

MOLECULAR BIOLOGY & GENETICS

Genes and speciation: is it time to abandon the biological species concept?

Xinfeng Wang^{1,†}, Ziwen He^{1,†}, Suhua Shi^{1,*} and Chung-I Wu^{1,2,*}

ABSTRACT

The biological species concept (BSC) is the cornerstone of neo-Darwinian thinking. In BSC, species do not exchange genes either during or after speciation. However, as gene flow during speciation is increasingly being reported in a substantial literature, it seems time to reassess the revered, but often doubted, BSC. Contrary to the common perception, BSC should expect substantial gene flow at the onset of speciation, not least because geographical isolation develops gradually. Although BSC does not stipulate how speciation begins, it does require a sustained period of isolation for speciation to complete its course. Evidence against BSC must demonstrate that the observed gene flow does not merely occur at the onset of speciation but continues until its completion. Importantly, recent genomic analyses cannot reject this more realistic version of BSC, although future analyses may still prove it wrong. The ultimate acceptance or rejection of BSC is not merely about a historical debate; rather, it is about the fundamental nature of species – are species (and, hence, divergent adaptations) driven by a relatively small number of genes, or by thousands of them? Many levels of biology, ranging from taxonomy to biodiversity, depend on this resolution.

Keywords: speciation, species concept, gene flow, geographical isolation, allopatry

INTRODUCTION

At a very basic level, species may be perceived either as discrete adaptive entities or as mutually exclusive gene pools. In his *Origin of Species*, Darwin [1] portrayed species almost exclusively as adaptive entities. He did devote Chapter 8 to hybrid sterility, which, he noted, is ‘not a specially acquired or endowed quality, but is incidental on other acquired differences.’ In the era of modern synthesis (e.g. [2–8]), the concept of species has shifted to the latter quality, emphasizing species mainly by their incompatibilities. The lament by Darwin that hybrid sterility ‘has been much underrated by late writers’ has been amply rectified (e.g. [9]).

Among the large number of species concepts compiled in recent times [10,11], a few indeed emphasize species being the adaptive units. These include the recognition and cohesion concepts [12,13], as well as the literature devoted to the adaptive aspect of species, for example, on augmenting the genetic variation for adaptation via hybridization [14–16]. Nevertheless, the main concept of species has been centered on the mutual exclusivity between

species. The biological species concept (BSC), the best example of the latter, has been the gold standard in the modern era [2–4].

In this article, we will review the modern literature that appears to reject BSC, both in concept and in empirical observations. Despite the overwhelming evidence against it, BSC remains the dominant species concept at present. Its dominance can be seen in the most popular textbooks read by the next generation of evolutionary biologists as if BSC had been proven beyond doubt [3,17]. Clearly, the contradictions need to be resolved. We should emphasize that BSC is far more than an unresolved issue for mere historical reasons. It is fundamental to our understanding of the genetic nature of species; hence, many key evolutionary issues depend on its resolution (see Discussion).

THE BIOLOGICAL SPECIES CONCEPT

In BSC, species are separated by various forms of reproductive isolation (RI) and do not exchange genes [2–8,18–24]. BSC also denotes a process that

¹State Key Laboratory of Biocontrol, Guangdong Key Lab of Plant Resources, School of Life Sciences, Sun Yat-Sen University, Guangzhou 510275, China and ²Department of Ecology and Evolution, University of Chicago, Illinois 60637, USA

*Corresponding authors. E-mails: ciwu@uchicago.edu; lssssh@mail.syu.edu.cn

†Equally contributed to this work.

Received 19 November 2019;
Revised 17 December 2019;
Accepted 31 December 2019

permits reproductively incompatible populations to emerge. In this process, the absence of gene flow is essential because genetic exchanges are perceived to be capable of reversing the divergence. BSC makes strong assumptions about the genetics of species divergence, postulating that almost the entire genome evolves as a cohesive unit [2,5]. It is essentially a genomic concept of species and speciation [5].

Once the process of speciation is completed, 'there can be a little genetic leakage between species through hybridization; [hence] the BSC does not require that species be 100 percent reproductively isolated' (p. 215 [3]). These exchanges are tolerated, for example, in hybrid zones (p. 217 [3] and [25,26]), *only* because such hybridizations have little consequence on the genetic integrity of the species *already formed*. The post-speciation 'genetic leakage' does not mean the tolerance of genetic exchanges during speciation under BSC. Genetic exchange during speciation, like cheating in a class, is delineated between zero and non-zero. (Otherwise, what does a little exchange, or a little cheating, mean?)

Here, we will use the term BSC broadly for both the process of speciation and the concept of species. As BSC demands a prolonged period of divergence during which there is no gene flow, geographical isolation appears to be a most effective mechanism. This mode is generally referred to as allopatric speciation.

Finally, there is a conspicuous but odd fact about BSC: under this concept, species status is often indeterminate when the lack of gene flow is a result of geographical distance. That is one reason why BSC is rarely used in the actual practice of delineating species. Instead, species delineation generally relies on phenotypic criteria such as morphology, ecological preferences, etc. Hence, BSC cannot be defended as a practical, albeit imperfect, concept. It is, instead, a perfectly impractical concept.

THE ALTERNATIVE 'GENIC VIEW' OF SPECIATION (VIS-À-VIS BSC)

In an alternative view, species are defined by a set of loci that govern the morphological, reproductive, behavioral and ecological characters. These 'speciation genes' may collectively account for no more than a fraction of the genome [5,27–29]. This fraction should be fitness-reducing on introgression [30], whereas the rest of the genome can be freely exchanged without a fitness consequence. In short, the diverging genomes comprise both introgressable and non-introgressable DNA segments. These non-introgressable segments are often referred to as 'genomic islands,' which are, in theory, more divergent than the rest of the genome [5,29–34]. This con-

cept has been referred to as the genic view of speciation [5], in contrast with the genomic perspective of BSC. The genic view of speciation differs from most species concepts as it merges the adaptive view and the isolation concept of species presented in the Introduction.

In the last 20 years, genomic islands have been identified in many speciation events [33–49] (see Table 1 for a compilation). The literature suggests that 'speciation with gene flow' is rather common in a wide array of taxa [35–49]. If DNA segments between the genomic islands indeed represent exchanges during speciation, it would seem timely to abandon BSC as the defining concept of species and speciation. Therefore, the issue is whether the genomic evidence on species divergence [35–49] is collectively incompatible with BSC.

AN EXPANDED BSC THAT PERMITS GENE FLOW IN A TRANSIENT EPISODE AT THE ONSET OF SPECIATION

In defense of BSC, we should first emphasize that its salient feature is 'a period of strict geographical isolation (or allopatry)' that is needed to *complete* the process of speciation, including the evolution of RI mechanisms. Other than that, BSC does not have specific requirements either before or after the period of allopatry. In any realistic scenario of allopatric speciation, BSC would also predict 'speciation with gene flow.' This is because geographical events leading to full allopatry often develop gradually. For example, the closure of the Isthmus of Panama, suggested to be the strongest evidence for allopatric speciation, takes a few million years to complete [50]. Other geographical and geological events, such as the rise of a mountain range or the advance of glaciers, would also take time. In other scenarios, it may take populations thousands of years to disperse and become geographically isolated. In short, the allopatry phase may generally be preceded by a period of diminishing gene flow.

In Fig. 1a, we depict this expanded model of BSC, which seems more realistic than generally portrayed. In this model, population B expands its range gradually between T0 and T2. Between T0 and T1, population B still overlaps with the parental population A and gene flow between the two populations only gradually diminishes. In the model conventionally portrayed, T0 – T1 would be compressed into a single time point. Such compression is neither realistic nor necessary for BSC. After T1, the two populations are truly allopatric without genetic exchanges.

Table 1. Publications on 'speciation with gene flow' since 2005.

Taxa	Species	Main research focus	Inferred stage of speciation	Reference	
1	<i>Anopheles</i>	<i>A. gambiae</i>	Genomic islands of speciation in <i>Anopheles gambiae</i>	Early stage. Reason: the materials are two sympatric, partially isolated subtaxa known as M form and S form of <i>A. gambiae</i>	Turner <i>et al.</i> , 2005
2	<i>Mus</i>	<i>M. musculus</i> and <i>M. m. domesticus</i>	Genomic islands of differentiation between house mouse subspecies	Early stage. Reason: the materials are subspecies, although there seems to be partial RI associated with the X	Harr, 2006
3	<i>Ficedula</i>	<i>F. albicollis</i> and <i>F. hypoleuca</i>	The genomic landscape of species divergence in <i>Ficedula</i> flycatchers	Uncertain. Reason: although the age of divergence is said to be > 1 Myrs, the fixed differences seem low (see their table 2)	Ellegren <i>et al.</i> , 2012
4	<i>Heliconius</i>	<i>H. melpomene</i> , <i>H. cydno</i> and <i>H. timareta</i>	Genome-wide evidence for speciation with gene flow in <i>Heliconius</i> butterflies	Uncertain. Reason: the case is commented in the main text. The density of speciation genes seems low	Martin <i>et al.</i> , 2013
5	<i>Helianthus</i>	<i>H. petiolaris</i> , <i>H. debilis</i> , <i>H. annuus</i> and <i>H. argophyllus</i>	Genomic islands of divergence are not affected by geography of speciation in sunflowers	Early stage. Reason: the materials are four recently diverged pairs of sunflower species with low pair F_{ST} . There are far fewer fixed changes than polymorphic ones (see their table 2)	Renaut <i>et al.</i> , 2013
6	<i>Mimulus</i>	<i>M. guttatus</i> and <i>M. nasutus</i>	Speciation and introgression between <i>Mimulus nasutus</i> and <i>Mimulus guttatus</i>	Uncertain. Reason: these two sister species are 200 000–500 000 years apart	Brandvain <i>et al.</i> , 2014
7	<i>Oryctolagus</i>	<i>O. cuniculus algirus</i> and <i>O. c. cuniculus</i>	The genomic architecture of population divergence between subspecies of the European Rabbit	Early stage. Reason: the materials are two subspecies of rabbits in the early stages of divergence	Carneiro <i>et al.</i> , 2014
8	<i>Anopheles</i>	<i>A. gambiae</i> species pair (<i>A. coluzzii</i> and <i>A. gambiae sensu stricto</i>)	Adaptive introgression between <i>Anopheles</i> sibling species eliminates a major genomic island but not reproductive isolation	Early stage. Reason: the materials are the M form and S form of <i>A. gambiae</i>	Clarkson <i>et al.</i> , 2014
9	<i>Corvus</i>	<i>C. (corone) corone</i> and <i>C. (corone) cornix</i>	The genomic landscape underlying phenotypic integrity in the face of gene flow in crows	Early stage. Reason: the two species show only a small number of narrow genomic islands across the whole genome	Poelstra <i>et al.</i> , 2014
10	<i>Anopheles</i>	<i>An. gambiae</i> complex	Extensive introgression in a malaria vector species complex revealed by phylogenomics	Early stage. Reason: it is evident that the introgressions in these species are earlier-stage events (see their Fig. 1C)	Fontaine <i>et al.</i> , 2015
11	Multiple taxa	All Darwin's finch species and two tanagers, <i>Tiaris bicolor</i> and <i>Loxigilla noctis</i>	Evolution of Darwin's finches and their beaks revealed by genome sequencing	Early stage. Reason: they find extensive evidence for interspecific gene flow throughout the radiation which represents the early stages of diversification when phenotypic transitions between species	Lamichhaney <i>et al.</i> , 2015
12	<i>Astatotilapia</i>	Two cichlid fish ecomorphs	Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake	Early stage. Reason: this study is about the discovery and detailed characterization of early-stage adaptive divergence of two cichlid fish	Malinsky <i>et al.</i> , 2015
13	<i>Vermivora</i>	<i>V. chrysoptera</i> and <i>V. cyanoptera</i>	Plumage genes and little else distinguish the genomes of hybridizing warblers	Early stage. Reason: the two species show extremely low differentiation: only six small genomic regions exhibit strong differences	Toews <i>et al.</i> , 2016
14	<i>Xiphophorus</i>	Three hybrid pops between sister species of <i>X. birchmanni</i> and <i>X. malinche</i>	Natural selection interacts with recombination to shape the evolution of hybrid genomes	Early stage. Reason: they studied three replicate hybrid populations that formed naturally between two sister swordtail fish species	Schumer <i>et al.</i> , 2018
15	<i>Heliconius</i>	Many <i>Heliconius</i> butterflies species	Genomic architecture and introgression shape a butterfly radiation	Early stage. Reason: introgressions happened during the process of adaptive radiation	Edelman <i>et al.</i> , 2019

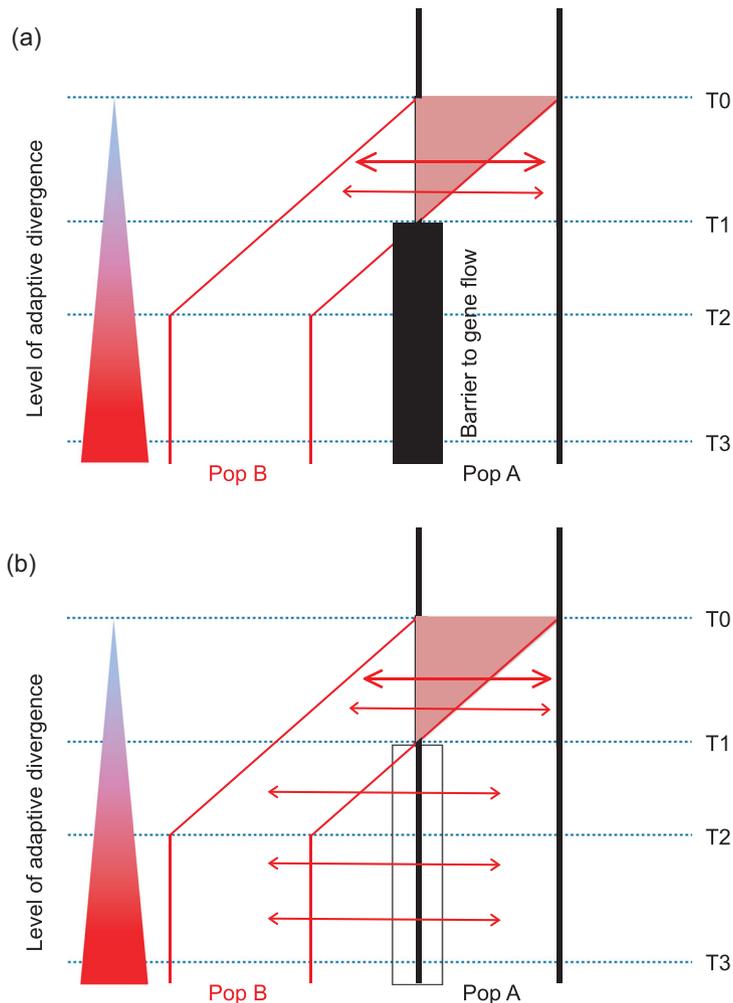


Figure 1. Two contrasting views of speciation. (a) The expanded model of biological species concept (BSC) that incorporates a phase of gene flow at the onset of speciation. This phase is associated with the gradual development of geographical isolation. The inference of speciation with gene flow needs to have a starting point, which is usually T0 in publications. The size of the arrow corresponds with the amount of gene flow. (b) The scenario that could reject the expanded BSC. In this scenario, gene flow continues, possibly diminished, all the way to the completion of speciation when traits of reproductive isolation have evolved.

Under BSC, the diverging populations may become good species at T2 by certain criteria and, by the time of T3, there is no ambiguity about their species status by most criteria. While the BSC model of Fig. 1a has a period of strict allopatry, an observer at T2 may still reach the conclusion that speciation between A and B has happened with gene flow. At T3 and beyond, the earlier signals of gene flow may dwindle to a level beyond detection. Hence, it may not be surprising that many genomic studies supporting the conjecture of speciation with gene flow are mostly about recent speciation events (see below).

THE EVOLUTION OF RI IN RELATION TO GENE FLOW AND ALLOPATRY

Between the genomic and genic view of speciation, the main difference is on how genes interact with the environments. However, the evolution of RI, in particular, of the postmating kind, depends on the interactions among genes within the same genome. Such interactions may change the dynamics of speciation. Two contrasting views on whether gene flow during speciation can disrupt the evolution of postmating RI are presented in this section.

Gene flow could disrupt the evolution of RI: the classical view of BSC

The BSC model portrayed in Fig. 1a leads back to the recurring question about BSC: why is geographical isolation necessary for speciation? Indeed, Darwin [1] placed much less emphasis on either geographical or RI than did Mayr [2]. By genetic reasoning, the genic view posits that speciation does not need strict allopatry as divergent selection can easily nullify the impact of gene flow [5]. (For that reason, the biological species concept should be more accurately renamed the isolation species concept.)

In Mayr's conception, geographical isolation is indispensable because the entire genome is perceived as a cohesive unit. Hence, any gene replacement by the genetic materials of a different species is bound to be deleterious. This conception may have originated in the interpretation of the genetics of postmating RI, or hybrid incompatibility (such as hybrid inviability or sterility).

The standard hypothesis is the Dobzhansky-Muller Incompatibility (DMI) model, which assumes the simplest genetic interactions possible [18,19]. In DMI, the ancestral population at the time of nascent speciation has two loci with alleles *a* and *b*, respectively. In population 1, (*a*, *b*) evolves to (*A*, *b*) and, in population 2, (*a*, *b*) evolves to (*a*, *B*). It is assumed that (*A*, *B*) causes hybrid incompatibility. By DMI, it is plausible that gene flow would impede the evolution of hybrid incompatibility. When the two populations are, respectively, accumulating the *A* and *B* alleles, any gene flow would mutually impede the spread of the new *A* and *B* alleles [18,19,34]. For example, if the immigrant (*a*, *B*) from population 2 mates with (*A*, *b*) individuals in population 1, many offspring of the new genotype (*A*, *B*) would die, making (*A*, *b*) less fit. It is true in the reciprocal direction as well. Accepting this simple argument, one may view geographical isolation to be necessary for the completion of speciation.

Gene flow could not disrupt the evolution of RI: counter arguments from the genetics of RI

The argument above that gene flow would disrupt the evolution of postmating RI is a qualitative one. Although it is intuitively appealing, the probability of evolving postmating RI should be a function of the mutation rate, the migration/hybridization rate, the selection intensity and the nature of fitness interactions (the number of loci, the degree of gene dominance, etc.) [51].

In a simulation study [51], a model shows that postmating RI can readily evolve as long as the amount of migration does not overwhelm the selective advantage of the resident allele in its own population. The model of Yang [51] further takes advantage of the recent MIM (for migration-isolation-migration) model of speciation. He *et al.* [22] identified the speciation mechanism along the Indo-Malayan coasts to be associated with the repeated openings and closures of the Strait of Malacca. In Yang's model [51], postmating RI can evolve readily if periods of migration are interspersed with periods of isolation.

The discussion so far relies on the DMI model with two loci. This simplest model (one locus per species) uses a genetic set-up that is least conducive for the evolution of RI and most susceptible to perturbation by gene flow among possible genetic models. In a series of fine-scale dissection of hybrid male sterility in *Drosophila* [27–29,52–56], it is shown that hybrid incompatibility usually involves several loci per species. The two-locus DMI has not been supported by any empirical study in the literature. As explained in Fig. 8 of Cabot *et al.* [54], even a simple extension of DMI to three loci would make the evolution of postmating RI much easier.

Gavrilets [57] generalizes this view into a 'holey landscape,' whereby the adaptive landscape is nearly flat with a few genotypes causing RI. These occasional genotypes represent adaptive 'holes' in the landscape. On this landscape, the evolution of hybrid incompatibility is relatively unimpeded as it only needs to avoid the adaptive holes. This view has extensive empirical support (see [28,29,52,55,56]; all reviewed in the last section of this perspective). Simulations by Yang [51] corroborate this view that the larger number of genes involved in RI indeed makes it easier to evolve postmating RI.

GENETIC COMPLEXITY UNDERLYING THE EVOLUTION OF RI

The discussions above point to a curious aspect of modeling speciation – the empirical observations

on the genetics of RI are often disregarded, as if they were mere details. Nevertheless, when postmating RI cannot easily evolve under a two-locus DMI model but can readily evolve under a multi-locus one, 'god is really in the details.' An equally instructive lesson of genetics is the modeling of premating isolation. Here, the tendency to evolve RI as a function of the number of loci is the exact opposite of evolving postmating isolation – the fewer the loci, the easier it is to evolve premating isolation. This can be demonstrated in the evolution toward the so-called 'Fisher's equilibrium' with only one locus.

Imagine a two-allele system with three genotypes – AA, Aa and aa. If AA and aa do not mate but all other combinations mate normally, then it is easy to see that any population would evolve in sympatry toward two isolated populations consisting solely of either AA or aa genotypes. (This is because Aa will disappear as a result of the absence of replacements from AA × aa mating.) In the theoretical models [58–61], the tendency is evident. A possible reason that premating and postmating isolation show opposite dependencies on the number of loci is the fitness reduction associated with the incompatible genotypes in the latter but not in the former.

GENE FLOW IN RELATION TO MODES OF GEOGRAPHICAL ISOLATION

The previous sections present the genetic arguments for, as well as against, the need for a phase of isolation with no gene flow, prior to the completion of speciation (Fig. 1a). The disagreement needs to be resolved empirically. In the convention, the inference of gene flow is based on the geographical distribution of the populations. It is assumed that allopatric distribution prohibits gene flow and sympatric distribution permits unimpeded exchanges. However, such a correspondence is inexact as allopatry can be bridged by long-distance migration and sympatry may often be associated with micro-allopatry [62–66]. In either case, geography is not a good indication of gene flow, or lack thereof.

No less important, if most speciation events are parapatric between adjacent populations [67,68], it would often be ambiguous whether gene flow has happened or not during speciation. Indeed, in the MIM cycles [22], the same geographical feature is associated with episodes of gene flow and full isolation. With all these considerations, we do not use the geographical mode of isolation to assess gene flow. Instead, studies that compare genomic sequences to infer past gene flow during speciation are the main literature in this review.

EVALUATION OF GENOMIC EVIDENCE FOR 'SPECIATION WITH GENE FLOW'

To reject BSC of Fig. 1a, it is no longer sufficient to show that gene flow happens between T0 and the observation time, say, T2. Instead, it is necessary to show that no period of strict isolation exists between T0 and T2. Furthermore, because the duration of [T0, T1] can be variable, depending on the strength of divergent selection and the rate of gene flow, the evolutionary dynamics of this period are expected to be highly variable as well. For that reason, gene flow in the transient period of (T0, T1) can be compatible with a wide range of observations at T2, or even at T3. In Table 1, we compile publications on speciation with gene flow, aiming to identify studies that can reject BSC. In the end, we could not find compelling evidence against BSC perhaps because no study is designed to test the model of Fig. 1a. Most reports are about closely related species or subspecies, often before evolving complete RI [35–49]. In these cases, the influence of gene flow in the transient episode between T0 and T1 would be significant. We select a few examples for discussion below; other examples are listed in Table 1.

- 1) A common approach to population divergence is the use of the F_{ST} statistic [35,42,46,47,49]. The F_{ST} statistic is primarily a measure of population differentiation, rather than species divergence. For that reason, F_{ST} is not appropriate for testing BSC. It has also been pointed out that the high F_{ST} values in many publications are not a result of high divergence between species, but low differentiation within populations [33].
- 2) Many Anopholes publications report gene flow during speciation [36,43,48]. Most of them are about subspecies, or different forms of the same species. Among them, Fontaine *et al.* [36] is about a more ancient system although RI remains incomplete. It is evident that the introgressions in these species are early-stage events (see their Fig. 1C).
- 3) Many studies on speciation gene flow state explicitly that these are early-stage events [35,38,47]. Other studies are not informative in this respect. Nevertheless, we find no case of large-scale introgression in late stages of speciation, when postmating RI is evident. These early-stage introgressions are reported in Darwin's finches throughout their adaptive radiation [38], cichlid fish with divergent mate preferences [35] and subspecies of rabbits [47].
- 4) Plants often show patterns of speciation that are different from animals, in particular, in their tendencies to hybridize [37,46]. Nevertheless, we still fail to find compelling evidence from plants

against the model of Fig. 1a. For example, the sunflower clades have been elegantly shown to exchange genes constantly. However, polymorphic mutations far outnumber fixed ones, suggesting the species to be in early stages of speciation (see table 2 of [46]).

- 5) In Martin *et al.*'s study [40] of *Heliconius*, a high incidence of genetic exchanges is reported and up to 40% of the genomes show signs of introgression across species. The authors also report long linkage disequilibrium (LD) blocks as evidence of recent introgressions. The long LD blocks indicate low density of 'speciation genes' in the introgression, which likely reflects nascent speciation. This topic is addressed in Fig. 2.

In a literature survey of this kind, crucial details are not always available to the readers. In cases where the authors present the diverging taxa as at an early stage, it is possible to stage them close to T1-T2. In cases where the taxa are close to T3 (e.g. when strong postmating RI has evolved), it is more difficult to see whether the model of Fig. 1a or b applies. In the next section, we attempt to make a distinction between the T3 stages of the two models.

CAN BSC BE REJECTED? STUDIES OF GOOD SPECIES WITH STRONG RI

The reviews of the literature above lead to the central question – what might constitute the necessary evidence against the BSC model of Fig. 1a, if BSC is indeed a faulty model?

A starting point should be the use of good species that are delineated by strong RI mechanisms and, hence, no longer in the nascent stage of speciation. In this section, we review the detailed genetic analysis of RI in the *Drosophila* literature, whereby functional tests of speciation genes have been done [28,29,52–56]. These studies show that, between sibling species that are morphologically indistinguishable, the number of 'speciation loci' is already high, in the hundreds or thousands [53–56]. Such a high density of adaptive differences would yield a distinct introgression pattern if gene flow continues toward the very end of speciation, as shown by the simulations in Fig. 2.

The number of speciation loci: from behavioral races to good species of *Drosophila*

This question can only be answered by detailed genetic analyses from incipient species all the way

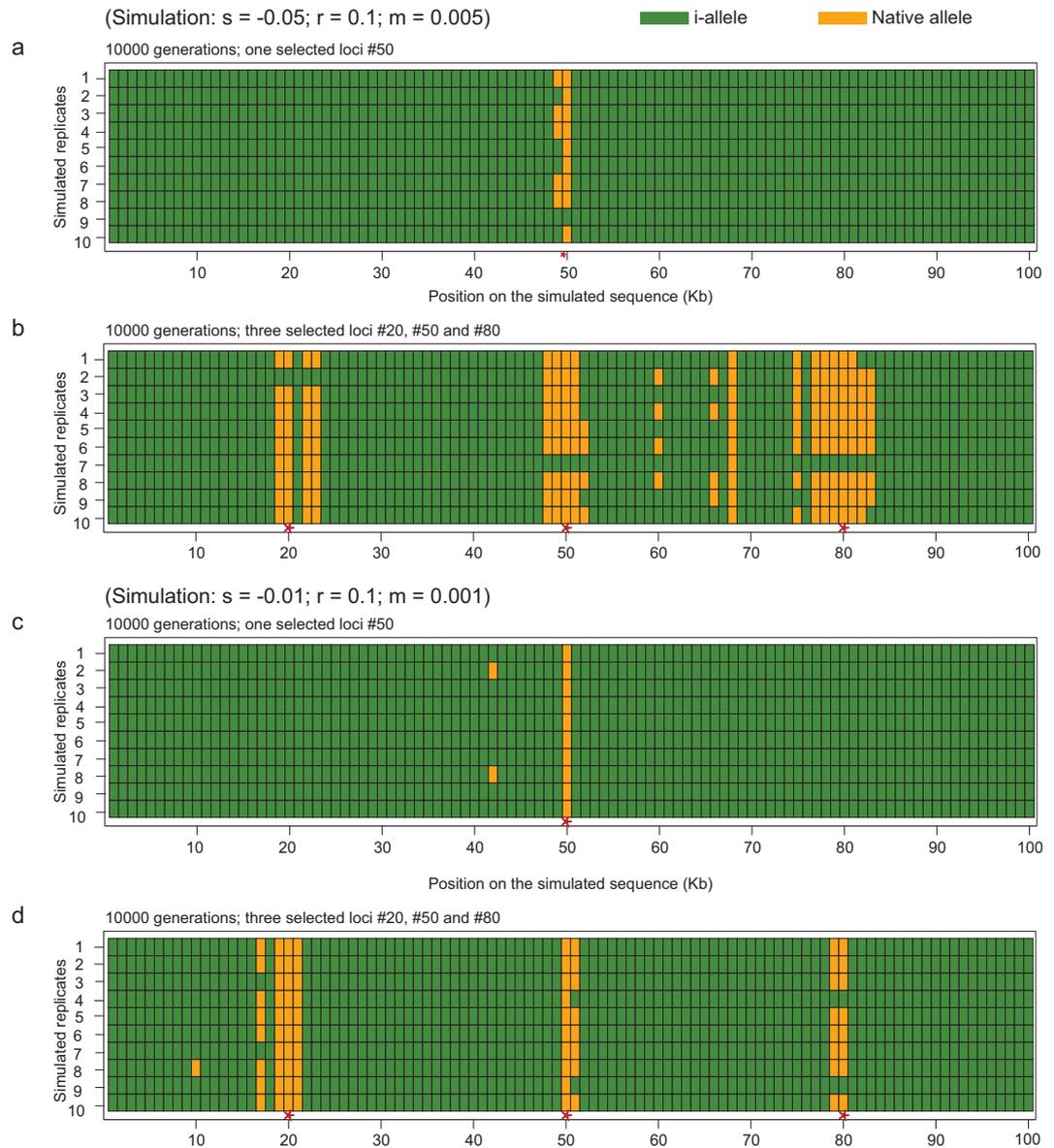


Figure 2. Simulated introgressions in haploid 100 kb genomes. The selected loci (or speciation genes) are marked by red stars at the bottom. Sites of introgression and non-introgression are marked green and orange, respectively. (a, b) These examples are done under strong selection ($s = -0.05$), low recombination ($r = 0.1$ for per 100 kb per generation) and high introgression ($m = 0.005$ per generation). (c, d) These examples are done under weak selection ($s = -0.01$), low recombination ($r = 0.1$ for per 100 kb per generation) and low introgression ($m = 0.001$ per generation). Note that very fine delineations of blocks are possible under the simulated conditions.

to ‘good species.’ Phenotypic observations suggest that even closely related species differ by a multitude of traits. When such traits were dissected at the genetic and molecular levels, each was found to be highly polygenic [42,49,69–73], implying extensive genetic divergence. A particularly instructive case is hybrid sterility in *Drosophila*. Among sibling species that show little morphological divergence, the number of genes involved in hybrid male sterility, that is spermatogenic failure, is in the hundreds [27,29,53–56]. For example, in the identification of hybrid sterility loci between *D. melanogaster* and *D.*

simulans, Sawamura *et al.* [53] used small chromosomal deficiencies to uncover the sterility loci. To their surprise, every single deficiency uncovered at least one such locus. Extrapolating to the whole genome, they estimate thousands of speciation genes between these two best-known sibling species.

The cloning of the component genes further reveals a complex web of interactions [28,52]. This level of divergence means that selection, presumably sexual in nature, drives spermatogenic programs to evolve extensively and rapidly [74] with hybrid sterility being the incidental byproduct. Other

‘speciation traits’ such as sexual isolation, genital morphology and neural development depict a genetic basis that is qualitatively similar [75–77]. Again, a detailed account of mating loci between the African and cosmopolitan behavioral races of *D. melanogaster* shows that the genetic differentiation in premating isolation may be even more extensive than hybrid male sterility [75,78]. The number of loci responsible for the differentiation in the genital morphology between *D. simulans* and *D. mauritiana* (closer to *D. simulans* than *D. melanogaster*) is estimated to be > 15 [76]. While the data are from the genus of *Drosophila*, which offer the rarely attained resolution, the genetic basis of hybrid sterility does follow rules applicable to a wide range of taxa [9] from vertebrates (birds and mammals) to invertebrates (*Drosophila* and butterflies). If we extend the definition of speciation loci to genes that cause fitness reduction upon introgression, a definition most appropriate for introgression analysis, the number of speciation loci becomes even larger. Fang *et al.* [30] showed that introgressions without any detectable phenotypes may often harbor strong fitness-reducing genes when the fitness is measured in long-term population experiments.

In conclusion, the extent of functional divergence at the genic level is very high, even among sibling species that look identical. The substantial genic divergence between good species may hold the key to the expected pattern of introgression in late stages of speciation.

Expected genomic patterns of introgressions when speciation is close to completion

The literature on gene flow during speciation is, by and large, compatible with Fig. 1a, and, hence, is also compatible with BSC. In contrast, the scenario of Fig. 1b, whereby gene flow continues until the completion of speciation, is incompatible with the core of BSC.

What might be the resultant genomic patterns, if gene flow continues as depicted in Fig. 1b? We carry out computer simulations with unceasing gene flow. The model clearly violates the core assumptions of BSC. In the simulation of a stretch of DNA 100 kb long, the populations are assumed to have diverged with one ‘speciation locus’ at T2 and three speciation loci at T3. For simplicity, gene flow occurs in only one direction. For introgression to be feasible, either the migration rate (m) has to be sufficiently high ($m = 0.005$, Fig. 2a and b) and/or selection against the speciation allele has to be weak ($s = 0.01$; Fig. 2c and d). With the genetic input from migration,

which is subjected to removal by natural selection, the genetic makeup would eventually reach a steady state. The steady-state results are presented in Fig. 2. As expected (but shown only in Supplementary Fig. S1), when migration is low ($m = 0.001$) and selection is strong ($s = -0.05$), stable introgression is not observed.

Figure 2 shows two large introgressions at T2 but, at T3, these two introgressions are broken into at least four much smaller segments. In other words, when gene flow and selection reach an equilibrium, the introgressions become much more fine-grained, effectively reflecting the number of speciation loci accrued. Such fine-grained introgressions have not been reported in the literature as most experimental designs rely on the statistical power inherent in large segments of DNA. Fine-grained introgressions are therefore not possible to detect in most studies. Furthermore, the fine-grained pattern represents a balance between gene flow and natural selection, but such a balance is also transient (and difficult to find) in nature. In a companion paper, Wang *et al.* [79] were able to find such a speciation event, thanks to several factors including the geographical distribution of populations and the duration of gene flow. The confluence of so many factors, necessary for testing BSC rigorously, may not be common.

THE IMPORTANCE OF A DEFINITIVE ACCEPTANCE, OR REJECTION, OF BSC

A realistic BSC model would require a period of strict isolation for speciation to complete its course. As shown in Fig. 1a, this BSC model must also incorporate a transient episode for genetic exchanges while geographical isolation is developing. Therefore, the proper test of BSC should be about how speciation ends (after a long period of allopatry with no gene flow), rather than how speciation begins (most likely with gene flow that diminishes in time). The extensive literature documenting gene flow during speciation is largely about speciation starting with a phase of gene flow [35–49]. In this sense, the collective evidence is akin to the ‘exceptions proving the rule (of BSC).’ The difficulty in rejecting BSC does not necessarily mean that strict allopatry is indeed the main mechanism of speciation. Genetic reasoning suggests that RI or ecological differentiation may easily evolve under continuous gene flow [5–7]. In Fig. 2, a possible test is proposed.

It would be wrong to perceive BSC (and its definitive acceptance or rejection) as no more than an issue of historical interest. The issue is fundamental because it is about the genetic nature of species. BSC is conceived on the assumption of strong

cohesiveness of the entire genome. By BSC, any genetic exchange between species should be maladaptive [2,5]. Modern genetics with all the transgenic experiments between species may seem to reject BSC as overly rigid as the transgenes usually function well in another species (unless the genes were chosen specifically for demonstrating interspecific differences; e.g. [28,52,77]). However, whether most randomly chosen genes have no fitness effects across species is unknown because the fitness decline only needs to be stronger than the strength of genetic drift, which is often $\ll 10^{-3}$. As demonstrated by Fang *et al.* [30], the absence of phenotypic effects associated with interspecific exchanges can have a fitness consequence measurable in long-term laboratory populations. If such observations are common, most genomes probably evolve with the broad cohesiveness beyond the resolution of modern molecular experiments that focus on large and measurable differences.

The resolution of the debate on BSC is important on many fronts of evolutionary biology. For example, in the taxonomic practice of naming species, BSC has not been highly relevant. This disconnect between adjacent fields has been a central dilemma in evolutionary biology [3–5,80,81]: do we or do we not accept the naming of new species that are not based on the (biological) species concept? Similarly, in the study of biodiversity, speciation mechanisms have not been relevant either. For example, a global hotspot of coastal biodiversity is in the Indo-western Pacific [82]. As pointed out by He *et al.* [22], the rejection that speciation mechanisms do not play a role in driving this spectacular biodiversity is the absence of geological features in the region that can impose the required geographical isolation. Ironically, the best-known feature of geographical isolation (i.e. the Isthmus of Panama) does not drive unusual biodiversity between the Pacific and Atlantic coasts. The disconnect between biodiversity and speciation analyses appears to be rooted in the conception of BSC on how species are formed.

There are many other discrepancies that reflect the dominance of BSC in modern evolutionary biology. The discrepancies can be resolved only when BSC is convincingly accepted or rejected. As stated, cases of sympatric speciation (e.g. [62–66]) would be a powerful rejection of BSC if micro-allopatry could be ruled out and gene flow could be demonstrated. Finally, the MIM cycle model cited above [22] offers a novel alternative, whereby phases of isolation and migration are interspersed during speciation. MIM cycles, a rejection of BSC, nevertheless retain some of its key features [22]. Many newer approaches to testing BSC will be needed, as a recent example shows [79].

SUPPLEMENTARY DATA

Supplementary data are available at [NSR](#) online.

FUNDING

This study was supported by the National Natural Science Foundation of China (91731301, 31600182, 31830005 and 31730046), the National Key Research and Development Plan (2017FY100705), Guangdong Basic and Applied Basic Research Foundation (2019A1515010752), the Fundamental Research Funds for the Central Universities (17lgpy99), and the Chang Hungta Science Foundation of Sun Yat-Sen University.

REFERENCES

1. Darwin C. *On the Origins of Species*. London: Murray, 1859.
2. Mayr E. *Animal Species and Evolution*. Cambridge: Harvard University Press, 1963.
3. Futuyma DJ and Kirkpatrick M. *Evolution*. Sunderland: Sinauer Associates, Inc., 2017.
4. Coyne JA and Orr HA. *Speciation*. Sunderland: Sinauer Associates, Inc., 2004.
5. Wu C.-I. The genic view of the process of speciation. *J Evol Biol* 2001; **14**: 851–65.
6. Van Alphen JJM and Seehausen O. Sexual selection, reproductive isolation and the genic view of speciation. *J Evol Biol* 2001; **14**: 874–5.
7. Rieseberg LH and Burke JM. A genic view of species integration. *J Evol Biol* 2001; **14**: 883–6.
8. Mayr E. Wu's genic view of speciation. *J Evol Biol* 2001; **14**: 866–7.
9. Wu C.-I and Davis AW. Evolution of postmating reproductive isolation: The composite nature of haldane's rule and its genetic bases. *Am Nat* 1993; **142**: 187–212.
10. Wilkins CJS. Summary of 26 species concepts. <https://ncse.ngo/species-kinds-and-evolution>
11. Futuyma DJ. *Evolution*. Sunderland: Sinauer Associates, Inc., 2013.
12. Paterson HEH. The Recognition Concept of Species. In: Vrba E (ed.) *Species and Speciation*. Pretoria: Transvaal Museum, 1985, 21–9.
13. Templeton AR. The meaning of species and speciation: a genetic perspective. In: Otte D and Endler JA (eds.) *Speciation and Its Consequences*. Sunderland: Sinauer, 1989, 3–27.
14. Marques DA, Meier JI and Seehausen O. A combinatorial view on speciation and adaptive radiation. *Trends Ecol Evol* 2019; **34**: 531–44.
15. Meier JI, Marques DA and Mwaiko S *et al.* Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat Commun* 2017; **8**: 14363.
16. Rieseberg LH. Hybrid origins of plant species. *Annu Rev Ecol Syst* 1997; **28**: 359–89.
17. Ridley M. *Evolution*. Malden: Blackwell Science, 2004.
18. Dobzhansky T. *Genetics and the Origin of Species*. New York: Columbia University Press, 1937.

19. Muller HJ. Isolating mechanisms, evolution and temperature. *Biol Symp* 1942; **6**: 71–125.
20. Knowlton N, Weigt LA and Solórzano LA *et al*. Divergence in proteins, mitochondrial DNA, and reproductive compatibility across the Isthmus of Panama. *Science* 1993; **260**: 1629–32.
21. Lessios HA. The Great American Schism: Divergence of Marine Organisms After the Rise of the Central American Isthmus. *Annu Rev Ecol Syst* 2008; **39**: 63–91.
22. He Z, Li X and Yang M *et al*. Speciation with gene flow via cycles of isolation and migration: Insights from multiple mangrove taxa. *Natl Sci Rev* 2019; **6**: 275–88.
23. Presgraves DC. The molecular evolutionary basis of species formation. *Nat Rev Genet* 2010; **11**: 175–80.
24. Haffer J. Speciation in Amazonian forest birds. *Science* 1969; **165**: 131–7.
25. Mallet J. Hybridization as an invasion of the genome. *Trends Ecol Evol* 2005; **20**: 229–37.
26. Abbott R, Albach D and Ansell S *et al*. Hybridization and speciation. *J Evol Biol* 2013; **26**: 229–46.
27. Wu C and Palopoli MF. Genetics of postmating reproductive isolation in animals. *Annu Rev Genet* 1994; **28**: 283–308.
28. Ting CT, Tsaur SC and Wu ML *et al*. A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* 1998; **282**: 1501–4.
29. Wu CI and Ting CT. Genes and speciation. *Nat Rev Genet* 2004; **5**: 114–22.
30. Fang S, Yukilevich R and Chen Y *et al*. Incompatibility and competitive exclusion of genomic segments between sibling *Drosophila* species. *PLoS Genet* 2012; **8**: e1002795.
31. Osada N and Wu C.-I. Inferring the mode of speciation from genomic data. *Genetics* 2005; **169**: 259–64.
32. Ting C-T, Tsaur S-C and Wu C.-I. The phylogeny of closely related species as revealed by the genealogy of a speciation gene, *Odysseus*. *Proc Natl Acad Sci USA* 2000; **97**: 5313–6.
33. Cruickshank TE and Hahn MW. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol Ecol* 2014; **23**: 3133–57.
34. Yang M, He Z and Shi S *et al*. Can genomic data alone tell us whether speciation happened with gene flow? *Mol Ecol* 2017; **26**: 2845–9.
35. Malinsky M, Challis RJ and Tyers AM *et al*. Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science* 2015; **350**: 1493–8.
36. Fontaine MC, Pease JB and Steele A *et al*. Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science* 2015; **347**: 1258524.
37. Brandvain Y, Kenney AM and Flagel L *et al*. Speciation and Introgression between *Mimulus nasutus* and *Mimulus guttatus*. *PLoS Genet* 2014; **10**: e1004410.
38. Lamichhaney S, Berglund J and Almén MS *et al*. Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* 2015; **518**: 371–5.
39. Schumer M, Xu C and Powell DL *et al*. Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science* 2018; **360**: 656–60.
40. Martin SH, Dasmahapatra KK and Nadeau NJ *et al*. Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Res* 2013; **23**: 1817–28.
41. Edelman NB, Frandsen PB and Miyagi M *et al*. Genomic architecture and introgression shape a butterfly radiation. *Science* 2019; **366**: 594–9.
42. Toews DPL, Taylor SA and Vallender R *et al*. Plumage Genes and Little Else Distinguish the Genomes of Hybridizing Warblers. *Curr Biol* 2016; **26**: 2313–8.
43. Turner TL, Hahn MW and Nuzhdin SV. Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biol* 2005; **3**: 1572–8.
44. Harr B. Genomic islands of differentiation between house mouse subspecies. *Genome Res* 2006; **16**: 730–7.
45. Ellegren H, Smeds L and Burri R *et al*. The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature* 2012; **491**: 756–60.
46. Renaut S, Grassa CJ and Yeaman S *et al*. Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nat Commun* 2013; **4**: 1827.
47. Carneiro M, Albert FW and Afonso S *et al*. The genomic architecture of population divergence between subspecies of the European Rabbit. *PLoS Genet* 2014; **10**: e1003519.
48. Clarkson CS, Weetman D and Essandoh J *et al*. Adaptive introgression between *Anopheles* sibling species eliminates a major genomic island but not reproductive isolation. *Nat Commun* 2014; **5**: 4248.
49. Poelstra JW, Vijay N and Bossu CM *et al*. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* 2014; **344**: 1410–4.
50. O'Dea A, Lessios HA and Coates AG *et al*. Formation of the Isthmus of Panama. *Sci Adv* 2016; **2**: e1600883.
51. Yang M. *Mangrove Speciation Under Intermittent Gene Flow and Mikania Micrantha Invasion Mechanism Research*. Guangzhou: Sun Yat-Sen University, 2016: 57–77.
52. Sun S, Ting CT and Wu C.-I. The normal function of a speciation gene, *Odysseus*, and its hybrid sterility effect. *Science* 2004; **305**: 81–3.
53. Sawamura K, Davis AW and Wu C.-I. Genetic analysis of speciation by means of introgression into *Drosophila melanogaster*. *Proc Natl Acad Sci* 2000; **97**: 2652–5.
54. Cabot EL, Davis AW and Johnson NA *et al*. Genetics of reproductive isolation in the *Drosophila simulans* clade: Complex epistasis underlying hybrid male sterility. *Genetics* 1994; **137**: 175–89.
55. Palopoli MF and Wu C.-I. Genetics of hybrid male sterility between *Drosophila* sibling species: A complex web of epistasis is revealed in interspecific studies. *Genetics* 1994; **138**: 329–41.
56. Perez DE and Wu C.-I. Further characterization of the *Odysseus* locus of hybrid sterility in *Drosophila*: One gene is not enough. *Genetics* 1995; **140**: 201–6.
57. Gavrilits S. Evolution and speciation on hole adaptive landscapes. *Trends Ecol Evol* 1997; **12**: 307–12.
58. Wu C.-I. A stochastic simulation study on speciation by sexual selection. *Evolution* 1985; **39**: 66–82.
59. Kondrashov AS. Deleterious mutations as an evolutionary factor: 1. The advantage of recombination. *Genet Res* 1984; **44**: 199–217.
60. Kondrashov AS. Deleterious mutations as an evolutionary factor. II. Facultative apomixis and selfing. *Genetics* 1985; **111**: 635–53.
61. Dieckmann U and Doebeli M. On the origin of species by sympatric speciation. *Nature* 1999; **400**: 354–7.
62. Barluenga M, Stölting KN and Salzburger W *et al*. Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature* 2006; **439**: 719–23.
63. Poelstra JW, Richards EJ and Martin CH. Speciation in sympatry with ongoing secondary gene flow and a potential olfactory trigger in a radiation of Cameroon cichlids. *Mol Ecol* 2018; **27**: 4270–88.
64. Schliewen UK, Kocher TD and McKaye KR *et al*. Evidence for sympatric speciation? *Nature* 2006; **444**: E12–3.
65. Pagan I and Holmes EC. Long-Term Evolution of the Luteoviridae: Time Scale and Mode of Virus Speciation. *J Virol* 2010; **84**: 6177–87.

66. Meyer JR, Dobias DT and Medina SJ *et al.* Ecological speciation of bacteriophage lambda in allopatry and sympatry. *Science* 2016; **354**: 1301–4.
67. Endler JA. *Geographic Variation, Speciation, and Clines*. Princeton: Princeton University Press, 1977.
68. Gavrillets S, Li H and Vose MD. Patterns of parapatric speciation. *Evolution* 2000; **54**: 1126–34.
69. Delmore KE, Toews DPL and Germain RR *et al.* The genetics of seasonal migration and plumage color. *Curr Biol* 2016; **26**: 2167–73.
70. Bay RA, Arnegard ME and Conte GL *et al.* Genetic coupling of female mate choice with polygenic ecological divergence facilitates stickleback speciation. *Curr Biol* 2017; **27**: 3344–9.
71. Tavares H, Whibley A and Field DL *et al.* Selection and gene flow shape genomic islands that control floral guides. *Proc Natl Acad Sci USA* 2018; **115**: 11006–11.
72. Van Belleghem SM, Rastas P and Papanicolaou A *et al.* Complex modular architecture around a simple toolkit of wing pattern genes. *Nat Ecol Evol* 2017; **1**: 52.
73. Küpper C, Stocks M and Risse JE *et al.* A supergene determines highly divergent male reproductive morphs in the ruff. *Nat Genet* 2016; **48**: 79–83.
74. Coulthart MB and Singh RS. High level of divergence of male-reproductive-tract proteins, between *Drosophila melanogaster* and its sibling species, *D. simulans*. *Mol Biol Evol* 1988; **5**: 182–91.
75. Hollocher H, Ting CT and Wu ML *et al.* Incipient speciation by sexual isolation in *Drosophila melanogaster*: Extensive genetic divergence without reinforcement. *Genetics* 1997; **147**: 1191–201.
76. Tao Y, Zeng ZB and Li J *et al.* Genetic dissection of hybrid incompatibilities between *Drosophila simulans* and *D. mauritiana*. II. Mapping hybrid male sterility loci on the third chromosome. *Genetics* 2003; **164**: 1399–418.
77. Shi L, Luo X and Jiang J *et al.* Transgenic rhesus monkeys carrying the human MCPH1 gene copies show human-like neoteny of brain development. *Natl Sci Rev* 2019; **6**: 480–93.
78. Ting CT, Takahashi A and Wu C.-I. Incipient speciation by sexual isolation in *Drosophila*: Concurrent evolution at multiple loci. *Proc Natl Acad Sci* 2001; **98**: 6709–13.
79. Wang X, Guo Z and Xu S *et al.* Genes and the species concept - How much of the genomes can be exchanged? *bioRxiv* 2019: 622019.
80. Chen JM, Poyarkov NA and Suwannapoom C *et al.* Large-scale phylogenetic analyses provide insights into unrecognized diversity and historical biogeography of Asian leaf-litter frogs, genus *Leptolalax* (Anura: Megophryidae). *Mol Phylogenet Evol* 2018; **124**: 162–71.
81. Liu Z, Chen G and Zhu T *et al.* Prevalence of cryptic species in morphologically uniform taxa – Fast speciation and evolutionary radiation in Asian frogs. *Mol Phylogenet Evol* 2018; **127**: 723–31.
82. Hoeksema BW. Delineation of the Indo-Malayan Centre of Maximum Marine Biodiversity: The Coral Triangle. In: Renema W (ed.) *Biogeography, Time, and Place: Distributions, Barriers, and Islands*. Dordrecht: Springer, 2007, 117–78.