● COMMENTARY

Taste for poison reevolves in fruit flies

Noah K. Whiteman[a,1] and Andrew D. Gloss[b]

Evolution is unpredictable. To illustrate this, Stephen J. Gould once proposed a thought experiment: rewind the tape of life to a point in the distant past, press the replay button, and observe the outcome. "Any replay of the tape," he supposed, "would lead evolution down a pathway radically different from the road actually taken" (1). However, nature is littered with examples of independently evolved organisms that share similar phenotypes (2). Reconciling the unpredictability of evolution with the fact that similar traits appear over and over again independently is an important task for evolutionary biologists.

The repeated evolution of traits, convergent evolution, can be the result of two or more lineages independently adapting to similar environments (2). However simple this fact, many critical questions remain, particularly at the genetic level. Chief among these is the extent to which cases of convergent evolution rely on molecular (genetic) parallelism: the process through which changes at the same genes lead to similar phenotypes in different lineages. In PNAS, Yassin et al. (3) leverage methods in evolutionary genomics, including a new statistical tool they invented, to investigate the molecular basis of convergent evolution in two fruit fly species that independently evolved to specialize on the same toxin-producing fruit. Importantly, this discovery

and future insights rest on our continued ability to explore and study our planet's threatened biosphere.

The story uncovered by Yassin et al. (3) begins around 30,000 y ago. At this time, a population of the rather ordinary fruit fly *Drosophila yakuba* invaded the island of Mayotte (Fig. 1*A*), situated in the Mozambique channel of the Indian Ocean roughly midway between northern Mozambique and northern Madagascar (3). To put this timeframe into perspective, *Homo sapiens* in Europe began painting the likenesses of the animals they hunted—and that hunted them—in caves near Chauvet, France around this same time (4). Seafaring humans, and the plant species they cultivated, would not yet colonize Mayotte for at least another 25,000 y (5). Sometime after the initial colonization by the flies, the Mayotte population of *D. yakuba* began adapting to life on the rotting fruit of the noni (Fig. 1*B*). The noni is a plant in the same family as coffee (Rubiaceae), called *Morinda citrifolia*, and is found in coastal areas throughout the Indian Ocean. Its fruits contain a toxin, octanonic acid, which is not tolerated by most insects. Yassin et al. (3) found that the Mayotte *D. yakuba* now not only tolerate the noni toxin, but prefer noni fruit over fruit from other plant species.

The findings of Yassin et al. (3) are an excellent example of rapid adaptation of organisms to novel, stressful environments. They also reveal a likely case of extreme diet specialization: *D. yakuba* feeds on rotting fruit from nearly 30 plant species on the African mainland, but feeds only on noni in Mayotte (3). Noteworthy here is the fact that specialization by fruit flies on the noni evolved twice: once on Mayotte and once on the Seychelles Islands.

Roughly 500,000 y ago, around the time *Homo erectus* created the first known geometric engravings on a shell on Java (6), a population of *Drosophila simulans* invaded the Seychelles Islands and subsequently evolved to specialize on the noni. The Seychelles population of *D. simulans* eventually diverged to the point where hybrids with *D. simulans* are relatively infertile, and this, along with other evidence, merited designation of the Seychelles lineage as a distinct species from *D. simulans*, called *Drosophila sechellia*, which is a well-studied model for ecological
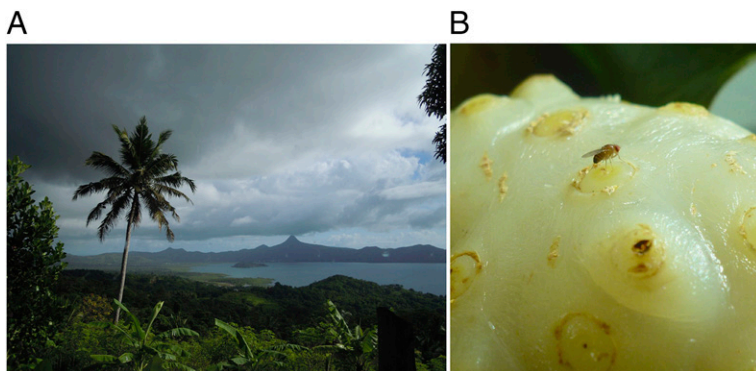

Fig. 1. (*A*) The island of Mayotte. Image courtesy of Nelly Gidaszewski. (*B*) *D. yakuba* female on noni fruit (*Morinda citrifolia*). Image courtesy of John Pool.

[a]Department of Integrative Biology, University of California, Berkeley, CA 94720; and [b]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721
Author contributions: N.K.W. and A.D.G. wrote the paper.
The authors declare no conflict of interest.
See companion article on page 4771.
[1]To whom correspondence should be addressed. Email: whiteman@berkeley.edu.

speciation. Notably, the genomic architecture of adaptation to toxins produced by the noni has been mapped in *D. sechellia* (7).

The system is therefore an excellent one to test if the convergently evolved noni-feeding *D. yakuba* from Mayotte and *D. sechellia* show evidence for molecular parallelism. A number of factors influence the probability that parallel genetic changes give rise to the same phenotype in different organisms (2). This complexity of factors makes it difficult to develop a priori expectations for the amount of molecular parallelism. Empirically characterizing the amount of molecular parallelism is typically difficult, particularly when phenotypic convergence occurred anciently. In exceptional cases, parallel changes are rampant across a gene sequence, and comparing gene sequences across an evolutionary tree of convergent and nonconvergent species lends strong support for molecular parallelism at candidate genes. One example is the gene *Prestin*, which is expressed in the outer hair cells in mammalian ears. The amino acid sequence of Prestin shows clear signatures of convergently evolved amino acids in bats and toothed whales, which both echolocate. Remarkably, the parallel changes found in *Prestin* actually confer sensitivity and selectivity to sounds at high frequency (ultrasound) in these species (8). However insightful these single-gene studies have been, searching for parallel changes does not lend itself well to quantifying the amount of parallel evolution across a genome driven by adaptation to similar environments. A genome-scale attempt to uncover parallel genetic changes linked to echolocation in bats and dolphins (9) lacked power to distinguish adaptive parallel changes from random ones (10). Moreover, these studies were limited to detection of protein-coding changes rather than regulatory changes. The latter influence when, where, and how much of a gene is transcribed, and regulatory changes are likely to be frequent targets of positive natural selection (11).

How can evolutionary biologists search entire genomes, in a relatively unbiased manner, for parallel genetic changes underlying phenotypic convergence? Population genomics, the study of DNA sequence variation across individuals in a population, offers a powerful first step. As a population adapts to a new environment, genetic variants ("alleles") conferring advantages in that environment increase in frequency in the population. This process leaves characteristic footprints in patterns of genetic variation across the genome (12). Searching for these footprints is a powerful way to uncover candidate genes involved in adaptation, which can be characterized in follow-up functional studies. Unfortunately, genomic footprints of adaptation fade away over time, so population genomics is a useful tool only to uncover the genetic basis of relatively recent adaptations (13, 14). However, cases of phenotypic convergence, in which one lineage evolved the trait recently, enables insight from population genomics to be extended to distantly related taxa with convergent traits.

Yassin et al. (3) compare allele frequencies between *D. yakuba* from Mayotte and the mainland using whole-genome sequencing of pools of individuals. They found low but statistically significant genome-wide differentiation between the island population and either mainland population (Kenya and Cameroon). One approach to identify regions of the genome contributing to local adaptation involves searching for regions of the genome with unusually dramatic allele frequency differences between populations, which may arise through positive natural selection (15). When three or more populations are compared, changes in allele frequency can be ascribed to each population through a metric known as the Population Branch Statistic (*PBS*) (16). High *PBS* values indicate that loci evolved under positive selection in a population of interest, but that is not necessarily a result of local adaptation to the habitat unique to

that population. It is important to verify that those loci do not also evolve similarly rapidly in populations associated with other habitats. To identify loci strongly associated with adaptation to noni, Yassin et al. (3) invented a modification of *PBS*, the Population Branch Excess (*PBE*) statistic, which specifically identifies loci evolving much more rapidly in the focal population than across other populations. In a clever advance, the authors identify peaks of genomic divergence specific to the Mayotte population. The *PBE* statistic should be a broadly useful tool in population genomics for associating adaptive evolution with environmental correlates in focal populations. A similar approach may be useful for detecting

Yassin et al.'s remarkable discovery on the island of Mayotte, at the ends of the Earth, serves as a timely reminder that naturalists—those who venture out to study organisms in their natural habitats—still occupy an essential place in biology.

balancing selection in experimental evolution studies because populations in a complex environment will show allele frequency stability relative to populations specializing on a subset of that environment. The *PBE* statistic is conservative for identifying loci underlying local adaptation in a population (e.g., to noni fruit): one is unlikely to find a *PBE* peak at a locus underlying adaptation to noni if natural selection favored standing variation, or if the locus was evolving quickly in all populations (e.g., is generally involved in local adaptation, to noni but also to other habitats).

To determine if adaptation to noni in *D. sechellia* and Mayotte *D. yakuba* relied on parallel genetic changes, Yassin et al. (3) quantified how often genomic regions known to underlie adaptation to noni in *D. sechellia* overlapped with regions of elevated *PBE* in the the Mayotte *D. yakuba* population. This point is critical because Yassin et al. (3) tested for molecular parallelism in a way that relied on independently derived data in the two species. Such quantitative and relatively unbiased approaches are necessary for genome-wide studies of molecular parallelism, given the propensity for false-positives in genome-wide studies. In this case, they found that genomic regions underlying adaptation to noni fruit in *D. sechellia* and regions containing *PBE* peaks in Mayotte *D. yakuba* overlapped more often than expected by chance, suggesting a parallel genetic basis for adaptation to noni in both lineages.

Specifically, three of the most salient peaks of genomic divergence in *D. yakuba* from Mayotte were, remarkably, in homologous regions—previously identified and using different methods—known to evolve rapidly or linked to adaptation to the noni in *D. sechellia*. This finding suggests that positive natural selection targeted the same regions of the genome in two flies that are distantly related, by 10 million y, because of the same environmental pressure (selective agent). It is also possible that such genome-wide studies can reveal functional information about poorly studied genes and gene families. Yassin et al. (3) found significant divergence specific to the Mayotte population within a region containing a cluster of genes encoding Osiris proteins (OSI), which are poorly studied in *Drosophila melanogaster*. In *D. sechellia*, this region was identified as a major quantitative trait locus for tolerance to octanoic acid (17). OSI proteins also appear to be involved in the physiological response to dietary mustard oils in the distantly related drosophilid *Scaptomyza flava* (18). Taken together, these findings suggest that a new function for OSI genes may have been revealed by a study of convergence across the Drosophilidae.

The *D. yakuba* system can be used to understand the early stages of noni adaptation. In the *D. yakuba* system, which is derived more recently than *D. sechellia*, Yassin et al.'s (3) demographic modeling clearly suggests that adaptation to the noni occurred in the context of ongoing gene flow with mainland populations of *D. yakuba*. It is less clear if adaptation to the noni was reliant on standing variation in the ancestral African population, new mutations, or a combination of the two in *D. yakuba*. Can the order in which mutations were fixed during adaptation to noni be uncovered? Haplotype-based analyses, which can more precisely uncover patterns of gene flow and even date the age of haplotypes and the timing of selective sweeps, may offer further insight into these questions (19).

The *D. yakuba* system is also ideal for determining the precise genetic changes underlying adaptation to noni. Specifically, powerful polymorphism-based mapping techniques may be able to uncover candidate loci underlying variation in fitness on the noni. If causative loci are polymorphic in the population, mapping studies could be conducted in the Mayotte population. Even if causative loci are fixed or nearly fixed, crossing the Mayotte and mainland populations could create a mapping population with causative loci at intermediate frequencies. Thus, the ability to easily cross noni-adapted and ancestral populations is an advantage in this system relative to *D. sechellia*, which requires crossing with a different species, *D. simulans*.

To validate the functional effects of putatively convergently selected regions in *D. yakuba*, and to understand the biochemical or physiological mechanisms, functional follow-up studies leveraging *Drosophila* genetic tools are needed. For example, it will be possible to interrogate the same genomic regions in *D. yakuba* and *D. sechellia* to assess how much convergence there really is from a functional perspective. Moreover, until this work is done, the patterns presented by Yassin et al. (3), however exciting, are correlational and will need to be verified using further mapping studies.

We have undeniably entered an era of big data biology, where possessing the skills to extract patterns from enormous datasets enables groundbreaking discoveries. For example, biologists interrogating thousands of human genome sequences have uncovered the genetic basis of local adaptation and disease. These are advances that Charles Darwin and Alfred Russell Wallace, the codiscoverers of the theory of evolution by natural selection, could have only imagined. Yassin et al.'s (3) remarkable discovery on the island of Mayotte, at the ends of the Earth, serves as a timely reminder that naturalists—those who venture out to study organisms in their natural habitats—still occupy an essential place in biology. Countless similar natural experiments remain to be uncovered by future generations of biologists, if we are able to preserve and protect the ecosystems in which they unfold (20).

1 Gould SJ (1989) *Wonderful Life: The Burgess Shale and the Nature of History* (W. W. Norton, New York).

2 Rosenblum EB, Parent CE, Brandt EE (2014) The molecular basis of phenotypic convergence. *Annu Rev Ecol Evol Syst* 45(1):203–226.

3 Yassin A, et al. (2016) Recurrent specialization on a toxic fruit in an island *Drosophila* population. *Proc Natl Acad Sci USA* 113:4771–4776.

4 Sadier B, et al. (2012) Further constraints on the Chauvet cave artwork elaboration. *Proc Natl Acad Sci USA* 109(21):8002–8006.

5 Msaidie S, et al. (2011) Genetic diversity on the Comoros Islands shows early seafaring as major determinant of human biocultural evolution in the Western Indian Ocean. *Eur J Hum Genet* 19(1):89–94.

6 Joordens JC, et al. (2015) *Homo erectus* at Trinil on Java used shells for tool production and engraving. *Nature* 518(7538):228–231.

7 Jones CD (2005) The genetics of adaptation in *Drosophila sechellia*. *Genetica* 123(1-2):137–145.

8 Liu Y, et al. (2010) Convergent sequence evolution between echolocating bats and dolphins. *Curr Biol* 20(2):R53–R54.

9 Parker J, et al. (2013) Genome-wide signatures of convergent evolution in echolocating mammals. *Nature* 502(7470):228–231.

10 Thomas GWC, Hahn MW (2015) Determining the null model for detecting adaptive convergence from genomic data: A case study using echolocating mammals. *Mol Biol Evol* 32(5):1232–1236.

11 Schlenke TA, Begun DJ (2004) Strong selective sweep associated with a transposon insertion in *Drosophila simulans*. *Proc Natl Acad Sci USA* 101(6):1626–1631.

12 Smith JM, Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genet Res* 23(1):23–35.

13 Kim Y, Stephan W (2002) Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* 160(2):765–777.

14 Kim Y, Nielsen R (2004) Linkage disequilibrium as a signature of selective sweeps. *Genetics* 167(3):1513–1524.

15 Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* 74(1):175–195.

16 Yi X, et al. (2010) Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* 329(5987):75–78.

17 Hungate EA, et al. (2013) A locus in *Drosophila sechellia* affecting tolerance of a host plant toxin. *Genetics* 195(3):1063–1075.

18 Whiteman NK, et al. (2012) Genes involved in the evolution of herbivory by a leaf-mining, Drosophilid fly. *Genome Biol Evol* 4(9):900–916.

19 Pool JE, Hellmann I, Jensen JD, Nielsen R (2010) Population genetic inference from genomic sequence variation. *Genome Res* 20(3):291–300.

20 Wilson EO (2016) *Half-Earth: Our Planet's Fight for Life* (W. W. Norton, New York), 1st Ed.