Speciation

Table 1. A proposed strategy for developing and validating models of bacterial evolution that might eventually be used to classify genetic diversity data and provide a firm foundation for a bacterial species concept.

1. Collect samples according to systematic ecological stratification. Focus on longitudinal studies, geographical studies, and measurement of physical and chemical gradients affecting bacterial growth. Consider biotic factors such as the presence of other competing bacteria or parasitic phage.
2. For each isolate, sequence as much as possible and affordable (16*S* rRNA, MLSA, auxiliary genes, full genomes, etc.).
3. Use empirical classification algorithms that use genetic and ecological data to jointly map isolates.
4. To guide model formulation, use population genetic tests on observed clusters, focusing on tests for selection, population structure, and gene flow.
5. Generate evolutionary models and simulate populations.
6. Test, then reject or adapt, evolutionary models according to agreement between simulations and real populations; if necessary, return to step 1.
7. For successful models, develop model-based methods for interpreting pure genetic data (without ecological covariates) and test on new data.
8. If one or more validated models emerge, use these to classify genetic data and to develop bacterial species concepts.

mental bacteria will be useful in dissecting the roles of the auxiliary and core genome in ecological differentiation. If after this process it emerges that some model or models are consistently validated for different study systems, these would inevitably form a good basis for identifying fundamental levels of clustering, or species.

In the foregoing we have emphasized ecotype and metapopulation models, but there are others that deserve consideration—notably the epidemic clonal model (*42*) and the impact of phage epidemics causing classic Lotka-Volterra boom-bust dynamics (*43*) illustrated in Fig. 2D—and it is possible, even likely, that more than one of these mechanisms may be relevant to any given problem in speciation and cluster formation. Distinguishing among these mechanisms is the bacterial species challenge (Table 1), described in 1991 by John Maynard Smith as follows: "Ecotypic structure, hitch-hiking, and localized recombination can explain the observed patterns of variation. The difficulty, of course, is that the model is sufficiently flexible to explain almost anything. To test the hypothesis of ecotypic structure, we need to know the distribution of electrophoretic types [i.e., genotypes] in different habitats" (*17*).

Much research on bacterial species to date has come from studies on pathogens, where the correct identification of species is crucial for accurate clinical diagnoses. However, for pathogens the identification of the multiple ecological niches within (for example) the nasopharynx or gut is difficult, and studies of the relationships between bacterial populations and ecology may be more fruitful for some environmental species where the categorization of niches is a more tractable enterprise. Hopefully, we will soon obtain richer data sets that map bacterial diversity onto ecology and provide a way to distinguish among various models of population differentiation and speciation, including those based on ecotypes or metapopulations.

References and Notes

1. J. Handelsman, *Microbiol. Mol. Biol. Rev.* **68**, 669 (2004).
2. W. P. Hanage, C. Fraser, B. G. Spratt, *Philos. Trans. R. Soc. London Ser. B* **361**, 1917 (2006).
3. M. Achtman, M. Wagner, *Nat. Rev. Microbiol.* **6**, 431 (2008).
4. F. M. Cohan, *Annu. Rev. Microbiol.* **56**, 457 (2002).
5. W. F. Doolittle, R. T. Papke, *Genome Biol.* **7**, 116 (2006).
6. D. Gevers et al., *Nat. Rev. Microbiol.* **3**, 733 (2005).
7. M. F. Polz, D. E. Hunt, S. P. Preheim, D. M. Weinreich, *Philos. Trans. R. Soc. London Ser. B* **361**, 2009 (2006).
8. D. B. Rusch et al., *PLoS Biol.* **5**, e77 (2007).
9. C. Darwin, *The Origin of Species* (Penguin Classics, London, 1985).
10. E. Stackebrandt et al., *Int. J. Syst. Evol. Microbiol.* **52**, 1043 (2002).
11. R. Facklam, *Clin. Microbiol. Rev.* **15**, 613 (2002).
12. J. C. Arbique et al., *J. Clin. Microbiol.* **42**, 4686 (2004).
13. D. E. Hunt et al., *Science* **320**, 1081 (2008).
14. See supplementary information in (*16*) for a statistical comparison of the ecotype model with an effective neutral model and an implicit estimate of N_e.
15. C. Fraser, W. P. Hanage, B. G. Spratt, *Science* **315**, 476 (2007).
16. A. Koeppel et al., *Proc. Natl. Acad. Sci. U.S.A.* **105**, 2504 (2008).
17. J. Maynard Smith, *Proc. R. Soc. London Ser. B* **245**, 37 (1991).
18. H. Ochman, A. C. Wilson, in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, F. C. Neidhart, Ed. (ASM Press, Washington, DC, 1987), pp. 1649–1654.
19. J. R. Thompson et al., *Appl. Environ. Microbiol.* **70**, 4103 (2004).
20. M. Kimura, *Trends Biochem. Sci.* **1**, N152 (1976).
21. R. Milkman, *Trends Biochem. Sci.* **1**, N152 (1976).
22. K. C. Atwood, L. K. Schneider, F. J. Ryan, *Proc. Natl. Acad. Sci. U.S.A.* **37**, 146 (1951).
23. B. R. Levin, *Genetics* **99**, 1 (1981).
24. F. M. Cohan, E. B. Perry, *Curr. Biol.* **17**, R373 (2007).
25. A. Rambaut et al., *Nature* **453**, 615 (2008).
26. R. A. Fisher, E. B. Ford, *Heredity* **1**, 143 (1947).
27. M. Slatkin, *Theor. Popul. Biol.* **12**, 253 (1977).
28. J. Majewski, F. M. Cohan, *Genetics* **152**, 1459 (1999).
29. E. J. Feil, B. G. Spratt, *Annu. Rev. Microbiol.* **55**, 561 (2001).
30. J. Majewski, F. M. Cohan, *Genetics* **153**, 1525 (1999).
31. J. Majewski, P. Zawadzki, P. Pickerill, F. M. Cohan, C. G. Dowson, *J. Bacteriol.* **182**, 1016 (2000).
32. S. T. Cowan, in *Microbial Classification, 12th Symposium of the Society for General Microbiology*, G. C. Ainsworth, P. H. A. Sneath, Eds. (Cambridge Univ. Press, Cambridge, 1962), pp. 433–455.
33. S. K. Sheppard, N. D. McCarthy, D. Falush, M. C. J. Maiden, *Science* **320**, 237 (2008).
34. A. C. Retchless, J. G. Lawrence, *Science* **317**, 1093 (2007).
35. K. Vetsigian, N. Goldenfeld, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 7332 (2005).
36. A. Tuanyok et al., *BMC Genomics* **9**, 566 (2008).
37. U. Dobrindt, B. Hochhut, U. Hentschel, J. Hacker, *Nat. Rev. Microbiol.* **2**, 414 (2004).
38. P. Wilmes, S. L. Simmons, V. J. Denef, J. F. Banfield, *FEMS Microbiol. Rev.* **33**, 109 (2009).
39. V. J. Denef et al., *Environ. Microbiol.* **11**, 313 (2008).
40. E. Bantinaki et al., *Genetics* **176**, 441 (2007).
41. R. A. Welch et al., *Proc. Natl. Acad. Sci. U.S.A.* **99**, 17020 (2002).
42. J. Maynard Smith, N. H. Smith, M. O'Rourke, B. G. Spratt, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 4384 (1993).
43. K. H. Hoffmann et al., *FEMS Microbiol. Lett.* **273**, 224 (2007).
44. We thank T. Connor and S. Deeny for useful discussions. Supported by University Research Fellowships from the Royal Society (C.F. and W.P.H.), a program grant from the Wellcome Trust (B.G.S.), grants from the U.S. Department of Energy Genomes to Life program (M.F.P. and E.J.A.), and the NSF/National Institute of Environmental Health Sciences Woods Hole Centre for Oceans and Human Health, the NSF Biological Oceanography Program, and the Moore Foundation (M.F.P.).

REVIEW

Is Genetic Evolution Predictable?

David L. Stern[1]* and Virginie Orgogozo[2]*

Ever since the integration of Mendelian genetics into evolutionary biology in the early 20th century, evolutionary geneticists have for the most part treated genes and mutations as generic entities. However, recent observations indicate that all genes are not equal in the eyes of evolution. Evolutionarily relevant mutations tend to accumulate in hotspot genes and at specific positions within genes. Genetic evolution is constrained by gene function, the structure of genetic networks, and population biology. The genetic basis of evolution may be predictable to some extent, and further understanding of this predictability requires incorporation of the specific functions and characteristics of genes into evolutionary theory.

One hundred and fifty years ago, Charles Darwin and Alfred Russell Wallace proposed that biological diversity results from natural selection acting on heritable variation in populations. Both Darwin and Wallace recognized the importance of heritable variation to evolutionary theory, but neither man knew the true cause of inheritance. Early in the 20th cen-

tury, the rediscovery of Mendel's studies allowed for a formal mathematical treatment of alleles in populations, generating the field of population genetics. Population geneticists treated genes and alleles as generic entities, particles that were inherited and somehow caused variation in the appearance, behavior, and physiology of organisms—what we call collectively the phenotype. This level of abstraction was appropriate given that a molecular understanding of gene function lay many decades in the future. Even with this rudimentary view of gene function, however, population genetics greatly clarified how real populations evolve, and this theoretical understanding spurred the New Synthesis, combining population genetics with ecology, systematics, and biogeography to explain and explore many questions in evolution.

In the past 40 years, molecular biologists have elucidated how genes regulate biological processes, but only the most basic mechanistic observations have been integrated into evolutionary biology. For example, evolutionary theory has effectively absorbed the distinction between coding (nonsynonymous) and silent (synonymous) substitutions in protein-coding regions, but other aspects of molecular biology currently contribute little to evolutionary thought. The time has now come to integrate the specifics of molecular and developmental biology into evolutionary biology. Over the past 15 years, many examples of the genes and mutations causing evolutionary change have been identified (*1*). Patterns in these data suggest that a synthesis of molecular developmental biology with evolutionary theory will reveal new general principles of genetic evolution.

Nonrandom Distribution of Evolutionarily Relevant Mutations

Recent studies suggest that the mutations contributing to phenotypic variation [evolutionarily relevant mutations (*2*)] are not distributed randomly across all genetic regions. The most compelling evidence comes from cases of parallel genetic evolution: the independent evolution of similar phenotypic changes in different species due to changes in homologous genes or sometimes in the same amino acid position of homologous genes.

Many cases of parallel evolution have been discovered across all of the kingdoms. At least 20 separate populations of the plant *Arabidopsis thaliana* have evolved null coding mutations (mutations that completely eliminate protein func-

tion) in the *Frigida* gene that cause early flowering (*3*). Resistance to DDT and pyrethroids has evolved in 11 insect species by mutations in either amino acid Leu^{1014} or Thr^{929} of the voltage-gated sodium channel gene *para* (*4*). Two virus populations independently subjected to experimental evolution in a novel host accumulated many of the same amino acid mutations (*5*). In total, about 350 evolutionarily relevant mutations have been found in plants and animals, and more than half of these represent cases of parallel genetic evolution (*1*).

One explanation for parallel genetic evolution is that most genes play specialized roles during development, and only some genes can evolve to generate particular phenotypic variants. For example, mutations in *rhodopsin* can alter light-wavelength sensitivity (*6*), and mutations in *lysozyme* may enhance enzyme activity at the particular pH of a fermenting gut (*7*). But the reverse would not be true. Mutations in *rhodopsin* are unlikely to enhance fermentation, and mutations in a digestive enzyme will not aid detection of a particular wavelength of light, even if each protein was expressed in the reciprocal organ.

Gene function explains part but not all of the observed pattern of parallel genetic evolution. In several cases, parallelism has been observed even though mutations in a large number of genes can produce similar phenotypic changes. For example, although more than 80 genes regulate flowering time (*8*), changes in only a subset of these genes have produced evolutionary changes in flowering time (*3*). Hundreds of genes regulate the pattern of fine epidermal projections, called trichomes, on *Drosophila melanogaster* larvae. But only one gene, called *shavenbaby*, has evolved to alter larval trichome patterns between *Drosophila* species, and this gene has accumulated multiple evolutionarily relevant mutations (*9*). What is special about these hotspot genes?

Developmental biology illuminates why hotspot genes such as *shavenbaby* exist. During development, multiple cell-signaling pathways and transcription factors act together to progressively divide the embryo into a virtual map that specifies when and where organs will form. The interactions between the genes encoding these signaling molecules and transcription factors can be represented as a genetic network. Gene interactions are modulated in large part by the cis-regulatory regions of patterning genes. (All genes are composed of two fundamentally different regions: a region encoding the gene product—a protein or an RNA—and adjacent *cis*-regulatory DNA that encodes the instructions governing when and where the gene product will be produced.) Transcription factors bind to *cis*-regulatory regions of target genes, and the summed effect of many such interactions at a target gene determines whether the gene is expressed or not. Patterning genes act within complex genetic net-

works, and usually each patterning gene contributes to the development of multiple cell types. For example, most patterning genes that are active during embryonic development of the epidermis contribute to the development of muscle-attachment sites, sensory organs, tracheal pits, trichomes, or other cell types.

The importance of regulatory networks in determining which genes may be evolutionary hotspots can be illustrated with the genetic network that governs larval trichome development in *D. melanogaster* (Fig. 1). In this network, developmental patterning genes first collaborate to divide the embryonic epidermis into domains expressing distinct transcription factors. These patterning genes then regulate the expression of the *shavenbaby* gene, a so-called input-output gene (*10*). Input-output genes integrate complex spatiotemporal information (the input) and trigger development of an entire program of cell differentiation (the output). The Shavenbaby protein activates expression of a battery of target genes that transform an epidermal cell into a trichome cell. Each target gene triggers a specific aspect of cell differentiation, and production of a differentiated trichome requires coordinated expression of all target genes. The pattern of trichomes over the body is thus determined by the distribution of Shavenbaby protein in the epidermis, which is controlled by the *cis*-regulatory region of the *shavenbaby* gene. The *shavenbaby* gene serves as a nexus for patterning information flowing in and for cell-fate information flowing out.

In the entire regulatory network governing development of the *Drosophila* embryo, only *shavenbaby*, with its specialized function to rally the entire module of trichome morphogenesis, can accumulate mutations that alter trichome patterns without disrupting other developmental processes. Genetic changes in upstream developmental genes will alter trichome production, but these mutations also disrupt other organs. Changes in any one of the downstream genes are not sufficient to create or eliminate a trichome; concerted changes in multiple downstream genes are required to build a trichome (*11*). Furthermore, all of the evolutionarily relevant mutations in *shavenbaby* that have been identified so far alter the *cis*-regulatory region and not the protein-coding region. Mutations in the protein-coding region would alter *shavenbaby* function in every cell that accumulates Shavenbaby protein, and this would alter every trichome produced in larvae and adults. Thus, a developmental perspective clarifies why *shavenbaby* is a hotspot for evolutionarily relevant mutations and why these mutations occur in the *cis*-regulatory region of the gene. We predict that the *cis*-regulatory regions of other input-output genes may be hotspots for other phenotypic characteristics.

The *shavenbaby* gene provides one example of a more general principle: that mutations af-

[1]Department of Ecology and Evolutionary Biology, Howard Hughes Medical Institute, Princeton University, Princeton, NJ 08544, USA. [2]CNRS, Université Pierre et Marie Curie, Bâtiment A, 5ème Étage, Case 29, 7 Quai Saint Bernard, 75005 Paris, France.

*To whom correspondence should be addressed. E-mail: dstern@princeton.edu (D.L.S.); virginie.orgogozo@normalesup.org (V.O.)
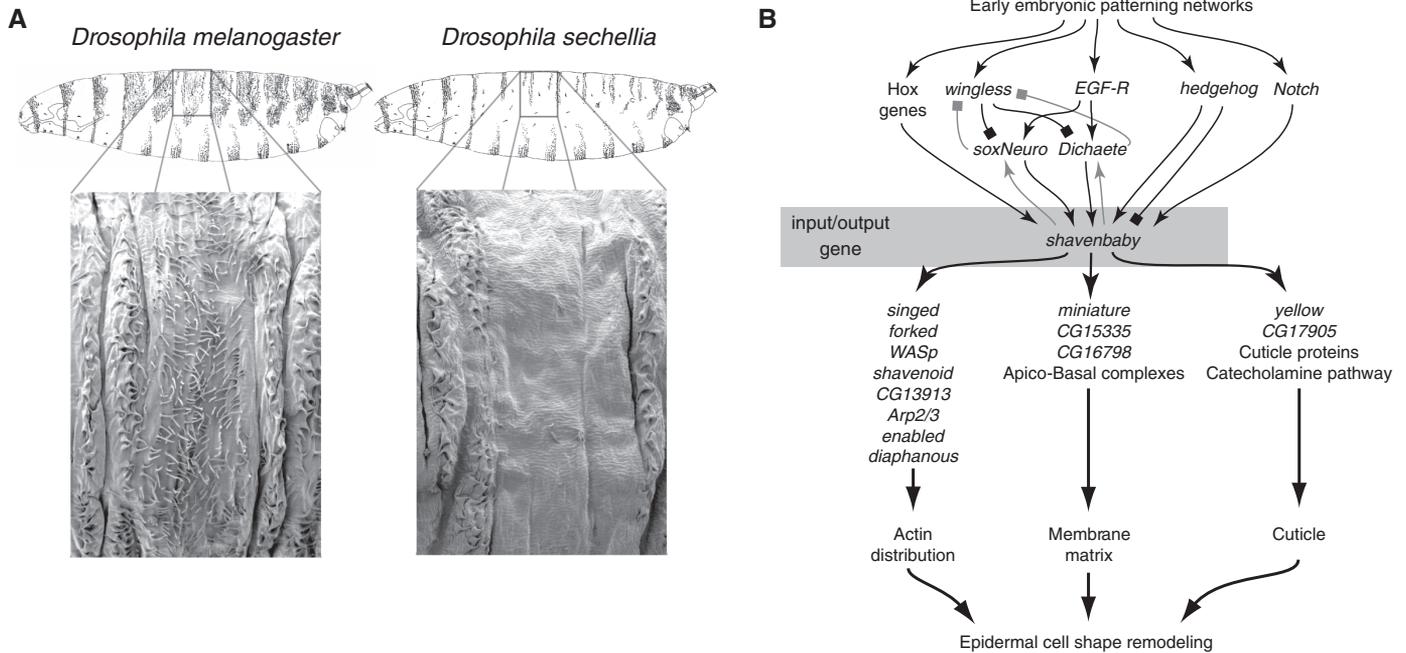
Speciation

A

Drosophila melanogaster   *Drosophila sechellia*

B

Fig. 1. Morphological divergence between species has been caused by repeated evolution at an input-output gene. (**A**) *D. melanogaster* and *D. sechellia* differ in the pattern of fine trichomes decorating the dorsal and lateral surfaces of the larvae. This difference is caused entirely by evolution of the *cis*-regulatory region of the *shavenbaby* gene (*9*). (**B**) The *cis*-regulatory region of the *shavenbaby* gene integrates extensive information from developmental patterning genes to generate a pattern of Shavenbaby protein expression that prefigures the pattern of trichomes on the first-instar larva. Cells accumulating Shavenbaby will differentiate a trichome because Shavenbaby protein regulates a large battery of genes that act together to transform an epithelial cell into a trichome (*11*).

fecting multiple phenotypic traits, so-called pleiotropic mutations, are unlikely to contribute to adaptive evolution. As we discuss next, pleiotropy and other genetic and population-genetic parameters seem to influence the distribution of evolutionarily relevant mutations.

The Factors Influencing the Distribution of Evolutionarily Relevant Mutations

Pleiotropy. Some mutations generate specific phenotypic changes, whereas pleiotropic mutations alter several seemingly unrelated traits. Two mutations that cause evolutionary increases in the number of thoracic bristles in *Drosophila* illustrate the difference between mutations with specific and pleiotropic effects (Fig. 2). A *cis*-regulatory change in the *scute* gene affects the number of sensory organs only on the thorax (*12*), whereas a coding mutation in the *poils au dos* gene increases the number of sensory organs on both the thorax and the wing (*13*). The *poils au dos* mutation is more pleiotropic than the *scute* mutation. *Scute*, like *shavenbaby*, is an input-output gene, whereas *poils au dos* is a patterning gene that, together with other pattering genes, regulates *scute* expression (Fig. 2). Mutations with pleiotropic effects will rarely change all phenotypic traits in a favorable way, and experimental evidence indicates that pleiotropic effects tend to reduce fitness (*14*). Selection may favor extra bristles on the thorax, but not extrasensory organs on the wing. Even if

one effect of a pleiotropic mutation provides a major improvement in fitness, the other effects may be deleterious and will reduce the likelihood that the mutation will become established in the population (*15*).

Epistasis. When examined in a single genetic background, a mutation may have a specific or a pleiotropic effect. But in another genetic background, the same mutation may produce a different phenotypic effect because of nonadditive interactions of alleles: so-called epistasis. For example, one allele in *A. thaliana* increases growth in one genetic background but reduces growth in a different genetic background (*16*). The second genetic background is not simply deleterious in general because a variant allele at a second locus causes higher growth in this background. Thus, the effects of one mutation can depend on the genetic variation present at other loci.

Epistasis is extremely common in natural populations and it may sometimes reduce the rate of evolution (*17*). Epistasis increases the phenotypic variance associated with a particular mutation, causing a mutation to have a fluctuating fitness effect dependent on the genetic background. Thus, in an *Arabidopsis* population containing multiple genetic backgrounds, we expect that selection for increased size will tend to favor nonepistatic alleles that increase growth in all backgrounds rather than epistatic alleles that increase growth in only one genetic background.

Plasticity. Populations exposed to repeated environmental changes may evolve genetic mechanisms that produce different phenotypes suited to different environmental conditions: so-called phenotypic plasticity. For example, aphids can produce multiple phenotypic forms in response to environmental conditions, including asexual forms that reproduce quickly and sexual forms that lay overwintering eggs. Mutations that eliminate sexual forms—that reduce plasticity—may provide a lineage with a short-term advantage, a much faster reproductive rate. But in the long term, aphid lineages that do not produce sexual forms tend to go extinct, perhaps because they fail to adapt to changing environmental conditions.

Similarly, in *A. thaliana* the *Frigida* gene controls plasticity for flowering time. *Frigida* responds to cold temperatures to induce flowering. In regions with warm winters, null *Frigida* mutations may provide a short-term advantage by consistently triggering flowering, even in the absence of a cold winter. But these mutations eliminate plasticity for flowering time, possibly preventing these plants from adapting to colder temperatures or from recolonizing areas in colder climates. Thus, the abundance of null *Frigida* mutations in *Arabidopsis* populations must result from factors that override the negative consequences of reduced plasticity.

Strength of selection. When an environmental change favors a phenotype that is vastly different

from the mean phenotype in a population, mutations causing large phenotypic changes toward the new optimum will be favored, at least initially (*18*). For example, recently domesticated races have probably experienced strong selection by farmers, and many recent domestication traits result from mutations that cause large phenotypic effects, including pleiotropic deleterious effects. As an example, six different null-coding mutations in the *myostatin* gene cause muscle hypertrophy in different breeds of cattle (*19*). *Myostatin* is a member of the *transforming growth factor*–β superfamily of growth factors and acts as a negative regulator of muscle development. Although null mutations of *myostatin* generate cattle with more and leaner meat, these cattle experience difficulties in calving and have reduced stress tolerance. Strong selection during domestication can obviously overcome the negative pleiotropic effects of null *myostatin* mutations.

Population history. The past and current sizes of a population also influence genetic evolution. Small population size increases the effects of random sampling of alleles, so-called genetic drift. In small populations, genetic drift will allow deleterious alleles to occasionally increase in frequency. For example, a small inbred population of Bedouins in Israel has evolved a high frequency of a recessive allele that causes deafness (*20*). With stronger genetic drift in small populations, natural selection will fail to promote the spread of adaptive mutations of small effect. Instead, in comparison with large populations, adaptive mutations of relatively large effect will tend to evolve by natural selection in small populations.

Small populations also have another critical effect on evolution: They limit the total number of new mutations introduced into the population each generation. Thus, small populations may end up selecting far-from-ideal mutations (those with pleiotropic consequences and epistatic effects) simply because potentially superior mutations occur at a lower rate.

The abundance of null *Frigida* mutations in populations of *A. thaliana* highlights the importance of population history in genetic evolution. Null *Frigida* mutations have the negative consequence of reducing
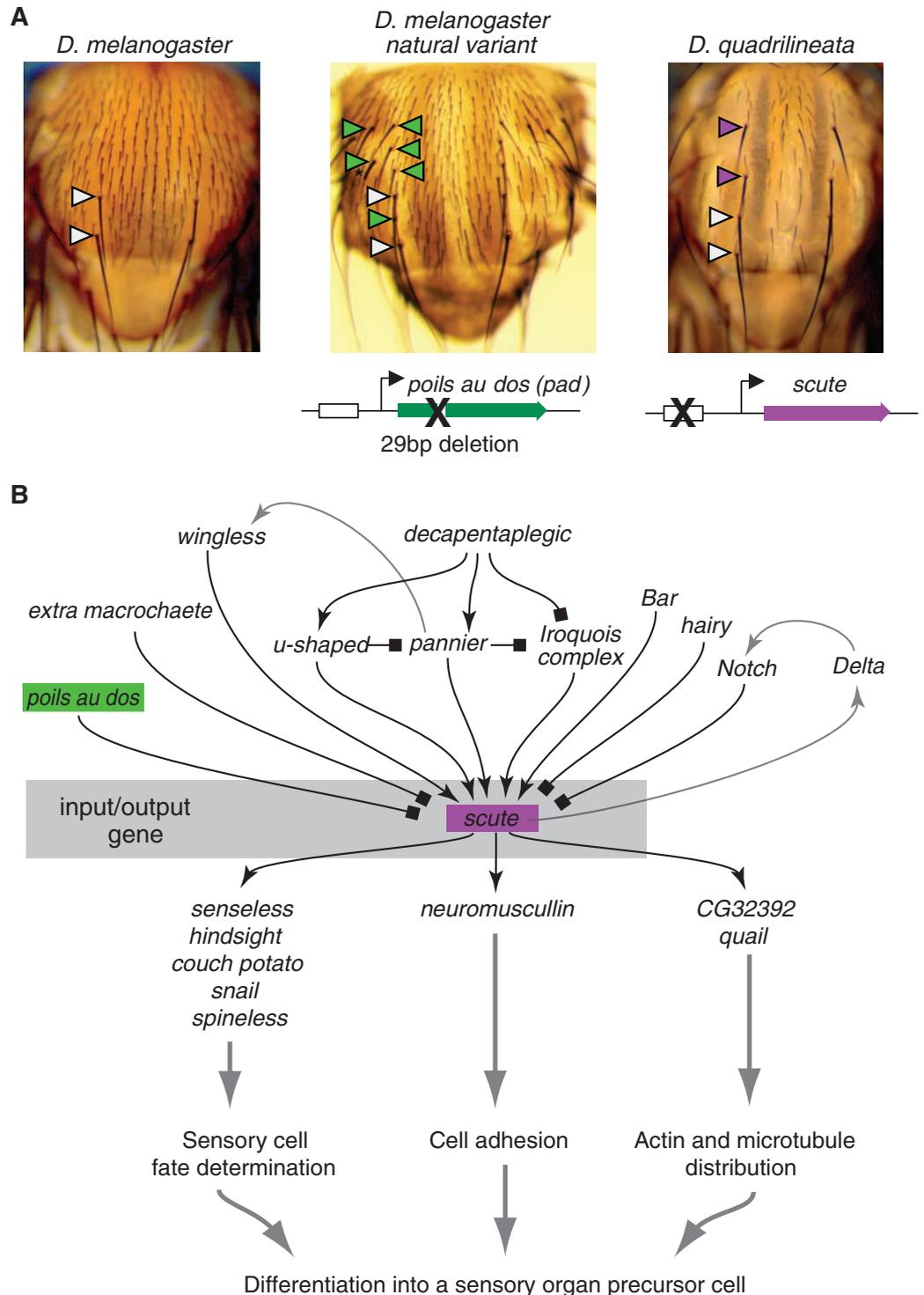

Fig. 2. Bristle patterns on the dorsal thorax of *Drosophila* species have evolved within species and between species because of different kinds of mutations. (**A**) A mutation generating a null allele of the *poils au dos* gene within a population of *D. melanogaster* increases the number of large bristles on the dorsal thorax (white triangles indicate normal bristles and green triangles indicate extra bristles) (*13*). In contrast, the increased number of bristles in *D. quadrilineata* results at least in part from changes in the *cis*-regulatory region of the *scute* gene (*12*). The extra bristles caused by the *poils au dos* mutation are not as precisely positioned as the extra bristles caused by the *scute* mutation (indicated by purple triangles). (**B**) The two evolving genes, *poils au dos* and *scute*, occupy different locations in the genetic network that generates the pattern of bristles. The *scute* gene is an input-output gene, whereas the *poils au dos* gene is a developmental patterning gene. The null mutation in *poils au dos* increases sensory organ numbers not only in the thorax but also in the wing.

plasticity for flowering time. These mutations also have pleiotropic effects [they reduce fruit production (*21*)] and display epistasis with respect to other genes that control flowering time (*22*). These observations suggest that null *Frigida* mutations are not ideal alleles for controlling flowering time. In fact, null *Frigida* mutations must only rarely, if ever, be involved in phenotypic divergence between species because homologs of the *Frigida* gene exist in diverse plant species. But natural selection has overcome the deleterious effects of null *Frigida* mutations to promote the spread of these mutations in small populations. *A. thaliana* has migrated from Scandinavia around the world in the footsteps of agriculture. Subpopulations have adapted to local conditions, including the relatively warm and short winters of more temperate regions. *A. thaliana* plants are self-fertile, so even a single plant can give rise to a new population. These small subpopulations provide fewer opportunities for beneficial mutations of specific effect to appear, and strong selection for rapid flowering has favored whatever mutations of strong effect arose in the population, such as null *Frigida* mutations. The abundance of null *Frigida* mutations probably reflects the fact that these mutations occur at a higher rate than mutations without associated deleterious consequences.

The Genetic Basis of Short-Term and Long-Term Evolution

The *Frigida* example is not unique. In many plants and animals, evolution over long periods (variation between species) appears to differ in several ways from evolution over shorter periods (variation between domesticated races and between individuals within a species) (*1*). Here are three general ways in which long-term and short-term genetic evolution differ.

First, epistasis is commonly found for the mutations that contribute to phenotypic variation within species, whereas it is rarely observed for the mutations that cause differences between species. Within *D. melanogaster*, variation in bristle number is caused by multiple loci of relatively small effect, and these loci have epistatic effects of the same order of magnitude as the additive effects (*23*, *24*). In contrast, morphological differences between *Drosophila* species result from multiple loci of intermediate-to-small effect that only rarely show epistasis (*25*, *26*). Studies of body size variation in chickens show a similar pattern,

with alleles segregating within species showing more epistasis than alleles differentiating species (*27*, *28*).

Second, null mutations, which arise frequently and often cause pleiotropic and epistatic effects, seem to contribute more to phenotypic variation within species than to phenotypic differences between species. About 55% of the 99 mutations known to cause domestication traits are null-coding mutations, whereas only 7% of the 75 mutations known to cause interspecific differences are null-coding mutations (Fig. 3). For example, although domesticated cattle stocks have evolved multiple null muta-
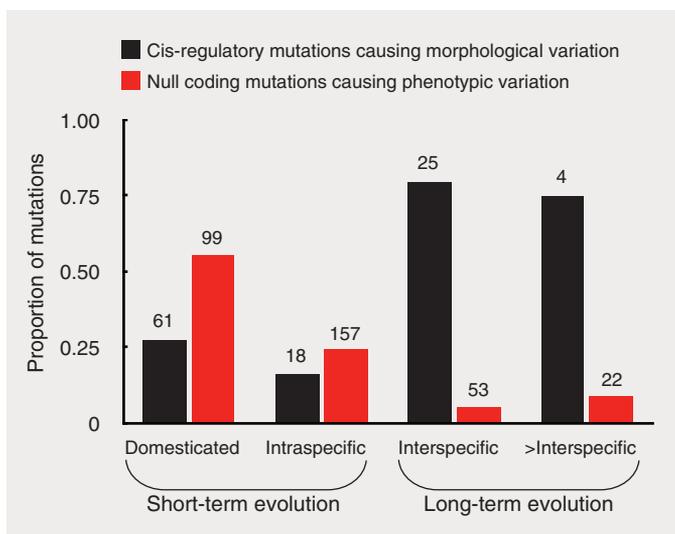

Fig. 3. Different kinds of mutations occur with different frequency during short-term and long-term evolution. Among all mutations causing morphological variation identified to date, the proportion of *cis*-regulatory mutations (black bars) is higher for long-term evolution than for short-term evolution. For all mutations that have been reported to cause phenotypic variation in either morphology or physiology, the proportion caused by null coding mutations (red bars) is higher for short-term evolution than for long-term evolution. The numbers above the bars refer to the total number of examples in each category. The number of cases of morphological evolution (black bars) is a subset of the number of cases of phenotypic evolution (red bars). Data are from (*1*).

tions of the *myostatin* gene, all mammal species investigated so far possess a functional *myostatin* gene.

Third, the frequency of *cis*-regulatory mutations causing morphological variation differs between taxonomic levels. Morphological changes may occur either through coding changes or through *cis*-regulatory changes (Fig. 2). Because mutations in *cis*-regulatory regions often have fewer pleiotropic effects than mutations in coding regions, morphological changes are expected to involve mainly *cis*-regulatory mutations (*1*, *2*, *29*). Within species, most mutations that cause morphological variation have been found in protein-coding regions (Fig. 3). In contrast, between species most mutations that

cause morphological differences have been found in *cis*-regulatory regions. Presumably, many of the coding mutations found within species fail to spread through populations, perhaps because of pleiotropic deleterious effects.

These striking and unexpected differences between short-term and long-term genetic evolution have emerged only recently with the accumulation of a sufficient number of case studies. These patterns are consistent with theoretical expectations of how the five parameters discussed earlier (pleiotropy, epistasis, plasticity, strength of selection, and population structure) should influence genetic evolution. Evolution over long periods, reflected in the differences between species, should result from mutations relatively devoid of pleiotropic and epistatic effects. In contrast, evolution over shorter periods, reflected in the differences between domesticated races and in the variation segregating within species, may often result from mutations that disrupt plasticity or that have pleiotropic and epistatic effects. In summary, differences between species are caused by a biased subset of the mutations that have appeared within natural populations (*1*).

Conclusions

Although mutations are thought to occur randomly in the genome, the distribution of mutations that cause biological diversity appears to be highly nonrandom. Gene function, gene structure, and the roles of genes and gene products in genetic networks all influence whether particular mutations will contribute to phenotypic evolution. Thus, for some phenotypic changes, evolutionarily relevant mutations are expected to accumulate in a few hot-spot genes and even in particular regions within single genes. In addition, population biology and ecology influence the spectrum of evolutionarily relevant mutations. Over short periods, adaptive mutations with deleterious pleiotropic effects may be selected because mutations without deleterious effects have not yet appeared. In contrast, over long periods adaptive mutations without pleiotropic deleterious effects have more opportunity to arise and be selected.

The genetic basis of phenotypic evolution thus appears to be somewhat predictable. These emerging patterns in the distribution of mutations causing phenotypic diversity derive, however, from a limited set of data culled from the published literature. It is possible that these patterns reflect biases in the way scientists have searched for evolutionarily relevant mutations (*1*). For

example, many researchers focus on candidate genes, which precludes the discovery of previously unknown genes. In the future, we expect that widespread adoption of unbiased experimental approaches, for example genetic mapping, will provide data for robust tests of the predictability of genetic evolution. Genetic mapping can be performed within species and, in rare cases, between closely related species. Alternatively, gene-by-gene replacement of all genes from one species into a second species, although experimentally tedious, may allow unbiased surveys for species that cannot be crossed. This approach would allow comparisons of distantly related taxa and provide a direct test of whether distantly related taxa have accumulated different kinds of evolutionarily relevant mutations than have closely related species.

More precise quantitative predictions about the mutations responsible for phenotypic evolution will probably result from further synthesis of molecular biology and population genetics. New theoretical models will encompass multiple population-genetic parameters within a genomic and developmental framework. These models may provide insight into how the distribution of spontaneously arising mutations is translated into the distribution of mutations segregating within populations and how these two distributions impact short-term and long-term evolution.

Finally, the fact that long-term genetic evolution may represent a biased subset of mutations has applied consequences, from the development of more efficient computer algorithms that utilize evolutionary search strategies to the improvement of agricultural crops and animals. Domestication often selects for mutations that have pleiotropic deleterious effects. Long-term evolution, in contrast, selects for mutations with specific phenotypic effects, and this class of mutations might be exploited to engineer domesticated races that possess desirable characteristics without associated unfavorable properties.

References and Notes

1. D. L. Stern, V. Orgogozo, *Evolution* **62**, 2155 (2008).
2. D. L. Stern, *Evolution* **54**, 1079 (2000).
3. C. Shindo *et al.*, *Plant Physiol.* **138**, 1163 (2005).
4. R. H. ffrench-Constant, B. Pittendrigh, A. Vaughan, N. Anthony, *Philos. Trans. R. Soc. B* **353**, 1685 (1998).
5. H. A. Wichman, M. R. Badgett, L. A. Scott, C. M. Boulianne, J. J. Bull, *Science* **285**, 422 (1999).
6. S. Yokoyama, *Annu. Rev. Genet.* **31**, 315 (1997).
7. C.-B. Stewart, A. C. Wilson, *Cold Spring Harb. Symp. Quant. Biol.* **52**, 891 (1987).
8. Y. Y. Levy, C. Dean, *Plant Cell* **10**, 1973 (1998).
9. A. P. McGregor *et al.*, *Nature* **448**, 587 (2007).
10. E. H. Davidson *et al.*, *Science* **295**, 1669 (2002).
11. H. Chanut-Delalande, I. Fernandes, F. Roch, F. Payre, S. Plaza, *PLoS Biol.* **4**, e290 (2006).
12. S. Marcellini, P. Simpson, *PLoS Biol.* **4**, e386 (2006).
13. J. M. Gibert, S. Marcellini, J. R. David, C. Schlotterer, P. Simpson, *Dev. Biol.* **288**, 194 (2005).
14. T. F. Cooper, E. A. Ostrowski, M. Travisano, *Evolution* **61**, 1495 (2007).
15. S. P. Otto, *Proc. Biol. Sci.* **271**, 705 (2004).
16. J. Kroymann, T. Mitchell-Olds, *Nature* **435**, 95 (2005).
17. R. Yukilevich, J. Lachance, F. Aoki, J. R. True, *Evolution* **62**, 2215 (2008).
18. H. A. Orr, *Evolution* **52**, 935 (1998).
19. R. H. Bellinge, D. A. Liberles, S. P. Iaschi, P. A. O'Brien, G. K. Tay, *Anim. Genet.* **36**, 1 (2005).
20. D. A. Scott *et al.*, *Am. J. Hum. Genet.* **57**, 965 (1995).
21. N. Scarcelli, J. M. Cheverud, B. A. Schaal, P. X. Kover, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 16986 (2007).
22. T. M. Korves *et al.*, *Am. Nat.* **169**, E141 (2007).
23. A. D. Long *et al.*, *Genetics* **139**, 1273 (1995).
24. M. C. Gurganus, S. V. Nuzhdin, J. W. Leips, T. F. C. Mackay, *Genetics* **152**, 1585 (1999).
25. Z.-B. Zeng *et al.*, *Genetics* **154**, 299 (2000).
26. S. J. Macdonald, D. B. Goldstein, *Genetics* **153**, 1683 (1999).
27. L. Jacobsson *et al.*, *Genet. Res.* **86**, 115 (2005).
28. S. Kerje *et al.*, *Anim. Genet.* **34**, 264 (2003).
29. S. B. Carroll, *Cell* **134**, 25 (2008).
30. We thank two referees for helpful comments, P. Simpson for providing photographs and advice for Fig. 2, and CNRS for research funding for O.V.

10.1126/science.1158997