2 Groups of cooperators do better than groups of defectors. For each of the 50 populations, we plot the benefit of cooperation, *i.e.* the average fitness drop individuals would

experience if cooperation was disabled, against the average amount secreted by an individual in the population. Except in the populations where no cooperation has evolved (red points), the benefit is always greater than zero. (PDF)

Figure S3 Individuals that cooperate more are the ones that benefit more from secretion. For each of the 50 populations, we plot the correlation between how much individuals secrete, and how much they benefit from secretion (*i.e.* the average fitness drop individuals would experience if cooperation was disabled), against the average amount secreted by an individual in the population. Except in the populations where no cooperation has evolved (red points), the correlation is significant and positive for all but 3 populations. (PDF)

Text S1 Secretion of a public good in Aevol satisfies the two usual requirements for a social dilemma. The fitness of an individual would always increase if it would stop cooperating (Supporting Figure S1), and in a group of cooperators the individuals have a higher fitness than in a group of cheater (Supporting Figure S2). (PDF)

Acknowledgments

We thank Tatiana Dimitriu, Julien Benard-Capelle and Ariel B. Lindner for useful discussions and comments on this study and David Parsons, Carole Knibbe and Guillaume Beslon for their help with the Aevol platform. We are grateful to the Center for the Study of Evolution in Action (BEACON, Michigan State University) and the CC-IN2P3 (Computation Center of the French National Institute for Nuclear and Particle Physics) for providing the computational resources needed by this study.

Author Contributions

Conceived and designed the experiments: AF DM FT. Performed the experiments: AF. Analyzed the data: AF DM. Contributed reagents/materials/analysis tools: AF DM. Wrote the paper: AF DM.

References

- Hamilton WD (1964) The genetical evolution of social behavior I+II. *Journal of Theoretical Biology* 7: 1–52.
- Axelrod R (1984) *The Evolution of Cooperation*. Basic Books, New York, NY.
- Buckling A, Harrison F, Vos M, Brockhurst M, Gardner A, et al. (2007) Siderophore-mediated cooperation and virulence in *Pseudomonas aeruginosa*. *FEMS Microbiology Ecology* 62: 135–141.
- Griffin AS, West SA, Buckling A (2004) Cooperation and competition in pathogenic bacteria. *Nature* 430: 1024–1027.
- Damore J, Gore J (2012) Understanding microbial cooperation. *Journal of Theoretical Biology* 299: 31–41.
- West S, Diggle S, Buckling A, Gardner A, Griffin A (2007) The social lives of microbes. *Annual Review of Ecology, Evolution, and Systematics* 38: 53–77.
- Brown S, West S, Diggle S, Griffin A (2009) Social evolution in microorganisms and a trojan horse approach to medical intervention strategies. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364: 3157–3168.
- Inglis R, Brown S, Buckling A (2012) Spite versus cheats: competition among social strategies shapes virulence in *Pseudomonas aeruginosa*. *Evolution* 66: 3472–3484.
- Grafen A (1984) Natural selection, kin selection and group selection. In: Krebs J, Davies N, editors, *Behavioural Ecology: An Evolutionary Approach*, Oxford, UK: Blackwell Scientific Publications. pp. 62–84.
- Lehmann L, Keller L, West S, Roze D (2007) Group selection and kin selection: Two concepts but one process. *Proceeding of the National Academy of Sciences USA* 104: 6736–6739.
- Axelrod R, Hamilton W (1981) The evolution of cooperation. *Science* 211: 1390–1393.
- Kerr B, Godfrey-Smith P (2002) Individualist and multi-level perspectives on selection in structured populations. *Biology and Philosophy* 17: 477–517.
- Doebeli M, Hauert C (2005) Models of cooperation based on the prisoner's dilemma and the snowdrift game. *Ecology Letters* 8: 748–766.
- Antal T, Traulsen A, Ohtsuki H, Tarnita C, Nowak M (2009) Mutation–selection equilibrium in games with multiple strategies. *Journal of Theoretical Biology* 258: 614–362.
- Iliopoulos D, Hintze A, Adami C (2010) Critical dynamics in the evolution of stochastic strategies for the iterated prisoner's dilemma. *PLoS Computational Biology* 6: e1000948.
- Foster K, Shaulsky G, Strassmann J, Queller D, Thompson C (2004) Pleiotropy as a mechanism to stabilize cooperation. *Nature* 431: 693–696.
- Dandekar A, Chugani S, Greenberg E (2012) Bacterial quorum sensing and metabolic incentives to cooperate. *Science* 338: 264–266.
- Jacob F, Monod J (1961) Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology* 3: 318–356.
- Lawrence J (1999) Selfish operons: the evolutionary impact of gene clustering in prokaryotes and eukaryotes. *Current Opinion in Genetics and Development* 9: 642–648.
- Normark S, Bergstrom S, Edlund T, Grundstrom T, Jaurin B, et al. (1983) Overlapping genes. *Annual Review of Genetics* 17: 499–525.
- Knibbe C, Mazet O, Chaudrier F, Fayard JM, Beslon G (2007) Evolutionary coupling between the deleteriousness of gene mutations and the amount of non-coding sequences. *Journal of Theoretical Biology* 244: 621–630.
- Knibbe C, Coulon A, Mazet O, Fayard JM, Beslon G (2007) A long-term evolutionary pressure on the amount of noncoding DNA. *Molecular Biology and Evolution* 24: 2344–2353.
- Nowak M, Bonhoeffer S, May R (1994) Spatial games and the maintenance of cooperation. *Proceedings of the National Academy of Sciences* 91: 4877–4881.
- Oliphant M (1994) Evolving cooperation in the non-iterated prisoner's dilemma: The importance of spatial organization. In: Brooks R, P M, editors, *Proceedings of the Fourth Artificial Life Workshop*. pp. 349–352. URL <http://mitpress.mit.edu/books/artificial-life-iv>.
- Frénoy A, Taddei F, Misevic D (2012) Robustness and evolvability of cooperation. In: Adami C, Bryson D, Ofria C, RT P, editors, *Artificial Life XIII: Proceedings of the Thirteenth International Conference on the Simulation*

- and Synthesis of Living Systems. pp. 53–58. doi:10.7551/978-0-262-31050-5-ch008.
26. Nogueira T, Rankin DJ, Touchon M, Taddei F, Brown SP, et al. (2009) Horizontal gene transfer of the secretome drives the evolution of bacterial cooperation and virulence. *Current Biology* 19: 1683–1691.
 27. Michod R (1996) Cooperation and conflict in the evolution of individuality. II. Conflict mediation. *Proceedings of the Royal Society of London Series B: Biological Sciences* 263: 813–822.
 28. Allen B, Traulsen A, Tarnita C, Nowak M (2012) How mutation affects evolutionary games on graphs. *Journal of Theoretical Biology* 299: 97–105.
 29. Altenberg L (2005) Evolvability suppression to stabilize far-sighted adaptations. *Artificial Life* 11: 427–443.
 30. Rice W (2002) Experimental tests of the adaptive significance of sexual recombination. *Nature Reviews Genetics* 3: 241–251.
 31. Sleight S, Bartley B, Lieviant J, Sauro H (2010) Designing and engineering evolutionary robust genetic circuits. *Journal of Biological Engineering* 4: 12.
 32. Tenaillon O, Taddei F, Radman M, Matic I (2001) Second-order selection in bacterial evolution: selection acting on mutation and recombination rates in the course of adaptation. *Research in Microbiology* 152: 11–16.
 33. Parsons DP, Knibbe C, Beslon G (2010) Importance of the rearrangement rates on the organization of transcription. In: Fellermann H, Dörr M, Hanczyc M, Laursen L, Maurer S, et al, editors, *Artificial Life XII: Proceedings of the Twelfth International Conference on the Synthesis and Simulation of Living Systems*. pp. 479–486. URL <http://mitpress.mit.edu/sites/default/files/titles/alife/0262290758chap92.pdf>.
 34. Misevic D, Frénoy A, Parsons D, Taddei F (2012) Effects of public good properties on the evolution of cooperation. In: Adami C, Bryson D, Ofria C, R.T.P., editors, *Artificial Life XIII: Proceedings of the Thirteenth International Conference on the Simulation and Synthesis of Living Systems*. pp. 218–225. doi:10.7551/978-0-262-31050-5-ch030.
 35. Knibbe C, Fayard JM, Beslon G (2008) The topology of the protein network influences the dynamics of gene order: From systems biology to a systemic understanding of evolution. *Artificial Life* 14: 149–156.