

*Review*

# What can patterns of differentiation across plant genomes tell us about adaptation and speciation?

Jared L. Strasburg<sup>1,\*</sup>, Natasha A. Sherman<sup>1</sup>, Kevin M. Wright<sup>2</sup>,  
Leonie C. Moyle<sup>1</sup>, John H. Willis<sup>2</sup> and Loren H. Rieseberg<sup>1,3</sup>

<sup>1</sup>*Department of Biology, Indiana University, Bloomington, IN 47405, USA*

<sup>2</sup>*Department of Biology, Duke University, Durham, NC 27708, USA*

<sup>3</sup>*Department of Botany, University of British Columbia, Vancouver, BC, Canada V6T 1Z4*

Genome scans have become a common approach to identify genomic signatures of natural selection and reproductive isolation, as well as the genomic bases of ecologically relevant phenotypes, based on patterns of polymorphism and differentiation among populations or species. Here, we review the results of studies taking genome scan approaches in plants, consider the patterns of genomic differentiation documented and their possible causes, discuss the results in light of recent models of genomic differentiation during divergent adaptation and speciation, and consider assumptions and caveats in their interpretation. We find that genomic regions of high divergence generally appear quite small in comparisons of both closely and more distantly related populations, and for the most part, these differentiated regions are spread throughout the genome rather than strongly clustered. Thus, the genome scan approach appears well-suited for identifying genomic regions or even candidate genes that underlie adaptive divergence and/or reproductive barriers. We consider other methodologies that may be used in conjunction with genome scan approaches, and suggest further developments that would be valuable. These include broader use of sequence-based markers of known genomic location, greater attention to sampling strategies to make use of parallel environmental or phenotypic transitions, more integration with approaches such as quantitative trait loci mapping and measures of gene flow across the genome, and additional theoretical and simulation work on processes related to divergent adaptation and speciation.

**Keywords:** genome scans; speciation; adaptation; reproductive isolation; genetic divergence; plants

## 1. INTRODUCTION

As research on speciation genetics broadens to include speciation genomics, a number of models have been proposed to explain genomic patterns of differentiation between populations or species. These include the notion of ‘genomic islands’ of speciation [1,2], in which differentiation is maintained via divergent selection or reproductive isolating barriers in portions of the genome while the remainder is permeable to gene flow; and divergence hitchhiking [3,4], in which divergently selected loci can act as nuclei for increased differentiation at linked neutral sites. More generally, information about the proportion of the genome contributing to divergent adaptation and isolation, as well as the nature of the genes and genomic regions involved, is fundamental to our understanding of the process of speciation. It informs theories of the

geography and temporal dynamics of speciation, the role of gene flow in preventing or enhancing adaptive divergence, and the relative contributions of selection and drift to divergence and reproductive isolation [5].

Empirical methods for examining patterns of genomic divergence have dramatically expanded in the last decade, on the heels of technological advances that have vastly increased the extent of genomic coverage possible in marker-based studies. Rapid declines in the costs of sequencing and genotyping technologies make it feasible to assay hundreds or thousands of loci for essentially any organism, and for better-characterized species sequence-based markers can be aligned to high-density genetic maps or sequenced genomes. One increasingly common approach is to compare genomic patterns of diversity within populations and/or divergence between populations at these markers, a process commonly called a genome scan, in order to identify genomic regions that do not conform to expectations based on some neutral demographic model. These regions are then considered candidates to contain one or more loci under selection, and they can be investigated in further detail in order to identify the precise target, mode and genomic or phenotypic consequences of selection.

\* Author for correspondence (jstrasbu@indiana.edu).

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rstb.2011.0199> or via <http://rstb.royalsocietypublishing.org>.

One contribution of 13 to a Theme Issue ‘Patterns and processes of genomic divergence during speciation’.

Here, we consider genome scan and related studies that examine genomic patterns of adaptation, divergence and reproductive isolation, drawing particularly on empirical examples in plant populations or species. The main goals of this review are to (i) discuss factors that may influence patterns of genomic divergence; (ii) consider how these factors may impact the efficacy of genome scan approaches, and how best to control them; and (iii) summarize genome scan and related studies that have been performed in plants, and discuss conclusions that can be drawn from them.

## 2. PATTERNS OF GENOME DIVERGENCE AND THEIR INTERPRETATION

The genome scan approach was first explicitly described by Lewontin & Krakauer [6], who noted that all loci should be similarly affected by demographic processes, whereas selection should only affect a subset of loci. Accordingly, given an expected neutral distribution of divergence values, outliers can be considered candidate markers for genomic regions under selection. However, theoretical and empirical work suggests that numerous factors can complicate this simple expectation, and the resulting inferences that can be drawn from genome scan studies [7–9]. These factors include the biogeographic history of contact between populations (especially whether and when gene flow occurred between them), the time frame and strength of selection acting on differentiating loci, the genetic architecture of traits under selection, and the population structure and demographic history of populations being analysed [10]. These factors can influence both the expected patterns of genomic divergence, and the ability of genome scan approaches to identify loci involved in adaptive differentiation or speciation. Although a complete review is beyond the scope of this paper, we briefly outline some of these considerations here.

### (a) *Expected patterns of genomic divergence*

The interpretation of genome scans depends, in part, on understanding the patterns of genomic differentiation that are expected under realistic biological scenarios. In some cases there are well-developed expectations; however, in others the geographical, temporal and selective details can make clear predictions difficult. For example, patterns of genomic divergence are expected to be strongly influenced by levels of gene flow as well as the strength of divergent selection at individual loci. When speciation begins in sympatry or parapatry, ongoing gene flow can homogenize most of the genome, resulting in genomic divergence at only the one or few loci experiencing strong divergent selection. In contrast, allopatric speciation does not require strong divergent selection at a small number of loci, and a larger portion of the genome may diverge through a combination of divergent selection, differential response to similar selective pressures and genetic drift [2].

In addition, in the presence of gene flow, it has been suggested that additional divergence may build up around the loci subjected to strong divergent selection because of reduced effective recombination rates around these loci [11]. This could potentially result in clustering of loci contributing to reproductive

isolation or adaptive divergence and an increase in the size of divergent regions through time, a phenomenon that has been referred to as divergence hitchhiking (hereafter DH; [4]). Using simulations, Feder & Nosil [3] found that DH could potentially play an important role under some circumstances, especially during intermediate stages of speciation when the number of loci under selection is fairly small. However, as they point out, their simulations dealt with equilibrium conditions in which accumulation of additional adaptive divergence near already selected sites—a key aspect of DH—was not considered. Feder *et al.* [12] expanded their simulations to include the possibility of new selected mutations accumulating, and again found that DH was restricted to a limited range of conditions. Instead, they suggested that additional adaptive divergence was more strongly influenced by interactions between selection and migration than by linkage to established adaptively divergent sites.

While these simulations are informative, it would be valuable to consider several additional factors that may influence patterns of genomic divergence, including the effects of a period of allopatry prior to secondary contact, population structure within species and asymmetric selective effects. We suspect that the latter two factors will reduce the efficacy of DH, especially with respect to divergence at linked neutral sites.

Empirical evidence for DH is currently limited. This is likely partly owing to the difficulty of distinguishing the effects of DH from other processes of genomic divergence. For example, recent admixture can yield large islands of divergence [13] that might be misinterpreted as evidence of DH. On the other hand, evidence for DH might be missed because genomic regions affected by DH may include low and high  $F_{ST}$  markers owing to variance in coalescence times [14]. Via [14] suggests the following two predictions of DH: (i) low- $F_{ST}$  markers in DH regions will have lower gene flow than low- $F_{ST}$  markers outside of DH regions and (ii) low- $F_{ST}$  markers in DH regions will more closely reflect species relationships than those outside of DH regions. However, it is not clear to us that either prediction distinguishes between a single large region of genomic isolation and multiple smaller independent regions that individually encompass one or more low  $F_{ST}$  markers. An alternative and potentially more powerful approach for estimating the impact of gene flow on genomic differentiation would be to compare sympatric and allopatric patterns of divergence in closely related taxa.

Expected patterns of genomic divergence between species that initially diverged in allopatry and later come back into secondary contact are likely to be even more complex. If the period of allopatry is brief, patterns of divergence are likely to be similar to initial divergence with gene flow [4]. However, if the period of allopatry is longer, and especially if the populations face divergent ecological conditions, patterns of introgression upon secondary contact are harder to predict and will depend critically on the strength and genomic architecture of adaptive divergence and isolation, as well as the duration and spatial pattern of secondary contact. Long-term equilibrium patterns of genomic divergence following secondary contact may be indistinguishable from those expected with divergence in the

presence of gene flow [15]. However, it may take many generations of contact to reach such equilibrium.

Finally, the degree to which adaptation occurs from standing variation versus newly arisen variants (sometimes referred to as soft selective sweeps versus hard selective sweeps) also affects patterns of genomic divergence. Adaptation from standing variation is likely to occur under a wide range of circumstances; in some cases, it may be the predominant source of adaptive variation, especially when population mutation rates are high [9]. Importantly, a newly arisen variant is initially found on a single genetic background, and if it is swept rapidly to fixation, there will be a strong signal of hitchhiking, with little variation at linked sites and high divergence between populations in different selective environments. In contrast, if a variant that has segregated long enough to recombine onto multiple genetic backgrounds is driven to fixation, the decrease in diversity at linked sites will be less pronounced. And if divergent adaptation occurs through moderate changes in allele frequencies at multiple sites, it is likely that none of these sites will exhibit substantial divergence between populations or species [16].

#### **(b) The efficacy of genome scans and their interpretation**

Clearly, patterns of genomic divergence are influenced by a range of geographical, temporal, demographic and selective factors. Because of these effects, analyses of genome scan data can potentially provide over-, under- or biased estimates of the number and type of loci involved in adaptation and speciation, and the kinds of selective processes or genetic architectures involved.

To date, the most attention has been paid to the effects of population structure on outlier detection, as it can potentially result in an increase in false-positives. Early simulations implied that genome scans should be robust to a number of equilibrium and non-equilibrium population structures and histories as well as among-locus mutation rate variation [17,18]. However, more recent simulation studies have presented a less-positive picture, suggesting that population structure should be considered in more detail. For example, Excoffier *et al.* [7] found that hierarchical population structure could cause a substantial increase in the rate of false-positives if not accounted for, and Foll & Gaggiotti [8] showed that the inclusion of isolated populations that have experienced strong bottlenecks can also result in increased false-positives.

In some cases, remedies for these demographic complications might be relatively straightforward. For example, sampling designs that minimize population structure and/or sample multiple habitat or trait transitions could reduce the likelihood of false-positives. Likewise, the development of analytical methods that are more robust to violations of demographic assumptions would be valuable; ideally, these methods would directly incorporate an estimated demographic history instead of having a simpler history imposed [8], and/or explicitly include demographic history in the estimation process [19]. Because of the difficulties involved when the demographic history of a population is unknown, in some instances, it may be preferable to take an

empirical approach [20] and rank the genes within the genome on the basis of pair-wise genetic diversity or genetic distance. This empirical approach is probably most useful in high-density scans, where there are many candidates and where the primary goal is the identification of genes with a recent history of selection.

In addition to demographic effects, the number of loci under selection can influence the interpretation of genome scan results. Most genome scan methods assume, at least implicitly, that relatively few loci are under selection. For example, the baseline  $F_{ST}$  upon which simulations are contingent is generally assumed to reflect neutral divergence. In many cases, this may be a reasonable assumption; but if selection is pervasive, it is clearly not. For example, Michel *et al.* [21] found that the majority of markers were linked to loci under divergent selection between host races of *Rhagoletis pomonella*, but very few were identified in an outlier analysis. Likewise, genomic admixture studies have shown that large proportions of the genome can be under selection (see below). The fact that 'neutral'  $F_{ST}$  estimates tend to be fairly low in plant systems studied so far (table 1 and electronic supplementary material, S1) may suggest that this phenomenon is rare; however,  $F_{ST}$  estimates between *R. pomonella* host races were also found to be quite low despite widespread divergent selection.

The prevalence of soft versus hard selective sweeps also impacts the efficacy of genome scans because the former leaves a less pronounced mark on patterns of genetic diversity and divergence (discussed above), and therefore are less likely to be detected with genome scan methods [16,49]. Traits are also expected to differ in their likelihood of being affected by soft versus hard sweeps depending on their genetic architecture [49]. Soft sweeps are more likely to involve polygenic traits or mutations of smaller effect [16,50]. In addition, for polygenic traits significant phenotypic evolution can occur without large allele frequency changes at any locus if covariances among loci contribute substantially to phenotypic variance [51]. Thus, loci identified in genome scans for selection probably reflect an unrepresentative subset of traits whose genetic architecture lends itself more easily to such detection [49]. A number of methods have been suggested to distinguish adaptation from standing variation versus new mutations, but they often require restrictive assumptions about population demography and the strength and origin of selective sweeps and are difficult to apply in many circumstances [50].

Another important consideration is the metric employed for outlier detection. Most analytical methods, theoretical treatments and empirical studies have focused on  $F_{ST}$ , which is a widely used index of differentiation. However,  $F_{ST}$  depends on both within- and between-population variation and thus the precise cause of  $F_{ST}$  outliers can be difficult to infer. Other metrics such as reduced genetic diversity or absolute allele frequency differences will better diagnose recent selection. The former can also provide insights into the direction of selection [52]. On the other hand, if the interest is in the genetic architecture of reproductive isolation, then coalescent estimates of unidirectional migration rates will be superior to  $F_{ST}$  in detecting regions of the

Table 1. Summary of representative studies taking genome scan approaches in plants. In some cases, different sets of outlier loci were identified based on different population comparisons, methodologies or significant cut-offs; we have reported the outliers the authors considered most relevant or promising where possible. Outliers are of high divergence unless otherwise specified; DOA, divergence outlier analysis; SFS, site frequency spectrum; AFLP, amplified fragment length polymorphism; QTL, quantitative trait locus; BDM, Bateson–Dobzhansky–Muller; SNP, single nucleotide polymorphism; CAPS, cleaved amplified polymorphic sequence; RAPD, random amplified polymorphic DNA. Additional details are in electronic supplementary material, S1.

system	markers	method/program	'neutral' $F_{ST}$	results	reference
Annual sunflowers, <i>Helianthus annuus</i> / <i>H. petiolaris</i>	108 mapped microsatellites	DOA (F <sub>DIST</sub> 2); association with QTLs for sterility, species differences	0.08	five outliers (4.6%); no association between genetic divergence and proximity to QTLs for sterility or species differences	[22]
<i>H. annuus</i> / <i>H. petiolaris</i>	77 mapped sequenced loci	DOA (BAYESFST), association with rearrangement status		26 outlier SNPs (3.2% of variable sites); borderline-significant association between genetic divergence and proximity to chromosomal breakpoints	[23]
<i>H. annuus</i> / <i>H. debilis</i>	88 mapped microsatellites	DOA (F <sub>DIST</sub> 2), diversity outlier analysis (lnRV, lnRH), interspecific gene flow (MIGRATE, BAYESASS)		three high-divergence outliers (3.4%) and one locus (1.1%) with significantly reduced variability in <i>H. annuus</i> ssp. <i>annuus</i> ; one locus possibly linked to a BDM incompatibility based on asymmetric pattern of gene flow	[24]
Loblolly pine, <i>Pinus taeda</i>	3059 SNPs, 23 microsatellites (some mapped)	DOA (F <sub>DIST</sub> 2), association with aridity	0.02	five loci (0.2%) associated with aridity and 24 outliers (0.8%); no overlap between two sets of loci	[25]
Western white pine, <i>Pinus monicola</i>	66 AFLP loci	DOA (DFDIST), association with climatic factors, bioclimatic mapping		eight outliers (12.1%); allele frequencies at four outliers (6.1%) significantly correlated with climatic factors relevant to species distribution based on bioclimatic models	[26]
White spruce, <i>Picea glauca</i>	534 mapped SNPs from 345 candidate genes	DOA (F <sub>DIST</sub> 2, NEWFST), qualitative association with climatic factors and quantitative traits	0.01	19 F <sub>DIST</sub> 2 outliers (5.5%), 2 NEWFST outliers (0.6%); 47 genes (13.6%) candidates for local adaptation based on locus $\times$ population parameter estimates in NEWFST (with relaxed criterion for significance)	[27]
European aspen, <i>Populus tremula</i>	70 mapped microsatellites	DOA (F <sub>DIST</sub> 2, DETSEL), diversity outlier analysis (lnRV, lnRH)	0.02–0.11 <sup>a</sup>	12 outliers (17.1%), on average 1.6 kb away from closest predicted gene	[28]
Campions, <i>Silene latifolia</i> / <i>S. dioica</i>	approximately 300 AFLPs	DOA (DFDIST)	0.39–0.58	29 (9.6%) and 19 (6.3%) high- and low-divergence outliers, respectively, in allopatric comparison; 34 (11.1%) and 15 (4.9%) high- and low-divergence outliers in sympatric comparison <sup>b</sup>	[29]
Oaks, <i>Quercus roberl</i> / <i>Q. petraea</i>	389 markers of various kinds; some mapped	DOA (F <sub>DIST</sub> 2)	0.36	47 outliers (12.1%), 20 of which were mapped; did not directly address co-localization of outliers and QTL for species differences [30]	[31]
Palms, <i>Howea belmoreana</i> / <i>H. forsteriana</i>	274 AFLPs	DOA (DFDIST)	0.31	four outliers (1.5%); interpreted results as supporting sympatric speciation driven by divergent selection at few loci	[32]
Violet, <i>Viola cazorlensis</i>	369 AFLPs	DOA (DFDIST, BAYESFST), association with floral traits	0.07	nine outliers (2.4%); variation in floral traits significantly associated with variation at outlier loci but not at neutral loci; strong concordance between F <sub>DIST</sub> and BAYESFST outliers	[33]
European beech, <i>Fagus sylvatica</i>	254 AFLPs	DOA (DFDIST), association with temperature	0.01–0.02	one outlier (0.4%); frequency of dominant allele significantly associated with altitude and temperature during establishment	[34]

(Continued.)



Table 1. (Continued.)

system	markers	method/program	'neutral' $F_{ST}$	results	reference
Mediterranean shrub, <i>Fumana thymifolia</i>	175–191 AFLPs	DOA (DFDIST), association with temperature and aridity	0.01–0.03	treatment plots exposed to varying water and temperature conditions; 5 (2.7%), 2 (1.1%), and 0 outliers in control versus drought, control versus warming, drought versus warming, respectively	[35]
Seagrass, <i>Zostera marina</i>	25 microsatellites	DOA (FDIST2, DETSEL)	0.02	three outliers (12.0%); no significant difference in frequency of outliers with EST-derived versus anonymous microsatellites	[36]
<i>Gladiolus carinatus</i> species complex	3 sequenced loci, 284 AFLPs	DOA (DFDIST), association with floral traits	0.03–0.30	eight (2.8%) and 11 (3.9%) high- and low-divergence outliers, respectively; of 11 loci considered best candidates, all but one are significantly associated with one or more floral traits	[37]
Japanese conifer, <i>Cryptomeria japonica</i>	148 CAPS loci (most mapped)	DOA (FDIST2, DETSEL)	0.06	five (3.4%) and two (1.1%) high- and low-divergence outliers, respectively, using FDIST2; 13 outliers (8.8%) using DETSEL; four of 13 were also FDIST2 outliers	[38]
Mountain rock-cress, <i>Arabis alpina</i>	140–712 AFLPs	association with climatic factors		associations at large (European), regional, and local scales; identified at least 10% of loci as significant at each scale, with minimum temperature and precipitation as major determinants	[39]
<i>A. alpina</i>	825 AFLPs	association with climatic factors		78 loci (9.5%) significantly associated with climatic factors, mainly mean annual minimum temperature; 61 loci in French Alps, 21 in Swiss Alps—only four loci shared between regions	[40]
Douglas fir, <i>Pseudotsuga menziesii</i> var. <i>menziesii</i>	121 sequenced loci	SFS- and divergence-based tests of selection		three to eight outliers (2.5–6.6%) with significantly reduced variation, possibly indicating selective sweep	[41]
<i>H. annuus</i> ssp. <i>texasus</i> / <i>H. debilis</i>	15 mapped AFLPs	association with QTLs for sterility, species differences		markers linked to sterility, phenotypic difference QTLs introgressed at lower than expected rate; other markers higher than expected rate, suggesting adaptive introgression	[42]
<i>Populus alba</i> / <i>P. tremula</i>	93 mapped markers of various kinds	DOA (FDIST2), genomic admixture analysis	0.37	33 outliers (35.5%); found significant variation in rates of introgression among replicate hybrid zones; evidence for strong pre- and post-zygotic isolation, epistasis	[43]
<i>H. annuus</i> / <i>H. petiolaris</i>	61 mapped RAPDs	genomic admixture analysis		reanalysis of Buerkle & Rieseberg [44] data; roughly half of loci deviated from expectations of neutral introgression, including evidence of adaptive introgression	[45]
<i>Arabidopsis lyrata</i>	2.85 million-probe microarray	tests for significant differentiation, association with soil type		compared divergence between serpentine and non-serpentine populations; far more significantly differentiated probes in comparisons between soil types; differentiated probes over-represented for ion transporters and membrane proteins	[46]
<i>A. lyrata</i>	next-generation population-level sequencing	DOA, association with soil type		compared replicate serpentine versus non-serpentine populations in USA and Scotland; polymorphisms associated with soil type over-represented for genes associated with ion transport and heavy metal detoxification	[47]
<i>Arabidopsis halleri</i>	820 AFLPs	DOA (DFDIST, DETSEL), association with heavy metal tolerance	0.07	compared tolerant and non-tolerant populations; proportion of outliers ranged from 1.4 to 5.6% depending on population comparison and method used	[48]

<sup>a</sup>Range of pairwise  $F_{ST}$  estimates from electronic supplementary material, table S3; authors do not explicitly say if they used trimmed  $F_{ST}$  estimates for outlier analyses.

<sup>b</sup>Authors reported results for both  $\alpha = 0.05$  and  $\alpha = 0.01$  cut-off; these are  $\alpha = 0.01$  values.

genome with low and/or asymmetric migration rates, as may be the case when hybridizing populations are isolated by Dobzhansky–Muller incompatibilities [24].

In addition to these sampling and analytical issues, the efficacy of genome scans can be impacted by the kinds and density of molecular markers employed and knowledge of the genomic location of these markers. The importance of marker density is obvious—many selected loci will be missed in low-density scans and the ability to identify the actual genes or sites that are under selection is greatly diminished. In addition, we find it worrisome that estimates of the proportion of outlier loci and the size of outlier regions tend to be larger in low-density than in high-density scans, perhaps implying that the former are not sufficiently conservative (electronic supplementary material, S2). The value of information on the genomic location of markers is self-evident as well because genome scans are likely to reveal multiple outliers corresponding to a single selective event. However, shared causation would not be evident without a high-density map or sequenced genome.

Genome scans are a relatively new approach, so it is to be expected that more realistic and appropriate models and analytical frameworks will continue to be developed. However, even with such advances, it is clear that the best inferences will be from those systems in which there are additional, independent, data on factors such as the genetic architecture of differentiation, the history of population contact and gene flow, the relevant selective forces, and/or the structure and demographic history of populations being examined [53]. Such data can be used, for example, to arbitrate on results of genome scans that would otherwise be ambiguous or potentially biased. With these considerations in mind, below we discuss the current literature on genome scans in plant systems, and then suggest further integrated approaches that can aid in understanding and interpreting genome scan data.

### 3. GENOME SCANS AND RELATED STUDIES IN PLANTS

Based on our summary of a number of recent plant genome scan studies (table 1 and electronic supplementary material, S1), the proportion of loci identified as outliers ranges from 0.4 to 35.5 per cent, with an average of 8.9 per cent. These numbers are comparable with an earlier survey that focused more on animal systems [2] and found that an average of 8.5 per cent of loci were outliers (range 0.4–24.5%). However, as Nosil *et al.* [2] point out, comparisons across studies are difficult because of the variety of analytical approaches used and the range of significance cut-offs used by different researchers. For example, De Carvalho *et al.* [28] used a relatively liberal false discovery rate of 10 per cent for identifying outliers, but only considered loci that were outliers according to two different methods that model different aspects of divergence; in contrast, Richardson *et al.* [26] considered loci significant outliers if they were above the 0.995 quantile. Likewise, multiple comparisons are often made among replicate populations or at different geographical scales; in some cases, only loci found to be outliers in more than one comparison are considered significant [28], whereas others view

unique loci as potentially reflecting local adaptation in specific population comparisons. In addition, researchers often remove some proportion of highest and lowest divergence loci before estimating the putatively neutral genome-wide  $F_{ST}$  value that is used in simulations [54]. However, the proportion trimmed appears somewhat arbitrary; different researchers trim different amounts (usually 20–30%, although as low as 0.5%) from each tail of the distribution, while some do not trim at all. Nosil *et al.* [55] found that varying the amount trimmed between 10 and 30 per cent had a modest impact on the estimated ‘neutral’  $F_{ST}$  and little to no impact on the outlier loci identified; however, Caballero *et al.* [56] found in a simulation study that the use of a trimmed mean  $F_{ST}$  resulted in substantially more outlier loci compared with non-trimmed mean  $F_{ST}$ . In any given study, the impact will depend on the overall distribution of  $F_{ST}$  values, and as such is difficult to predict. Nonetheless, to the extent that it is possible to compare across studies, we note that results in plant systems are quite similar to those reported in animal systems in terms of the numbers of outlier loci identified.

In addition to identifying outlier loci, researchers in a number of studies have attempted to identify associations between these loci and climatic or phenotypic traits thought to potentially be adaptively relevant. For example, a number of studies have found that outliers among populations in several tree species are often significantly associated with aridity or temperature, especially minimum seasonal temperatures [26,41].

Comparisons between potentially adaptive phenotypic divergence among populations and divergence at putatively selected versus neutral loci may be a useful source of additional confirmation. For example, Herrera & Bazaga [33] identified nine loci as high-divergence outliers among populations of the violet species *Viola cazorlensis*, and found that divergence in floral traits potentially under pollinator-mediated selection was significantly associated with divergence at the outlier loci but not the remaining putatively neutral loci; they interpreted this as evidence of adaptive divergence for floral traits at these loci. Alternatively, these loci may reflect local adaptation at some other unknown trait. A similar approach has been taken in comparing associations with different habitats to which populations are potentially adapted at outlier versus neutral loci (reviewed in [2]).

The ability to associate differentiated loci with potentially functionally important variation also allows a more concrete set of conclusions to be drawn from genome scan data. In the most genomically comprehensive scans to date in plants, Turner *et al.* [46] used an *Arabidopsis thaliana* tiling array and high-throughput sequencing of pooled population samples [47] to investigate the genomic basis of adaptation to serpentine soils in *Arabidopsis lyrata*. Using the *A. thaliana* genome as a reference, they were able to show that outlier loci between populations on serpentine versus non-serpentine soils were enriched for genes involved in ion transport and metal detoxification spread throughout the genome. They also found that differentiation between serpentine and non-serpentine soils involves several large (10–80 kb) duplications or deletions on multiple chromosomes, containing numerous genes

with known or suspected roles in protection from toxic compounds. Finally, using replicate serpentine and non-serpentine populations in Scotland and the USA, they found evidence for both parallel and convergent evolution in three highly differentiated genes involved in ion transport and metal detoxification [47].

Other genome scan studies have focused on genomic patterns of reproductive isolation and species differences. For example, Yatabe *et al.* [22] used 108 mapped microsatellites to examine patterns of divergence between broadly sympatric, hybridizing sunflower species, *Helianthus annuus* and *Helianthus petiolaris*. They identified five outlier loci (4.6%), but found no significant association between levels of genetic divergence and proximity to previously identified quantitative trait loci (QTLs) for either species differences or sterility. Based on this, and on the overall low levels of genomic differentiation between the two species, they concluded that the unit of isolation between the two species is probably quite small, with much of the genome permeable to gene flow; other studies using different datasets have confirmed this result [23,57]. However, a study of a parapatric species pair, *H. annuus* and *Helianthus debilis*, showed that markers near QTLs for phenotypic differences and hybrid sterility introgressed at lower than expected rates. The exceptions were two markers near QTLs for traits at which admixed populations had converged phenotypically; these markers introgressed at higher than expected rates. This result is consistent with experimental work indicating adaptive introgression of both biotic and abiotic traits in *H. annuus* ssp. *texanus* [58,59]. In a larger genome scan study involving these species, Scascitelli *et al.* [24] found that three of 88 (3.4%) loci were high-divergence outliers but found no evidence of clustering of outliers (all outliers were on different linkage groups).

Other studies of genomic patterns of reproductive isolation have considered genomic scan data in the context of introgression patterns in hybrid zones [60]. A recently described 'genomic clines' method [45] compares introgression patterns at individual loci relative to the genomic background pattern of introgression to detect several different forms of selection. Using this approach, Gompert & Buerkle [45] reanalysed a dataset from two independent *H. annuus*/*H. petiolaris* hybrid zones and found that 16 of 61 loci differed significantly between zones, compared with one in the original analysis [44]. These results suggest greater intraspecific variation for isolating factors than originally reported. Another study taking this genomic clines approach considered three hybrid zones between the European aspen species *Populus alba* and *Populus tremula* [43]. Circa one-third (33/93) of loci were high-divergence outliers, and two-locus epistasis was common. These results, combined with the fact that genomic patterns of isolation are similar across three hybrid zones, suggest strong intrinsic post-zygotic barriers between the two species.

Both sunflower and aspen hybrid zone analyses found that a large proportion of the genome displayed non-neutral patterns of introgression. However, as with many genome scan approaches, the method of Gompert & Buerkle [45] assumes that a small proportion of the markers analysed are under selection.

As discussed above, if this assumption is violated, the number of loci actually under selection may be under-estimated. In addition, one implementation of the method requires that allele frequencies in the parental populations be similar across all loci. Gompert & Buerkle [45] suggest that this can be achieved by choosing markers with fixed or nearly fixed differences between the parental populations. While this seems reasonable for populations in recent secondary contact, for populations with a longer history of hybridization, selecting for the most differentiated loci will bias the estimated genomic background level of admixture, making it more difficult to detect low-admixture outliers and excluding loci that have experienced adaptive introgression from the analysis entirely.

Another promising approach that to our knowledge has very rarely been tried (but see [24]) is to combine genome scan and QTL mapping approaches with comparisons of demographic parameters such as gene flow rates, effective population sizes and divergence times [61,62] among classes of loci (for example, comparing outliers versus non-outliers or candidate genes versus non-candidate genes). Genes associated with reproductive isolation or species differences should exhibit lower levels of gene flow and potentially smaller local effective population sizes. Variation in divergence time estimates among loci might also be used to suggest specific modes of divergence. For example, when populations are initially strongly isolated at one or a few loci while gene flow throughout the rest of the genome is unimpeded, we might expect to see older divergence time estimates in analyses of datasets containing just these loci compared with datasets containing only neutral loci; however, if this initial isolation results in a substantial genome-wide decrease in gene flow between incipient species, variation in divergence time estimates among datasets would be very limited. Current methods for estimating these parameters [61,62] assume selective neutrality, and it is not clear how selection may affect parameter estimation. A recent simulation study in which a single locus that had undergone a divergent selective sweep was analysed together with neutral loci showed that current effective population size and gene flow estimates were biased downwards as expected, while divergence time estimates were largely unchanged [63]. However, more detailed studies, including studies analysing neutral and selected loci separately, are required to more thoroughly validate this approach. Moreover, estimates of gene flow timing among loci might help determine the order in which reproductive barriers or species differences emerged. Current methods for estimating long-term gene flow timing are not informative [64], but this would be a valuable avenue for future work.

#### 4. CONCLUSIONS

Above, we have tried to highlight examples of plant genome scan studies that go beyond the simple identification of anonymous outliers. While these more enriched approaches are still relatively few, they might already provide some insights into probable modes and models of species divergence. For example, cases where



loci identified as outliers have been mapped to genomic locations suggest two emerging patterns. First, initial results indicate that there is often a poor match between genomic regions containing high-divergence outliers and regions containing QTLs for phenotypic differences or reproductive isolation. There are a number of possible explanations for this lack of congruence across methodologies, both biological and technical. For example, the traits measured in QTL studies might be unimportant as barriers to gene flow in natural populations. Alternatively, if genomic regions of differentiation are small, genomic scan data may be too coarse to detect QTL effects on divergence. In addition, given the many potential sources of false-positives in genome scans, it is possible that some uncorroborated outliers may be artefacts.

Second, genome scans of hybridizing populations or species thus far provide little evidence for large genomic islands of divergence (see electronic supplementary material, S1) as predicted by DH [14]. However, much higher density scans, and comparisons of genomic patterns of divergence of both sympatric and allopatric populations, will be required to address the evidence for or against this and other such models of genomic divergence.

Realizing the promise of genome scan methods clearly requires more theoretical, methodological and empirical development. We have already highlighted the need for clearer expectations about patterns of genomic divergence under specific temporal, selective and genetic scenarios, for more careful consideration of alternative hypotheses in the interpretation of genome scan studies, and for more nuanced models and methods with which to analyse these data [3,12]. Finally, if (as we hope) genome scans are to contribute to our understanding of the genomic architecture of adaptation and speciation, and the nature of the genes involved in these processes, they must move beyond the identification of outlier regions to the dissection and characterization of their underlying genes. To do so will require integration with complementary approaches such as QTL mapping and sequence and functional analyses, which can reveal the phenotypes and physiological responses associated with these loci and uncover the exact genetic changes and functions of the genomic regions involved [5].

This work was supported by a National Institutes of Health Ruth L. Kirschstein Postdoctoral Fellowship (5F32GM072409-02) to J.L.S. and a NSERC grant (327475) to L.H.R.

## REFERENCES

- Turner, T. L., Hahn, M. W. & Nuzhdin, S. V. 2005 Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biol.* **3**, 1572–1578. (doi:10.1371/journal.pbio.0030285)
- Nosil, P., Funk, D. J. & Ortiz-Barrientos, D. 2009 Divergent selection and heterogeneous genomic divergence. *Mol. Ecol.* **18**, 375–402. (doi:10.1111/j.1365-294X.2008.03946.x)
- Feder, J. L. & Nosil, P. 2010 The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution* **64**, 1729–1747. (doi:10.1111/j.1558-5646.2009.00943.x)
- Via, S. & West, J. 2008 The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Mol. Ecol.* **17**, 4334–4345. (doi:10.1111/j.1365-294X.2008.03921.x)
- Nosil, P. & Feder, J. L. 2012 Genomic divergence during speciation: causes and consequences. *Phil. Trans. R. Soc. B* **367**, 332–342. (doi:10.1098/rstb.2011.0263)
- Lewontin, R. C. & Krakauer, J. 1973 Distribution of gene frequency as a test of theory of selective neutrality of polymorphisms. *Genetics* **74**, 175–195.
- Excoffier, L., Hofer, T. & Foll, M. 2009 Detecting loci under selection in a hierarchically structured population. *Heredity* **103**, 285–298. (doi:10.1038/hdy.2009.74)
- Foll, M. & Gaggiotti, O. 2008 A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**, 977–993. (doi:10.1534/genetics.108.092221)
- Hermisson, J. & Pennings, P. S. 2005 Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* **169**, 2335–2352. (doi:10.1534/genetics.104.036947)
- Gompert, Z., Parchman, T. L. & Buerkle, C. A. 2012 Genomics of isolation in hybrids. *Phil. Trans. R. Soc. B* **367**, 439–450. (doi:10.1098/rstb.2011.0196)
- Charlesworth, B., Nordborg, M. & Charlesworth, D. 1997 The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet. Res.* **70**, 155–174. (doi:10.1017/S0016672397002954)
- Feder, J. L., Gejji, R., Yeaman, S. & Nosil, P. 2012 Establishment of new mutations under divergence and genome hitchhiking. *Phil. Trans. R. Soc. B* **367**, 461–474. (doi:10.1098/rstb.2011.0256)
- Rieseberg, L. H., Whitton, J. & Gardner, K. 1999 Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* **152**, 713–727.
- Via, S. 2012 Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Phil. Trans. R. Soc. B* **367**, 451–460. (doi:10.1098/rstb.2011.0260)
- Barton, N. H. & Hewitt, G. M. 1985 Analysis of hybrid zones. *Annu. Rev. Ecol. Syst.* **16**, 113–148. (doi:10.1146/annurev.es.16.110185.000553)
- Pritchard, J. K., Pickrell, J. K. & Coop, G. 2010 The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr. Biol.* **20**, R208–R215. (doi:10.1016/j.cub.2009.11.055)
- Beaumont, M. A. & Nichols, R. A. 1996 Evaluating loci for use in the genetic analysis of population structure. *Proc. R. Soc. Lond. B* **263**, 1619–1626. (doi:10.1098/rspb.1996.0237)
- Balding, D. J. 2003 Likelihood-based inference for genetic correlation coefficients. *Theor. Popul. Biol.* **63**, 221–230. (doi:10.1016/S0040-5809(03)00007-8)
- Nielsen, R. et al. 2009 Darwinian and demographic forces affecting human protein coding genes. *Genome Res.* **19**, 838–849. (doi:10.1101/gr.088336.108)
- Voight, B. F., Kudaravalli, S., Wen, X. Q. & Pritchard, J. K. 2006 A map of recent positive selection in the human genome. *PLoS Biol.* **4**, 446–458. (doi:10.1371/journal.pbio.0040072)
- Michel, A. P., Sim, S., Powell, T. H. Q., Taylor, M. S., Nosil, P. & Feder, J. L. 2010 Widespread genomic divergence during sympatric speciation. *Proc. Natl Acad. Sci. USA* **107**, 9724–9729. (doi:10.1073/pnas.1000939107)
- Yatabe, Y., Kane, N. C., Scotti-Saintagne, C. & Rieseberg, L. H. 2007 Rampant gene exchange across a strong reproductive barrier between the annual sunflowers, *Helianthus annuus* and *H. petiolaris*. *Genetics* **175**, 1883–1893. (doi:10.1534/genetics.106.064469)



- 23 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. (doi:10.1093/molbev/msp043)
- 24 Scascitelli, M., Whitney, K. D., Randell, R. A., King, M., Buerkle, C. A. & Rieseberg, L. H. 2010 Genome scan of hybridizing sunflowers from Texas (*Helianthus annuus* and *H. debilis*) reveals asymmetric patterns of introgression and small islands of genomic differentiation. *Mol. Ecol.* **19**, 521–541. (doi:10.1111/j.1365-294X.2009.04504.x)
- 25 Eckert, A. J., van Heerwaarden, J., Wegrzyn, J. L., Nelson, C. D., Ross-Ibarra, J., Gonzalez-Martinez, S. C. & Neale, D. B. 2010 Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae). *Genetics* **185**, 969–982. (doi:10.1534/genetics.110.115543)
- 26 Richardson, B. A., Rehfeldt, G. E. & Kim, M. S. 2009 Congruent climate-related genecological responses from molecular markers and quantitative traits for western white pine (*Pinus monticola*). *Int. J. Plant Sci.* **170**, 1120–1131. (doi:10.1086/605870)
- 27 Namroud, M. C., Beaulieu, J., Juge, N., Laroche, J. & Bousquet, J. 2008 Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Mol. Ecol.* **17**, 3599–3613. (doi:10.1111/j.1365-294X.2008.03840.x)
- 28 De Carvalho, D. *et al.* 2010 Admixture facilitates adaptation from standing variation in the European aspen (*Populus tremula* L.), a widespread forest tree. *Mol. Ecol.* **19**, 1638–1650. (doi:10.1111/j.1365-294X.2010.04595.x)
- 29 Minder, A. M. & Widmer, A. 2008 A population genomic analysis of species boundaries: neutral processes, adaptive divergence and introgression between two hybridizing plant species. *Mol. Ecol.* **17**, 1552–1563. (doi:10.1111/j.1365-294X.2008.03709.x)
- 30 Saintagne, C., Bodenes, C., Barreneche, T., Pot, D., Plomion, C. & Kremer, A. 2004 Distribution of genomic regions differentiating oak species assessed by QTL detection. *Heredity* **92**, 20–30. (doi:10.1038/sj.hdy.6800358)
- 31 Scotti-Saintagne, C., Mariette, S., Porth, I., Goicoechea, P. G., Barreneche, T., Bodénès, C., Burg, K. & Kremer, A. 2004 Genome scanning for interspecific differentiation between two closely related oak species [*Quercus robur* L. and *Q. petraea* (Matt) Liebl.]. *Genetics* **168**, 1615–1626. (doi:10.1534/genetics.104.026849)
- 32 Savolainen, V. *et al.* 2006 Sympatric speciation in palms on an oceanic island. *Nature* **441**, 210–213. (doi:10.1038/nature04566)
- 33 Herrera, C. M. & Bazaga, P. 2008 Population-genomic approach reveals adaptive floral divergence in discrete populations of a hawk moth-pollinated violet. *Mol. Ecol.* **17**, 5378–5390. (doi:10.1111/j.1365-294X.2008.04004.x)
- 34 Jump, A. S., Hunt, J. M., Martinez-Izquierdo, J. A. & Penuelas, J. 2006 Natural selection and climate change: temperature-linked spatial and temporal trends in gene frequency in *Fagus sylvatica*. *Mol. Ecol.* **15**, 3469–3480. (doi:10.1111/j.1365-294X.2006.03027.x)
- 35 Jump, A. S., Penuelas, J., Rico, L., Ramallo, E., Estiarte, M., Martinez-Izquierdo, J. A. & Lloret, F. 2008 Simulated climate change provokes rapid genetic change in the Mediterranean shrub *Fumana thymifolia*. *Glob. Change Biol.* **14**, 637–643. (doi:10.1111/j.1365-2486.2007.01521.x)
- 36 Oetjen, K. & Reusch, T. B. H. 2007 Genome scans detect consistent divergent selection among subtidal vs. intertidal populations of the marine angiosperm *Zostera marina*. *Mol. Ecol.* **16**, 5156–5167. (doi:10.1111/j.1365-294X.2007.03577.x)
- 37 Rymer, P. D., Manning, J. C., Goldblatt, P., Powell, M. P. & Savolainen, V. 2010 Evidence of recent and continuous speciation in a biodiversity hotspot: a population genetic approach in southern African gladioli (*Gladiolus*; Iridaceae). *Mol. Ecol.* **19**, 4765–4782. (doi:10.1111/j.1365-294X.2010.04794.x)
- 38 Tsumura, Y., Kado, T., Takahashi, T., Tani, N., Ujino-Ihara, T. & Iwata, H. 2007 Genome scan to detect genetic structure and adaptive genes of natural populations of *Cryptomeria japonica*. *Genetics* **176**, 2393–2403. (doi:10.1534/genetics.107.072652)
- 39 Manel, S., Poncet, B. N., Legendre, P., Gugerli, F. & Holderegger, R. 2010 Common factors drive adaptive genetic variation at different spatial scales in *Arabis alpina*. *Mol. Ecol.* **19**, 3824–3835. (doi:10.1111/j.1365-294X.2010.04716.x)
- 40 Poncet, B. N. *et al.* 2010 Tracking genes of ecological relevance using a genome scan in two independent regional population samples of *Arabis alpina*. *Mol. Ecol.* **19**, 2896–2907. (doi:10.1111/j.1365-294X.2010.04696.x)
- 41 Eckert, A. J., Wegrzyn, J. L., Pande, B., Jermstad, K. D., Lee, J. M., Liechty, J. D., Tarse, B. R., Krutovsky, K. V. & Neale, D. B. 2009 Multilocus patterns of nucleotide diversity and divergence reveal positive selection at candidate genes related to cold hardiness in coastal Douglas fir (*Pseudotsuga menziesii* var. *menziesii*). *Genetics* **183**, 289–298. (doi:10.1534/genetics.109.103895)
- 42 Rieseberg, L. H., Kim, S. C., Randell, R. A., Whitney, K. D., Gross, B. L., Lexer, C. & Clay, K. 2007 Hybridization and the colonization of novel habitats by annual sunflowers. *Genetica* **129**, 149–165. (doi:10.1007/s10709-006-9011-y)
- 43 Lexer, C., Joseph, J. A., van Loo, M., Barbara, T., Heinze, B., Bartha, D., Castiglione, S., Fay, M. F. & Buerkle, C. A. 2010 Genomic admixture analysis in European *Populus* spp. reveals unexpected patterns of reproductive isolation and mating. *Genetics* **186**, 699–712. (doi:10.1534/genetics.110.118828)
- 44 Buerkle, C. A. & Rieseberg, L. H. 2001 Low intraspecific variation for genomic isolation between hybridizing sunflower species. *Evolution* **55**, 684–691. (doi:10.1554/0014-3820(2001)055[0684:LIVFGI]2.0.CO;2)
- 45 Gompert, Z. & Buerkle, C. A. 2009 A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. *Mol. Ecol.* **18**, 1207–1224. (doi:10.1111/j.1365-294X.2009.04098.x)
- 46 Turner, T. L., Von Wettberg, E. J. & Nuzhdin, S. V. 2008 Genomic analysis of differentiation between soil types reveals candidate genes for local adaptation in *Arabidopsis lyrata*. *PLoS ONE* **3**, e3183. (doi:10.1371/journal.pone.0003183)
- 47 Turner, T. L., Bourne, E. C., Von Wettberg, E. J., Hu, T. T. & Nuzhdin, S. V. 2010 Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nat. Genet.* **42**, 260–263. (doi:10.1038/ng.515)
- 48 Meyer, C. L., Vitalis, R., Saumitou-Laprade, P. & Castric, V. 2009 Genomic pattern of adaptive divergence in *Arabidopsis halleri*, a model species for tolerance to heavy metal. *Mol. Ecol.* **18**, 2050–2062. (doi:10.1111/j.1365-294X.2009.04159.x)
- 49 Teshima, K. M., Coop, G. & Przeworski, M. 2006 How reliable are empirical genomic scans for selective sweeps? *Genome Res.* **16**, 702–712. (doi:10.1101/gr.5105206)
- 50 Barrett, R. D. H. & Schluter, D. 2008 Adaptation from standing genetic variation. *Trends Ecol. Evol.* **23**, 38–44. (doi:10.1016/j.tree.2007.09.008)
- 51 Latta, R. G. 2004 Gene flow, adaptive population divergence and comparative population structure across loci.

- New Phytol.* **161**, 51–58. (doi:10.1046/j.1469-8137.2003.00920.x)
- 52 Nachman, M. W. & Payseur, B. A. 2012 Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. *Phil. Trans. R. Soc. B* **367**, 409–421. (doi:10.1098/rstb.2011.0249)
  - 53 Renaut, S., Maillet, N., Normandeau, E., Sauvage, C., Derome, N., Rogers, S. M. & Bernatchez, L. 2012 Genome-wide patterns of divergence during speciation: the lake whitefish case study. *Phil. Trans. R. Soc. B* **367**, 354–363. (doi:10.1098/rstb.2011.0197)
  - 54 Bonin, A., Taberlet, P., Miaud, C. & Pompanon, F. 2006 Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Mol. Biol. Evol.* **23**, 773–783. (doi:10.1093/molbev/msj087)
  - 55 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. (doi:10.1111/j.1558-5646.2007.00299.x)
  - 56 Caballero, A., Quesada, H. & Rolan-Alvarez, E. 2008 Impact of amplified fragment length polymorphism size homoplasy on the estimation of population genetic diversity and the detection of selective loci. *Genetics* **179**, 539–554. (doi:10.1534/genetics.107.083246)
  - 57 Kane, N. C., King, M. G., Barker, M. S., Raduski, A., Karrenberg, S., Yatabe, Y., Knapp, S. J. & Rieseberg, L. H. 2009 Comparative genomic and population genetic analyses indicate highly porous genomes and high levels of gene flow between divergent *Helianthus* species. *Evolution* **63**, 2061–2075. (doi:10.1111/j.1558-5646.2009.00703.x)
  - 58 Whitney, K. D., Randell, R. A. & Rieseberg, L. H. 2006 Adaptive introgression of herbivore resistance traits in the weedy sunflower *Helianthus annuus*. *Am. Nat.* **167**, 794–807. (doi:10.1086/504606)
  - 59 Whitney, K. D., Randell, R. A. & Rieseberg, L. H. 2010 Adaptive introgression of abiotic tolerance traits in the sunflower *Helianthus annuus*. *New Phytol.* **187**, 230–239. (doi:10.1111/j.1469-8137.2010.03234.x)
  - 60 Payseur, B. A. 2010 Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Mol. Ecol. Resour.* **10**, 806–820. (doi:10.1111/j.1755-0998.2010.02883.x)
  - 61 Becquet, C. & Przeworski, M. 2007 A new approach to estimate parameters of speciation models with application to apes. *Genome Res.* **17**, 1505–1519. (doi:10.1101/gr.6409707)
  - 62 Hey, J. 2010 Isolation with migration models for more than two populations. *Mol. Biol. Evol.* **27**, 905–920. (doi:10.1093/molbev/msp296)
  - 63 Strasburg, J. L. & Rieseberg, L. H. 2010 How robust are ‘isolation with migration’ analyses to violations of the IM model? A simulation study. *Mol. Biol. Evol.* **27**, 297–310. (doi:10.1093/molbev/msp233)
  - 64 Strasburg, J. L. & Rieseberg, L. H. 2011 Interpreting the estimated timing of migration events between hybridizing species. *Mol. Ecol.* **20**, 2353–2366. (doi:10.1111/j.1365-294X.2011.05048.x)