JOURNAL OF Evolutionary Biology

doi: 10.1111/jeb.12009

COMMENTARY

Hybridization and the build-up of genomic divergence during speciation

J. L. FEDER*, S. M. FLAXMAN†, S. P. EGAN* & P. NOSIL‡

*Department of Biological Sciences, University of Notre Dame, South Bend, IN, USA
†Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO, USA
‡Department of Animal and Plant Sciences, University of Sheffield, Sheffield, UK

Abbott *et al.* (2013) extensively review our understanding of the roles of hybridization in speciation. The topic is discussed in terms of both the homogenizing effects hybridization may have impeding population divergence (e.g. Hendry *et al.*, 2001; Taylor *et al.*, 2006; Seehausen *et al.*, 2008) and its creative potential for generating new biodiversity via the production of novel genotypes/phenotypes (e.g. Rieseberg *et al.*, 2003; Seehausen, 2004). It is the latter topic that is the focus of the review and is an aspect of hybridization that is garnering increased attention and acceptance in recent years, not just in plants, but also in animals.

As the creative aspects of hybridization are admirably covered in Abbott *et al.*'s (2013) study, we instead concentrate on the conditions under which the homogenizing effects of hybridization can be overcome to result in population divergence. This is a fundamental question regardless of the source of variation underlying divergence or the geographical mode of differentiation (initial differentiation in allopatry, parapatry or sympatry). The term 'speciation-with-gene-flow' circumscribes the issue and focuses attention on whether at a given level of hybridization – usually defined in terms of the gross migration rate (m) – it is possible for increased reproductive isolation to evolve. From this perspective, speciation-with-gene-flow reduces emphasis on how populations initially evolved to reach a level of reduced m and instead focuses on whether they will diverge further, remain at an equilibrium state, or reverse course and fuse.

Selective processes promoting genomic divergence

There are three selective processes that generally act to further population divergence in the face of gene flow (reviewed in Feder *et al.*, 2012b). The first is selection acting directly on a gene region. When the selection coefficient (s) favouring an allele in one population vs. another is ~$m/2$ or greater, there is a non-negligible chance that direct selection on the locus alone will act to establish the variant and increase population divergence, and when $s \gg m$, there is a high probability (Feder *et al.*, 2012b). Thus, if the supply of new mutations with large selective effects is not strongly limited, then there is little impediment to speciation-with-gene-flow. However, when this is not the case, two forms of genetic hitchhiking can help facilitate divergence.

The first form, termed 'divergence hitchhiking' (DH), is due to local reductions in effective gene flow (m_e) for sites physically linked to genes subject to divergent selection. This occurs because divergent selection can reduce gene flow at sites linked to the direct targets of selection before alleles at these sites have a chance to recombine away and introgress into the other population. This reduction in m_e means that a linked site does not require as large an s value to overcome the homogenizing effects of gene flow.

Divergence can also be facilitated by 'genome hitchhiking' (GH). In this case, the combined effect of all genes causing reproductive isolation reduces average m_e across the genome. Although the degree of reduced m_e will vary across the genome according to the distribution of selected sites, if average m_e is reduced enough globally, then divergence is not limited to just strongly selected genes and those tightly physically linked. Rather, divergence can accumulate genome-wide. DH and GH are not mutually exclusive, and both are consequences of divergent selection. The key difference is the focus on physical linkage to selected sites vs. more widespread divergence arising from global reductions in gene flow.

The three processes enhancing population divergence have been recently integrated into a four-phase model for speciation-with-gene-flow (Feder *et al.*, 2012a). In this framework, populations successively pass through 'phases' where direct selection on individual genes, DH and GH sequentially assume greater relative importance for generating increased genomic differentiation (Fig. 1). We stress that these different 'phases' are not discrete and do not have strong boundaries. Rather, they describe different parts of a continuum of divergence where the *relative* importance of DS, DH and GH changes for facilitating differentiation.

A key point that we make here not covered by Feder *et al.* (2012a) is that these phases can be associated with different species concepts (Fig. 1). Moreover, although the model was developed for *de novo* speciation in sympatry, it is also applicable for divergence initiated in parapatry or allopatry. Indeed, Wu (2001) described a complementary model involving secondary contact based on different *patterns* of genomic divergence rather than on the evolutionary *processes* generating them. A challenge is therefore to equate patterns observed in

Correspondence: Jeffrey L. Feder, Department of Biological Sciences, University of Notre Dame, South Bend, IN 46556, USA.
Tel.:+1 574 631 4159; fax: +1 574 631 7413; e-mail: feder.2@nd.edu

genome-wide scans of population divergence with different processes promoting divergence.

As discussed above, direct selection alone can readily promote speciation-with-gene-flow if many new mutations with $s > (m/2)$ arise. As this may often not be the case, discerning how effective DH and GH are at facilitating divergence becomes important. In the remainder of this commentary, we discuss DH and GH in more detail, culminating in a genomics perspective of how new species might be recognized. We also explore differences associated with speciation under primary vs. secondary contact.

Divergence hitchhiking and physical linkage

DH may make it possible for differentiation to build and spread from a few initially strongly diverged gene regions to encompass larger regions of the genome (Via, 2009, 2012). However, theory predicts that physical linkage is most effective at reducing m_e for sites within a recombination distance of < 1cM from a strongly divergently selected gene, limiting the scope for DH in much of the genome (Feder & Nosil, 2010; Feder *et al.*, 2012b). Most aspects of DH pertaining to selection strength, recombination rate and migration rate have been reviewed elsewhere (Feder & Nosil, 2010; Feder *et al.*, 2012b). We focus here on a few novel issues.

One concerns mutation. DH could be important when m is relatively high and the mutation process produces new alleles of mostly very minor effect; it would be difficult for these new alleles to diverge by DS alone unaided by DH. Note that many such small s mutations must establish to generate much progress towards speciation because individually each will contribute little to increasing reproductive isolation. In contrast, if only a modest number of moderately to strongly selected genes diverge by direct selection, then GH can become enabled and predominate as the hitchhiking process aiding divergence (Feder & Nosil, 2010; Feder *et al.*, 2012b). Thus, the mutation distribution can affect whether or not DH is critical for speciation.

Another important contributing factor for an increased role of DH in speciation may be structural features of the genome, such as chromosomal inversions and centromeres that reduce recombination rates below 1cM for large regions of the genome (Noor *et al.*, 2001; Rieseberg, 2001; Noor & Bennett, 2009). Certain forms of epistasis, such as those involved in the coupling of performance and preference for a habitat, may also place a premium on tight linkage and accentuate the importance of DH (Via, 2001). Finally, under prolonged periods of stabilizing selection in patches with different optima, slight advantages due to physical linkage can result in the evolution of 'concentrated' genetic architectures with tighter clustering of locally adaptive alleles and/or alleles of major effect replacing those of minor effect (Yeaman & Whitlock, 2011). In this regard, aspects of genomic architecture such as the recombination rate itself and the distribution of effect sizes for diverged loci can evolve.

However, empirical tests for DH should be tempered by considerations beyond just theoretical expectations. For example, it is conceivable that large genomic regions of divergence attributed to DH are artefacts stemming from inadequate marker density and undetected QTL. In addition, one-allele genetic systems in which the same variant generates reproductive isolation between populations (c.f. Felsenstein, 1981) can inflate perceptions of how clustered divergence is in the genome (e.g. see Ortíz-Barrientos & Noor, 2005).

Physical linkage can also have inhibitory consequences for speciation (for review see Ortiz-Barrientos *et al.*, 2002). First, when favourable mutations arise frequently and many are segregating in populations, Hill–Robertson effects associated with linkage can impede the selective establishment of new variants and thus speciation. Second, optimal combinations of alleles at multiple loci need to be assembled in the first place, and physical linkage can hamper such assembly (Kirkpatrick *et al.*, 2002). This might be a major issue when initial colonization of a new habitat is dependent on standing variation and involves complexes of alleles at multiple loci. Third, when divergence is predicated on new mutations, novel variants often arise in nonfavoured habitats or genetic backgrounds. In such cases, tight physical linkage can impede these mutations from recombining out of these unfavourable backgrounds, thus decreasing their chances of establishing (Feder *et al.*, 2012b). Fourth, recent theory has shown that the origin of intrinsic hybrid incompatibilities via divergent selection can occasionally increase the rate of neutral gene flow (Agrawal *et al.*, 2012). This occurs because hybrid incompatibilities can impede ecological divergence between populations and thus weaken barriers to gene flow arising from extrinsic selection. This impeding effect, and the associated increase in gene flow, is enhanced by physical linkage between the loci underlying incompatibilities. More work is needed on the balance between the constraining and promoting effects of physical linkage on speciation, rather than assuming physical linkage always promotes speciation.

Genome hitchhiking, multilocus clustering and taxonomic status

Taking these considerations concerning DH into account, it may often be the case that taxa rapidly progress from a phase dominated by DS to one where GH plays a role in facilitating divergence across the genome. Due to reduced gene flow across the genome, gene regions with lower s values, and eventually even neutral ones, can diverge. This process can generate a pattern of isolation-by-adaptation (IBA), that is, a gen-

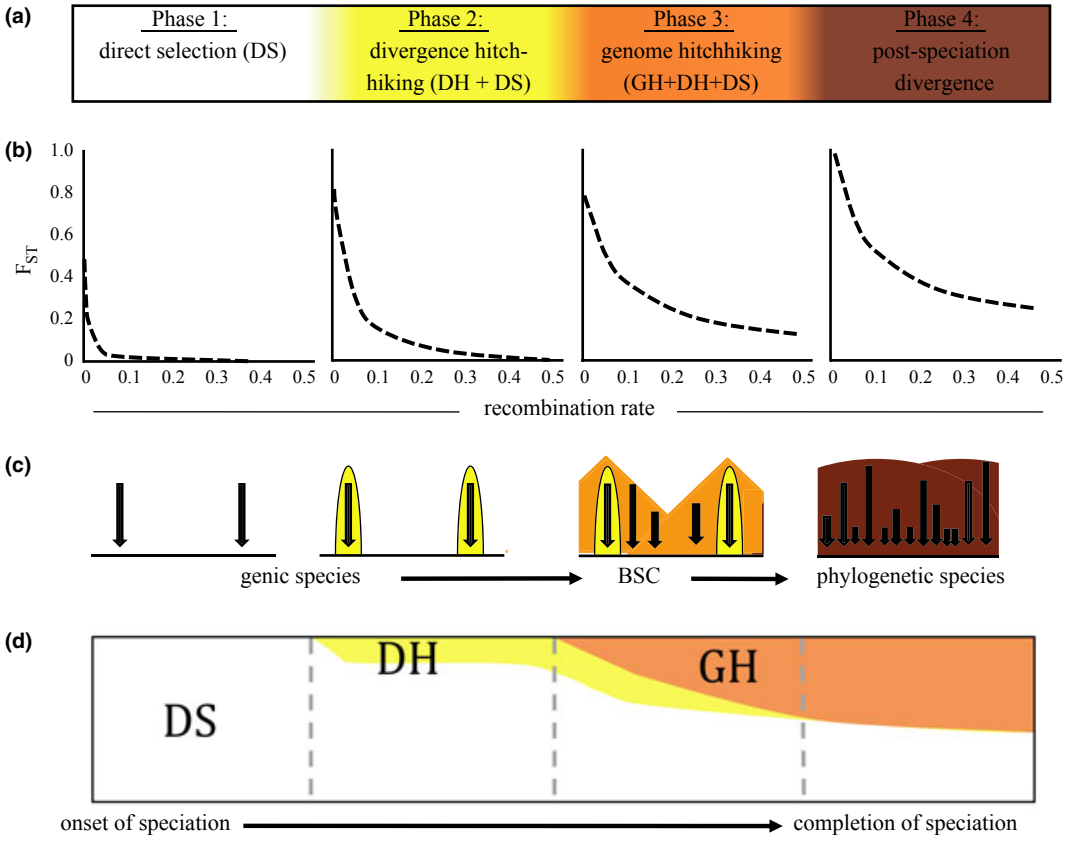

Fig. 1 Four-phase model of speciation-with-gene-flow. (a) Phases of differentiation as populations evolve increasing levels of reproductive isolation (RI) along the continuum from partially isolated races to completely isolated species. (b) Characteristic patterns of divergence (F_{ST}) that may be observed at equilibrium for linked neutral genes at varying recombination distances from loci under divergent selection as speciation proceeds through the four-phase model. (c) The phases do not form discrete boundaries. Rather, they signify more diffuse transitions representing differences in the relative importance of different processes in facilitating the establishment of new selected mutations and furthering speciation. Phase 1 = direct action of divergent selection (DS) on loci predominates (indicated by arrows). Phase 2 = divergence hitchhiking (DH) resulting from local reductions in effective migration rate for sites near selected loci (indicated by yellow-shaded areas surrounding arrows) assumes greater potential significance for aiding in the establishment of new selected mutations. Phase 3 = genome hitchhiking (GH) is enabled as the combined strength of selection acting on all loci reduces the effective migration rate sufficiently across the entire genome (orange-shaded area) for new mutations having modest to low *s* values to establish genome-wide. Phase 4 = populations evolve as if they are essentially allopatric with essentially no gene flow (red-shaded area). The four phases can be equated to different species concepts (phases 1 and 2 = genic species; phase 2 and 3 = biological species; phase 4 = phylogenetic species, where high-frequency private alleles and/or fixed diagnostic, derived differences are observed. (d) Hypothesized relative roles of each process as speciation progresses. DS contributes throughout, with the other processes joining in. As post-speciation divergence is defined by zero effective migration, DH becomes less important in phase 4. Figure modified from Feder *et al.* (2012b) study with permission from Elsevier.

eral association between the degree of ecological divergence between populations and genetic differentiation (Nosil *et al.*, 2008, 2009). We note that IBA does not necessarily predict genetic divergence will be strong across many regions of the genome. In fact, most regions might still be very weakly differentiated. Rather, IBA simply predicts that genetic differentiation will be greater for ecologically divergent population pairs than for ecologically similar ones, as is often observed (Nosil *et al.*, 2009 for review).

Abbott *et al.* (2013) draw attention to well-established cline theory (Barton, 1979, 1983) to highlight cases that when different endogenous clines meet and overlap, they are expected to become coupled, and then these coupled multilocus clines may move together in space. Such moving tension zones can be trapped by natural barriers to dispersal (Barton, 1979) or coupled with local adaptation clines that are geographically stabilized by selection. These aspects of cline theory point to how different barriers to gene flow will accumulate together in the genome and have synergistic consequences for reproductive isolation. As a result, the genome begins to 'congeal' with reproductive isolation increasingly becoming a characteristic of the gen-

ome rather than of individual loci, as initially conceptualized by Ernst Mayr in the biological species concept (Coyne & Orr, 2004), although without the need for extensive co-adapted networks.

These arguments concerning the coupling of geographical clines might also be applied to the coupling of locus-specific rates of admixture within single geographical localities, that is, 'genomic clines' (Gompert & Buerkle, 2011; Gompert *et al.*, 2012). This conceptualization of congealing genomic clines therefore equates spatial aspects of hybrid zone theory with the nonspatial aspects of sympatric speciation, providing a potential means for operationally identifying species. Different species may form genetically diagnosable clusters in distance networks across their geographical ranges, as effective gene flow across the genome is more strongly lowered between local, co-occurring populations of different species than it is among different localities within species. This pattern of clustering across the geographical range for species contrasts with intraspecific varieties or ecotypes that may diverge locally in sympatry, but not show genetic continuity across their geographical ranges.

As mentioned above, an empirical challenge is to equate patterns observed in genome-wide scans of divergence with the phases of speciation (where DS, DH and GH assume varying potential importance) and thus with the taxonomic status of populations. One component of the challenge is that the genome is made of interspersed sites that directly experience divergent selection and neutral sites that do not. There is a lag in the time it takes for neutral divergence to accumulate between populations dictated by the interaction of mutation, drift, indirect effects of selection acting on them and migration. Hence, patterns of neutral divergence can either under- or overestimate m_e for a given region of the genome, thereby biasing our characterization of species. In principle, if we knew the intensity of divergent selection at every site, we could approximate how the entire genome might eventually equilibrate and more accurately assess the phases of divergence and taxonomic status. Although this is unlikely to be realized in most systems, evidence from selection experiments, transplant experiments and QTL mapping of traits will still aid this endeavour. Another challenge concerns the continuous nature of speciation. Speciation is usually not an event distinguishable by a point in time; although the genome might increasingly congeal as speciation proceeds, there may not be an event horizon where full genome-wide reproductive isolation becomes inevitable. Indeed, if conditions change, formerly 'good species' can fuse into a single taxon (Seehausen *et al.*, 2008 for review). Consequently, the identification of geographically consistent genetic clusters does not signify a unique breakpoint in the timeline of speciation, but only a means to operationally recognize that divergence and the processes acting to facilitate it have reached a phase where we may consider populations to represent species.

Aspects of primary and secondary contact

We highlight here some important differences between secondary and primary contact modes of divergence. First, reproductive isolation due to intrinsic genetic incompatibilities, antagonistic sexual selection, meiotic drive and genetic drift are more likely to evolve during initial periods of allopatry and secondary contact than for primary contact scenarios. The consequences of these forms of reproductive isolation and any associated genome structure for speciation-with-gene-flow following secondary contact need to be further investigated.

Second, during the initial stage of divergence in allopatry, genome structure is usually not a major consideration because selection is not opposed by gene flow and numerous mutations of small effect can more easily accumulate across the genome than is possible in sympatry. Following secondary contact and introgression, local concentrations of these loci in the genome could effectively act as a large effect gene and serve as focal points to propagate further divergence by DH. Thus, secondary contact models could be more likely to generate populations in phase two of divergence (Fig. 1), in which DH plays a key role, than when speciation-with-gene-flow is initiated *de novo*.

Third, secondary vs. primary modes of divergence have implications for chromosomal inversions that are predicted to facilitate genomic divergence, and are increasingly being described across a wide range of taxa including *Mimulus* monkey flowers, *Drosophila* flies, *Rhagoletis* flies, *Gasterosteus* stickleback fish, *Heliconius* butterflies and *Helianthus* sunflowers (Noor *et al.*, 2001; Feder *et al.*, 2003; Strasburg *et al.*, 2009; Lowry & Willis, 2010; Joron *et al.*, 2011; Jones *et al.*, 2012). How do inversions spread to high frequency? One possibility is that hybridization and gene flow allow for the spread of inversions by natural selection. The idea is that when chromosomal rearrangements originate, they might capture locally favourable combinations of genes. These combinations are kept together by suppressed recombination within inversions, and thus, inversions are favoured and spread by selection. As highlighted by Kirkpatrick & Barton (2006), this scenario could occur under primary sympatric divergence. However, conditions for the spread of inversions may be more favourable when inversions pre-exist at low frequencies in allopatric taxa and become favoured following secondary contact and gene flow (Feder *et al.*, 2011). The reasons for an increased probability of establishment in secondary contact are the following: (i) in initial allopatry, the inversions will likely capture all favoured locally adapted alleles together, which can be a limita-

tion to their spread in sympatry, and (ii) multiple standing copies of the inversion can exist in allopatry, which makes the rearrangements less likely to be lost stochastically compared with a new inversion arising in a single copy in sympatry. This 'mixed mode' model is a novel and creative aspect of hybridization promoting speciation not fully explored by Abbott *et al.* (2013), although they do cover the general topic of standing variation extensively.

Conclusions

As highlighted by Abbott *et al.* (2013), advances in mass sequencing will reveal new insights into the reticulate nature of speciation and a greater appreciation for the often recycled nature of variation fuelling population divergence. However, the greatest contribution of the new technology may be in the area of genome architecture: *understanding how the genes involved in reproductive isolation are distributed and arrayed across the genome and how their frequency and physical linkage affect the potential for speciation.* Much remains to be done. However, the foundation exists to build a unified theory of speciation genomics in which the causal basis for reproductive isolation is connected to underlying physiology and genetic mechanisms, and the consequences of these associations for the organization and evolution of patterns of genome-wide divergence can be resolved.

Acknowledgments

The work was supported, in part, by grants to JLF from NSF and the USDA and by a European Research Council (ERC) Starter Grant to PN. We also acknowledge the support of the Advanced Diagnostic and Therapeutics initiative and the Environmental Change Initiative at the University of Notre Dame to JLF and SPE.

References

Abbott, R., Albach, D., Ansell, S., Arntzen, J.W., Baird, S.J.E., Bierne, N. *et al.* 2013. Hybridization and speciation. *J. Evol. Biol.* **26**: 229–246.

Agrawal, A.A., Feder, J.L. & Nosil, P. 2012. Ecological divergence and the evolution of intrinsic postmating isolation with gene flow. *Int. J. Ecol.* **2012**, doi: 10.1155/2012/456374.

Abbott, R., Albach, D., Ansell, S., Arntzen, J.W., Baird, S.J.E., Bierne, N. *et al.* 2013. Hybridization and speciation. *J. Evol. Biol.* **2011**: doi: 10.1155/2011/435357.

Barton, N.H. 1979. Gene flow past a cline. *Heredity* **43**: 333–339.

Barton, N.H. 1983. Multilocus clines. *Evolution* **37**: 454–471.

Coyne, J.A. & Orr, H.A. 2004. *Speciation*. Sinauer Associates, Inc., Sunderland, MA.

Feder, J.L. & Nosil, P. 2010. The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution* **64**: 1729–1747.

Feder, J.L., Berlocher, S.H., Roethele, J.B., Dambroski, H., Smith, J.J., Perry, W.L. *et al.* 2003. Allopatric genetic origins for sympatric host-plant shifts and race formation in Rhagoletis. *Proc. Natl Acad. Sci. USA* **100**: 10314–10319.

Feder, J.L., Geji, R., Powell, T.H.Q. & Nosil, P. 2011. Adaptive chromosomal divergence driven by mixed geographic mode of evolution. *Evolution* **65**: 2157–2170.

Feder, J.L., Egan, S.P. & Nosil, P. 2012a. The genomics of speciation-with-gene-flow. *Trends Genet.* **28**: 342–350.

Feder, J.L., Geiji, R., Yeaman, S. & Nosil, P. 2012b. Establishment of new mutations under divergence and genome hitchhiking. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**: 461–474.

Felsenstein, J. 1981. Skepticism towards Santa Rosalia, or why are there so few kinds of animals? *Evolution* **35**: 124–138.

Gompert, Z. & Buerkle, C.A. 2011. Bayesian estimation of genomic clines. *Mol. Ecol.* **20**: 2111–2127.

Gompert, Z., Parchman, T.L. & Buerkle, C.A. 2012. Genomics of isolation in hybrids. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**: 439–450.

Hendry, A.P., Day, T. & Taylor, E.B. 2001. Population mixing and the adaptive divergence of quantitative traits in discrete populations: a theoretical framework for empirical tests. *Evolution* **55**: 459–466.

Jones, F.C., Grabherr, M.G., Chan, Y.F., Russell, P., Mauceli, E., Johnson, J. *et al.* 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**: 55–61.

Joron, M., Frezal, L., Jones, R.T., Chamberlain, N.L., Lee, S.F., Haag, C.R. *et al.* 2011. Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* **477**: 203–U102.

Kirkpatrick, M. & Barton, N. 2006. Chromosome inversions, local adaptation and speciation. *Genetics* **173**: 419–434.

Kirkpatrick, M., Johnson, T. & Barton, N. 2002. General models of multilocus evolution. *Genetics* **161**: 1727–1750.

Lowry, D.B. & Willis, J.H. 2010. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biol.* **8**: e1000500.

Noor, M.A.F. & Bennett, S.M. 2009. Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity* **103**: 439–444.

Noor, M.A.F., Grams, K.L., Bertucci, L.A. & Reiland, J. 2001. Chromosomal inversions and the reproductive isolation of species. *Proc. Natl Acad. Sci. USA* **98**: 12084–12088.

Nosil, P., Egan, S.P. & Funk, D.J. 2008. Heterogeneous genomic differentiation between walking-stick ecotypes: "Isolation by adaptation" and multiple roles for divergent selection. *Evolution* **62**: 316–336.

Nosil, P., Funk, D.J. & Ortíz-Barrientos, D. 2009. Divergent selection and heterogeneous genomic divergence. *Mol. Ecol.* **18**: 375–402.

Ortíz-Barrientos, D. & Noor, M.A.F. 2005. Evidence for a one-allele assortative mating locus. *Science* **310**: 1467–1467.

Ortiz-Barrientos, D., Reiland, J., Hey, J. & Noor, M.A.F. 2002. Recombination and the divergence of hybridizing species. *Genetica* **116**: 167–178.

Rieseberg, L.H. 2001. Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* **16**: 351–358.

Rieseberg, L.H., Raymond, O., Rosenthal, D.M., Lai, Z., Livingstone, K., Nakazato, T. *et al.* 2003. Major ecological

transitions in wild sunflowers facilitated by hybridization. *Science* **301**: 1211–1216.

Seehausen, O. 2004. Hybridization and adaptive radiation. *Trends Ecol. Evol.* **19**: 198–207.

Seehausen, O., Takimoto, G., Roy, D. & Jokela, J. 2008. Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Mol. Ecol.* **17**: 30–44.

Strasburg, J.L., Scotti-Saintagne, C., Scotti, I., Lai, Z. & Rieseberg, L.H. 2009. Genomic patterns of adaptive divergence between chromosomally differentiated sunflower species. *Mol. Biol. Evol.* **26**: 1341–1355.

Taylor, E.B., Boughman, J.W., Groenenboom, M., Sniatynski, M., Schluter, D. & Gow, J.L. 2006. Speciation in reverse: morphological and genetic evidence of the collapse of a three-spined stickleback (Gasterosteus aculeatus) species pair. *Mol. Ecol.* **15**: 343–355.

Via, S. 2001. Sympatric speciation in animals: the ugly duckling grows up. *Trends Ecol. Evol.* **16**: 381–390.

Via, S. 2009. Natural selection in action during speciation. *Proc. Natl Acad. Sci. USA* **106**: 9939–9946.

Via, S. 2012. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**: 451–460.

Wu, C. 2001. The genic view of the process of speciation. *J. Evol. Biol.* **14**: 851–865.

Yeaman, S. & Whitlock, M.C. 2011. The genetic architecture of adaptation under migration-selection balance. *Evolution* **65**: 1897–1911.

Received 13 August 2012; revised 4 September 2012; accepted 7 September 2012