The Molecular Diversity of Adaptive Convergence

Olivier Tenaillon, *et al.*

Science **335**, 457 (2012);
DOI: 10.1126/science.1212986

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by clicking here.

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines here.

The following resources related to this article are available online at www.sciencemag.org (this infomation is current as of January 27, 2012):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
http://www.sciencemag.org/content/335/6067/457.full.html

Supporting Online Material can be found at:
http://www.sciencemag.org/content/suppl/2012/01/25/335.6067.457.DC1.html

This article **cites 44 articles**, 16 of which can be accessed free:
http://www.sciencemag.org/content/335/6067/457.full.html#ref-list-1

This article appears in the following **subject collections:**
Evolution
http://www.sciencemag.org/cgi/collection/evolution

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. Copyright 2012 by the American Association for the Advancement of Science; all rights reserved. The title *Science* is a registered trademark of AAAS.

The Molecular Diversity of Adaptive Convergence

Olivier Tenaillon,[1,2,3]† Alejandra Rodríguez-Verdugo,[1] Rebecca L. Gaut,[1] Pamela McDonald,[1] Albert F. Bennett,[1] Anthony D. Long,[1]* Brandon S. Gaut[1]*†

To estimate the number and diversity of beneficial mutations, we experimentally evolved 115 populations of *Escherichia coli* to 42.2°C for 2000 generations and sequenced one genome from each population. We identified 1331 total mutations, affecting more than 600 different sites. Few mutations were shared among replicates, but a strong pattern of convergence emerged at the level of genes, operons, and functional complexes. Our experiment uncovered a set of primary functional targets of high temperature, but we estimate that many other beneficial mutations could contribute to similar adaptive outcomes. We inferred the pervasive presence of epistasis among beneficial mutations, which shaped adaptive trajectories into at least two distinct pathways involving mutations either in the RNA polymerase complex or the termination factor *rho*.

Beneficial mutations fuel adaptation, yet little is known about their potential diversity. If identical populations adapted to a fixed environment, would adaptation occur via identical convergent mutations or via numerous alternative pathways? How might such pathways be shaped by interactions among beneficial mutations? Studies have shown that both the rate and effect of beneficial mutations varies during the course of adaptation because of epistatic interactions that modify the amplitude of fitness effects. However, these studies have been either statistical, without knowledge of affected sites and functions (*1*), or based on interactions between only a handful of beneficial mutations (*2*–*5*).

The potential number and diversity of beneficial mutations can be assessed experimentally under tight control and replication, but assessment requires a large number of replicates for statistical power; complete genome sequencing, so that mutations are identified unambiguously; and a complex biological system, to ensure that the number of potential adaptive solutions is not trivial. Thus far, no studies have fulfilled all three criteria. Experimental evolution of *Drosophila*, yeast, and *Escherichia coli* has been particularly limited by the level of replication. The most prominent long-term studies consist of only 5 (*6*) and 12 (*7*) lines per treatment, and only a subset of lines has been sequenced (*8*, *9*).

We coupled whole-genome sequencing with 115 replicate populations to characterize the genetic response of *E. coli* to high temperature (42.2°C). Temperature is a complex environmental variable because it governs the rates of biological reactions that underlie activities such as respiration, growth, and reproduction. Adaptation to temperature is also pertinent for understanding the biological consequences of ongoing global climate change. Finally, previous experiments have documented a rapid adaptive response of *E. coli* to high temperature (*10*), with hints that the genetic response may be diverse (*11*, *12*). However, these studies lacked sufficient replication or genetic characterization to assess fully the diversity of genetic changes underlying adaptation.

We serially propagated more than 100 *E. coli* asexual lines for 2000 generations in Davis minimal medium, supplemented with 25 mg/l glucose, at a constant temperature of 42.2°C. These lines originated from a common *E. coli* B REL1206 ancestral clone that had been adapted to the same medium for 2000 generations at 37°C (*7*). Selection was strong, because the temperature was the maximum for population persistence in the experimental environment. The duration of the experiment was chosen to encompass most of the expected fitness gain (*13*), but adaptation was likely not complete. At the end of the experiment, a single clone of each population was sequenced and analyzed for relative fitness and yield at 42.2°C (table S1). Fitness was estimated by competition with the ancestor (*14*) and increased markedly in all lines to a mean of 1.42 (±0.024 95% confidence level). Yield, a measure of absolute fitness, was on average 1.94-fold that of the ancestor (paired *t* test, $P < 0.001$).

To investigate the genetic changes underlying adaptation, we sequenced all 115 clones to an average of 90× coverage, for a total data set of 50 gigabase pairs (Gbp) (*14*). We developed a computational pipeline to identify all de novo mutations relative to the *E. coli* B reference genome (*14*, *15*). On the basis of both phenotypic and genotypic analysis, one of the lines had an increased mutation rate, because of a nonfunctional *mut*L gene. This line had higher fitness but many more mutations (73) than the other lines and was excluded from additional analyses. For the remaining 114 lines, we detected a total of 1258 molecular changes (Fig. 1), with an average of 11.0 events per clone: 6.9 point mutations, 2.3 short insertions and deletions, 1.0 large deletions (>30 bp), 0.6 insertional sequence (IS) element integrations, and 0.2 large duplications. There was no detectable correlation with a clone's fitness and its number of mutations ($r = -0.11$; $P = 0.63$), but the point mutations emitted a strong signal of adaptive evolution. Over 114 genomes, the ratio of nonsynonymous to synonymous mutations per site was 5.75, as was the ratio of intergenic to synonymous mutations. We thus estimate that ~80% of intergenic and nonsynonymous mutations were beneficial (*14*).

Convergent mutations provide additional evidence of adaptive events (*16*), but convergence varied by mutational type. Among point mutations, none of the 36 synonymous and 157 of the 634 nonsynonymous mutations were shared among two or more lines, but some were shared extensively. Eighteen lines contained a nonsynonymous mutation in codon 966 of the RNA polymerase (RNApol) β subunit (*rpoB*), and 17 lines contained a nonsynonymous mutation in codon 15 of the *rho* gene.

In contrast to point mutations, 69% (82 of 119) of large (>30-bp) deletions were identical between at least two lines. Of these deletions, 43% had end points overlapping IS elements, consistent with deletion mediated by homologous recombination between elements. These results demonstrate that IS elements can be the basis for adaptive mutation events (*17*), even if the insertions themselves are selectively neutral or even deleterious. Duplications, indels, and IS element integrations were shared among lines at proportions intermediate to those of point mutations and large deletions (Fig. 1, C and D).

Previous *E. coli* studies suggested that the gene level is more appropriate to assess convergence (*18*). In some cases, this was obviously true. For example, the gene *yba*L, a predicted potassium promoter, was mutated in 65 lines because of 38 nonsynonymous mutations, 6 in-frame indels, 17 frameshift mutations, 3 large deletions, and one IS integration. The targeting of 65 different mutations to one gene is highly nonrandom ($P < 10^{-100}$). *yba*L inactivation may be advantageous in our experimental setting (*14*) and also occurred in an independent high-temperature study (*12*). Overall, convergence is much higher among genes than among sites: On average, two strains shared 2.6% of mutations (excluding synonymous mutations) but shared 20.2% of modified genes and 24.5% of affected operons (Fig. 2A).

Selection may ultimately target multigenic functional groups, such as protein complexes. Focusing on genes with >5 mutations, we clustered them into 10 functional units containing 35.7% of the mutations (*14*) (Table 1 and table S2). Among these, the RNApol complex was an obvious target, with 205 mutations and the most-mutated gene (*rpo*B). The other functional units accrued mutations more frequently than expected (Table 1) including proteins that regulate the *rpo*S stress response (RSS), which had been identified previously as a target of selection under high temperature (*11*). At this functional level, two lines shared an average of 31.5% of affected units (Fig. 2A).

---

[1] Department of Ecology and Evolutionary Biology, University of California, Irvine, 321 Steinhaus Hall, Irvine, CA 92697, USA. [2] INSERM, UMR-S 722, 75018 Paris, France. [3] Université Paris Diderot, Sorbonne Paris Cité, UMR-S 722, Site Xavier Bichat, 75018 Paris, France.

*These authors contributed equally to the manuscript.

†To whom correspondence should be addressed. E-mail: olivier.tenaillon@inserm.fr (O.T.); bgaut@uci.edu (B.S.G.)

The difference in convergence between point mutations (2.6%) and functional units (31.5%) suggests that we have not explored the diversity of possible adaptive mutations. To illustrate this result qualitatively, we plotted the number of different beneficial mutations at various levels (i.e., mutation, gene, operon, or functional unit) as a function of the number of sequenced lines (*14*). This exercise indicates that we are far from detecting all possible beneficial mutations (Fig. 2B). However, the discovery of affected genes, operons, and functional units was nearly saturated, which suggested that fewer replicates may have recovered the major targets of selection.

To estimate the number of sites that contribute to an adaptive response, we developed a simple model of mutation sampling analogous to the coupon collector's problem (*19*). Assuming that beneficial mutations are sampled from a set of L mutations, all with an equal mutation rate (μ) and selective coefficient (s), we fit the model to the saturation curves in Fig. 2B (*14*). For genes with >3 point mutations, we estimate that L = 850 possible sites of beneficial mutations are required to yield our 400 observed point mutations (Fig. 2C). L = 850 is a minimum, because our approach assumes no variance in μ and s among sites. With the addition of variance (*20*), the estimated number of sites increases, potentially reaching several thousand sites (Fig. 2C). We conclude that a large number of potentially beneficial sites are clustered within a few operational units. This was expected for the case of gene inactivation, for which different mutations lead to the same phenotype, but the diversity of possible solutions in essential functions, like RNApol, is more surprising (table S2).

Do interactions among beneficial mutations shape the adaptive trajectory? By examining all combinations of a small number of beneficial mutations, recent studies have demonstrated negative epistasis between beneficial mutations in different genes (*3*–*5*) and sign epistasis—in which a

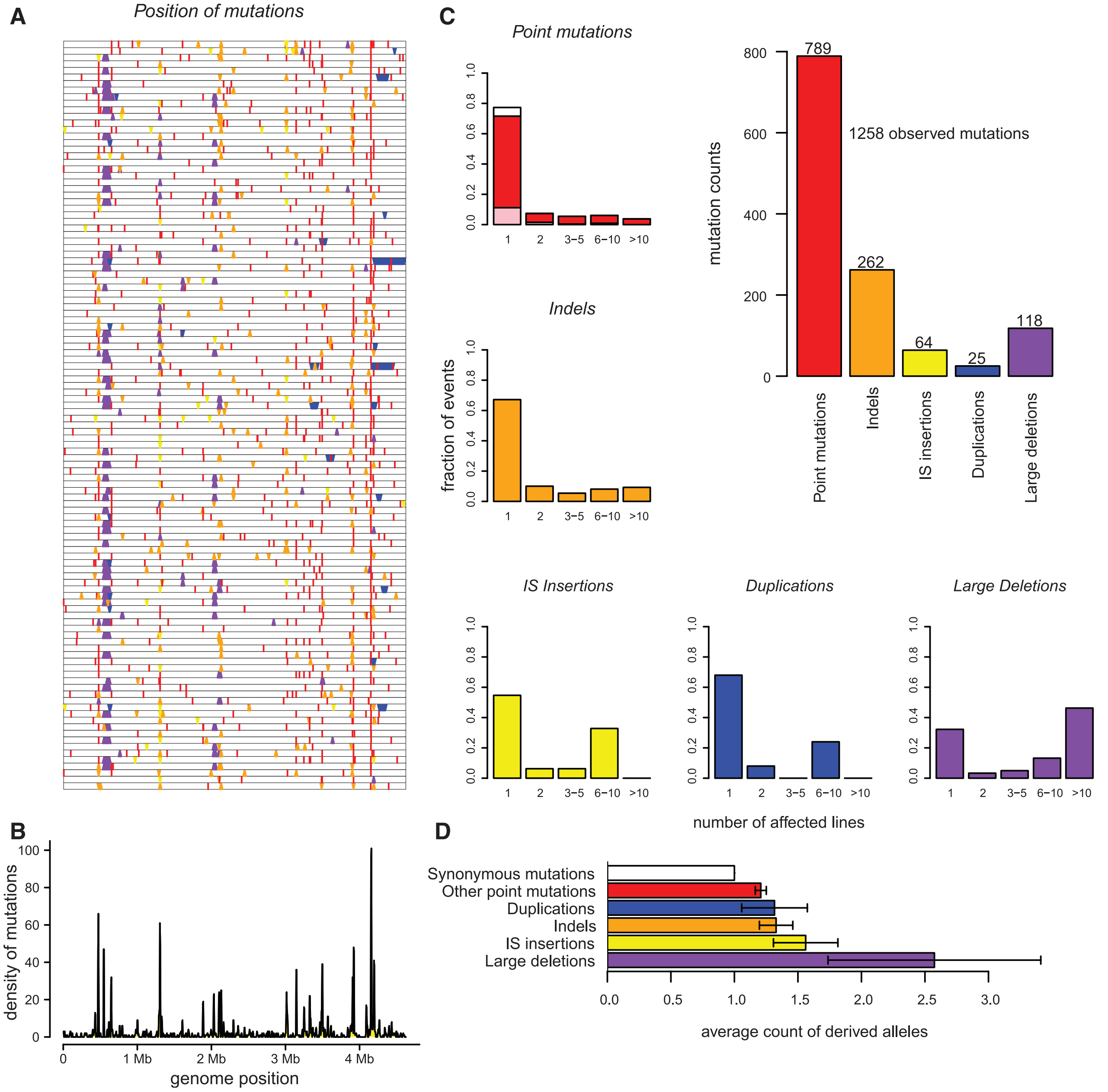

Fig. 1. (**A**) Mutations in 114 independently evolved clones represented along the *E. coli* B chromosome (*15*). Downward and upward triangles are insertions and deletions, respectively. Mutational types are colored as in (**C**). (**B**) The density of mutations along the genome in 5-kb sliding windows. (**C**) The distribution of events according to mutational type. Point mutations are split into nonsynonymous mutations (red), synonymous (white), and intergenic (orange). (**D**) The number of lines sharing mutational types (means ± SEM). All synonymous mutations were singletons.

mutation becomes either deleterious or beneficial depending on the genetic background—within a gene (*2*). We examined epistasis statistically using a resampling procedure (*14*) and also measured associations among mutations using *D*′ (*21*) and the correlation coefficient (*r*). We focused analyses on operational units with >25 mutations to have reasonable statistical power (Table 1).

Our data contain striking signals of associations. Within units, strains harbored dramatically fewer multiple hits than expected ($P < 1e^{-23}$ within genes and $P < 1e^{-14}$ within functional units); for

Table 1. Operational units with >25 mutational events. Class: Categories correspond to levels, where 1 is a mutation; 2 is the gene; 3 is the operon; 4 is a functional unit. Genes involved are separated by a line when physically separated on the chromosome. Mutations per gene are in parentheses. *P* value testing of whether the number of accumulations in the unit is random. Mutational types: NS, nonsynonymous mutations and short in-frame indels; inactivation, gene inactivations; Reg, mutations in putative regulatory regions or genes.

Unit	Class	Function	Mutations (no.)	Genes involved (no.)	Log (*P* value)	Mutational type
RNApol	4: Protein complex	Extended RNApol complex	205	*rpo*B (87) *rpo*C (21) *rpo*A (4) *rpo*D (36) *nus*A (12) *rho* (45)	−267	NS and Reg
YBAL	2	Putative CPA-2 proton antiporter	65	*yba*L (65)	−113	Inactivation
CLS	2	Cardiolipin synthetase	64	*cls* (64)	−114	Inactivation
ROD	4: Protein structure	Protein complex involved in the rod shape of the cell	54	*mrd*A (22) *mrd*B (8) *mre*B (12) *mre*C (5) *mre*D(5) *yfg*A (2)	−57	NS
DEL_1	1: Large deletion	Deletion of the region including fep, ent, and cus operons	45	From 547 to 619 kbp	NA	Large deletion bounded by IS elements
ICLR	2	Transcriptional repressor of the glyoxylate bypass operon	39	*icl*R	−68	Inactivation
GLG	3	Glycogen synthesis and degradation	36	*glg*A (2) *glg*C (6) *glg*P (28)	−36	NS
GLP	4: Metabolic and regulatory connection	Uptake of glycerol and *sn*-glycerol 3-phosphate from periplasm to cytoplasm and its regulation	36	*glp*E (2) *glp*G (2) *glp*R (9) *glp*F (13) *glp*K (2) *glp*T (8)	−34	Inactivation
FRU	4: Metabolic proximity	Fructose 1,6-bisphosphate synthesis and degradation	35	*fba*B (25) *fbp* (7) *fru*K (3)	−47	NS and Reg
KPS	4: Protein complex	Secretion complex of capsule	33	*kps*D (13) *kps*E (11) *kps*M (7) *kps*T (2)	−31	Inactivation
ILVL	2	Leader peptide regulating isoleucine and valine biosynthesis operons	29	*ilv*L (29)	−58	Reg
RSS	4: Regulatory connection	General stress response regulation	27	*csp*C (13) *rss*B (11) *rpo*S (3)	−34	NS and Reg

example, ICLR, CLS, *rho*, *rpo*D, KPS, YBAL, and GLP (defined in Table 1) never had >1 mutation within a single line (Fig. 3A). This pattern is explained by negative epistasis, for which the fixation of a single beneficial mutation transforms tens to hundreds of potentially adaptive mutations to be either neutral or deleterious (*3*). The cause is clear for gene inactivation: Once a gene like *yba*L is inactivated, little fitness benefit can be achieved by further mutations. However, we observed similar patterns for genes that retained function. For example, none of the six ROD genes were inactivated, but no genome harbored more than one mutation in this gene set ($P < 1e^{-3}$), even though ROD was a repeated target of mutation (Table 1).

The exception was RNApol, which accumulated >1 mutation within a single line (Fig. 3A). In fact, *rpo*B differed from all other genes in having two or more mutations in a single genome, occurring in 11 lines. Because the *rpo*BC operon is pleiotropic (*22*), the occurrence of multiple mutations is compatible with selection for compensatory mutations or for multiple phenotypes. Although a different set of mutations in RNApol were recovered after adaptation to 37°C in slightly different conditions (*22*), none were observed over >20,000 generations in the population from which our ancestor was derived (*8*), which suggested that our RNApol mutations are temperature-specific adaptations.

We also detected epistasis between units, including negative associations. For example, all 114 lines had at least one mutation in the *rpo*BC operon or one mutation in *rho* but these tended to be in repulsion (Fisher's exact test: $P < 3e^{-6}$; Fig. 3B). In fact, the two codons with the most mutations in the data set (17 mutations in *rho* codon 15 and 18 mutations in *rpo*B codon 966) were in complete repulsion (Fisher's exact test: $P < 0.01$ for *rpo*B Ile^{966}, and $P < 2e^{-}7$ for *rho* Ile^{15}). Similarly, a previous thermal adaptation experiment detected an early mutation in *rho* without ensuing *rpo*BC mutations (*12*). Other associations between units were consistent with positive epistasis. For example, mutations in *rpo*D (Fisher's exact test: $P < 1e^{-4}$), ILV ($P < 0.0004$), KPS ($P < 0.0001$), and RSS ($P < 0.002$) occurred only in an *rpo*BC mutant background, which suggested that the sign or amplitude of effects depends on the *rpo*BC background.

The overall pattern of associations suggests at least two competing evolutionary trajectories (Fig. 3B). In the first, mutations in *rpoBC* are in positive epistasis with changes in *rpo*D, ILV, KPS, RSS, and ROD. In the second, a mutation in *rho* deters the acquisition of mutations in *rpoBC* and favors selection for mutations in *cls* and *icl*R. There is no difference in average fitness between clones that traversed the different pathways (*t* test; $P = 0.96$), but the two trajectories suggest physiological interactions that have yet to be deciphered functionally. Moreover, these pathways suggest dynamic interactions between

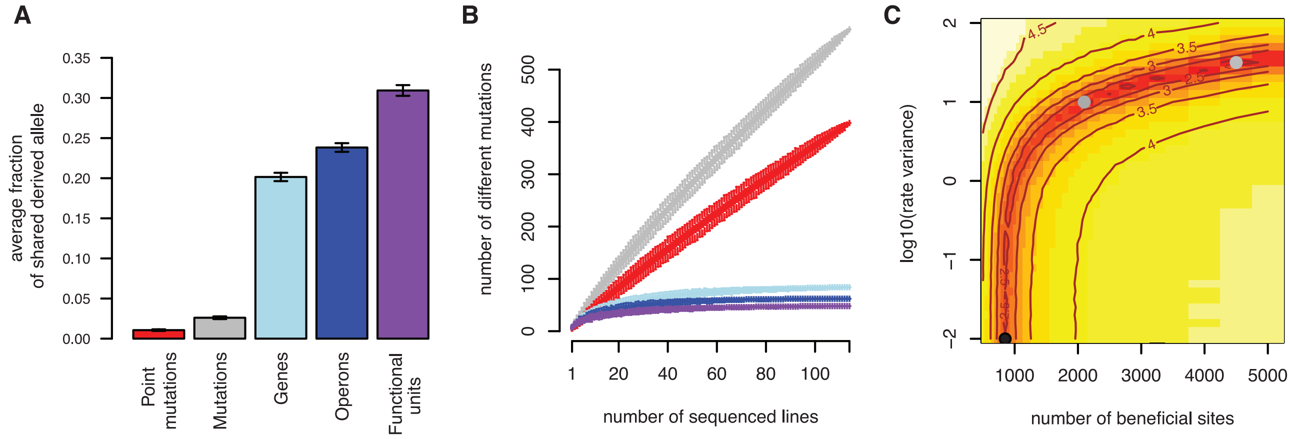

Fig. 2. (**A**) The pairwise fraction of shared events as a function of organizational level (means ± SEM). (**B**) The number of different mutations for subsamples of lines. The colors represent the levels of organization, as defined in (A). (**C**) The fit of the model used to estimate the number of potential beneficial sites *L* for point mutations; hotter colors represent a better fit to the data. The black dot represents the estimate ($L = 850$) under the coupon collector model; the gray dots represent solutions when the sampling probability of beneficial mutations varies among sites.

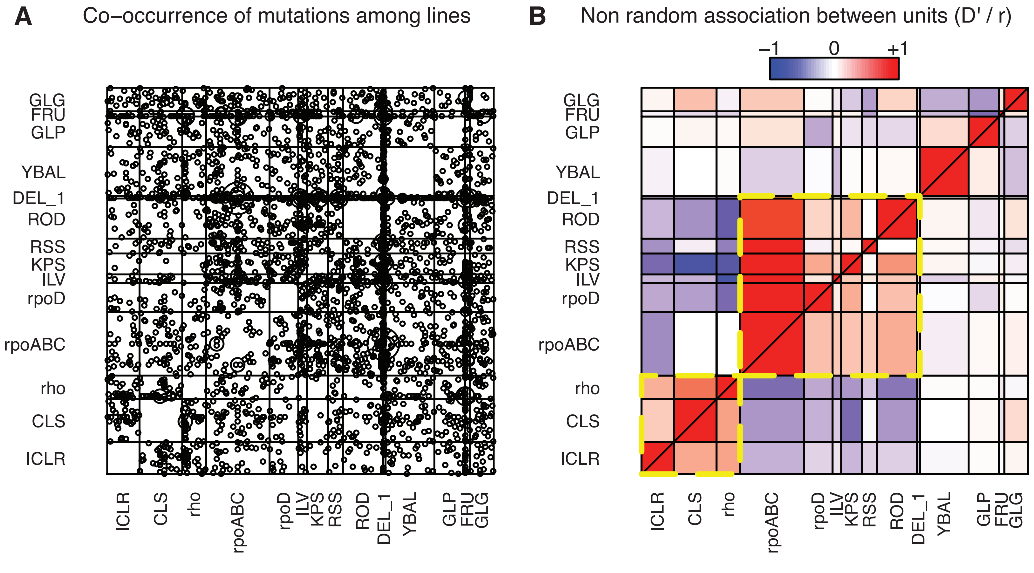

Fig. 3. (**A**) A plot of the co-occurrence of mutations among lines. The labels represent units defined in Table 1, except RNApol, which was split to emphasize unique patterns of accumulation. Each circle represents an instance of a mutation in the units on both the *y* and *x* axes; the size of the circle represents the number of lines sharing both mutations. Most diagonal units had no more than one mutation in a line, and so have no corresponding circles (e.g., YBAL). (**B**) Both *D*′ (top left) and the correlation coefficient, *r* (bottom right), measure associations between mutations. The color of the boxes corresponds to the scale, with hot colors representing positive associations and cooler colors negative associations. The broken yellow lines encompass two alternative evolutionary strategies.

a single beneficial mutation and entire "blocks" of potential beneficial mutations. In this model, one mutation affects its own block, so that potentially beneficial mutations usually become neutral or deleterious, and also opens new blocks to selection, which affects the amplitude and sign of mutational effects within the new block.

References and Notes

1. O. K. Silander, O. Tenaillon, L. Chao, *PLoS Biol.* **5**, e94 (2007).
2. D. M. Weinreich, N. F. Delaney, M. A. Depristo, D. L. Hartl, *Science* **312**, 111 (2006).
3. D. R. Rokyta *et al.*, *PLoS Genet.* **7**, e1002075 (2011).
4. A. I. Khan, D. M. Dinh, D. Schneider, R. E. Lenski, T. F. Cooper, *Science* **332**, 1193 (2011).
5. H. H. Chou, H. C. Chiu, N. F. Delaney, D. Segrè, C. J. Marx, *Science* **332**, 1190 (2011).
6. H. Teotónio, I. M. Chelo, M. Bradić, M. R. Rose, A. D. Long, *Nat. Genet.* **41**, 251 (2009).
7. R. E. Lenski, M. Rose, S. Simpson, S. Tadler, *Am. Nat.* **138**, 1315 (1991).
8. J. E. Barrick *et al.*, *Nature* **461**, 1243 (2009).
9. M. K. Burke *et al.*, *Nature* **467**, 587 (2010).
10. A. F. Bennett, K. M. Dao, R. E. Lenski, *Nature* **346**, 79 (1990).
11. M. M. Riehle, A. F. Bennett, A. D. Long, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 525 (2001).
12. T. Kishimoto *et al.*, *PLoS Genet.* **6**, e1001164 (2010).
13. A. Bennett, R. Lenski, J. Mittler, *Evolution* **46**, 16 (1992).
14. Materials and methods are available as supporting material on *Science* Online.
15. H. Jeong *et al.*, *J. Mol. Biol.* **394**, 644 (2009).
16. P. A. Christin, D. M. Weinreich, G. Besnard, *Trends Genet.* **26**, 400 (2010).
17. L. Chao, C. Vargas, B. B. Spear, E. C. Cox, *Nature* **303**, 633 (1983).
18. R. Woods, D. Schneider, C. L. Winkworth, M. A. Riley, R. E. Lenski, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 9107 (2006).
19. K. Read, *Am. Stat.* **52**, 175 (1998).
20. A. Poon, B. H. Davis, L. Chao, *Genetics* **170**, 1323 (2005).
21. R. C. Lewontin, *Genetics* **49**, 49 (1964).
22. T. M. Conrad *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 20500 (2010).

Acknowledgments: We thank L. Chao, C. Herman, and D. Schneider for discussion. This work was supported by NSF grant DEB-0748903 to A.F.B., A.D.L., and B.S.G. and by Agence Nationale de la Recherche, Programme Génomique, grant ANR-08-GENM-023-001 to O.T. Raw sequence data are available for download at http://wfitch.bio.uci.edu/~tdlong/Coli_EE_RawData.tgz.

Supporting Online Material
www.sciencemag.org/cgi/content/full/335/6067/457/DC1
Materials and Methods
Tables S1 to S3
References (*23*–*45*)

22 August 2011; accepted 2 November 2011
10.1126/science.1212986

Centrosome Loss in the Evolution of Planarians

Juliette Azimzadeh,[1]* Mei Lie Wong,[1] Diane Miller Downhour,[2]
Alejandro Sánchez Alvarado,[3] Wallace F. Marshall[1]*

The centrosome, a cytoplasmic organelle formed by cylinder-shaped centrioles surrounded by a microtubule-organizing matrix, is a hallmark of animal cells. The centrosome is conserved and essential for the development of all animal species described so far. Here, we show that planarians, and possibly other flatworms, lack centrosomes. In planarians, centrioles are only assembled in terminally differentiating ciliated cells through the acentriolar pathway to trigger the assembly of cilia. We identified a large set of conserved proteins required for centriole assembly in animals and note centrosome protein families that are missing from the planarian genome. Our study uncovers the molecular architecture and evolution of the animal centrosome and emphasizes the plasticity of animal cell biology and development.

The centrosome controls essential cellular processes such as cell division, migration, and polarity by anchoring microtubule-nucleating factors and cell cycle regulators. In animals, it also plays a crucial role during development (including organogenesis) through its ability to control nucleus positioning and cell division orientation, and by nucleating primary cilia. The centrosome is found in all animals examined and is essential to the development of all animal species studied so far. However, centrosomes are dispensable for cell division in some instances, in particular during mouse early embryogenesis and the later stages of *Drosophila* development, raising the question of how central the centrosome really is to fundamental cell biology and development (*1*, *2*).

Animal centrosome duplication relies on duplication of its core components, the centrioles, once per cell cycle. A second pathway, called acentriolar pathway, allows assembly of large numbers of centrioles in multiciliated cells undergoing terminal differentiation (*3*). In vertebrates,

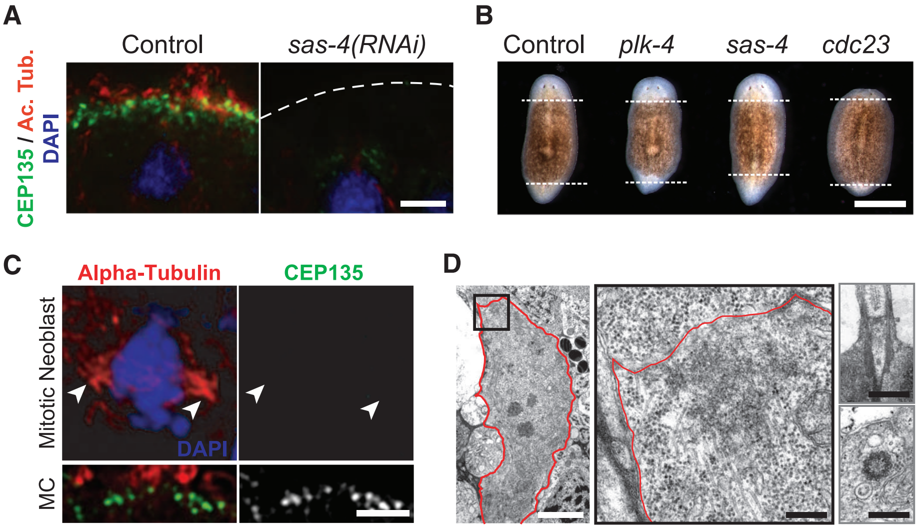

Fig. 1. Centrioles in planarians are present in multiciliated cells but not proliferating cells. (**A**) Immunofluorescence staining of ventral epidermis in control or *sas-4(RNAi)* animals; centrioles (green, anti–SMED-CEP135), cilia (red, anti–acetylated tubulin), nuclei [blue, DAPI (4′,6-diamidino-2-phenylindole)]. Dashed line: apical cell boundary. (**B**) Regenerating trunk fragments. Dashed lines: boundary between newly formed and preexisting tissues. (**C**) Immunofluorescence staining of control mitotic neoblast; centrioles (green, anti–SMED-CEP135), mitotic spindle (red, anti–alpha-tubulin), chromosomes (blue, DAPI). Arrowheads: spindle poles. Lower panel: centrioles in a multiciliated cell (MC) from the same section. (**D**) Left: TEM view of control mitotic neoblast (3200×). Red line: plasma membrane. Middle: high-magnification view (15,000×) of boxed area. Right: High-magnification views (21,000×) of centrioles from the ventral multiciliated epidermis are shown for comparison. Top: longitudinal section through a centriole and the base of the corresponding cilium; bottom: cross section. Bar: 5 μm in (A), (C), and (D) (left panel); 1 mm in (B); and 0.25 μm in (D) (middle and right panels).

---

[1]Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143, USA. [2]University of Utah School of Medicine, Department of Genetics, Salt Lake City, UT 84132–3401, USA. [3]Stowers Institute for Medical Research and Howard Hughes Medical Institute, Kansas City, MO 64110, USA.

*To whom correspondence should be addressed. E-mail: juliette.azimzadeh@ucsf.edu (J.A.); wallace.marshall@ucsf.edu (W.F.M.)