SOURCE DATA

Thomas Lemberger

Chief Editor, Molecular Systems Biology
Deputy Head, Scientific Publications, EMBO
The publication of scientific information is intended to move science forward. More specifically, the act of publishing is a *quid pro quo* in which authors receive credit and acknowledgment in exchange for disclosure of their scientific findings.
move science forward
credit
disclose findings
data
open data
data integrity
Big Data
Structured data

structures
sequences
functional genomics
proteomics
metabolomics
genotype
phenotype
‘Publishing’ papers

‘Depositing’ datasets
“The two vital components of the scientific endeavor – the idea and the evidence – are too frequently separated.”
Journals publish data!

Introduction

Cooperation is a widespread phenomenon in nature. However, costly cooperative strategies are vulnerable to exploitation by cheats that do not cooperate but benefit from the benefits produced by the cooperating individuals (Axelrod and Hamilton, 1981; Nowak, 2006). Therefore, the persistence of cooperators in nature has been a puzzling question for evolutionary biologists and has been much theoretical and experimental research trying to elucidate the mechanisms underlying this phenomenon (Trivers, 1971; Maynard Smith, 1976; West et al., 2002; Nowak et al., 2006). Microbial studies have shown that cooperation can be maintained in nature by mechanisms such as group-beneficial (Queller et al., 2001; Mesibov et al., 2003) and indirect reciprocity (Taylor and Nowak, 2007; Macaluso and Gavrilets, 2009; Eguiluz et al., 2009), and social evolution (Chang et al., 2009). Recently, it has become increasingly clear that in addition to population dynamics, cooperation also affects the evolution of cooperation (Brown et al., 2010).

One such important ecological factor is interspecific interactions (Lieberman et al., 2008). However, almost all laboratory experiments aimed at understanding cooperation have relied on pairwise interactions, ignoring the complexity of interactions that are in the wild and exist within complex communities where they interact with other species (Chapin et al., 1994).

Molecular Systems Biology 2012 1

Competition between species can stabilize public-goods cooperation within a species

Hector Callen1 and Jeff Gore2

1 Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA and 2 Department of Physics, Massachusetts Institute of Technology, Cambridge, MA, USA

Corresponding author: Department of Physics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, 39-120, Cambridge, MA 02139, USA; E-mail: callen@mit.edu; gore@mit.edu

Competition between species in a major ecological force that can drive evolution. Here, we test the effect of this force on the evolution of cooperation within a species. We use stochastic simulation of host–parasite coevolution models, as a model cooperation system that is subject to selection by other species, and we find that when cooperators are the dominant species, the frequency of cooperators increases over time. However, when competing with other species with a lower frequency of cooperators, the frequency of cooperators decreases over time. In both cases, the frequency of cooperators in the population remains stable over time. These findings suggest that cooperation in nature may be maintained by the presence of species that do not cooperate and that the frequency of cooperators in a species population may be determined by the frequency of cooperators in the population.
Scientific publishing

• Dominant channel for the dissemination of peer-reviewed data.
• Journals function as a proxy for quality in research assessment.
• The rate of publishing keeps increasing.
• Papers are human-readable but poorly machine-readable.
search
HOW SCIENCE GOES WRONG
Title

Abstract

Synopsis

Main paper

'Expanded view'

Datasets & code
mTORC1 signaling is positively regulated by growth factors through the PI3K-AKT pathway (Growth Factor/Nutrient module in Figures 2 and 3). The binding of insulin to its cell surface receptor leads to the recruitment and phosphorylation of IRS-1, which promotes the recruitment and activation of PI3K at the cell surface membrane. Active PI3K converts phosphatidylinositol-4,5-phosphate to phosphatidylinositol-3,4,5-phosphate (PIP3), a process antagonized by the lipid phosphatase PTEN. When produced at the plasma membrane, PIP3 recruits both PDK1 and AKT, resulting in the phosphorylation and partial activation of AKT. Whereas PDK1 phosphorylates AKT at Thr308, additional phosphorylation at Ser473 by mTORC2 (see below) is necessary for optimal activation of AKT in vitro (Sarbassov et al., 2005). mTORC1 is thought to be activated in part by AKT through the tuberous sclerosis complex proteins, TSC1 and TSC2. The TSC1-TSC2 complex is a critical negative regulator of mTORC1 (Huang and Manning, 2008b). Because of its central role in regulating mTORC1, 34 species depicting extensive details about the TSC1-TSC2 complex (post-translation modifications, interactors, cellular locations) were represented in the comprehensive mTOR map. In response to growth factors, TSC2 is phosphorylated and functionally inactivated by AKT (Inoki et al., 2002; Manning et al., 2002). ERK1/2 and RSK1/2 were also shown to phosphorylate and inactivate TSC2 in response to growth factors (Roux et al., 2004; Ballif et al., 2005; Ma et al., 2005), suggesting that PI3K and Ras/MAPK pathways collaborate to inhibit TSC1-TSC2 function in response to growth factors. Whereas TSC2 functions as a GAP toward the small Ras-related GTPase Rheb, TSC1 is required to stabilize TSC2 and prevent its proteasomal degradation (Huang and Manning, 2008b). While the active GTP-bound form of Rheb was shown to directly interact with mTOR to stimulate its catalytic activity (Long et al., 2005), Rheb may also promote substrate recognition by mTORC1 (Sancak et al., 2007; Sato et al., 2009). Nutrients, such as amino acids, regulate mTORC1 signaling via different mechanisms. Amino acid availability regulates mTORC1 in a TSC2-independent but Rheb-dependent manner (Smith et al., 2005; Gulati and Thomas, 2007), but the exact mechanism remains poorly understood (Growth Factor/Nutrient module in Figures 2 and 3). Two complementary studies have provided compelling evidence that the Rag family of small GTPases is both necessary and sufficient to transmit a positive signal from amino acids to mTOR (Kim et al., 2008; Sancak et al., 2008). The current model proposes that amino acids induce the movement of mTORC1 to lysosomal membranes, where Rag proteins reside. More precisely, a complex encoded by the MAPKSP1, ROBLD3 and c11orf59 genes, interacts with the Rag GTPases, recruits them to lysosomes, and was shown to be essential for mTORC1 activation (Sancak et al., 2010). mTORC1 activity is sensitive to oxygen deprivation, and one pathway by which this occurs involves activation of the TSC1-TSC2 complex by REDD1, a hypoxia-inducible protein (Hypoxia module in Figures 2 and 31; Brugarolas et al., 2004). Newly synthesized REDD1 was found to interact with 14-3-3 and relieve TSC2 from 14-3-3-dependent repression (DeYoung et al., 2008). mTORC1 also senses insufficient cellular energy levels through AMPK, a protein kinase activated in response to a low ATP/AMP ratio (Inoki et al., 2003) and by LKB1-mediated phosphorylation (Low Energy module in
mTORC1 signaling is positively regulated by growth factors through the PI3K-AKT pathway (Growth Factor/Nutrient module in Figures 2 and 3). The binding of insulin to its cell surface receptor leads to the recruitment and phosphorylation of IRS-1, which promotes the recruitment and activation of PI3K at the cell surface membrane. Active PI3K converts phosphatidylinositol-4,5-phosphate to phosphatidylinositol-3,4,5-phosphate (PIP3), a process antagonized by the lipid phosphatase PTEN. When produced at the plasma membrane, PIP3 recruits both PDK1 and AKT, resulting in the phosphorylation and partial activation of AKT. Whereas PDK1 phosphorylates AKT at Thr308, additional phosphorylation at Ser473 by mTORC2 (see below) is necessary for optimal activation of AKT in vitro (Sarbassov et al, 2005). mTORC1 is thought to be activated in part by AKT through the tuberous sclerosis complex proteins, TSC1 and TSC2. The TSC1-TSC2 complex is a critical negative regulator of mTORC1 (Huang and Manning, 2008b). Because of its central role in regulating mTORC1, 34 species depicting extensive details about the TSC1-TSC2 complex (post-translation modifications, interactors, cellular locations) were represented in the comprehensive mTOR map. In response to growth factors, TSC2 is phosphorylated and functionally inactivated by AKT (Inoki et al, 2002; Manning et al, 2002). ERK1/2 and RSK1/2 were also shown to phosphorylate and inactivate TSC2 in response to growth factors (Roux et al, 2004; Ballif et al, 2005; Ma et al, 2005), suggesting that PI3K and Ras/MAPK pathways collaborate to inhibit TSC1-TSC2 function in response to growth factors. Whereas TSC2 functions as a GAP toward the small Ras-related GTPase Rheb, TSC1 is required to stabilize TSC2 and prevent its proteasomal degradation (Huang and Manning, 2008b). While the active GTP-bound form of Rheb was shown to directly interact with mTOR to stimulate its catalytic activity (Long et al, 2005), Rheb may also promote substrate recognition by mTORC1 (Sancak et al, 2007; Sato et al, 2009). Nutrients, such as amino acids, regulate mTORC1 signaling via different mechanisms. Amino acid availability regulates mTORC1 in a TSC2-independent but Rheb-dependent manner (Smith et al, 2005; Gulati and Thomas, 2007), but the exact mechanism remains poorly understood (Growth Factor/Nutrient module in Figures 2 and 3). Two complementary studies have provided compelling evidence that the Rag family of small GTPases is both necessary and sufficient to transmit a positive signal from amino acids to mTOR (Kim et al, 2008; Sancak et al, 2008). The current model proposes that amino acids induce the movement of mTORC1 to lysosomal membranes, where Rag proteins reside. More precisely, a complex encoded by the MAPKSP1, ROBLD3 and c11orf59 genes, interacts with the Rag GTPases, recruits them to lysosomes, and was shown to be essential for mTORC1 activation (Sancak et al, 2010). mTORC1 activity is sensitive to oxygen deprivation, and one pathway by which this occurs involves activation of the TSC1-TSC2 complex by REDD1, a hypoxia-inducible protein (Hypoxia module in Figures 2 and 31; Brugarolas et al, 2004). Newly synthesized REDD1 was found to interact with 14–3–3 and relieve TSC2 from 14–3–3-dependent repression (DeYoung et al, 2008). mTORC1 also senses insufficient cellular energy levels through AMPK, a protein kinase activated in response to a low ATP/AMP ratio (Inoki et al, 2003) and by LKB1-mediated phosphorylation (Low Energy module in...
mTORC1 signaling is positively regulated by growth factors through the PI3K-AKT pathway (Growth Factor/Nutrient module in Figures 2 and 3). The binding of insulin to its cell surface receptor leads to the recruitment and phosphorylation of IRS-1, which promotes the recruitment and activation of PI3K to its cell surface membrane. Active PI3K converts phosphatidylinositol-4,5-phosphate to phosphatidylinositol 3,4,5-triphosphate (PIP3), a process antagonized by the lipid phosphatase Pten. When produced at the plasma membrane, PIP3 recruits both PIK3 and PDK1 resulting in the phosphorylation and partial activation of AKT. Whereas PDK1 phosphorylates AKT at Thr308, additional phosphorylation at Ser473 by mTORC2 (see below) is necessary for optimal activation of AKT in vitro (Sarbassov et al., 2005). Because of its central role in regulating mTORC1, 34 species depicting extensive details about the TSC1-TSC2 complex (post-translation modifications, interactors, cellular locations) were represented in the comprehensive mTOR map. In response to growth factors, TSC2 is phosphorylated and functionally inactivated by AKT (Inoki et al., 2002; Manning et al., 2003). ERK1/2 and RSK1/2 were also shown to phosphorylate and inactivate TSC2 in response to growth factors (Roux et al., 2004; Ballif et al., 2005; Ma et al., 2005), suggesting that PI3K and Ras/MAPK pathways collaborate to inhibit TSC1-TSC2 function in response to growth factors. Whereas TSC2 functions as a GAP toward the small Ras-related GTPase Rheb, TSC1 is required to stabilize TSC2 and prevent its proteasomal degradation (Huang and Manning, 2008b). While the active GTP-bound form of Rheb was shown to directly interact with mTOR to stimulate its catalytic activity (Long et al., 2005), Rheb may also promote substrate recognition by mTORC1 (Sancak et al., 2007; Sato et al., 2009). Nutrients, such as amino acids, regulate mTORC1 signaling via different mechanisms. Amino acid availability regulates mTORC1 in a TSC2-independent but Rheb-dependent manner (Smith et al., 2005; Gulati and Thomas, 2007), but the exact mechanism remains poorly understood (Growth Factor/Nutrient module in Figures 2 and 3). Two complementary studies have provided compelling evidence that the Rag family of small GTPases is both necessary and sufficient to transmit a positive signal from amino acids to mTOR (Kim et al. 2008; Sancak et al., 2008). The current model proposes that amino acids induce the movement of mTORC1 to lysosomal membranes, where Rag proteins reside. More precisely, a complex encoded by the MAPKSP1, ROBLD3 and c11orf59 genes, interacts with the Rag GTPases, recruits them to lysosomes, and was shown to be essential for mTORC1 activation (Sancak et al., 2010). mTORC1 activity is sensitive to oxygen deprivation, and one pathway by which this occurs involves activation of the TSC1-TSC2 complex by REDD1, a hypoxia-inducible protein (Hypoxia module in Figures 2 and 31; Brugarolas et al., 2004). Newly synthesized REDD1 was found to interact with 14-3-3 and relieve TSC2 from 14-3-3-dependent repression (DeYoung et al., 2008). mTORC1 also senses insufficient cellular energy levels through AMPK, a protein kinase activated in response to a low ATP/AMP ratio (Inoki et al., 2003) and by LKB1-mediated phosphorylation (Low Energy module in Figures 2 and 3).
What is a figure?

A scientific result converted into a collection of pixels…
Tools to publish figures as structured digital objects that link the human-readable illustrations with machine-readable metadata and ‘source data’ in order to

• improve data transparency;
• make published data useable;
• enable data-oriented search.
SourceData

- Figure source data files hosted by the journals
- Link to ‘unstructured data’ repositories

Data

- Focus on the biological content
- Use standard identifiers and existing controlled vocabularies

Metadata

- Data-oriented semantic search of the literature.
- Overcome some of the limitations of keyword-based search

Search
Figure 1

Systematic analysis of PDGF-stimulated Erk phosphorylation kinetics. (A) Immunoblots, representative of five or six independent experiments, used to quantify relative amounts of phosphorylated Erk (p-Erk1/2) and total Erk (t-Erk1). NIH 3T3 fibroblasts were modulated by retroviral induction of dominant-negative (S17N) or constitutively active (G12V) H-Ras expression or incubation with inhibitors of PI3K (100 µM LY294002) or MEK (50 µM PD98059). The respective controls are empty pBM-puro vector or 0.2% DMSO. Lysates were prepared from cells that were unstimulated or stimulated with PDGF-BB for 5, 15, 30, 60, or 120 min. (B–E) Quantification of Erk phosphorylation, normalized as described under Materials and methods, comparing either S17N Ras expression (B; n=6), PI3K inhibition (C; n=5), G12V Ras expression (D; n=6), or MEK inhibition (E; n=5) with their respective controls. Values are reported as mean±s.e.m., and comparisons to control in (B, C) are by Student’s t-test: *P<0.05; **P<0.01. Source data is available for this figure at www.nature.com/msb.

Full figure and legend (660K)
Source data for figure 1BD (6K)
Source data for figure 1CE (5K)
Figures & Tables Index
Silencing of HDAC6 impairs embryonic vessel formation in zebrafish. (A) Aberrant splicing of Danio rerio HDAC6 mRNA after HDAC6 splice-blocking Mo injection by PCR. Injection of the HDAC6 SB-Mo generated at 24 h post fertilization a morphant signal of 333 bp, whereas the HDAC6 wt signal completely disappeared (253 bp), showing the functionality of the Mo. Whole-zebrafish embryo mRNA was isolated 24 h after Mo injection and subjected to RT-PCR. Actin mRNA expression serves as loading control. (B) HDAC6 protein expression was analysed in whole-zebrafish embryo lysate at 24 h after injection of HDAC6 translation-blocking or splice-blocking Mo. Protein lysates were subjected to western blotting with HDAC6-specific antibody. Actin was used as loading control. C–F phenotyping of HDAC6 morphants 48 h post fertilization. (C) Representative confocal fluorescence pictures of vessel in the anterior part of tg(lfli1:EGFP) zebrafish embryos after injection of HDAC6 translation-blocking or control Mo. Arrows indicate vessel defects. (D–F) For quantification of vessel defects, HDAC6 Mo- or control Mo-treated zebrafish embryos were stained for GFP using anti-GFP antibody. (D) Representative overview pictures and higher magnification of two regions of the anterior part of control-Mo-injected and HDAC6 TB-Mo-injected embryos are shown. Arrows indicate vessel defects. (E) Quantification of defects in ISVs and DLAbs for HDAC6 and control morphants. Statistical significance was calculated for
- Data archival service
- Data ‘transparency’
- Data reuse
- Data-oriented search
Actionable data

Clustering phenotype populations by genome-wide RNAi and multiparametric imaging

Florian Fuchs, Grégoire Pau, Dominique Kranz, Oleg Sklyar, Christoph Budjan, Sandra Steinbrink, Thomas Horn, Angelika Peda, Wolfgang Huber & Michael Boutros.
SourceData

- Figure source data files hosted by the journals
- Link to ‘unstructured data’ repositories

Data

- Focus on the biological content
- Use standard identifiers and existing controlled vocabularies

Metadata

Search

- Data-oriented semantic search of the literature.
- Overcome some of the limitations of keyword-based search
Structured metadata: ‘perturbation-observation-assay’

(Level 0: metadata associated to individual panels.)

Level 1: ‘object-oriented’ representation of experimental variables as a list of chemical and biological components.

Level 2: represent the causality of the experimental design: “Measurement of Y as a function of A, B, C, using assay P in biological system S.”

Level 3: machine-readable representation with standard identifiers.
‘Data copy editors’
Data workflow
SourceData

Data
- Figure source data files hosted by the journals
- Link to ‘unstructured data’ repositories

Metadata
- Focus on the biological content
- Use standard identifiers and existing controlled vocabularies

Search
- Data-oriented semantic search of the literature.
- Overcome some of the limitations of keyword-based search
Data-oriented search

Query:

More-like-this:

Survey (n=441)

98% of respondents find this function “useful” (42% rate it as “Fantastic!”)
Data-oriented search
Resulting hypothesis: test drug Z in disease D.
TGFβ, Smad3

Rad51

Nuclear complexes
Further opportunities

- Context- and network-based ranking metrics
- Linking to external semantic-aware resources
- High level (visual) summarization
- Structured representation of assays
- Use same metadata for data series in datasets
Publishing as a distributed infrastructure

Research data → Database → Users

Research data → Journals → Users
Publishing as a distributed infrastructure

Research data → Journals → Database → Users
A global view of pleiotropy and phenotypically derived gene function in yeast

Johanna Marie Dudley1, Daniel Woyke Jr1,2, Arvind Tomp1, J. Shotkit3 and George Church1,2

1 Department of Biochemistry, Harvard Medical School, Boston, MA, USA and 2 School of Computer Science, Yeshiva University, Naval Air, Yonkers, New York, USA
2 Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA, USA
3 Department of Computer Science, Harvard University, Cambridge, MA, USA

Introduction

Pleiotropy occurs when a mutation in a single gene produces effects on more than one characteristic, such as multiple molecular phenotypes. It has been found to be widespread in microbial genomes and is thought to play a role in the evolution of complex organisms. In yeast, pleiotropic mutations are relatively common and have been extensively studied. However, the mechanisms underlying pleiotropy are not well understood. Understanding these mechanisms could lead to new insights into the evolution of complex organisms and the development of novel therapeutic strategies.

Techniques and resources developed in the field of functional genomics and computational biology have the potential to help address these challenges. For example, the use of unbiased genetic screens can identify novel genes involved in complex traits, such as drug resistance or disease susceptibility. These screens can also be used to discover new targets for drug development.

In this study, we performed a comprehensive analysis of the phenotypic effects of thousands of mutations in yeast. We identified a large number of pleiotropic phenotypes, including changes in growth, morphology, and metabolism. These results suggest that pleiotropy is a widespread phenomenon in yeast and that it plays a key role in the evolution of complex organisms. Our findings also have implications for genetic disease, as pleiotropy is thought to be a common feature of many genetic disorders in humans.

By understanding the mechanisms underlying pleiotropy, we can gain new insights into the evolution of complex organisms and the development of novel therapeutic strategies.