



A genetic region that controls coloration generates morphs of Midas cichlid fish, but speciation involves traits controlled by a number of different genes.

EVOLUTION

How many genetic changes create new species?

Speciation in plants and animals can involve differentiation across many genetic regions

By Patrik Nosil^{1,2}, Jeffrey L. Feder³,
Zachariah Gompert²

The formation of new species generates biodiversity and is often driven by evolution through natural selection. However, the number of genetic changes involved in speciation is largely unknown. Many theoretical models predict that if speciation occurs without geographic isolation, it will be driven by a small number of genes. The logic is that only the few genes that experience the strongest natural selection can overcome the homogenizing effect of ge-

netic mixing (i.e., gene flow) to diverge between populations. However, empirical studies in plants and animals now suggest that speciation—even with gene flow—involves differentiation in surprisingly many genetic regions. This is thought possible because the effects of selection can become coupled across correlated genes such that the selection each gene experiences is much stronger than it would receive in isolation. Thus, the potential for genes to evolve collectively because of coupling may be a key to understanding speciation.

Over recent decades, major strides have been made in understanding the speciation process, which is characterized by the evolution of reproductive isolation (i.e., barriers to interbreeding) and eventually widespread differentiation across the genome. For example, it is now known that natural selection often drives speciation (1).

Moreover, it has been shown that selection can stem from the ecological environment, as proposed by Darwin, or can occur within genomes through conflict and competition between genetic elements.

In contrast to improved knowledge of the role of selection in speciation, the genetic changes involved are less understood (1). Filling this gap in understanding is important because genetic details, such as the number of genes affected by selection and their genomic organization, can influence the dynamics by which new species form (2, 3). For example, if only a few genes that are found together on the same chromosome drive speciation, the process could be highly constrained. By contrast, if many different genes and types of genetic changes drive speciation, the process may be more flexible, but less predictable (3). Although specific forms of reproductive isolation are

¹Centre d'Ecologie Fonctionnelle et Evolutive, CNRS, EPHE, IRD, University Paul Valéry Montpellier 3, Montpellier, France. ²Department of Biology, Utah State University, UT 84322, USA. ³Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556, USA. Email: patrik.nosil@cefe.cnrs.fr

known to be controlled by many genes, the overall genome-wide architecture of differentiation during the active phases of speciation remains poorly understood (1).

A key consideration is whether geographic factors allow interbreeding between diverging populations (i.e., gene flow) (2–5). When divergence occurs with strong geographic isolation—for example between mountain ranges—-independent populations can readily diverge in many genetic regions by the combination of natural selection and chance. The situation with gene flow is different. This is because gene flow acts as a homogenizing force that mixes genes between populations, keeping them similar and preventing them from diverging into separate species. For speciation to occur, natural selection must act in contrasting directions in different populations, to oppose this mixing effect and generate differences. Amounts of gene flow, and thus the selection required to counter them, form a continuum ranging from absent or low to high and can vary over time and space during the speciation process.

Theoretical models predict that speciation with high gene flow is promoted by concentrated genetic architectures comprising one or few genes of large effect (2–4). This allows these few genes to experience strong natural selection to overcome gene flow. That is, selection needs to be concentrated on a few regions, rather than being spread thin among many genes. For example, with a gene flow rate of 0.10 (10% of individuals are immigrants), divergence occurs more readily if two genes each experience a selection intensity of 0.20 each (20% difference in expected fitness, higher than the gene flow rate) than if 10 genes experience an intensity of 0.04 (weaker than gene flow). The total selection intensity of 0.40 is the same in these two cases, but per gene, selection is stronger than gene flow only in the former (5). Thus, models and some findings have led to the concept that speciation with gene flow is driven by a few, isolated genomic islands of divergence. The remainder of the genome is overwhelmed by gene flow and cannot diverge.

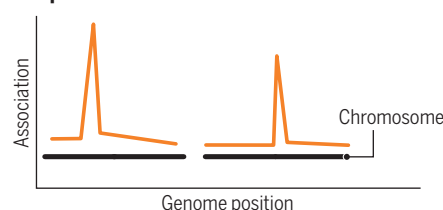
In contrast to these theoretical predictions, emerging data suggest that speciation with gene flow can involve many genetic regions. For example, population differentiation and reproductive isolation based on numerous regions across the genome have now been reported in systems diverging with gene flow, including insects (e.g., *Anopheles* mosquitoes, *Rhagoletis* flies, *Timema* stick insects) (6–8), fish (cichlids and stickleback) (9, 10), and plants (sunflowers) (11), among others. Moreover, when divergence does occur in few genetic

regions, this is often associated with discrete phenotypic morphs (as in mimetic butterflies) (12), rather than the strong reproductive isolation and genome-wide differentiation that characterize distinct species. In some cases, different outcomes have even been tied to variation in the genetic architecture of diverging traits. For example, differentiation in a genetic region controlling coloration generates morphs of Midas cichlid fish, but speciation and stable genome-wide differentiation involve traits controlled by a greater number of genes, such as jaw morphology and body shape (9). Thus, divergence in polygenic traits and multiple genetic regions seems key to progress toward speciation in some systems (see the figure).

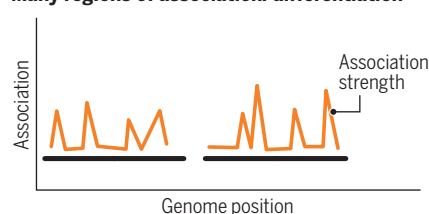
Genetic changes that differentiate morphs and species

Patterns illustrate when few versus many genetic regions differentiate taxa. Black lines above the x axis represent two different chromosomes; the y axis represents the strength of genotype-phenotype association or the degree of population-genetic differentiation, represented by the orange traces.

Few peaks of association/differentiation



Many regions of association/differentiation



Given that models predict that speciation with gene flow will involve few genes, can these recent empirical findings be reconciled with theory? The answer is yes, but this requires a look at a different body of theory, focused on geographic clines, which describes how allele frequencies change over space. This theory was largely developed in the 1980s by Nick Barton and colleagues to understand the dynamics of hybrid zones between species (13). The models show how statistical associations between different genetic regions (called “linkage disequilibrium”) can cause selection on one genetic region to be transferred to other correlated genetic regions. In essence, selec-

tion spills over, through association, from one genetic region to others. In this manner, the effects of selection can spread and become coupled across the genome, rather than being isolated to single genes (4, 13). Such coupling means that the total selection that each genetic region experiences is much stronger than the direct selection it experiences in isolation. Thus, numerous genetic regions can evolve collectively as a unit to overcome gene flow (4).

Notably, the geographic configuration of populations can change during the speciation process, as species ranges expand and contract over time. This means that gene flow itself can be dynamic and change over time. Rather than occurring entirely in the presence or absence of gene flow, speciation likely involves periods of each. In turn, this can result in episodic hybridization, with diverse implications for evolution. For example, periods of geographic isolation might allow a pool of standing genetic variation to build up, including in structural features such as chromosomal inversions, which kick-starts the coupling process and later aids divergence with gene flow (4, 14). Moreover, occasional bouts of gene flow can play a creative role in evolution by allowing for introgression of genes that facilitate adaptation, as reported in tropical butterflies (12), cichlids (14), sunflowers (11), and Darwin’s finches (15). Thus, gene flow can have creative as well as homogenizing effects, and speciation likely reflects a balance between these effects.

Ideas concerning few versus many genes driving speciation are not in conflict. This is exemplified by *Helianthus* sunflowers (11) and *Rhagoletis* flies (7). In both systems, differentiation occurs genome-wide on many chromosomes but is accentuated in regions of reduced recombination, such as those harboring inversions. Thus, even with widespread genomic differentiation, by no means do all genetic regions always diverge equally; those under strong selection or experiencing low recombination may diverge more readily. Moreover, interactions between genes can amplify how strongly genes cause reproductive isolation, again making some genes more critical for speciation than others. Indeed, even speciation involving many genetic regions likely involves a finite number of building blocks, which can be combined in different ways to create diversity, as reported in sunflowers (11) and cichlids (14).

Despite this emerging evidence for the potential importance of genomic coupling in speciation, much work remains to be done. For example, almost all studies to date are purely correlational and often not highly replicated. Thus, experiments are needed to

test causal effects of differentiated regions on reproductive isolation and coupling, as are larger studies examining numerous taxon pairs. Such experiments could comprise experimental evolution studies in the lab or field transplants in the wild. Moreover, natural history and molecular studies, including those using ancient DNA, are required to establish whether divergence really occurred with gene flow. This is critical because false inference of gene flow could lead to erroneous evidence for coupling, because numerous genetic regions diverge between isolated populations through other processes (5). Moreover, genome assembly errors, particularly in regions of low recombination, could lead to a single genetic region being falsely inferred as multiple regions of differentiation. Nonetheless, initial evidence points to coupling as a consideration for understanding speciation. In turn, this suggests that sets of genes might exhibit emergent properties not seen by individual genes, implying the potential for sudden tipping points in the ability for selection to overcome gene flow (4).

By combining theory and data, evolutionary biologists are now poised to better understand how new species are created. A large part of this understanding will involve discovering how the effects of genes and phenotypes become coupled to cause a transition from polymorphism or geographic variation within species to genome-wide differences between distinct species. A major outstanding question is the extent to which the microevolutionary process of coupling can explain broader macroevolutionary patterns of biodiversity, as observed in radiations such as those in cichlid fish (9, 14). ■

REFERENCES AND NOTES

1. D. Schluter, *Science* **323**, 737 (2009).
2. S. Gavrilets, *Fitness Landscapes and the Origin of Species* (Princeton Univ. Press, 2004).
3. S. Yeaman, M. C. Whitlock, *Evolution* **65**, 1897 (2011).
4. S. M. Flaxman, A. C. Wacholder, J. L. Feder, P. Nosil, *Mol. Ecol.* **23**, 4074 (2014).
5. T. E. Crickshank, M. W. Hahn, *Mol. Ecol.* **23**, 3133 (2014).
6. M. K. N. Lawnczak et al., *Science* **330**, 512 (2010).
7. A. P. Michel et al., *Proc. Natl. Acad. Sci. U.S.A.* **107**, 9724 (2010).
8. R. Riesch et al., *Nat. Ecol. Evol.* **1**, 82 (2017).
9. A. F. Kautt et al., *Nature* **588**, 106 (2020).
10. F. C. Jones et al., *Nature* **484**, 55 (2012).
11. M. Todesco et al., *Nature* **584**, 602 (2020).
12. N. B. Edelman et al., *Science* **366**, 594 (2019).
13. N. H. Barton, *Evolution* **37**, 454 (1983).
14. M. D. McGee et al., *Nature* **586**, 75 (2020).
15. S. Lamichanay et al., *Science* **359**, 224 (2018).

ACKNOWLEDGMENTS

The work was funded by a grant from the European Research Council (EE-Dynamics 770826, <https://erc.europa.eu/>) to P.N., from NSF (DEB-1638997) and the U.S. Department of Agriculture–National Institute of Food and Agriculture program (2015-67013-23289) to J.L.F., and from the NSF (DEB 1844941) to Z.G. We thank D. Schluter for ongoing discussions concerning evolutionary genetics.

10.1126/science.abf6671

GENETICS

New genes from borrowed parts

Vertebrate genes acquire new capabilities by capturing parasitic genomic elements

By Aaron Wacholder^{1,2}
and Anne-Ruxandra Carvunis^{1,2}

The vast phenotypic diversity of life is in part a consequence of a continual process of genetic innovation. New genes, with distinct structures and capabilities, emerge regularly throughout evolutionary history. Making use of genomics technologies, researchers are beginning to form an understanding of the details of the processes by which new genes arise. On page 797 of this issue, Cosby *et al.* (1) provide clarity for one such process. Transposons are parasitic genomic elements that replicate by inserting copies of themselves in the host genome. Cosby *et al.* report how vertebrate genes have captured DNA transposon domains, generating new genes that encode new fusion proteins with distinct domain architectures. Fusion of transposon domains with host genes appears to be frequent, with 94 fusion events identified over tetrapod evolution. Transposon domain capture may be a common source of new genes and molecular innovation across the tree of life.

Cosby *et al.* expand on and generalize previous work that has characterized a small sample of host-transposon fusion proteins (2–4). For example, the mamma-

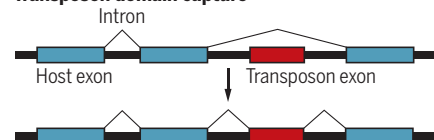
lian gene *GTF2IRD2* (GTF2I repeat domain containing 2) (2), implicated in the rare genetic disorder Williams-Beuren syndrome, was previously discovered to be a carboxyl-terminal fusion of an ancestral transcription factor with a transposon domain. By using large-scale genomic comparison, Cosby *et al.* were able to reconstruct the general process by which DNA transposon domains are captured by host genes. Some DNA transposon families have been highly successful at replicating within tetrapod genomes, inserting themselves in proximity with numerous host genes. Capture occurs through exon shuffling: An exon from the transposon is spliced into the transcript of a nearby host gene, usually making use of the native splice sites within the transposon. The fusion product emerges initially as an alternative splice isoform to the main prefusion product and then becomes the dominant isoform through DNA sequence changes over evolutionary time.

DNA transposons appear especially well-suited for generating transcription factors through exon-shuffling. DNA transposons encode transposases that must recognize and bind to the transposon itself to replicate it. Cosby *et al.* found that most domains captured from DNA transposons perform this self-recognition role. Thus, the captured DNA-binding domains in

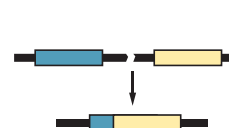
Origins of new protein-coding genes

A gene can acquire new molecular capacities by capturing a domain from a transposon or from a different host gene. Fusion among host genes can similarly assemble preexisting domains into distinct combinations. Alternatively, a gene can evolve new molecular capacities through a series of major sequence changes. This can be facilitated by gene duplication, which creates a copy that can diversify. In *de novo* gene birth, a fully new gene evolves from a previously noncoding sequence.

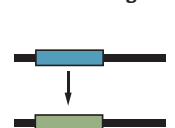
Transposon domain capture



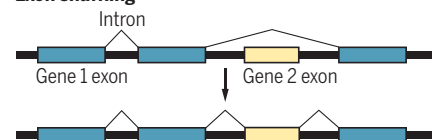
Gene fusion



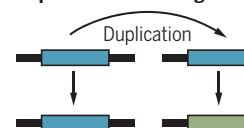
Extreme divergence



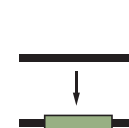
Exon shuffling



Duplication and divergence



De novo birth



How many genetic changes create new species?

Patrik Nosil, Jeffrey L. Feder and Zachariah Gompert

Science **371** (6531), 777-779.
DOI: 10.1126/science.abf6671

ARTICLE TOOLS

<http://science.sciencemag.org/content/371/6531/777>

REFERENCES

This article cites 14 articles, 5 of which you can access for free
<http://science.sciencemag.org/content/371/6531/777#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

Copyright © 2021, American Association for the Advancement of Science