References

1. Zheng, Y. *et al.* (2016) Rice domestication revealed by reduced shattering of archaeological rice from the Lower Yangtze valley. *Sci. Rep.* 6, 28136
2. Spengler, R.N., III (2020) Anthropogenic seed dispersal: rethinking the origins of plant domestication. *Trends Plant Sci.* 25, 340–348
3. Berg, G. and Raaijmakers, J.M. (2018) Saving seed microbiomes. *ISME J.* 12, 1167–1170
4. Bartlett, M. and Patterson, E. (2019) Many ways to drop a fruit: the evolution of abscission zones in the grasses. *New Phytol.* 225, 1407–1409
5. Yu, Y. *et al.* (2019) Divergent gene expression networks underlie morphological diversity of abscission zones in grasses. *New Phytol.* 225, 1799–1815
6. Dong, Y. and Wang, Y.Z. (2015) Seed shattering: from models to crops. *Front. Plant Sci.* 6, 476
7. Humphreys, A.M. *et al.* (2011) Ecology and evolution of the diaspore "burial syndrome". *Evolution* 65, 1163–1180
8. Nelson, E.B. (2018) The seed microbiome: origins, interactions, and impacts. *Plant Soil* 422, 7–34
9. Fang, K. *et al.* (2019) Tissue-specific and geographical variation in endophytic fungi of *Ageratina adenophora* and fungal associations with the environment. *Front. Microbiol.* 10, 2919
10. Wampach, L. *et al.* (2018) Birth mode is associated with earliest strain-conferred gut microbiome functions and immunostimulatory potential. *Nat. Commun.* 9, 5091
11. Van Dongen, J.T. *et al.* (2003) Structure of the developing pea seed coat and the post-phloem transport pathway of nutrients. *Ann. Bot.* 91, 729–773
12. Flannigan, B. (1974) Distribution of seed-borne microorganisms in naked barley and wheat before harvest. *Trans. Br. Mycol. Soc.* 62, 51–58

Forum

Functional Genomics Offers New Tests of Speciation Hypotheses

David P. Hopkins,[1,4,*] Venera I. Tyukmaeva,[1,4,*] Zach Gompert,[2] Jeff Feder,[3] and Patrik Nosil[1,*]

Speciation is a fundamental process shaping biodiversity. However, existing empirical methods often cannot provide key genetic and functional details required to validate speciation theory. New gene modification technologies can verify the causal functionality of genes with astonishing accuracy, helping resolve questions about how reproductive isolation evolves during speciation.

Genomic Tools to Study Speciation

Understanding the process of speciation is one of the main goals of evolutionary biology. Theories concerning speciation are among the best developed in biology (e.g., [1–3]), and throughout the **-omics** (see Glossary) era, expectations have been high for validating theories about how reproductive isolation (RI) evolves to create new species. However, while an abundance of empirical work has raised our understanding of various aspects of speciation [2,3], there is still a paucity of experimental studies that explicitly distinguish alternative hypotheses. This is because many studies lack the fine genetic detail, functional link to RI, and direct manipulation experiments to substantiate theory. Motivated by the dawn of a technological revolution in functional genomics (Box 1), we outline how testing of classical hypotheses in speciation could benefit from advances in gene modification and move beyond current 'candidate gene validation' studies, to help spur new research and discovery in the field.

Pleiotropy and Tight Linkage in Speciation Studies

One of the main goals of speciation research is to ascertain the genetic basis of RI. Modern approaches utilize methods such as **genome scans**, **quantitative trait loci (QTL) mapping**, or **genome-wide association studies (GWAS)** to identify the regions of the genome with high population divergence or association with traits causing RI. However, depending on the frequency of recombination and density of markers, such genomic regions can still contain multiple, sometimes hundreds, of genes and regulatory elements. Such low resolution means that distinguishing effects of **pleiotropy** from those of tight **genetic linkage** of several genes is difficult

Glossary

Chromosomal inversion: a structural rearrangement of DNA sequence where the inverted sequence is reversed relative to the collinear sequence.
Coupling: collective effects of different traits or factors involved in reproductive isolation, which strengthen the barrier to gene flow between the diverging populations.
Dobzhansky–Muller incompatibilities: negative epistatic interactions between different genes, often in hybrids.
Ecological speciation: evolution of RI between populations as a result of ecologically based divergent natural selection.
Epistasis: a phenomenon in which the effect of an allele at one locus is dependent on an allele (or alleles) at one or more other loci (i.e., a between-locus interaction).
Genetic linkage: a nonrandom association of alleles at different loci. Also, a term used by classical geneticists to refer to genes that reside on the same chromosome.
Genome scans: method of comparison between populations/species across the genome to identify differentiated genetic regions across the genome.
Genome-wide association studies (GWAS): studies associating genotypic and phenotypic variation, generally using segregating natural genetic variation.
Heterosis: hybrid vigor, or a phenomenon of enhanced function of a trait in hybrids.
Knockdown: artificial reduction of gene expression by blocking its transcription or breaking down mRNA.
Knock-in: artificial insertion of the nucleotide sequence into a genome.
Knockout: artificial permanent deactivation of the gene with loss of its functionality.
Magic traits: a trait subject to divergent ecological adaptation, which has a pleiotropic effect, causing premating isolation.
Mutation order speciation: accumulation of different incompatible mutations in separate populations subject to the same selective regime.
-Omics: fields of biology with names ending in -omics and that aim to collect large data sets of biological molecules that translate into the structure, function, and dynamics of an organism. Examples include genomics, transcriptomics, proteomics, or metabolomics.
One-allele mating mechanism: a scenario when RI forms due to the same allele spreading in both diverging populations (e.g., a single allele that induces assortative mating with a self-referenced phenotype).
Pleiotropy: a phenomenon when a gene affects more than one phenotype.
Quantitative trait loci (QTL) mapping: a statistical method of associating genetic and phenotypic variation via establishing sets of recombinants via genetic crosses.
Reproductive isolation (RI): genetically based barriers to gene flow between populations/species.

Box 1. Functional Genetics Manipulation Methods

RNAi

RNAi is an antiviral eukaryotic cell pathway that, after recognizing double stranded (ds)RNA in the cytoplasm, targets and digests the corresponding mRNA strand, therefore temporarily knocking down gene expression [6]. It is used as a molecular method to alter gene expression by injecting RNA molecules into organisms to neutralize complementary targeted mRNA molecules. RNAi silencing machinery is present in many, though not all, eukaryotic organisms. Major advantages of using RNAi in evolutionary biology applications are: (i) the possibility of studying the functions of essential genes when knockout causes lethality, and (ii) application to study species that are difficult to work with at the embryonic (egg) stage, a prerequisite for some alternative methods (including those discussed later).

CRISPR/Cas9 Gene Editing

CRISPR/Cas9 is a genome engineering tool adopted from the bacterial immune system, which consists of a guide RNA and Cas9 protein (recent ongoing developments utilize various Cas proteins), capable of cleaving double-strand breaks (DSB) at the specified target sites. The produced genetic changes are heritable if they occur in the germline. The CRISPR/Cas9 editing toolbox has been used for gene (or nucleotide) insertion and deletion, structural changes, replacement of the genes, precise nucleotide editing, and gene regulation (Figure I) [12].

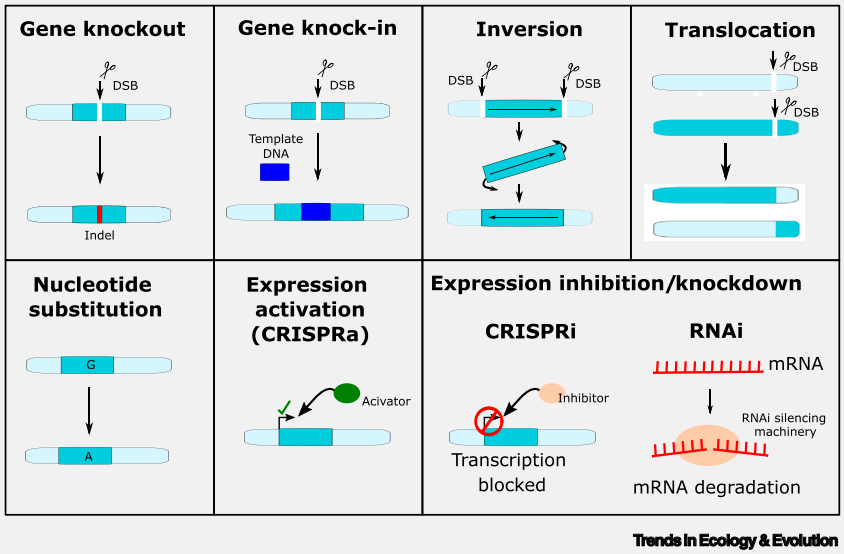

Figure I. Gene Editing Toolbox. Abbreviations: DSB, double-strand break; Indel, insertion/deletion.

[4]. Pleiotropy is relevant for speciation as it explains how multiple traits needed for RI can resist dissociation via recombination. By contrast, models of speciation via genetic linkage require genes to be first positioned correctly in a genome to coordinate their effects and recombination can still dissociate traits controlled by different genes. Accordingly, although pleiotropy and linkage might act similarly over short time scales, both mutation and recombination affect them differently over longer time periods [5].

Distinguishing pleiotropy from linkage is relevant in the '**magic traits**' concept [4,5], where RI is generated as a pleiotropic by-product of divergent selection and is thought to increase the chance of speciation-with-gene-flow [4]. A number of studies suggest putative examples, but very few achieve the genetic resolution to exclude linkage as an alternative [5].

Another classic hypothesis where a single gene enables speciation-with-gene-flow is the **one-allele mating mechanism**. In this model, RI is achieved in diverging populations via the spread of a single allele (e.g., an allele inducing preference for a self-referenced phenotype or an allele causing individuals to 'stay where they were born') [4]. Investigating this concept requires finer mapping of putative RI loci than has been conducted to date.

Snowball theory: a speciation theory predicting that the number of genetic incompatibilities that reduce hybrid fitness grows at a faster than linear rate over time (i.e., snowballs).

New gene-editing methods provide means to test the aforementioned hypotheses. For example, manipulating candidate loci with **knockdown** via a temporary suppression of its expression (RNAi or CRISPRa; Box 1) or permanent gene **knockout** with the CRISPR/Cas system has been shown to work successfully in a variety of organisms [6] (Box 1). This could be applied to determine if a locus affects one or more traits associated with RI or if different loci have discrete functionality (Figure 1A).

Inversions

Chromosomal inversions are common and can be involved in local adaptation and speciation because they strongly suppress recombination among blocks of linked genes [7]. For populations diverging-with-gene-flow, recombination suppression presents a powerful mechanism to allow selected and RI genes to be inherited together. Indeed, patterns of **heterosis** and **Dobzhansky–Muller incompatibilities** of genes linked by inversions could have important consequences on the likelihood of speciation [7]. In addition, when inversions form, they generate new mutations at their breakpoints that could serve both as a source of genetic variation and a time stamp of the inversion's age. Thus, comparing the age of causal variants to breakpoint mutations could reveal if divergent alleles pre-date the inversion or accumulated after its formation.

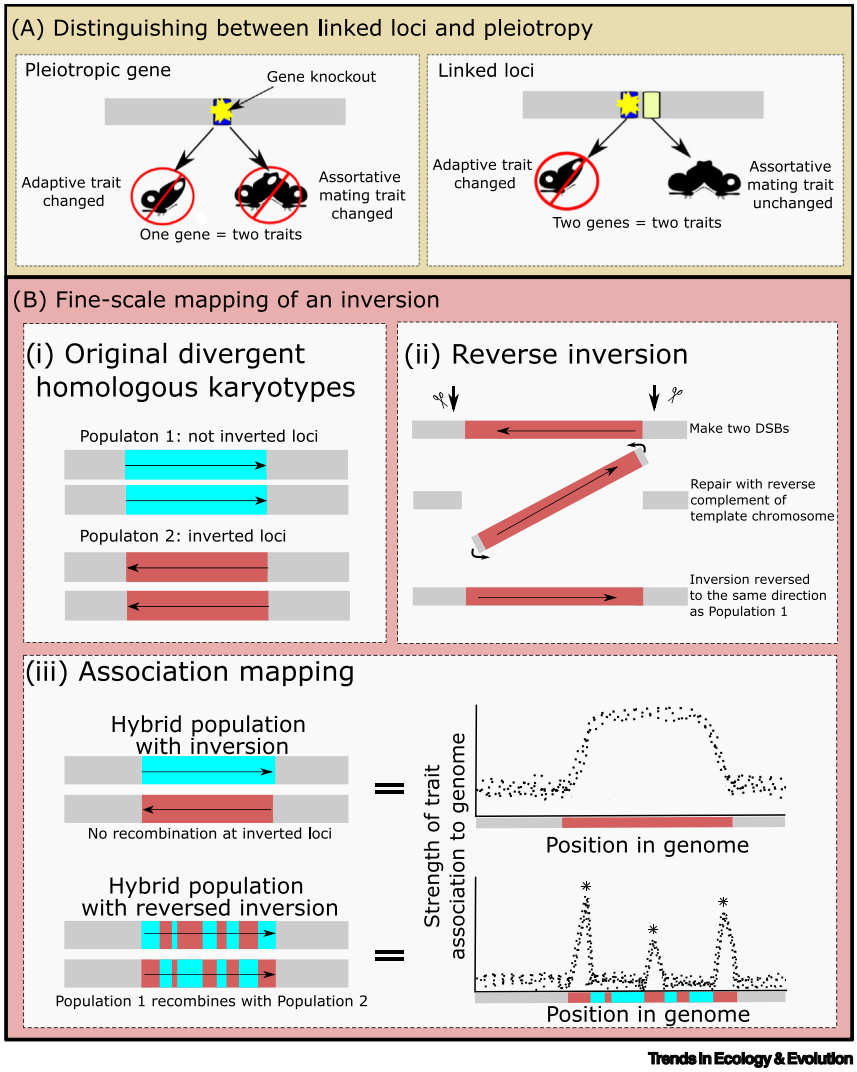

Figure 1. Experimental Design and Predictions for Speciation Hypotheses Tests with Use of Functional Genetics Methods. (A) Distinguishing between pleiotropy and linked loci, (B) fine map of the causal loci across inversion. In the population with two inversion karyotypes (i) reversing an inversion (ii) would allow the recombination and, therefore, association mapping possible (iii). Asterisks (*) indicate loci associated with trait of interest. Abbreviations: DSB, double-strand break.

Traditional methods, while useful for identifying inversions, are less helpful for resolving their causal genetic relationship to RI. Specifically, inversions suppress the recombination in hybrids that is needed to dissect how the multiple genes they can contain contribute and interact to generate RI (Figure 1B). As inversions often contain dozens to hundreds of genes, studies of individual loci within inversions are not feasible. Excitingly, CRISPR/Cas9 has been used in both animal and plant models to experimentally create targeted and large inversions (e.g., [8]). In principle, such precise inversion engineering could also be used to reverse an existing inversion. Genetic crossing using this genetically produced 'reverted' locus could then be used to create recombinants with a noninverted variant, thus allowing fine-mapping of traits and their causal loci across the inversion (Figure 1B). Once the causal variants are identified, their age and relationships within the inversion could be determined.

Genes in a Genomic Context

Earlier, we discussed how genetic modification methods can be used to explore the role of pleiotropy and linkage in RI evolution, however, it could also be used to investigate how genes function in new genomic contexts (i.e., different genetic backgrounds). For instance, genetic modification could be used to test how hybrid dysfunction evolves during speciation; this often involves negative epistatic interactions between genes. Quantifying the strength of **epistasis** could help validate classical theoretical concepts that remain poorly empirically supported, such as the **snowball theory** of hybrid incompatibility [1], and lead to interesting new routes of research via systems biology approaches [9].

Previously, hybrid dysfunction has proved to be difficult to study because of the numerous gene–gene interactions possible and the multilevel nature of genes-to-phenotype expression. However, making controlled genetic **knock-ins** (Box 1) to swap divergent genes between populations (or knockouts, in case of gene duplication) could be a first step towards exploring how epistasis in diverging loci functions and evolves. For example, gene knock-ins could be used to quantify exactly how much epistasis results from a specific number of controlled allele swaps. This could provide important extensions to existing results that assume only pairwise gene interactions, likely an underestimation of reality.

Knock-ins of candidate RI genes into otherwise identical genomes could also be a way to precisely explore the role of selection in speciation. Such 'common genome experiments' could help, for example, distinguish **ecological speciation** from **mutation order speciation** [2]. In addition, a critical question concerns how different selected genes may

become coupled together to accentuate their consequences for RI, accelerating rates of divergence [1,10]. Knock-ins and knockouts of candidate RI genes for taxa at various stages of divergence or of different degrees of hybrid ancestry could aid in studying the **coupling** process and its role in speciation. Notably manipulation of genome structure itself, controlling for gene content, has been shown to cause RI [11], demonstrating the need to also consider the structural genetic contexts behind speciation.

Engineering Holistic Approaches

Advancements in science are often achieved by matching new technology with novel or classic ideas. In this context, now may be a moment when speciation research can be moved forward by cutting-edge molecular methods. However, one should keep in mind that no gene exists in isolation and, while describing the functionality of individual genes is essential, it is important to address how genes evolve in their genomic contexts. Therefore, we expect the future of speciation research to be increasingly holistic and incorporate systems biology approaches, where functional gene manipulation methods could be an important key to unlocking new discoveries and answering seminal questions about the origins of biodiversity.

Acknowledgments

This work was funded by a grant from the European Research Council (EE-Dynamics 770826, https://erc.europa.eu/).

[1]Centre d'Ecologie Fonctionelle et Evolutive, Centre National de la Recherche Scientifique, Montpellier, 34293, France
[2]Department of Biology, Utah State University, Logan, UT 84322, USA
[3]Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556, USA
[4]These authors contributed equally to this work

*Correspondence:
david.hopkins.boa@gmail.com (D.P. Hopkins),
vtyukmaeva@gmail.com (V.I. Tyukmaeva), and
p.nosil@sheffield.ac.uk (P. Nosil).

https://doi.org/10.1016/j.tree.2020.08.001

References

1. Wolf, J.B.W. and Ellegren, H. (2017) Making sense of genomic islands of differentiation in light of speciation. *Nat. Rev. Genet.* 18, 87–100
2. Campbell, C.R. *et al.* (2018) What is speciation genomics? The roles of ecology, gene flow, and genomic architecture in the formation of species. *Biol. J. Linn. Soc.* 124, 561–583
3. Ravinet, M. *et al.* (2017) Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *J. Evol. Biol.* 30, 1450–1477
4. Kopp, M. *et al.* (2018) Mechanisms of assortative mating in speciation with gene flow: connecting theory and empirical research. *Am. Nat.* 191, 1–20
5. Archambeault, S.L. *et al.* (2020) Adaptation via pleiotropy and linkage: association mapping reveals a complex genetic architecture within the stickleback *Eda* locus. *Evol. Lett.* 4, 282–301
6. Sioud, M. (2020) RNA and CRISPR interferences: past, present, and future perspectives. In *RNA Interference and CRISPR Technologies* (Sioud, M., ed.), pp. 22, Springer
7. Faria, R. *et al.* (2019) Evolving inversions. *Trends Ecol. Evol.* 34, 239–248
8. Shou, J. *et al.* (2018) Precise and predictable CRISPR chromosomal rearrangements reveal principles of Cas9-mediated nucleotide insertion. *Mol. Cell* 71, 498–509
9. Tavassoly, I. *et al.* (2018) Systems biology primer: the basic methods and approaches. *Essays Biochem.* 62, 487–500
10. Butlin, R.K. and Smadja, C.M. (2018) Coupling, reinforcement, and speciation. *Am. Nat.* 191, 155–172
11. Yadav, V. *et al.* (2020) Centromere scission drives chromosome shuffling and reproductive isolation. *Proc. Natl. Acad. Sci. U. S. A.* 117, 7917–7928
12. Doudna, J.A. (2020) The promise and challenge of therapeutic genome editing. *Nature* 578, 229–236