Trends in Ecology & Evolution

CellPress REVIEWS

Opinion

Sympatric Speciation in the Genomic Era

Andrew D. Foote[1],*

Sympatric speciation has been of key interest to biologists investigating how natural and sexual selection drive speciation without the confounding variable of geographic isolation. The advent of the genomic era has provided a more nuanced and quantitative understanding of the different and often complex modes of speciation by which sympatric sister taxa arose, and a reassessment of some of the most compelling empirical case studies of sympatric speciation. However, I argue that genomic studies based on contemporary populations may never be able to provide unequivocal evidence of true primary sympatric speciation, and there is a need to incorporate palaeogenomic studies into this field. This inability to robustly distinguish cases of primary and secondary 'divergence with gene flow' may be inconsequential, as both are useful for understanding the role of large effect barrier loci in the progression from localised genic isolation to genome-wide reproductive isolation. I argue that they can be of equivalent interest due to shared underlying mechanisms driving divergence and potentially leaving similar patterns of coalescence.

Trends

Recent studies have highlighted that novel genomics approaches and genome-wide data can provide new insights into the geographical context and underlying processes of sympatric speciation.

Sympatric speciation is the evolutionary divergence and reproductive isolation of sister species arising from a single ancestral species in the absence of any barriers to gene flow. As such, this apparently rare phenomenon offers important insights into the role of selection in driving speciation.

Genomic studies have reassessed what were thought to be compelling empirical examples of sympatric speciation. In some cases, these studies have found evidence for multiple colonisations and homogenisation of the genomes of the two waves of colonisers upon secondary contact. In other examples, the findings have strengthened the case for divergence in sympatry. Understanding the biogeography and evolutionary history of the genomic regions underlying ecological adaptation and sexual selection is fundamental to understanding how speciation can progress when driven by natural and sexual selection without any period of physical separation.

A Century of Contention over Sympatric Speciation

Primary sympatric speciation is the evolution of reproductive isolation without geographic barriers, in which new species arise from a single ancestral population [1–5]. As these criteria do not allow for any physical separation between the incipient species, the potential for interbreeding and gene flow remains throughout the speciation process, from inception to completion. **Recombination** (see Glossary) can therefore break up linkage between alleles beneficially associated with environmental variation and alleles associated with incompatibilities and reproductive isolation [6]. As such, it is the most extreme, restrictive and arguably the most controversial scenario of **divergence with gene flow** [7–11]. Thus, the existence and relevance of this mode of speciation in nature has been hotly debated for over a century [1–11]. The continued great interest for evolutionary biologists in sympatric speciation is understanding the seemingly rare conditions and processes under which natural and sexual selection can drive ecological divergence and reproductive isolation in a continuously distributed population [4,7], as compared with allopatric speciation, in which geographic barriers initiate reproductive isolation and population divergence follows [2–8]. Under the latter scenario, it can be difficult to establish the extent of the role of selection due to ecological variation relative to intrinsic barriers developed through drift during geographic isolation in promoting reproductive isolation [12].

After over a century of debate, and despite its theoretical plausibility and some apparently compelling empirical examples, many facets of sympatric speciation remain controversial. Given this, a recent review on speciation argued that the debate over allopatric versus sympatric speciation was unproductive and should not be a significant part of the future research agenda [13]. However, as per the oft-quoted prediction by Mayr: 'Sympatric speciation is like the Lernaean hydra which grew two new heads whenever one of its old heads was cut off . . . the issue will be raised again at regular intervals' [8]. The advent of high-throughput sequencing, coupled with the development and application of population genomic methods

[1]Molecular Ecology and Fisheries Genetics Laboratory, School of Biological Sciences, Bangor University, Bangor, Gwynedd LL57 2UW, UK

*Correspondence: a.foote@bangor.ac.uk (A.D. Foote).

https://doi.org/10.1016/j.tree.2017.11.003

that allow the inference of complex evolutionary histories, has led to a resurgent interest in sympatric speciation and a reassessment of some of the most compelling empirical case studies [14–16].

In the genomic era, we can now quantify the genetic contribution of one or more ancestral populations to contemporaneously sampled sympatric daughter species. These advances have led to some of the most compelling examples of primary sympatric speciation being reconsidered as a product of multiple colonisations and secondary contact. Other examples appear to be robust. However, here I argue that such hindcasting approaches have limited ability to distinguish between periods of spatial overlap, but the absence of gene flow (i.e., when no coalescence take place between the ancestral incipient species), and the absence of gene flow during periods of spatial separation. I propose that a forward-in-time approach utilising palaeogenomics may be a complementary approach that could leverage additional information in some contexts. Lastly, I consider whether primary and secondary sympatric speciation represent a mechanistic dichotomy. I suggest that primary and secondary contact can leave a similar genomic signature when speciation is driven by tightly clustered or large effect loci. Arguably, population genomic studies should place less focus on whether study systems result from primary or secondary contact and instead focus on the mechanistic aspects of the genomic architecture; thereby facilitating progress in identifying the conditions and processes under which natural and sexual selection can drive speciation, without extrinsic barriers to gene flow [13].

Genomic Insights into the Ancestral Context of Sympatric Speciation

A compelling empirical case study of primary sympatric speciation requires the robust inference of past biogeography; specifically, that the present-day sympatric daughter species arose from a common ancestral population, with no period of geographic isolation (Box 1). Prior to the genomic era, empiricists used phylogenetics and assumed that the geographic distribution of the ancestral population was the same as the present-day daughter species, if they formed **monophyletic** species pairs or flocks in geographically isolated ‘island’ habitats [17–24]. However, a major limitation of the inference of sympatric speciation from the monophyletic relationship among sympatric species is that monophyly may result from several processes other than true sympatric speciation (Figure 1). Modelling speciation as a bifurcating tree presents a point estimate of this evolutionary process [14] and does not consider the possibility that species derived ancestry from multiple source populations [25–27]. This is a key flaw with the criteria of Coyne and Orr [4]; monophyly of sympatric sister species is consistent with, but not exclusive to, a scenario of sympatric speciation. It does not provide conclusive evidence that present-day sympatric sister species emerged from a single colonisation, nor does it reject the alternative scenario of multiple colonisations in which monophyly results from introgression upon secondary contact [7,28].

However, these different scenarios do typically generate different patterns of genome-wide ancestry that can be used to distinguish between them. Under a scenario of sympatric speciation from a single source population, the daughter species will share a common ancestry, with segregating alleles being mainly those that are recently derived or were at low frequency in the ancestral population [14–16]. Alternatively, if sympatric sister species are the result of multiple colonisations and gene flow upon secondary contact, then each species should share differing proportions of ancestry with source outgroups (Figure 1). We can consider this as a continuum, from a single **panmictic** colonising population (Figure 1A) to colonisation by an admixed population (Figure 1B) and lastly multiple colonisations and secondary contact following periods of geographic isolation (Figure 1C). This is a

Glossary

Barrier loci: genetic loci that cause reduced gene flow between speciating taxa at a localised region of the genome.
Coalescent: when two lineages sampled from different populations merge back in time in a commonly shared ancestral lineage.
Disruptive selection: selection that favours extreme phenotypes over intermediate phenotypes within a population.
Divergence with gene flow: the build-up of genetic and phenotypic differences, despite ongoing exchange of genes. This differentiation is typically driven by disruptive natural selection. The term has been used inclusive of scenarios of divergence under ongoing gene flow upon secondary contact, and thus does not exclusively refer to sympatric speciation.
Ecomorph: a population that has distinctive ecological and morphological features.
Genomic islands: a region of the genome that is highly differentiated (estimated using F_{ST}) between taxa compared with the genome-wide mean level of differentiation.
Magic trait: a trait subject to divergent selection and a trait contributing to mate choice, which are pleiotropic expressions of the same gene(s).
Monophyletic: belonging to a clade containing all the descendants of a single ancestor.
Panmixia: random mating within a population.
Parapatric speciation: the evolution of reproductive isolation in the absence of geographical barriers to gene flow, in which the diverging populations have adjacent ranges.
Pleiotropic: an allele that has an effect on more than one trait.
Polymorphisms: genetic loci that have more than one allele.
Quantitative trait loci: genetic markers that are correlated with phenotype. These markers contain, or are linked to, genes and regulatory regions associated with quantitative phenotypic variation.
Recombination: the process by which genomic regions are exchanged and broken up, producing new combinations of alleles at different loci.

Recombination occurs during meiosis in eukaryotic cells.

Box 1. Pregenomic Era Criteria for Identifying Sympatric Speciation

In their classic view of speciation, Coyne and Orr [4] proposed four criteria that would need to be met in order for compelling case studies of sympatric speciation to be established. Given the restrictive conditions under which sympatric speciation is theoretically possible, these criteria for assessing empirical examples are equally stringent. Following the argument of Mayr [8], they place the burden of proof on sympatric speciation and assume allopatric speciation as the null hypothesis. The four criteria can arguably be split into two components: one specifying the biogeographic conditions and the other specifying the genetic criteria under which an empirical case study would make a compelling example of sympatric speciation (Figure I).

Biogeographic Component

(i) Species must have largely or completely overlapping geographic range (Figure IA).

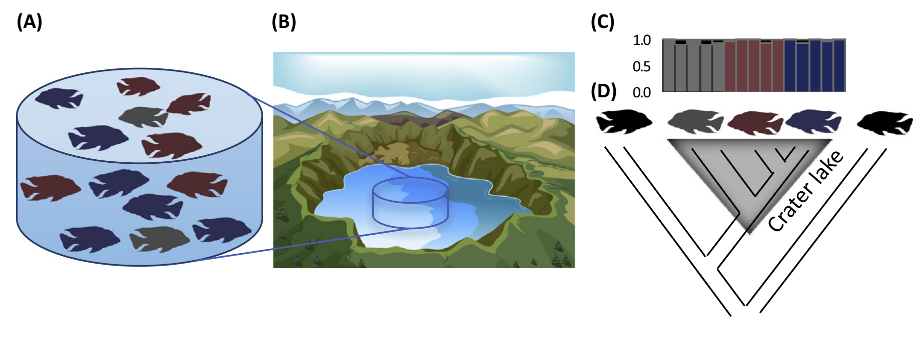
(ii) The biogeographic and evolutionary history of the groups must make the existence of an allopatric phase very unlikely (Figure IB).

Genetic Component

(iii) Speciation must show substantial reproductive isolation (Figure IC).

(iv) Sympatric species must be endemic sister species or an endemic monophyletic species flock (Figure ID).

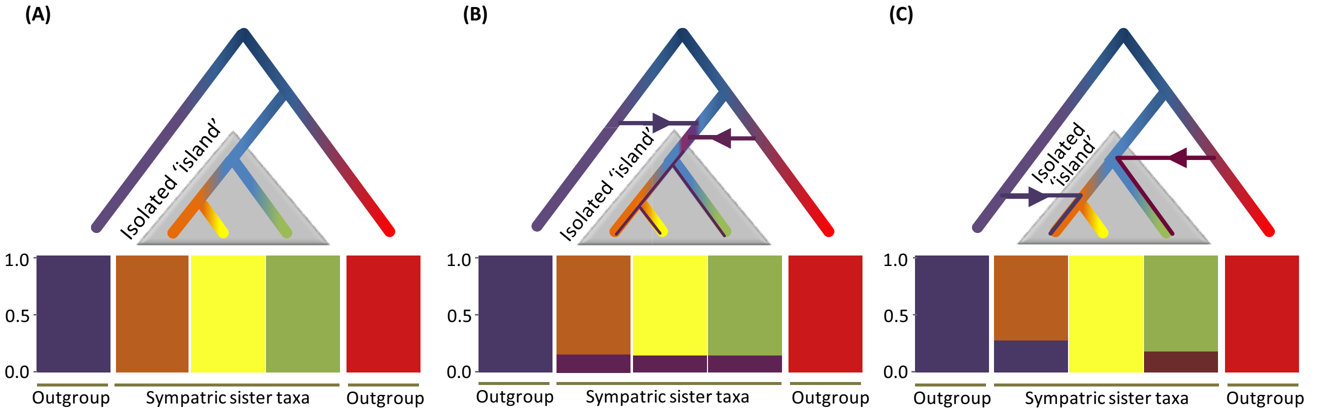
As with most aspects of the study of sympatric speciation, these criteria have been a point of contention. See Bolnick and Fitzpatrick [7] for an in-depth discussion and review of these conditions.



Trends in Ecology & Evolution

Figure I. Biogeographic and Genetic Criteria for Sympatric Speciation. Empirical case studies on crater lake cichlids were among the first to be considered as compelling examples of primary sympatric speciation [17–19]. Cichlid species in these studies had distributions that overlapped and different species were in ‘cruising range’ *sensu* Mayr [8]. The high rim of the caldera of these craters isolates the lake from neighbouring rivers, and the conical shape of the lake bottom prevents the formation of separate basins during periods of low water level [17]. Thus, there are no geographical barriers to gene flow within the crater lake. (C) Analyses of nuclear DNA markers suggest that gene flow occurs predominantly within rather than between species (illustrated here with an admixture plot) [19]. (D) Phylogenetic analyses show that cichlid species within each lake form a monophyletic clade with respect to outgroups from neighbouring river systems, suggesting that they radiated *in situ* from a single shared ancestral population [17–19].

representative, but not an exhaustive list of possible scenarios that could generate the same consensus phylogenetic pattern as sympatric speciation. Recently developed genomics methods can provide robust evidence of admixture and estimate ancestry proportions, even if gene flow events had occurred hundreds of generations ago and under scenarios of incomplete lineage sorting and demographic change [29–32]. For example, the closely related *D*-statistic (ABBA-BABA) and *f*-statistic tests identify taxa that share an excess of ancestry (measured as derived alleles and allele frequencies, respectively) with an outgroup [29,30]. The tract length of genomic regions inferred to have introgressed during secondary contact can



Trends in Ecology & Evolution

Figure 1. Evolutionary Histories That Could Result in a Monophyletic Relationship among Sympatric Sister Species. Schematic tree figures (top) are coloured to indicate changes in allele frequencies during divergence and introgression (indicated by horizontal arrows). Schematic ancestry palettes (bottom) are coloured to indicate the differences in ancestry proportions shared between the sympatric sister species and outgroups under each scenario. (A) Speciation follows a single colonisation of an isolated 'island' habitat and divergence during sympatry. Under this scenario, the three sympatric sister species would share a similar proportion of their ancestry with outgroups. (B) Colonisation of an isolated 'island' habitat preceded by admixture with the outgroups followed by a period of panmixia could also result in the three sympatric sister species sharing a similar proportion of their ancestry with outgroups; however, colonisation by a structured metapopulation could result in the amount of shared ancestry with outgroups differing among ecotypes. (C) Multiple independent colonisations of an isolated 'island' habitat over time, and episodic admixture upon secondary contact would result in the introgressed species sharing more of their ancestry with the outgroups most closely related to the source population of this secondary colonisation. These three examples are not meant to be exhaustive, but simply illustrative of how different evolutionary histories can result in the same majority-rule topology if evolutionary history is modelled as a single bifurcating tree. This figure is adapted from [14].

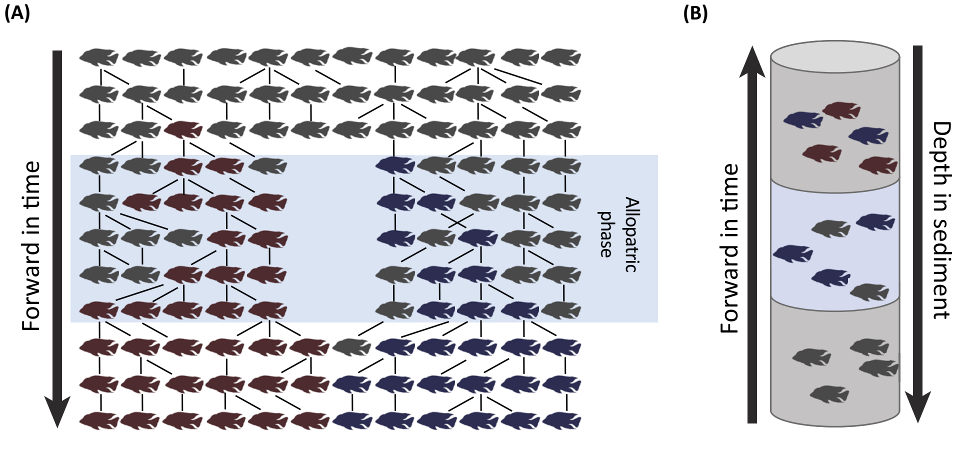
provide further information on the timing of gene flow, and whether introgression pre- or post-dated sympatric diversification [33,34].

The application of such a population genomics approach has reassessed the sympatric origins of arguably some of the most compelling empirical examples of sympatric speciation: monophyletic species pairs and flocks of cichlids found in small uniform crater lakes in Cameroon, Nicaragua and Tanzania [14–16]. The lakes were argued to be sufficiently small in size, ecologically monotonous with no microgeographical barriers and isolated from outside riverine populations by the crater rim, that sympatric speciation appeared to be the most likely biogeographical scenario under which these sister species had diverged [17,18]. In each case, cichlid species within the lakes have diverged in ecologically associated morphological traits, and show evidence of reproductive isolation and monophyly, consistent with sympatric origins [15,17–21,35]. However, analyses of genome-wide ancestry have revealed varying complexity in the evolutionary history of cichlids within each study area. These range from genomic ancestries that are best explained by multiple colonisations of Cameroon crater lakes and secondary gene flow following periods of allopatry [14]; to divergence in sympatry in Nicaraguan crater lakes, but following secondary colonisation events and admixture prior to the radiations within each lake [16]; to what appears to be speciation following a single colonisation in a Tanzanian crater lake, albeit with some gene flow from the lake to nearby outgroup populations [15].

These descriptive results can then be developed into demographic models, allowing the estimation of ancestral divergence times, effective population sizes and migration rates, and the testing of alternative evolutionary scenarios (e.g., [16,36,37]). However, modelling whether sympatric populations diverged with gene flow, or whether migration took place sometime after the populations had diverged, consistent with secondary contact, requires the estimation of the timing and the number of migration events [38–40]. These parameters can be intractable, as genomic data from present-day populations can be consistent with many migration and admixture scenarios, which result in the same coalescent times [39,40]. More general caveats also apply, for example, most models are oversimplified representations of biological reality, and only inputted models are tested. Model-based approaches are therefore best accompanied with model-free methods to identify a range of estimates for parameters, and scenarios to test. In additional, there is a need to exclude non-neutral loci and account for genome-wide variation in effective migration and recombination [37,41].

The biological realism and relevance of the classification of the mode of speciation into the discrete geographic categories such as sympatric, **parapatric** and allopatric have been questioned. Almost all candidate case studies of sympatric speciation have some degree of spatiotemporal differentiation between sister taxa, for example, due to the patchy distribution of preferred habitat [14,42–44]. To countenance this, some have suggested that the relationship between taxa during the speciation process may be better quantified in a population genetics framework that quantifies key parameters such as migration rate [42]. This approach, and modelling sympatric speciation in general, relies on assuming a starting point of panmixia in the ancestral population [5]. Yet this assumption of ancestral panmixia has been difficult or impossible to prove or reject in empirical case studies prior to the genomic era [42]. Others have argued for retaining a spatial component of sympatric speciation, in accordance with Mayr's definition [8]: that speciating sister taxa should be in 'cruising range' of each other throughout the speciation process [44]. However, in each case, the geographic context of speciation is divided into artificially discrete categories, whether they be based on spatial or genetic measures of separation [11]. Instead, the geographic context of speciation is perhaps best viewed as a graded continuum [10,11]. The genomic approaches outlined above estimate the contribution of the shared ancestral population and any other contributing outgroup populations to the ancestry of the daughter species, thereby providing a continuous and quantitative measure of the context and mode of speciation. This still does not fully resolve the uncertainty in the geographic context of divergence. For example, even among sympatric taxa with no detectable contribution from ancestral outgroups, as in Figure 1A, there may have been periods of spatial segregation among currently sympatric sister taxa. Ultimately, our ability to reconstruct the evolutionary history of sympatric sister taxa back to the shared ancestral population using backward-in-time genomic approaches is constrained to being able to identify periods of gene flow through **coalescent** events, but this does not distinguish periods of spatial overlap without gene flow from periods of spatial isolation.

Because of the timescales over which evolutionary processes such as adaptation and speciation take place, forward-in-time approaches are rarely utilised due to the limitations on the number of generations that can be sampled. However, the advent of palaeogenomics is expanding the scope of timescales over which we can sample genomes and look at genetic change from an ancestral population going forward in time to daughter species, and can complement hindcasting from contemporaneously sampled genomes. For example, sediment cores from postglacial lakes can be used to sample lineages from the time the glaciers retreated to the present day (Figure 2). Such an approach has recently been applied to extract DNA from sediment of two lakes in Sweden, spanning the past 10 000 years, to reconstruct the colonisation and connectivity between whitefish (*Coregonus lavaretus*) ecotypes [45]. Whilst only very low concentrations of DNA are found in sediments, the sequencing of hard parts within the different layers of the sediment core (e.g., bones or spines) can yield genomic sequences that allow the tracking of genomic changes at **quantitative trait loci** (QTL) forwards in time.



Trends in Ecology & Evolution

Figure 2. Palaeogenomic Sampling of Divergent Speciating Lineages from Sediment Cores. (A) An isolated lake is founded by a single lineage (grey). During a period of spatial separation within the lake, two daughter lineages are derived (red and blue) and are adapted to local ecological conditions and associated mate choice. Upon secondary contact, mate choice maintains this segregation of the two lineages. Sampling the contemporary lineages from the lake, one would reconstruct an ancestral history similar to that portrayed in Figure 1A, and would be unable to distinguish whether reproductive isolation had become established despite lineages having remained spatially overlapped throughout their postcolonisation history, or, as in this case, whether reproductive isolation had developed during a period of spatial isolation. (B) Sampling sediment cores of lakes and sequencing the sediment layers, or hard body parts within them, provides a time series of genomic data that can elucidate the temporal patterns of spatial overlap, in addition to the chronology and tempo of genomic changes associated with adaptation and speciation (i.e., the onset of selection). In the example shown, the sediment core has been drilled in an area used exclusively by the blue lineage during the allopatric phase. Sampling multiple cores would establish the approximate distribution of both lineages through space and time.

The Genomic Architecture of Sympatric Speciation

The genomic architecture of a trait can be summarised as the number of underlying loci, their effect size and additivity and their physical spacing across the genome. In addition to being shaped by recent and ongoing selection, this genomic architecture can be influenced by processes that include demographic history, linked selection in the ancestral population, recent and ongoing selection, and recombination rate [46].

Key questions in the study of sympatric speciation are how a genomic architecture shaped by gradual, incremental changes that occur under natural selection can account for rapid bursts of adaptive divergence; how localised genomic changes result in genome-wide reproductive isolation; and how they can overcome the homogenising effect of ongoing gene flow [47–49]. Over the past decade genomic studies of adaptation have progressed from investigating single or a few candidate genes to genome-wide studies, and have highlighted how divergence linked to adaption can be widespread across the genome. Yet the chronology of genic change during speciation, and how this progresses from individual '**barrier loci**', through to genome-wide differentiation (and how to study these processes), is still contentious and widely debated (see [49] and associated commentaries).

One of the primary approaches to exploring these questions has been to compare genome-wide variation in differentiation (F_{ST}) of allele frequencies across the 'speciation continuum', that is, between multiple pairs of sympatric and allopatric sister taxa that are at different stages of divergence [47,48]. This approach has been applied to multiple taxa with varied results. While most such studies to date have shown a progressive increase in the build-up of mean genome-wide differentiation across the speciation continuum [50–53], and some have highlighted important barrier loci that reduce localised effective migration within some genomic regions due to being associated with adaptation and/or reproductive isolation [54,55], many of these studies have identified alternative underlying causes of heterogeneity in the landscape of genomic differentiation [50–52]. These include reduced diversity from linked selection in the ancestral population, for example, due to background selection removing deleterious variants [56]; background selection is in-turn associated with variation in recombination rate and gene density in regions such as centromeres [57,58]; and selection on genome-wide smaller effect loci underlying polygenic traits. The genomic background of these different processes can then mask any potential signal from barrier loci associated with adaptation or reproductive isolation. However, young examples of sympatric speciation may generate rare exemplar study systems, in which there are clear '**genomic islands**', which contain barrier loci associated with reproductive isolation and ecological diversification.

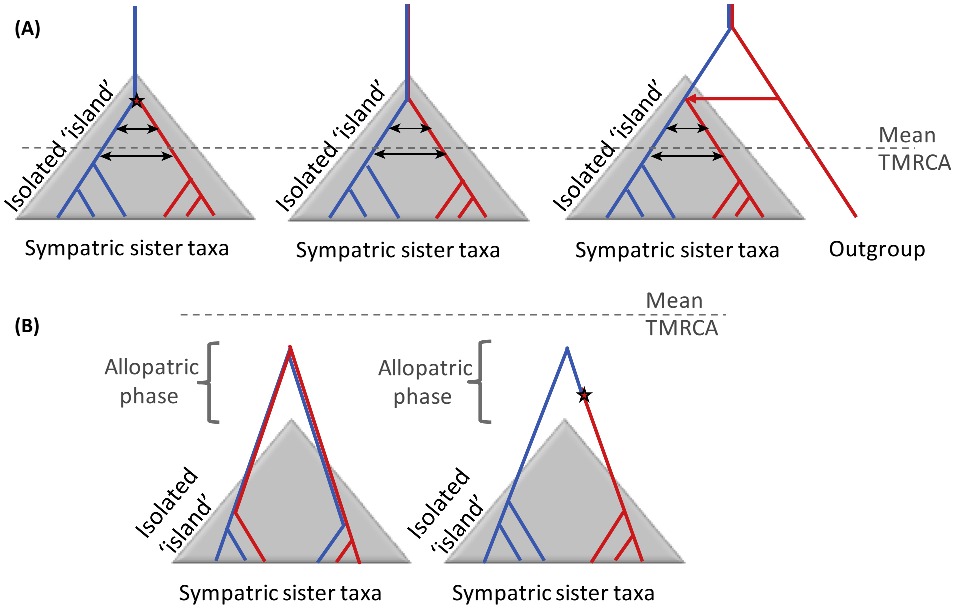
The effect size of a locus on a phenotypic trait has a positive correlative relationship with **pleiotropy** and deleterious effects [59]; therefore, adaptation is predicted to typically progress due to small changes in frequency across many alleles, each with a small additive phenotypic effect [60]. However, as noted earlier, in scenarios of ongoing gene flow during sympatry, recombination would be expected to break up linkage between loci associated with ecological adaptation and those associated with mate preference, thus counteracting ecologically driven speciation [6–9]. In addition, the strength of selection on a locus is not only a function of its effect size and its interaction with the environment, but also a function of effective population size (*N*e). The more robust examples of primary sympatric speciation are typically those that have colonised a remote, or closed, ecosystem prior to diverging, for example, Lord Howe Island flora [23,24] and crater lake cichlids [17–19]. Thus, it seems realistic that only a small number of initial colonisers founded these island or closed ecosystems. This founder effect is expected to greatly lower selection coefficients at loci of small effect that act additively on traits. Therefore, traits associated with ecological variation or mate choice that diverge during sympatric speciation are more likely to be determined by loci tightly linked to each other in genomic regions of low recombination such as inversions [46,61], or be synergistically pleiotropic, that is, so-called **magic traits**, which have a role in both ecological adaptation and assortative mating [62]. Therefore, these study systems are those that we expect barrier loci of large effect to be differentiated against a homogenous genome-wide background.

Recent genomic studies investigating QTL in model systems for speciation with gene flow have largely validated these predictions. For example, in Midas cichlids in Nicaraguan crater lakes, the highest effect size QTL for body shape and pharyngeal jaw morphology, both traits that show ecological-associated variation [20,21], are tightly clustered on a single chromosome and allele frequencies at these loci segregate in sympatric sister species [63]. Comparison of the genomes of benthic and littoral **ecomorphs** of *Astatotilapia* cichlids from a Tanzanian crater lake found regions of high differentiation and high divergence clustered mainly in five linkage groups harbouring genes associated with morphology and optical sensitivity, and therefore ecological variation and mate choice [15]. A recent study on sympatric populations of monkey flower species *Mimulus laciniatus* and *Mimulus guttatus* found that a few large effect size QTL explained much of the variance in flowering time and flower size traits [64]. Differences in flowering time are thought to be locally adaptive: *M. laciniatus* is found on dry exposed rocky outcrops and flowers earlier than *M. guttatus* to avoid the seasonal drought; and act as a prezygotic barrier to gene flow, therefore qualifying as a 'magic trait' [64]. Allochrony also plays a role in reproductive isolation between sympatric hawthorn and apple-infesting host races of the *Rhagoletis pomonella* fly, which differ in the intensity and timing of diapause [65]. Singlenucleotide polymorphism (SNP) loci associated with the timing of diapause onset and diapause intensity were in several tightly linked clusters, thought to be within inversions [66].

The findings of these empirical studies are highly concordant with the predictions of most theoretical models of sympatric speciation, which require linkage between loci associated with reproductive isolation and loci associated with ecological adaptation, or pleiotropy in which ecological adaptation and reproductive isolation evolve simultaneously [67–69]. This contrasts with empirical examples in which a period of allopatry was important in segregating alleles associated with ecological variation. In examples of the latter scenario, intrinsic barriers can build up in many widespread genomic regions without recombination breaking them up during this allopatric phase. Thus, in many examples of sympatric speciation we anticipate large changes in allele frequencies at single or a few loci, while the rest of the genome is homogenised, until complete genome-wide isolation is established. Therefore, the coalescent times of the barrier loci are expected to pre-date the genome-wide time to most recent common ancestor (TMRCA) [70] (Figure 3). By contrast, if genome-wide polygenic adaptation and reproductive incompatibilities have evolved in allopatry, prior to secondary contact, then the TMRCA of the loci associated with reproductive isolation will be within the genome-wide range and need not be associated with large changes in allele frequencies, making them cryptic to genome-wide scan methods.

Strict primary divergence with gene flow may not be needed for studying the evolution of large effect barrier loci against a homogenous genomic background. In theory, this pattern could also be expected even if the genetic underpinning of divergent ecological adaptation and reproductive isolation develops during allopatry, and alleles then segregate again after an initial period of mixing upon secondary contact, provided there is genome-wide homogenisation upon secondary contact (Figure 3). An allopatric phase and/or introgression events can facilitate speciation by intensifying **disruptive selection** and introducing new genomic variation that can act as a substrate for segregating **polymorphisms** under natural and sexual selection. Guerrero and Hahn [71] recently suggested that balanced polymorphisms in the ancestral population could sort upon splitting into daughter species, either due to ecological variation selecting for alternate alleles or through selectively neutral sorting. They highlighted that such a process could explain the high absolute genetic divergence (D_{XY}), suggestive of an ancient divergence, in the few genomic islands found when comparing the littoral and benthic ecomorphs of the Tanzanian crater lake Massoko. The two ecomorphs are estimated to have diverged only 500–1000 years ago, having diverged from the putative source population 10 000 years ago in a crater lake that formed about 50 000 years ago [15]. Guerrero and Hahn [71] highlight that these regions containing putative balanced polymorphisms would form ‘genomic islands’ even without background F_{ST} and D_{XY} being lowered due to genome-wide homogenisation from gene flow. However, it is not hard to imagine that these cichlid ecomorphs could have arisen and collapsed multiple times since colonising the crater, for example, due to episodic changes in water depth. If negative frequency-dependent selection maintained ecologically adaptive polymorphisms even when the two forms collapse into an otherwise homogenous population, such a process of repeated collapse and vicariance could mask any genomic signature of divergent origins in the present-day populations, with the exception of balanced polymorphisms, which would coalesce much further back in time than the genome-wide mean TMRCA (Figure 3).

Lineage sorting and high genomic differentiation are also found at loci of large effect size in the partially sympatric benthic–limnetic species pairs of three-spined sticklebacks found in several lakes in British Columbia, Canada, which are hypothesised to have originated from a secondary



Trends in Ecology & Evolution

Figure 3. Patterns of Genomic Differentiation due to Sympatric and Allopatric Divergence. (A) Schematic tree figures (top) are coloured to indicate changes in allele frequencies at a large effect barrier locus during divergence and introgression (indicated by red horizontal arrow). During divergence with gene flow in sympatry, there is genome-wide homogenisation due to ongoing gene flow (indicated by black horizontal arrows). The segregation of alleles in different incipient species at large effect barrier loci associated with ecological adaptation and reproductive isolation will pre-date the mean genome-wide coalescent time. This should be true whether the segregating alleles in barrier loci result from *de novo* mutations (indicated by star) during sympatry (left), standing variation that was present prior to the sympatric phase, including from balanced polymorphisms (middle), introgression and secondary contact (right). Thus, such loci should stand out against a background of homogenised loci in genome-wide scans. (B) In many scenarios where genome-wide incompatibilities have evolved during allopatry, which preclude gene flow upon secondary contact, TMRCA of alleles at incompatibility loci will fall within the range of the genome-wide mean TMRCA, and both will pre-date secondary contact. This pattern may not be ubiquitous. For example, balanced polymorphisms that segregated upon speciation would still have a TMRCA that pre-dated the genome-wide mean [71]. Abbreviation: TMRCA, time to most recent common ancestor.

invasion [72]. A principal component analysis of genome-wide neutrally evolving SNPs found a pattern of clustering by lake [73], which would be consistent with independent divergence of the benthic and limnetic forms of stickleback within each lake. However, SNPs evolving under natural selection grouped individuals by ecological niche, with further clustering of the older benthic form with geographically proximate single-form freshwater populations, whilst the younger limnetic form clustered more closely with marine populations [73]. These results are consistent with reuse of standing genetic variation from a second marine-to-freshwater colonisation, which then provided the raw genetic material for divergence within each lake driven by disruptive selection. Thus, the adaptation and speciation loci coalesce much further back in time than the mean TMRCA of unlinked neutral loci. A further example is the sympatric hawthorn and apple-infesting races of *R. pomonella* fruit fly, in which the inversion polymorphism influencing diapause traits evolved during an allopatric phase greater than a million years ago [74].

Concluding Remarks

In the genomic era, sympatric speciation continues to be a controversial and much-debated phenomenon. The exemplar study systems, such as crater lake cichlids of Cameroon, which had convinced even the most hardened sceptics [4], have been called into question. Genome sequences provide the unprecedented means to reconstruct the ancestry of contemporary populations; for example, identifying where sympatric sister taxa that were thought to represent a monophyletic group are instead derived from multiple ancestral source populations [14]. However, there remains a bias towards being able to disprove primary sympatric speciation, whilst generating conclusive evidence in support of primary sympatric speciation based on hindcasting using modern genomes remains elusive. I suggest that palaeogenomics may have a complementary role to play in future studies; for example, the sequencing of DNA from sediment cores can identify the temporal patterns of spatial overlap between two speciating lineages, even in the absence of gene flow. Lastly, the great interest of biologists in sympatric speciation has been how two lineages can diverge and become reproductively isolated in the absence of extrinsic barriers. In the genomic era, we can study this process at the genic level. In this opinion article, I have highlighted several characteristics of the genomic underpinning of sympatric speciation, and that these can be found in examples of primary and secondary sympatric speciation. I therefore contend that it is the investigation of the process of sympatric speciation rather than a dogmatic search for true primary sympatric speciation that will be most valuable to our understanding of speciation and adaptation at the genomic level (see Outstanding Questions).

Outstanding Questions

Genomic reassessments of some of the compelling case studies have identified a signature of secondary contact and admixture. However, they have not always identified if these are causal to subsequent radiation, and whether the genetic components underlying eco-phenotypic associations and reproductive isolation result from alleles that segregated during allopatry and recombined during secondary contact. There are also compelling cases of sympatric speciation, for example, Lord Howe palms, that have yet to be compared with a wide range of candidate outgroups using genomic approaches to test for the possibility of divergence arising from secondary colonisations.

Is sympatric speciation rare in nature? Genomic approaches allow the investigation of the ancestry of species, which would not have been considered as compelling cases of sympatric speciation under the restrictive criteria of Coyne and Orr. Recent studies have shown that genomic analyses of ancestry to test for a signature of secondary contact can be applied to organisms such as killer whales that live in large ocean basins and have a huge dispersal capability. Investigation of the ancestry of a wider range of taxa should elucidate how common or rare a phenomenon sympatric speciation really is.

A key focus of future studies should be to investigate the concordance of empirical studies with prediction of theoretical models of sympatric speciation. In particular, the role of pleiotropy and so-called magic traits.

Acknowledgements

I would like to thank the editor, Paul Craze, and Jeff Feder and one anonymous reviewers for their constructive feedback, and Alex Papadopulos for useful discussions on this topic which greatly improved this manuscript. Financial support was provided by the Welsh Government and Higher Education Funding Council for Wales through the Sêr Cymru National Research Network for Low Carbon, Energy and Environment, and from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 663830.

References

1. Poulton, E.B. (1904) What is a species? *Trans. Entomol. Soc. Lond.* 1903, 77–116
2. Jordan, D.S. (1905) The origin of species through isolation. *Science* 22, 545–562
3. Mayr, E. (1942) *Systematics and Origin of Species,* Columbia University Press
4. Coyne, J. and Orr, H. (2004) *Speciation,* Sinauer Associates
5. Gavrilets, S. (2003) Models of speciation: what have we learned in 40 years? *Evolution* 57, 2197–2215
6. Slatkin, M. (1987) Gene flow and the geographic structure of natural populations. *Science* 236, 787–792
7. Bolnick, D.I. and Fitzpatrick, B.M. (2007) Sympatric speciation: models and empirical evidence. *Annu. Rev. Ecol. Evol. Syst.* 38, 459–487
8. Mayr, E. (1963) *Animal Species and Evolution,* Belknap
9. Felsenstein, J. (1981) Skepticism towards Santa Rosalia, or why are there so few kinds of animals? *Evolution* 35, 124–138
10. Jiggins, C.D. (2006) Sympatric speciation: why the controversy? *Curr. Biol.* 16, R333–R334
11. Butlin, R.K. *et al.* (2008) Sympatric, parapatric or allopatric: the most important way to classify speciation? *Philos. Trans R. Soc. Lond. B Biol. Sci.* 363, 2997–3007
12. Bierne, N. *et al.* (2011) The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Mol. Ecol.* 20, 2044–2072
13. Marie Curie Speciation Network (2012) What do we need to know about speciation? *Trends Ecol. Evol.* 27, 27–39
14. Martin, C.H. *et al.* (2015) Complex histories of repeated gene flow in Cameroon crater lake cichlids cast doubt on one of the clearest examples of sympatric speciation. *Evolution* 69, 1406–1422
15. Malinsky, M. *et al.* (2015) Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science* 350, 1493–1498
16. Kautt, A.F. *et al.* (2016) Multispecies outcomes of sympatric speciation after admixture with the source population in two radiations of Nicaraguan crater lake cichlids. *PLoS Genet.* 12, e1006157
17. Schliewen, U.K. *et al.* (1994) Sympatric speciation suggested by monophyly of crater lake cichlids. *Nature* 368, 629–632
18. Schliewen, U.K. and Klee, B. (2004) Reticulate sympatric speciation in Cameroonian crater lake cichlids. *Front. Zool.* 1, 5
19. Barluenga, M. *et al.* (2006) Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature* 439, 719–723
20. Elmer, K.R. *et al.* (2010) Rapid sympatric ecological differentiation of crater lake cichlid fishes within historic times. *BMC Biol.* 8, 60
21. Elmer, K.R. *et al.* (2010) Local variation and parallel evolution: morphological and genetic diversity across a species complex of neotropical crater lake cichlid fishes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365, 1763–1782
22. Filchak, K.E. *et al.* (2000) Natural selection and sympatric divergence in the apple maggot *Rhagoletis pomonella*. *Nature* 407, 739–742
23. Savolainen, V. *et al.* (2006) Sympatric speciation in palms on an oceanic island. *Nature* 441, 213–213

24. Papadopulos, A.S.T. *et al.* (2011) Speciation with gene flow on Lord Howe Island. *Proc. Natl. Acad. Sci. U. S. A.* 108, 13188–13193
25. Cavalli-Sforza, L.L. (1973) Analytic review: some current problems of human population genetics. *Am. J. Hum. Genet.* 25, 82–104
26. Cavalli-Sforza, L.L. and Piazza, A. (1975) Analysis of evolution: evolutionary rates, independence and treeness. *Theor. Popul. Biol.* 8, 127–165
27. Felsenstein, J. (1982) How can we infer geography and history from gene frequencies? *J. Theor. Biol.* 96, 9–20
28. Schliewen, U.K. *et al.* (2006) Evolutionary biology – evidence for sympatric speciation? *Nature* 444, E12–E13
29. Green, R.E. *et al.* (2010) A draft sequence of the Neandertal genome. *Science* 328, 710–722
30. Durand, E.Y. *et al.* (2011) Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* 28, 2239–2252
31. Patterson, N. *et al.* (2012) Ancient admixture in human history. *Genetics* 192, 1065–1093
32. Peter, B.M. (2016) Admixture, population structure and F-statistics. *Genetics* 202, 1485–1501
33. Harris, K. and Nielsen, R. (2013) Inferring demographic history from a spectrum of shared haplotype lengths. *PLoS Genet.* 9, e1003521
34. Lawson, D.J. *et al.* (2012) Inference of population structure using dense haplotype data. *PLoS Genet.* 8, e1002453
35. Schliewen, U.K. *et al.* (2001) Genetic and ecological divergence of a monophyletic cichlid species pair under fully sympatric conditions in Lake Ejagham, Cameroon. *Mol. Ecol.* 10, 1471–1488
36. Meier, J.I. *et al.* (2016) Demographic modelling with whole-genome data reveals parallel origin of similar *Pundamilia* cichlid species after hybridization. *Mol. Ecol.* 26, 123–141
37. Rougeaux, C. *et al.* (2017) Modeling the multiple facets of speciation-with-gene-flow towards inferring the divergence history of Lake Whitefish species pairs (*Coregonus clupeaformis*). *Genome Biol. Evol.* 9, 2057–2074
38. Strasburg, J. and Rieseberg, L. (2011) Interpreting the estimated timing of migration events between hybridizing species. *Mol. Ecol.* 20, 2353–2366
39. Sousa, V.C. *et al.* (2011) On the nonidentifiability of migration time estimates in isolation with migration models. *Mol. Ecol.* 20, 3956–3962
40. Juric, I. *et al.* (2016) The strength of selection against Neanderthal introgression. *PLoS Genet.* 12, e1006340
41. Sousa, V.C. *et al.* (2013) Identifying loci under selection against gene flow in isolation-with-migration models. *Genetics* 194, 211–233
42. Fitzpatrick, B.M. *et al.* (2008) What, if anything, is sympatric speciation? *J. Evol. Biol.* 21, 1452–1459
43. Babik, W. *et al.* (2009) How sympatric is speciation in the *Howea* palms of Lord Howe Island? *Mol. Ecol.* 18, 3629–3638
44. Mallet, J. *et al.* (2009) Space, sympatry and speciation. *J. Evol. Biol.* 22, 2332–2341
45. Olajos, F. *et al.* (2017) Estimating species colonization dates using DNA in lake sediment. *Methods Ecol. Evol.* Published online September 25, 2017. http://dx.doi.org/10.1111/2041-210X.12890
46. Lynch, M. and Walsh, B. (2007) *The Origins of Genome Architecture,* Sinauer Associates
47. Feder, J.L. *et al.* (2012) The genomics of speciation-with-gene-flow. *Trends Genet.* 28, 342–350
48. Seehausen, O. *et al.* (2014) Genomics and the origin of species. *Nat. Rev. Genet.* 15, 176–192
49. 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
50. Renaut, S. *et al.* (2013) Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nat. Commun.* 4, 1827
51. Martin, S.H. *et al.* (2013) Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Res.* 23, 1817–1828
52. Foote, A.D. *et al.* (2016) Genome-culture coevolution promotes rapid divergence of killer whale ecotypes. *Nat. Commun.* 7, 11693
53. Vijay, N. *et al.* (2016) Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. *Nat. Commun.* 7, 13195
54. Nadeau, N.J. *et al.* (2012) Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 343–353
55. Jones, F.C. *et al.* (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484, 55–61
56. Cruickshank, T.E. and Hahn, M.W. (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol. Ecol.* 23, 3133–3157
57. Ellegren, H. *et al.* (2012) The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature* 491, 756–760
58. Burri, R. *et al.* (2015) Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Res.* 25, 1656–1665
59. Wagner, G.P. *et al.* (2008) Pleiotropic scaling of gene effects and the ‘cost of complexity’. *Nature* 452, 470–472
60. Fisher, R.A. (1930) *The Genetical Theory of Natural Selection,* Oxford University Press
61. Yeaman, S. and Whitlock, M.C. (2011) The genetic architecture of adaptation under migration–selection balance. *Evolution* 65, 1897–1911
62. Servedio, M.R. *et al.* (2011) Magic traits in speciation: ‘magic’ but not rare? *Trends Ecol. Evol.* 26, 389–397
63. Fruciano, C. *et al.* (2016) Genetic linkage of distinct adaptive traits in sympatrically speciating crater lake cichlid fish. *Nat. Commun.* 7, 12736
64. Ferris, K.G. *et al.* (2017) The genetic architecture of local adaptation and reproductive isolation in sympatry within the *Mimulus guttatus* species complex. *Mol. Ecol.* 26, 208–224
65. Bush, G.L. (1969) Sympatric host race formation and speciation in frugivorous flies of the genus *Rhagoletis* (Diptera Tephritidae). *Evolution* 23, 237–251
66. Ragland, G.J. *et al.* (2017) A test of genomic modularity among life-history adaptations promoting speciation with gene flow. *Mol. Ecol.* 26, 3926–3942
67. Dieckmann, U. and Doebeli, M. (1999) On the origin of species by sympatric speciation. *Nature* 400, 354–357
68. Fry, J.D. (2003) Multilocus models of sympatric speciation: Bush vs Rice vs Felsenstein. *Evolution* 57, 1735–1746
69. Gavrilets, S. (2004) *Fitness Landscapes and the Origin of Species,* Princeton University Press
70. Yang, M. *et al.* (2017) Can genomic data alone tell us whether speciation happened with gene flow? *Mol. Ecol.* 26, 2845–2849
71. Guerrero, R.F. and Hahn, M.W. (2017) Speciation as a sieve for ancestral polymorphism. *Mol. Ecol.* 26, 5362–5368
72. Taylor, E.B. and McPhail, J.D. (2000) Historical contingency and ecological determinism interact to prime speciation in sticklebacks, *Gasterosteus*. *Proc. Biol. Sci.* 267, 2375–2384
73. Jones, F.C. *et al.* (2012) A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. *Curr. Biol.* 83–90
74. Feder, J.L. *et al.* (2003) Allopatric genetic origins for sympatric host-plant shifts and race formation in *Rhagoletis*. *Proc. Natl. Acad. Sci. U. S. A.* 100, 10314–10319