(Fig. 4A). VopSΔ30 did not modify DA-Rac-T35A, confirming that the AMP modification is specific for Thr[35] (Fig. 4A). As expected, DA-Rac incubated with VopS-H348A was not modified (Fig. 4A). To confirm that VopS modifies other members of the Rho family GTPases, we repeated the in vitro labeling assay using Rho, Rac, and Cdc42. In the presence of VopS, all of the GTPases were modified with AMP, whereas they were not modified in the presence of ^{32}P-α–labeled ATP alone or by VopS-H348A (Fig. 4B). Thus, VopS modifies Rho GTPases with AMP. We now refer to this activity as AMPylation and the enzyme as an AMPylator. VopS uses this posttranslational modification of AMPylation to hinder signaling between Rho GTPases and their downstream effectors by blocking the effector binding site on the switch I region of the GTPase with AMP.

Both VopS and protein kinases use ATP to modify substrates, but the phosphate attached to the substrate is distinct. Kinases use the γ phosphate of ATP to modify their substrates on tyrosine, threonine, and serine residues, whereas VopS uses the α phosphate linked to adenosine to modify its substrate on a threonine residue. This type of posttranslational modification on eukaryotic proteins has not previously been observed. However, it has been observed for bacterial glutamine synthetase, albeit autocatalytically on a tyrosine residue, resulting in the sensitization of end-product inhibition (*13*, *14*). Because bacterial type III secreted effectors often mimic eukaryotic mechanisms, the observation of AMPylation by a bacterial effector prompted us to investigate whether eukaryotes use this posttranslational modification. Incubation of S100 HeLa cell lysates with ^{32}P-γ–labeled ATP predictably revealed many phosphorylated protein substrates (Fig. 4C). This modification, phosphorylation, was labile in the presence of a phosphatase (Fig. 4C). To test whether the same type of experiment would reveal AMPylated protein substrates, we incubated S100 lysate with ^{32}P-α–labeled ATP. A number of radiolabeled proteins were observed but were insensitive to phosphatase treatment (Fig. 4C). The addition of purified recombinant VopSΔ30 to the reaction using ^{32}P-α–labeled ATP, but not ^{32}P-γ–labeled ATP, specifically increased labeling at the predicted size of the Rho GTPases (Fig. 4C). Thus, VopS is not a promiscuous AMPylator but rather targets the Rho family of GTPases. Consistent with this observation, phosphorylation and AMPylation did not occur in the presence of denatured protein (Fig. 4C). Thus, eukaryotic proteins can use ATP to modify proteins by AMPylation.

VopS contains a C-terminal Fic domain, and mutation of an invariant histidine residue within this domain led to the discovery of the catalytic activity of modifying proteins with AMP. The conserved histidine is critical for the AMPylation activity. The limited eukaryotic distribution of Fic resembles that of other components of signal transduction machinery and might support a role for AMPylation by eukaryotic Fic domains in signaling. Structures of Fic domains place the conserved polar residues of this motif within a cleft that could represent an active site, with conserved side chains (from E and N) forming polar contacts with a phosphate in one structure (fig. S3A) (*15*). A β hairpin located near the motif binds peptide in another structure, placing a side chain of the peptide within van der Waals contact of the motif histidine (fig. S3B). Although enzymes, such as an activated E1, form AMP-bound covalent enzyme intermediates to drive chemical ligation reactions (*16*), AMP has not previously been shown to be used as a stable posttranslational modification for a protein. This activity represents an ideal posttranslational modification because it (i) uses a highly abundant high-energy substrate, ATP; (ii) results in the formation of a reversible phosphodiester bond; (iii) is bulky enough to bind to an adaptor protein and be used in dynamic multidomain signaling complexes; and (iv) alters the activity of the protein it modifies. It is intriguing that we observed this modification on threonine because this residue is used in many other modifications that might compete with AMPylation. The identification of the substrates and enzymes involved in eukaryotic AMPylation will undoubtedly add a new layer to the expanding complexity of our information about cellular signal transduction.

References and Notes

1. N. A. Daniels *et al.*, *J. Infect. Dis.* **181**, 1661 (2000).
2. K. Makino *et al.*, *Lancet* **361**, 743 (2003).
3. K. S. Park *et al.*, *Microbiol. Immunol.* **48**, 313 (2004).
4. P. Ghosh, *Microbiol. Mol. Biol. Rev.* **68**, 771 (2004).
5. D. L. Burdette, M. L. Yarbrough, A. Orvedahl, C. J. Gilpin, K. Orth, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 12497 (2008).
6. T. Casselli, T. Lynch, C. M. Southward, B. W. Jones, R. DeVinney, *Infect. Immun.* **76**, 2202 (2008).
7. K. S. Park *et al.*, *Infect. Immun.* **72**, 6659 (2004).
8. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; X, any amino acid; and Y, Tyr.
9. R. Utsumi, Y. Nakamoto, M. Kawamukai, M. Himeno, T. Komano, *J. Bacteriol.* **151**, 807 (1982).
10. T. Hakoshima, T. Shimizu, R. Maesaki, *J. Biochem.* **134**, 327 (2003).
11. J. B. Bliska, K. L. Guan, J. E. Dixon, S. Falkow, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 1187 (1991).
12. N. Abdul-Manan *et al.*, *Nature* **399**, 379 (1999).
13. P. B. Chock, S. G. Rhee, E. R. Stadtman, *Annu. Rev. Biochem.* **49**, 813 (1980).
14. M. S. Brown, A. Segal, E. R. Stadtman, *Proc. Natl. Acad. Sci. U.S.A.* **68**, 2949 (1971).
15. M. E. Cuff *et al.*, Midwest Center for Structural Genomics; structure has been deposited in the Protein Data Bank (www.rcsb.org) with the identification number 2F6S.
16. A. L. Haas, J. V. Warms, I. A. Rose, *Biochemistry* **22**, 4388 (1983).

17. We thank N. Alto, R. Taussig, P. Sternweis, S. Mukherjee, M. Rosen, E. Olson, J. Goldstein, M. Brown, T. Iida, T. Honda, L. McCarter, and the Orth lab for insightful discussions, critical reading, and/or generous supply of reagents. K.O. and M.L.Y. are supported by grants from NIH–Allergy and Infectious Disease (R01-AI056404) and the Welch Foundation (I- 1561). L.N.K. and N.G. are supported by the Welch Foundation (I-1505) and Howard Hughes Medical Institute. K.O. is a Beckman Young Investigator, Burroughs Wellcome Investigator, and W. W. Caruth Biomedical Scholar.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1166382/DC1
Materials and Methods
Figs. S1 to S3
References

25 September 2008; accepted 14 November 2008
Published online 27 November 2008;
10.1126/science.1166382
Include this information when citing this paper.

Simpson's Paradox in a Synthetic Microbial System

John S. Chuang,* Olivier Rivoire, Stanislas Leibler

The maintenance of "public" or "common good" producers is a major question in the evolution of cooperation. Because nonproducers benefit from the shared resource without bearing its cost of production, they may proliferate faster than producers. We established a synthetic microbial system consisting of two *Escherichia coli* strains of common-good producers and nonproducers. Depending on the population structure, which was varied by forming groups with different initial compositions, an apparently paradoxical situation could be attained in which nonproducers grew faster within each group, yet producers increased overall. We show that a simple way to generate the variance required for this effect is through stochastic fluctuations via population bottlenecks. The synthetic approach described here thus provides a way to study generic mechanisms of natural selection.

A simple general principle has emerged from theoretical and experimental studies of common-good producer–nonproducer interactions: For producers to be selected and maintained, they have to be the privileged recipients of the common good (*1*–*18*). Producers may become privileged recipients by virtue of kinship or spatial proximity or discrimination through reciprocity or some distinctive feature (*1*–*5*, *9*, *10*, *14*).

In this work, we considered the scenario in which producers and nonproducers are distributed heterogeneously into subpopulations of varying composition, some starting with higher

Center for Studies in Physics and Biology and Laboratory of Living Matter, The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA.

*To whom correspondence should be addressed. E-mail: chuangj@rockefeller.edu

and some with lower fractions of producers. As each of these subpopulations grows, its fraction of producers falls. Yet, it is possible for the overall fraction of producers to rise. This phenomenon, usually analyzed in terms of kin selection or group selection (*3*), can be distilled into an elementary mathematical feature known as Simpson's paradox (*5*, *19*, *20*) (Fig. 1). Because the global producer proportion is a weighted sum of the subpopulation proportions, when there is sufficient covariance between growth rate of a subpopulation and its fraction of producers, a counterintuitive behavior of the whole population can be observed: Producers increase overall.

A similar situation may arise in natural microbial systems (*7*, *8*, *13*), but the effects associated with population heterogeneities are difficult there to assess quantitatively. In addition, in natural systems, evolutionary pressures may have selected for genetic backgrounds that limit the spread of nonproducer mutants and ensure that producers have the growth advantage (*11*). We have taken a synthetic approach (*21*–*25*), which circumvents the difficulties of quantifying the interactions in natural populations. Having no evolutionary history, a synthetic system can furthermore be engineered to minimize pleiotropic effects, enabling measurement and control of other parameters.

We constructed two *Escherichia coli* strains that recapitulate the interaction of producers and nonproducers (*26*) (fig. S1). The common good in this system is a membrane-permeable (*27*) Rhl autoinducer molecule (fig. S1), rewired to activate antibiotic (chloramphenicol; Cm) resistance gene expression. Otherwise isogenic, green fluorescent protein (GFP)–marked producers synthesize the Rhl autoinducer constitutively, whereas nonfluorescent nonproducers do not. The system, denoted *rhl-catLVA* (fig. S1), exhibited the expected properties for public-good producers and nonproducers. First, in antibiotic-containing media, producers grew in a density-dependent manner (fig. S2, left) that was abolished when a synthetic autoinducer was exogenously supplied (fig. S2, right), indicating that autoinducer production was limiting. Second, when started from the same initial density, pure cultures of nonproducers grew slower than pure cultures of producers in antibiotic (fig. S3). However, addition of either synthetic autoinducer or cell-free conditioned medium (containing autoinducer made by producers) increased nonproducer growth in antibiotic-containing media (fig. S3).

When cultures containing preincubated mixtures of producers and nonproducers were diluted into antibiotic-containing medium, both the relative growth of the two strains within a mixture and the growth of the whole mixture depended on the initial proportion of producers (Fig. 2A). In our system, nonproducers proliferated faster than producers within each mixture, independently of the initial composition, as verified by flow cytometry (Fig. 2A). The slower growth of producers can be attributed to the cost of production of the autoinducer, together with the linked GFP reporter. Furthermore, cultures with a higher proportion of producers grew to larger population sizes during the course of the selection phase (12 to 13 hours, 30°C) in antibiotic, and ending subpopulation size was strongly correlated with the initial producer proportion ($r = 0.988$ in 21 trials).

The relative growth advantage of nonproducers versus producers within a mixture, combined with the observed correlation between the overall growth of a mixture and the initial proportion of producers, implied that a situation described by Simpson's paradox could be observed (Fig. 1). To test this, we measured the global proportion of producers in the 10 nonpure cultures

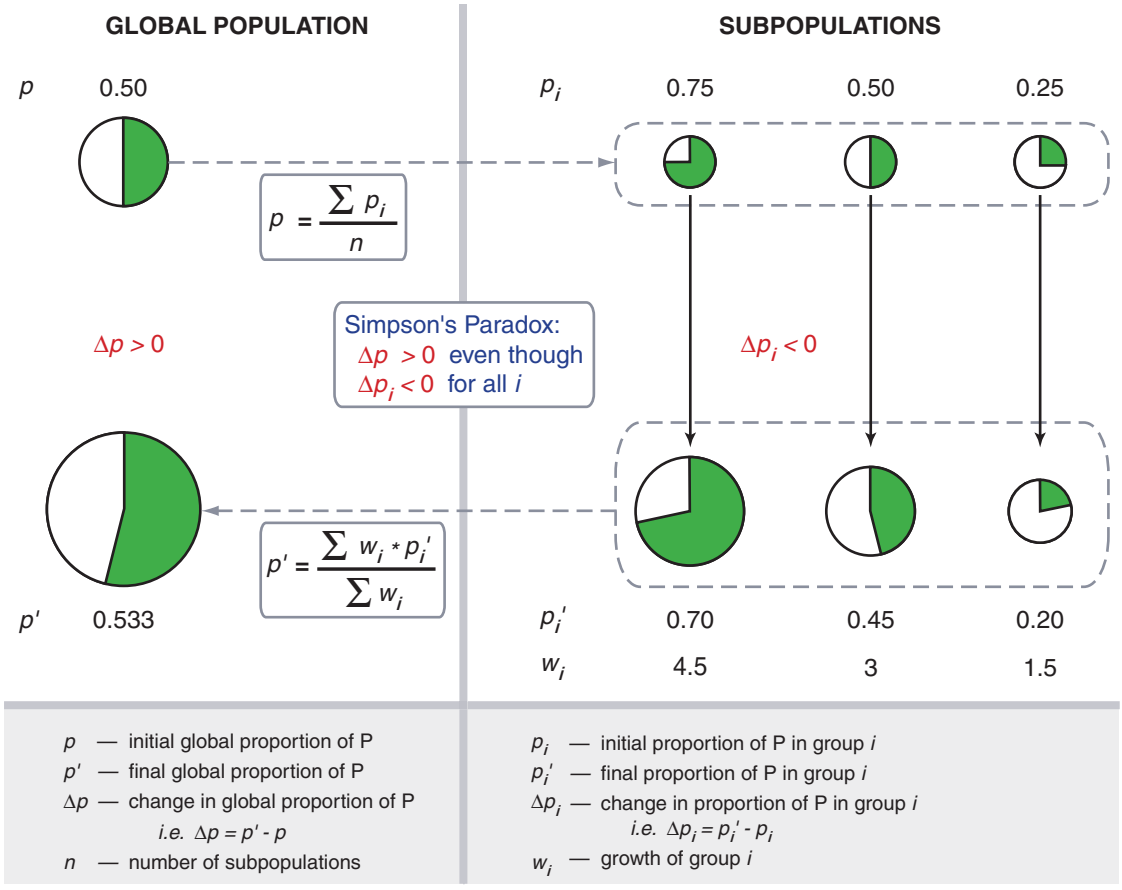

Fig. 1. Principle of Simpson's paradox. Simpson's paradox (*5*, *19*, *20*) refers to a situation in which several groups, composed of two types of elements, P and NP, evolve so that the proportion of P elements decreases within each group but nevertheless increases in average overall. The right side shows the evolution of three hypothetical subpopulations represented by pie charts of P (green) and NP (white) slices; the initial and final subpopulations are connected by solid black arrows. The left side shows the corresponding composition of the initial and final global population formed by these three subpopulations (dotted lines). As a whole, the figure illustrates the paradox of P decreasing in each subpopulation ($\Delta p_i < 0$ for every group i) but increasing overall ($\Delta p > 0$). This "paradox" is a purely statistical effect, based on the fact that the global proportion of P is a group size–weighted average that differs from the nonweighted average. In general, if we start from n groups of equal sizes and proportions p_i of P ($i = 1,\ldots, n$), whether Simpson's paradox is observed depends on the changes in proportion of P within each group, Δp_i, and on the overall growth of each group, w_i. The global variation of the proportion of P, Δp, satisfies (*29*) $\langle w_i \rangle \Delta p = \mathrm{cov}(w_i, p_i) + \langle w_i \Delta p_i \rangle$, where $\langle . \rangle$ denotes a nonweighted average over the groups, $\langle x_i \rangle = (\sum x_i)/n$, and $\mathrm{cov}(w_i, p_i) = \langle w_i p_i \rangle - \langle w_i \rangle \langle p_i \rangle$ represents the covariance between w_i and p_i. Simpson's paradox corresponds to having $\Delta p_i < 0$ for all groups i, but $\Delta p > 0$ globally, and requires sufficient correlation between w_i and p_i, viz, $\mathrm{cov}(w_i, p_i) > -\langle w_i \Delta p_i \rangle$.

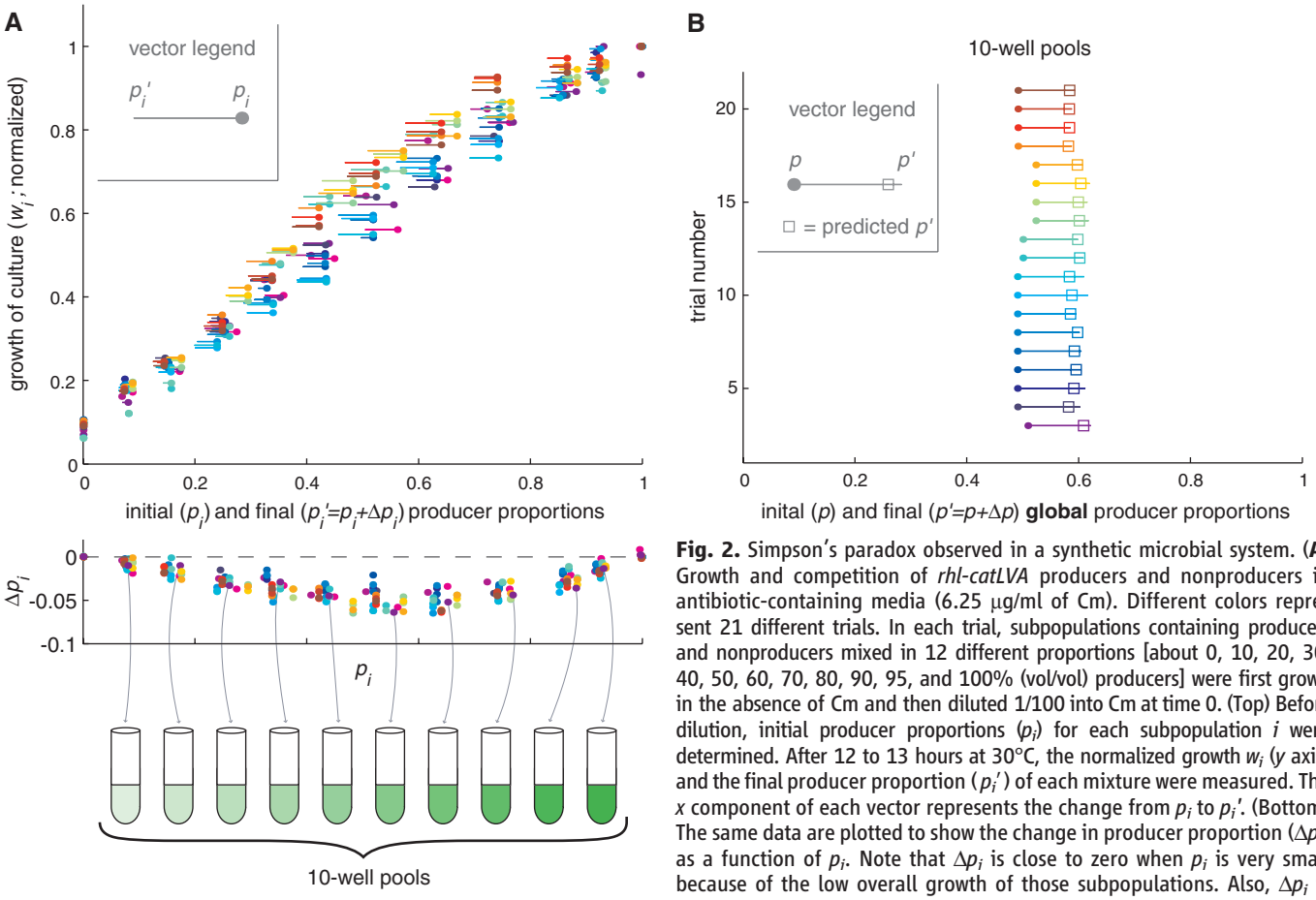

Fig. 2. Simpson's paradox observed in a synthetic microbial system. (**A**) Growth and competition of *rhl-catLVA* producers and nonproducers in antibiotic-containing media (6.25 μg/ml of Cm). Different colors represent 21 different trials. In each trial, subpopulations containing producers and nonproducers mixed in 12 different proportions [about 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 95, and 100% (vol/vol) producers] were first grown in the absence of Cm and then diluted 1/100 into Cm at time 0. (Top) Before dilution, initial producer proportions (p_i) for each subpopulation i were determined. After 12 to 13 hours at 30°C, the normalized growth w_i (y axis) and the final producer proportion (p_i') of each mixture were measured. The x component of each vector represents the change from p_i to p_i'. (Bottom) The same data are plotted to show the change in producer proportion (Δp_i) as a function of p_i. Note that Δp_i is close to zero when p_i is very small because of the low overall growth of those subpopulations. Also, Δp_i is zero when $p_i = 1$ because those cultures are pure producer populations. The curved arrows joining purple points and individual test tubes depict how the 10 nonpure subpopulations from each trial were pooled to form the global populations (10-well pools) in (B). (**B**) Simpson's paradox satisfied by *rhl-catLVA* producers and nonproducers. Initial (p) and final (p') global proportions of producers were determined in 19 of the 21 trials from (A) (the first two trials were not measured at a global level). Each vector plots the change from p to p' (mean change = 0.105, SD = 0.011, n = 19). Colors indicate the correspondence between each trial of subpopulations in (A) (leftward vectors) and each global pool in (B) (rightward vectors). Square symbols show predicted p' values calculated by using the data from individual subpopulations of (A).

with different initial proportions of producers by mixing equal volumes of these cultures and then determining the composition with flow cytometry. As predicted by straightforward calculations (Fig. 1) based on the data obtained for the growth of independent mixtures (Fig. 2A), the global proportion of producers increased relative to its initial value (Fig. 2B), despite the fact that within each mixture, producers did not increase in proportion (Fig. 2A). These results are an illustration of Simpson's paradox in growing microbial populations.

The question of the maintenance of producers has also been addressed in studies of bacterial quorum-sensing mutants (*16*) and siderophore nonproducers (*12*) using simplified laboratory conditions and quantitative measurements. However, in our experiments, there is no direct benefit of being a producer, specifically, no negative frequency dependence ($\Delta p_i > 0$ for lower values of p_i). More importantly, producers and nonproducers are grown here as nonclonal mixtures under selective conditions in which the public good confers an advantage to nonproducers [in contrast, see (*18*)].

The conditions for observing Simpson's paradox are not only based on the properties of the two strains. First, the time scale of the antibiotic selection phase is important; if cultures are allowed to reach stationary phase, those differing greatly in producer proportion (e.g., 20% compared with 80%) will have comparable group sizes, thus preventing Simpson's paradox from occurring. Second, Simpson's paradox also depends on the variance of initial group composition. Thus, long-term maintenance of producers requires an ecological process that enables repeated formation of new groups with a sufficiently large composition variance (*1*). The degree of dispersal can be controlled by varying the "spatial scale of competition" (*6*), that is, the degree to which the groups are mixed at each step. From this point of view, Simpson's paradox corresponds to the observation that "global competition" (complete mixing) may have an effect opposite to "local competition" (no mixing).

Notably, a simple possible mechanism for generating sufficiently large composition variance is an extreme dilution of the group populations (*1*, *2*). Such Poisson dilution gave rise to stochastic fluctuations of group compositions with sufficiently large variances to generate Simpson's paradox. We diluted a 50:50 mixture of producers and nonproducers into 288 wells (three 96-well plates), so that the number of founder cells in these groups was Poisson-distributed with a mean λ (estimated by counting the number of wells with no cells). As before, all founding groups of bacteria were first grown in the absence of antibiotic and then diluted into the antibiotic-containing medium for 12 hours at 30°C. The 288 wells were subsequently pooled together, and the overall proportion of producers was measured. Figure 3A shows the change in producer proportion in a single round: For small enough λ, the generated variance was sufficient to reach the Simpson's paradox regime.

Subsequently, we tested whether this process could be iterated over several rounds, starting with a 10% producer mixture. After each round, the Poisson dilution–antibiotic selection protocol was repeated with use of the pooled population as the new starting mixture for the next round; as a result the overall proportion of producers increased over several consecutive rounds to >95% producers (Fig. 3B). The stochastic fluctuations constitute, therefore, a possible mechanism for

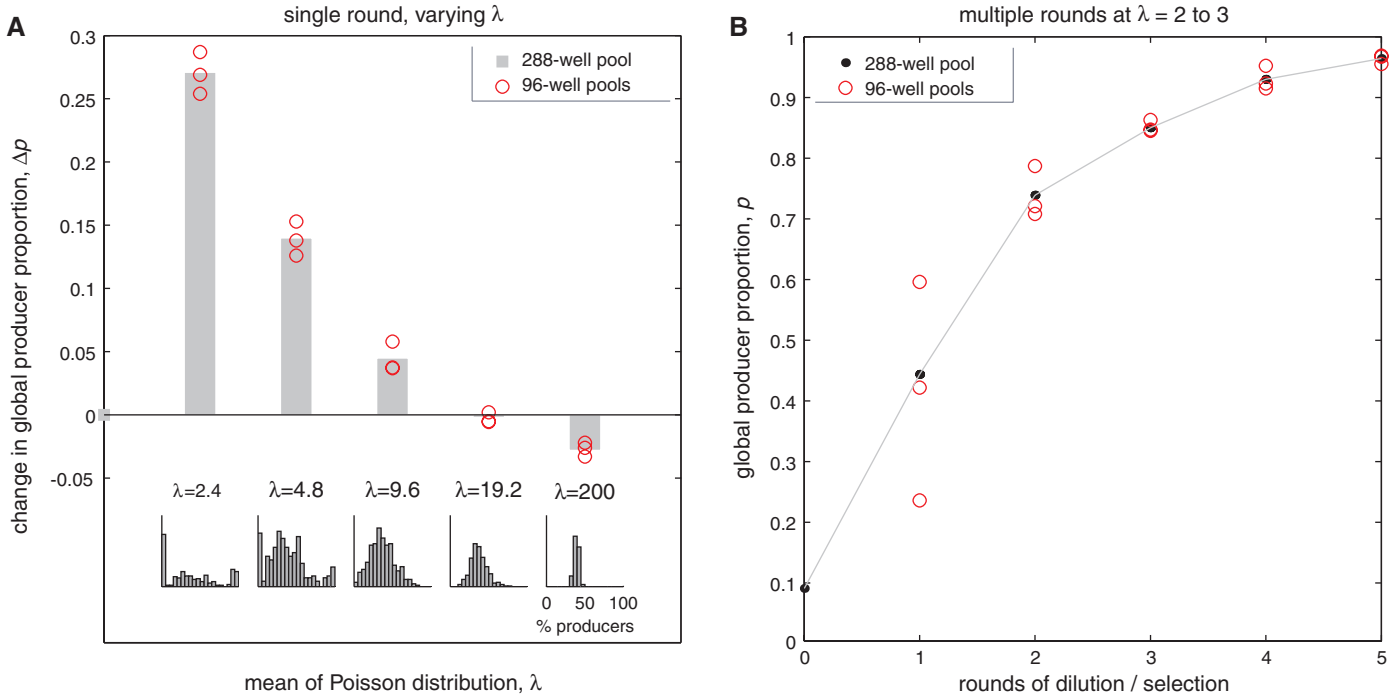

Fig. 3. Poisson dilution conditions can generate sufficient variance to satisfy Simpson's paradox. (**A**) Effect of Poisson mean on the global change in producer proportion. A 50:50 mixture of producers and nonproducers was strongly diluted into three 96-well plates per treatment and first grown without Cm. Then, at time 0, the mixtures were diluted 1/100 into 6.25 μg/ml of Cm for 12 hours at 30°C. The change (Δp) in producer proportion for the global population (288-well pool) is shown for different dilutions, which resulted in the indicated values of λ, the mean number of founder cells per well. Note that the *x* axis is not to scale. Red circles, showing the observed change in separate subpools of each of the three individual plates (96-well pools), provide an estimate of the variation arising from pooling smaller numbers of wells. (Insets) Histogram show the distributions of initial (time 0) producer proportions (p_i) resulting from Poisson dilution conditions. (**B**) Iteration of Poisson dilution conditions and growth in Cm. A 10% producer mixture underwent five successive rounds of the dilution and growth described in (A). After each round, the three 96-well plates were pooled (288-well pool), analyzed for global producer proportion (p), strongly rediluted, and then aliquoted to three new 96-well plates for the subsequent round. The Poisson mean λ ranged from 2 to 3 in each round. Red circles are as described in (A). Resulting from stochastic effects, the larger variation observed in earlier rounds (red circles) is expected because p is small. Note that, in general, the variation also depends on λ and the number of pooled wells.

reversing the direction of selection in growing populations [as noticed in (*1*) and (*2*)]. Although others (*15*) have demonstrated the effects of small group size on producer dynamics within a single population, in this study small founder group size is shown to indirectly favor producers through its effect on population distributions.

From a biological standpoint, our synthetic system provides a simple example of conflicting levels of selection (*28*): Producer cells are beneficial in terms of growth at the higher level of populations (whether single groups or a set of groups), but they are nevertheless not necessarily favored by natural selection, which primarily operates at the lower level of individuals (producers cells have thus a disadvantage within any single group). Simpson's paradox corresponds to a situation where this conflict is overcome, and the trait beneficial to the population is selected in spite of its individual cost. The synthetic system presented here is amenable to further analysis. For example, both the cost of being a producer and the benefit of receiving the public good could in principle be experimentally manipulated. Our work thus shows that one can analyze quantitatively how the direction of selection can be influenced by a combination of cellular and ecological parameters.

References and Notes

1. W. D. Hamilton, *Am. Nat.* **97**, 354 (1963).
2. J. B. S. Haldane, *The Causes of Evolution* (Longmans Green, London, 1932).
3. W. D. Hamilton, in *ASA Studies 4: Biosocial Anthropology*, R. Fox, Ed. (Malaby, London, 1975), pp. 133–153.
4. R. Axelrod, W. D. Hamilton, *Science* **211**, 1390 (1981).
5. E. Sober, D. S. Wilson, *Unto Others: The Evolution and Psychology of Unselfish Behavior* (Harvard Univ. Press, Cambridge, MA, 1998).
6. S. A. Frank, *Foundations of Social Evolution* (Princeton Univ. Press, Princeton, NJ, 1998).
7. P. B. Rainey, A. Buckling, R. Kassen, M. Travisano, *Trends Ecol. Evol.* **15**, 243 (2000).
8. B. J. Crespi, *Trends Ecol. Evol.* **16**, 178 (2001).
9. L. Lehmann, L. Keller, *J. Evol. Biol.* **19**, 1365 (2006).
10. D. C. Queller, E. Ponte, S. Bozzaro, J. E. Strassmann, *Science* **299**, 105 (2003).
11. K. R. Foster, G. Shaulsky, J. E. Strassmann, D. C. Queller, C. R. L. Thompson, *Nature* **431**, 693 (2004).
12. A. S. Griffin, S. A. West, A. Buckling, *Nature* **430**, 1024 (2004).
13. L. Keller, M. G. Surette, *Nat. Rev. Microbiol.* **4**, 249 (2006).
14. B. Kerr, C. Neuhauser, B. J. M. Bohannan, A. M. Dean, *Nature* **442**, 75 (2006).
15. M. A. Brockhurst, *PLoS One* **2**, e634 (2007).
16. S. P. Diggle, A. S. Griffin, G. S. Campbell, S. A. West, *Nature* **450**, 411 (2007).
17. K. M. Sandoz, S. M. Mitzimberg, M. Schuster, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 15876 (2007).
18. G. M. Dunny, D. J. Brickman, M. Dworkin, *Bioessays* **30**, 296 (2008).
19. C. R. Blyth, *J. Am. Stat. Assoc.* **67**, 364 (1972).
20. P. J. Bickel, E. A. Hammel, J. W. O'Connell, *Science* **187**, 398 (1975).
21. J. W. Chin, *Nat. Chem. Biol.* **2**, 304 (2006).
22. W. Weber, M. Daoud-El Baba, M. Fussenegger, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 10435 (2007).
23. W. Shou, S. Ram, J. M. G. Vilar, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 1877 (2007).
24. K. Brenner, D. K. Karig, R. Weiss, F. H. Arnold, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 17300 (2007).
25. F. K. Balagaddé *et al.*, *Mol. Syst. Biol.* **4**, 187 (2008).
26. Information on materials and methods is available on *Science* Online.
27. J. P. Pearson, C. Van Delden, B. H. Iglewski, *J. Bacteriol.* **181**, 1203 (1999).
28. L. Keller, Ed., *Levels of Selection in Evolution* (Princeton Univ. Press, Princeton, NJ, 1999).
29. G. R. Price, *Nature* **227**, 520 (1970).

30. We are grateful to C. Bargmann, C. Guet, L. Keller, E. Kussell, R. Losick, and A. Murray for comments on the manuscript and members of our laboratory for advice and discussion. We also thank the Yale Coli Genetic Stock Center and the National BioResource Project (NIG, Japan) for strains and the Rockefeller Flow Cytometry Resource Center for help. J.S.C. was partly supported by a NIH fellowship and O.R. by a Human Frontier Science Program fellowship.

Supporting Online Material

www.sciencemag.org/cgi/content/full/323/5911/272/DC1
Materials and Methods
Figs. S1 to S3
References

9 July 2008; accepted 19 November 2008
10.1126/science.1166739