Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Insect Science

# Genomic transitions during host race and species formation

Meredith M Doellman and Jeffrey L Feder

Darwin recognized species as discontinuous, yet considered them to be formed by an incremental process of natural selection. Recent theoretical work on 'genome-wide congealing' is bridging this gap between the gradualism of divergent selection and rapid genome-wide divergence, particularly during ecological speciation-with-gene-flow. Host races and species of phytophagous insects, displaying a spectrum of divergence and gene flow among member taxa, provide model systems for testing predicted non-linear transitions from 'genic' divergence at a few uncoupled loci to 'genomic' divergence with genome-wide coupling of selected loci and strong reproductive isolation. Integrating across natural history, genomics, and evolutionary theory, emerging research suggests a tipping point from 'genic' to 'genomic' divergence between host races and species, during both sympatric speciation and secondary contact.

**Address**
Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556, USA

Corresponding author: Doellman, Meredith M (mdoellma@nd.edu)

**Current Opinion in Insect Science** 2019, **31**:84–92

This review comes from a themed issue on **Insect genomics**

Edited by **John G Oakeshott, Charles Robin** and **Karl HJ Gordon**

For a complete overview see the Issue and the Editorial

Available online 26th November 2018

**https://doi.org/10.1016/j.cois.2018.11.006**

## Introduction

Speciation is a continuous process with taxa often envisioned as diverging along a continuum, as they evolve from freely interbreeding populations to partially differentiated races to fully reproductively isolated species [1,2]. In the Darwinian world view, new species result from natural selection causing differences, usually of relatively small effect on fitness, to gradually accumulate between populations and generate reproductive isolation (RI) [3]. However, new forms sometimes appear suddenly in the geological record [4]. As a result, some have cast the process of speciation as more abrupt, involving non-Darwinian processes other than natural selection. In punctuated equilibrium, genetic drift associated with the founding of small peripheral populations has been hypothesized to cause rapid transitions to new species [4,5]. Can these seemingly different views concerning the dynamics of speciation be reconciled?

Recent theoretical and empirical work on the coupling of barriers to gene flow during speciation have helped to address this question and paradoxically imply that non-linear temporal dynamics involving rapid transitions from races to species are actually predicted by gradual Darwinian natural selection [6••,7,8••,9••]. Research on insects and, in particular, phytophagous specialists that form host races on different plants, is helping shape this emerging view about how the accumulation of small changes may, at certain points in time, push populations past tipping points and lead to the rapid origin of new species [8••]. Here, we discuss recent theoretical models describing non-linear transition dynamics, review empirical work consistent with these models, and lay out a plan for comprehensive testing of these models with empirical data, with an emphasis on host races and speciation in phytophagous insects.

## The idea of rapid genome-wide congealing

Phytophagous insects have been studied as models of ecological speciation-with-gene-flow, with partially reproductively isolated host races representing early stages of divergence [1,10–13]. Whether and when host races diverge into species depends on the antagonism between gene flow and divergent natural selection, usually due to host-associated ecological or environmental fitness trade-offs [14]. When the magnitude of divergent selection ($s$) acting on a gene between host races is less than the migration rate ($m$), then little allele frequency divergence will occur for the locus [6]. Hence, the majority of the genome is expected to be homogenous between host races of phytophagous insects experiencing high gene flow. Only those mutations having large effects on fitness greater than the migration rate ($s > m$) will establish, attain high levels of allele frequency differentiation between races, and generate RI [15]. This has led to a 'genic' view of divergence, with the majority of the genetic differences between host races reflecting large effect genes that establish due to strong direct selection overpowering $m$ [16]. If the rate at which such large effect differences arise due to mutation is roughly uniform through time, then host races will also diverge and become reproductively isolated from one another at a generally constant rate.

However, genomic data have implied that host races are often also differentiated by polygenic traits encoded by many small effect loci [17•,18••,19,20,21]. When gene flow between host races is high and each small effect locus is subject to relatively weak selection ($s < m$), then genetic divergence may be expected to be generally low and dispersed through the genome. However, theory implies that as many small effect mutations gradually accumulate between populations through time, a threshold or tipping point can be reached where their effects combine to collectively reduce the effective rate of gene flow ($m_e$) below their individually small $s$ values. At this point, a rapid non-linear transition can occur from low to high genome-wide divergence for selected loci and RI termed 'genome-wide congealing' (GWC) [6••,7]. Essentially, the genome increasingly becomes the unit of divergent selection between populations, as the indirect effects of selection on small effect loci become coupled and whole migrant and hybrid genotypes are selected against rather than just individual loci. As a consequence, allele frequency differences for these selected loci dramatically increase and, accordingly, so do the levels of linkage disequilibrium (LD) they display among each other both within and between populations [6••,7,9••]. In this regard, LD within and between populations can be a particularly informative metric of the extent to which the genome as a whole is affected differentially by selection [6••]. Natural selection on polygenic adaptations encoded by many small effect loci therefore actually predicts rapid evolutionary transitions from races to species under certain conditions of divergence with gene flow (see below for discussion about allopatric divergence).

It is important to note that not all loci need be of small effect to see such rapid transitions. A number of large effect loci, as well as structural features of the genome such as inversions and translocations that can reduce recombination between genes and enhance their combined indirect effects on reducing gene flow, may facilitate initial race formation [22–25]. By increasing LD among loci under divergent selection, these genome rearrangements can promote progress toward the tipping point, which may otherwise be a slow process if predicated only on the accumulation of new, small effect mutations [26]. However, when many subsequent changes are based on mutations of small effect, temporally non-linear dynamics will still be observed in patterns of genomic divergence as races evolve into new species. After a lag period, neutral mutations can also show non-linear patterns of temporal divergence much like selected sites, as RI increasingly becomes a genome-wide property and $m_e$ reduces significantly below the inverse of twice the effective population size ($1/2n_e$), elevating the role of genetic drift in population divergence [9••].

It is also worth stressing that GWC theory is not restricted to cases where population divergence is initiated in the face of gene flow in sympatry or parapatry. Similar considerations likely also apply to secondary contact following allopatric divergence. Although divergence in allopatry is expected to proceed linearly in time (but see Refs. [27,28]), non-linear temporal dynamics can be observed in the outcomes of gene flow upon secondary contact. Fusion versus persistence can depend on whether or not populations have reached a threshold of coupling among loci underlying RI sufficient to prevent large scale genomic introgression [8••,18••]. Moreover, RI in allopatry may also be the result of inherent genomic incompatibilities [29]. Alternate alleles could be fixed under genetic drift, directional selection in similar ecological conditions or, most common in cases of speciation-with-gene-flow, divergent selection between differing habitats. The prediction of a rapid transition from races to species, under scenarios of both primary divergence in the face of gene flow and maintenance of species boundaries upon secondary contact may ultimately help to explain punctuated patterns of biodiversity in nature and the fossil record, and are not anomalous with respect to Darwinian theory.

## Genome-wide congealing: theoretical predictions and empirical support

The GWC framework provides-specific predictions regarding the expected genome-wide patterns of population genetics statistics, for both selected and neutral sites (see Box 1 for roadmap for testing GWC). During the 'genic' phase characteristic of partially isolated host races, a comparatively long period of build-up of genome-wide standing variation can occur [7]. Genome-wide divergence is expected to be weak, with genetic differentiation concentrated primarily around selected loci and little to no between or within deme (population) LD among physically unlinked loci [6,9••] (Figure 1). Apple and hawthorn host races of *Rhagoletis pomonella* (see Box 2) [17•,30•,31,32], *Adenostoma* and *Ceanothus* attacking populations of *Timema cristinae* stick insects [18••,33], and alfalfa (*Medicago sativa*), clover (*Trifolium pratense*), and pea (*Pisum sativum*) races of pea aphids (*Acyrthosiphon pisum*) [34,35,36•,37] appear to reside within this 'genic' phase, showing relatively high ongoing gene flow and heterogeneous patterns of differentiation at selected sites, while background genomic divergence remains low. *R. pomonella* host races also display low LD among selected loci on different chromosomes [32], and both *R. pomonella* and *Timema* spp. host races fail to cluster globally by host affiliation [18] (see Table 1).

As the rapid transition from host race to species occurs, emergent genomic processes can negate gene flow and foster rapid genome-wide divergence through positive feedback mechanisms. Coupling among selected loci occurs, as a combination of direct and indirect selection facilitates accelerated differentiation and the build-up of both between and within deme LD for these loci [6••,7,9••] (Figure 1). Consistent with classic spatial

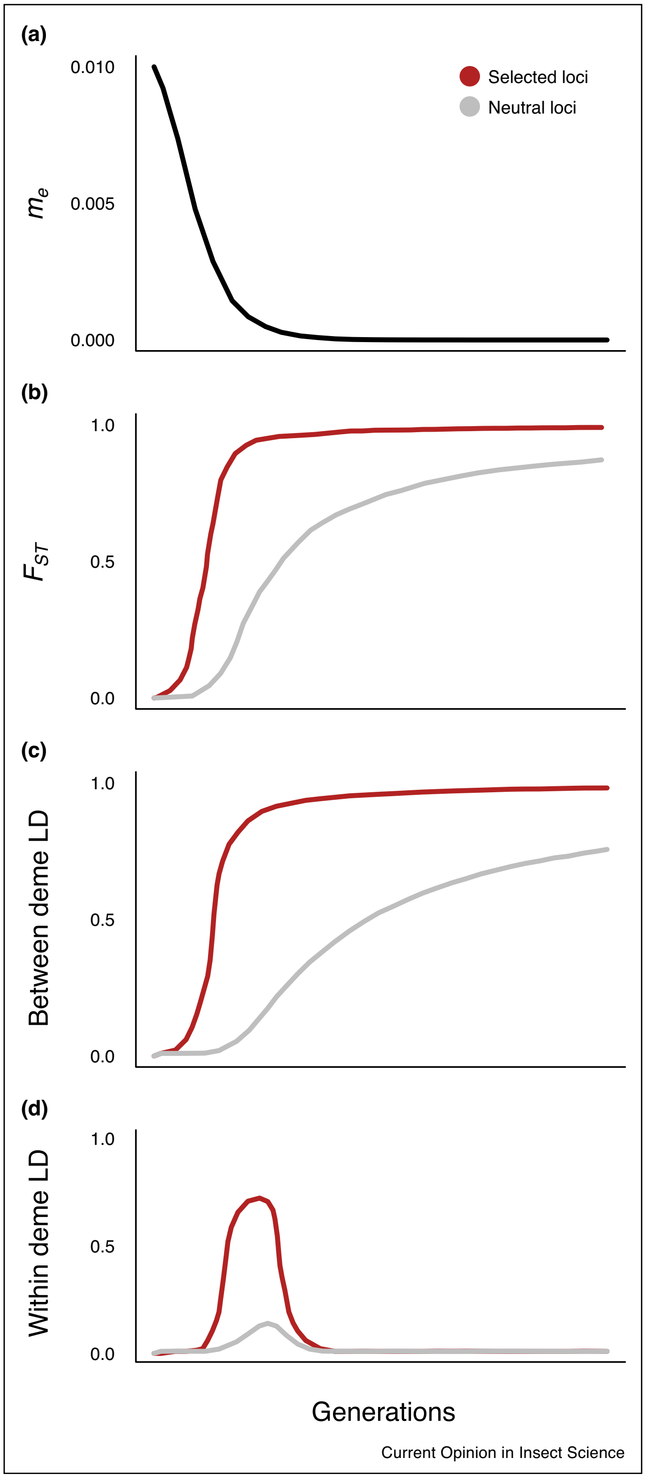
**Box 1 Road map for testing GWC theory.**

One challenge in linking theoretical predictions of GWC to empirical data lies in finding tractable systems for testing the hypothesis. The ideal system would allow for sampling the full speciation continuum, all within one clade of related populations with members in spatial contact currently and throughout their divergence. This system should have a rich history of ecological and evolutionary research, in which the important reproductive isolating barriers and axes of divergence are well characterized and history of gene flow is known [7]. Putative loci under selection can be identified in a variety of ways, through outlier analysis (although see Refs. [44,48,49]), scans for signatures of selection [36,50], genome-wide association studies (GWAS) [17,18••,51], quantitative trait locus (QTL) mapping [52], and genomic and geographic cline analyses [43]. Manipulative experiments should also be used both to confirm that identified loci do indeed change in response to selection as predicted and to estimate the strength of selection (*s*) for model parameterization [19,21,30•,33,53,54•,55]. Gene flow (*m*) can be estimated in field studies and from genomic data, and demographic modeling can be used to reconstruct the history of gene flow in the system to establish its suitability for testing GWC theory [56].

Using empirical parameter estimates (*m* and *s*), system-specific GWC predictions can be generated, and diagnostic patterns of population genetic statistics, for both selected and neutral classes of loci, established (see discussion of theoretical predictions). Genome-wide patterns of divergence can be described by scanning the genome with traditional $d_{xy}$ and $F_{ST}$ approaches and hidden Markov models (HMM) to identify regions of enhanced differentiation [18••,57].

Comparisons can then be drawn between theoretical and empirical distributions of these statistics as well as genome-wide LD [9••], for both selected and neutral classes of loci. Finally, species that have transitioned past the GWC threshold and are in the 'genomic' phase of divergence should cluster distinctly across their ranges. In contrast, while differing locally at sympatric sites host races still in the 'genic' phase of speciation, where the direct effects of selection on individual loci are mainly responsible for divergence, will tend not to cluster globally across the entirety of their ranges. Both model-based and model free clustering analyses, such as STRUCTURE (and related models [58–60]) and discriminant analysis of principal components (DAPC) can be used to assess the level of clustering from host races to species [32,61].

**Figure 1**



Current Opinion in Insect Science

Predicted non-linear transitions during GWC in **(a)** effective migration rate ($m_e$), **(b)** $F_{ST}$, **(c)** between deme LD, and **(d)** within deme LD. Expected patterns for selected loci are depicted in red, while neutral loci are grey. Figures adapted from Refs. [6••,9••], based on GWC simulations with $s = 0.005$ and $m = 0.01$.

coupling models developed for geographic clines, these non-linear temporal transitions occur only if loci are allowed to build-up LD (i.e. not if loci are selected in a 'bean bag genetics' fashion) [6••,38,39•]. Higher gene flow requires stronger coupling for GWC [9••]; however, if the ratio of migration to selection is too high, coupling and divergence will not occur [6••]. As GWC is rapid on an evolutionary or geological scale, taxa actually in the transitional stage may be relatively rare in nature, and therefore, not often observed [8••,18••].

In the 'genomic' phase of divergence RI has transitioned from being a property of specific loci to an increasing property of the genome. Crossing this tipping point can signify the formation of a new species [14,40]. A congealed genome is expected to show genome-wide elevated allele frequency differences, $F_{ST}$ and between deme LD (including between loci on different

**Box 2 Case study: Rhagoletis pomonella species complex.**

The *R. pomonella* species complex provides an appropriate model for testing GWC theory. This well-characterized ongoing radiation displays a spectrum of divergence and gene flow among member taxa; in addition to the well-known apple and hawthorn host races, the *R. pomonella* species complex includes several native hawthorn races, three other named species, *Rhagoletis mendax*, *Rhagoletis zephyria*, and *Rhagoletis cornivora*, several undescribed species, and an isolated hawthorn-infesting *R. pomonella* population in the Mexican highlands [62–65]. Microsatellite work has shown that levels of genome-wide differentiation do indeed span the speciation continuum; named species cluster distinctly across their entire range, while host races do not, and divergence between *R. pomonella* and the undescribed flowering dogwood fly likely lies near the transition point from host races to species [66,67]. Recent selection experiments and GWAS have identified loci underlying key life history traits under differential selection between apple and hawthorn host races of *R. pomonella* [17•,30•,32]. Future work should evaluate LD relationships among these loci and between these and putatively neutral loci within and between the host races and species in the complex for evidence of rapid transitions between 'genic' and 'genomic' phases of ecological divergence supporting the GWC hypothesis (Figure 3).

chromosomes) for selected loci [6••,7] (Figure 1). In contrast, within deme LD is predicted to transiently elevate as migrants with increasingly diverged genomes are exchanged between host races but then fall as effective gene flow of whole genomes is curtailed by selection (note that this does not consider LD from structural variants) [9••]. Early in the 'genomic' phase, neutral sites will show less genome-wide differentiation and lower LD than selected sites [9••]. Thus, from the example above, alfalfa, clover, and pea attacking host races of the aphid *A. pisum* display hot spots of increased genetic divergence distributed throughout the genome that contain many candidate chemosensory genes [34,35,36•,37]. If strong LD were to be shown among candidate genes in any host race comparison, those aphids might be early in the 'genomic' phase of divergence despite low background (neutral) divergence. Neutral loci are expected to show non-linear temporal dynamics eventually but after a lag period, with rapid divergence occurring 10 s–100 s–1000 s of generations after selected sites [9••]. In this regard, a recent targeted sequencing study demonstrated that *A. pisum* associated with *Lathyrus pratensis* (meadow vetchling) displays genome-wide elevated divergence among putatively selected chemosensory genes, as well as putative neutral loci [41•], suggesting a transition into the 'genomic' phase of divergence. Combined with evidence of low gene flow (very few hybrids) with the other *A. pisum* host races, this reinforces the likelihood that *L. pratensis*-associated *A. pisum* is nearing complete speciation [34,35].

However, caution is warranted because neutral processes can, in some cases, produce similar non-linear divergence patterns, especially with long periods of reduced gene flow between allopatric populations [42•]. However,

**Table 1**

**Model phytophagous insect systems discussed in the text for testing GWC.**

| Taxon | Levels of divergence | Divergence with Gene flow | Genome-wide divergence | Patterns of LD | Putative selected loci | Citations |
|---|---|---|---|---|---|---|
| *Heliconius:* melpomene-clade and cydno-clade | Geographic races to species | Likely history of sympatry, parapatry, and allopatry | Low heterogenous divergence among races; elevated divergence between species, especially on Z chromosome | High between loci on same and different chromosomes for species not race comparisons | Wing patterning and mate preference (tightly linked) | [46,47,72–75] |
| *Timema* spp. | Ecotypes to species | Ongoing sympatry between ecotypes; likely allopatry and secondary contact among species | Low heterogenous divergence among ecotypes, including one major site on LG8; elevated genome-wide divergence between sympatric species | Unknown | Color pattern (LG 8); cuticular hydrocarbons (highly polygenic) | [18,33,51] |
| *Acyrthosiphon pisum* | Host races to putative species | Likely evolved via host shifts in sympatry | Divergence within hotspots but not background among races; elevated divergence in hotspots and background between putative species | Unknown | Chemosensory genes (gustatory and olfactory receptors; genome-wide distribution of hotspots, more common on autosomes) | [34–37,41] |
| *Rhagoletis pomonella* species complex | Host races to species | Likely evolved via host shifts in sympatry | Low heterogeneous divergence among races; accentuated divergence between species | When considering host races together, no LD among selected loci on different chromosomes | Eclosion timing (highly polygenic, concentrated in inversions on chr. 1–3; Pre-winter diapause depth (highly polygenic) | [17,30–32] |

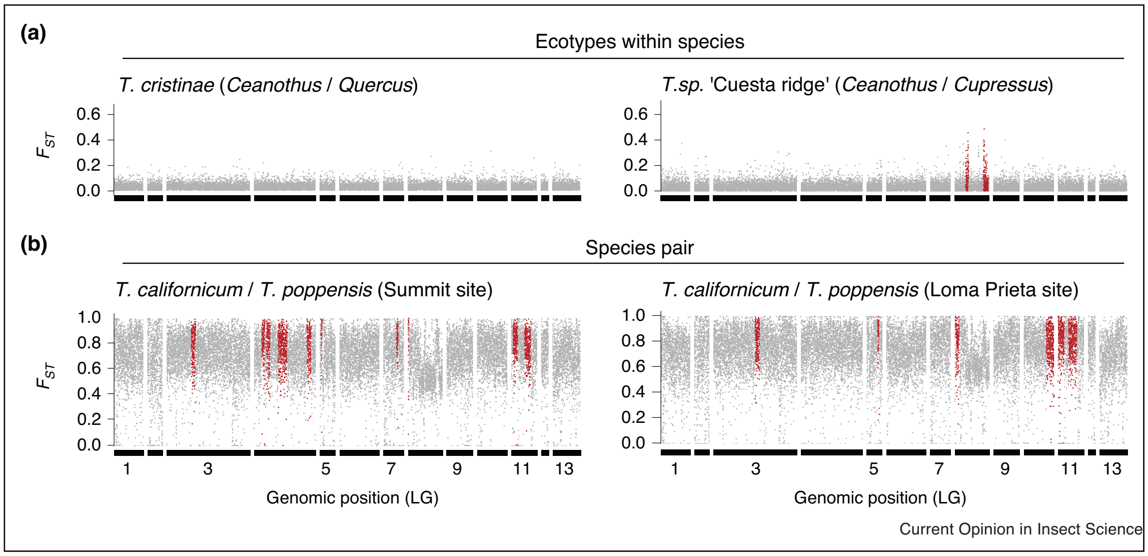
when a history of gene flow is known, non-linear transitions in divergence are good indicators of GWC dynamics. This highlights the importance of knowing the demographic histories of populations, the historical role of gene flow in divergence, and the loci or gene regions under divergent selection (ideally corroborated by manipulative experiments and GWAS) before interpreting patterns of differentiation and the significance of outlier loci generated in genome scan studies of natural populations (see Box 1).

## Genome-wide congealing: beyond strict sympatry

The genome-wide congealing framework extends to other modes of speciation beyond sympatry, informing predictions in scenarios of parapatry and secondary contact following allopatric divergence, as well [8••,18••] (see Table 1). Taxa that may require a degree of geographic separation to speciate, such as many species of *Timema* stick insects, show a range of levels of differentiation, in which intermediate stages of divergence may be transient or remain intact when challenged with increased levels of gene flow in nature or through manipulative field transplant experiments [18••]. The shape of spatial and genomic clines can be used to infer the degree to which coupling has occurred and genome-wide divergence displays non-linear patterns in space [43]. However, non-linear dynamics in time are also predicted by GWC theory, given the threshold of divergence and coupling required to prevent fusion [8••,18••]. Riesch *et al.* [18••] showed discontinuities in $F_{ST}$ values in between host race and species level population comparisons, and conducted selection and association experiments to identify loci putatively under selection. Putatively selected loci showed higher $F_{ST}$ in sympatric host race comparisons in the 'genic' phase of divergence, while $F_{ST}$ was elevated genome-wide for both selected and neutral loci, as expected for congealed genomes, in sympatric species level comparisons (Figure 2).

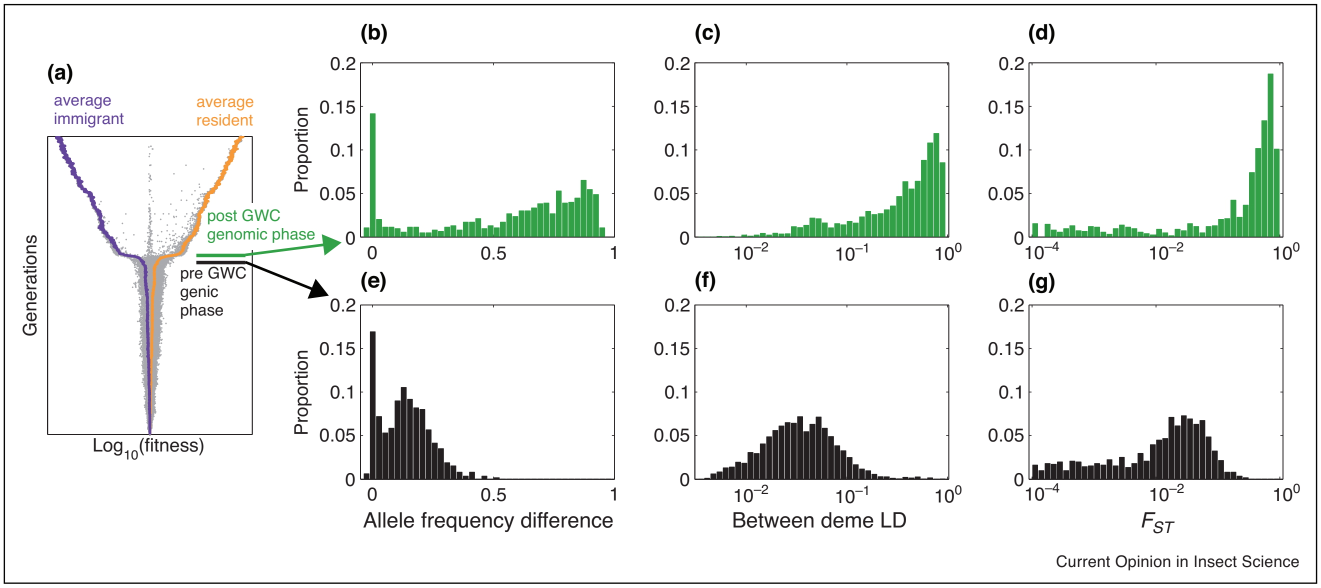
Divergence among *Heliconius* butterflies, which form geographic and mimicry-related rather than host-associated races and species, may also have involved periods of past allopatry. Recent work has shown that $F_{ST}$ outliers (putatively selected loci) between the strongly divergent sympatric species *Heliconius melpomene* and either *Heliconius cydno galanthus* or *Heliconius pachinus* have increased intrachromosomal within deme LD compared to within the more weakly divergent *H. cydno galanthus* and *H. pachinus* [9••]. *H. melpomene* also showed evidence of coupling among outlier loci across chromosomes, supported by a peak of higher pairwise LD among some loci. No coupling was observed among chromosomes in the weakly diverged parapatric pair *H. cydno galanthus* and *H. pachinus*. Thus, *H. melpomene* has possibly transitioned into the 'genomic' phase of divergence from *H. cydno galanthus* and *H. pachinus*, while the latter less

**Figure 2**



Sample $F_{ST}$ scans for ecotype and species level divergence in *Timema* spp. stick insects, showing regions of accentuated differentiation (red loci) relative to genomic background differentiation (grey loci). Highlighted regions (red) identified by HMM analysis likely contain genes differentially selected or contributing to RI between host associations or species. **(a)** Comparisons between host-associated ecotypes within species ('genic' phase of divergence) display low divergence genome-wide, with limited accentuated divergence around color pattern genes on linkage group 8 (LG8). **(b)** Genome-wide differentiation is high for both regions of accentuated divergence and the genomic background between species *T. californicum* and *T. poppensis* sampled at two sympatric sites ('genomic' phase of divergence). Divergent selection on highly polygenic traits, such as cuticular hydrocarbons, likely drives genome-wide divergence in these comparisons. Figure modified from Ref. [18••].

**Figure 3**



Non-linear divergence dynamics during GWC, modeled based on empirical *R. pomonella* data ($s = 0.03$, $m = 0.05$). **(a)** The fitness of the average immigrant (purple) and average resident (orange) within a deme diverge rapidly during the transition from 'genic' to 'genomic' phases of GWC. **(b–g)** Predicted distributions of allele frequency differences, between deme LD, and $F_{ST}$ for selected loci for populations in the 'genic' and 'genomic' phases of divergence. (b–d) Comparisons between *R. pomonella* and other species, such as the blueberry maggot *R. mendax*, should resemble the 'genomic' phase of divergence (green). (e–g) Patterns of differentiation among apple and hawthorn host races of *R. pomonella* should resemble the 'genic' phase of divergence (black). Figure modified from Ref. [7].

divergent pair appears to be in the 'genic' phase of divergence. Such patterns may prove to be common in nature, as genome-wide data accumulate more support for complex histories of alternating allopatry and gene flow or mixed modes of speciation [44–47].

## Conclusions

Both theoretical and empirical challenges remain concerning the role GWC plays in population divergence and speciation. Future modeling work on GWC should incorporate the recombination landscape more fully, examine the effects of pleiotropy and epistasis on the process, extend existing models more formally to scenarios of secondary contact (but see Ref. [6••]), and investigate alternative forms of selection and reproductive isolation (e.g. inherent genomic incompatibilities, sexual selection, and reinforcement). When assessing empirical data generated by genomic sequencing studies, the combination of forward-time simulations, coalescent simulations, and model-based parameter inference necessary for good null models may pose a computational burden to testing the predictions of GWC. Finally, much empirical work is still based solely on genome scans [57,68–71], and elucidating the demographic history of a population and characterizing the timing and extent of past and ongoing gene flow remain non-trivial challenges [56].

In conclusion, host races and species of phytophagous insects offer a wealth of opportunities for investigating speciation. The increasing ease with which high throughput DNA sequence data can be cheaply and quickly generated no longer greatly impedes connecting genomic patterns of differentiation with the population genetic processes driving divergence and speciation. Broad support for non-linear dynamics and tipping points between host races and species of phytophagous insects would imply that emergent processes of selection acting on the genome beyond its direct effects on individual genes may contribute greatly to speciation. Such non-linear temporal dynamics of the coupling of reproductive barriers to gene flow would link micro- and macro-evolutionary processes and help further elucidate how Darwinian theory of evolution by natural selection generates perceived gaps between species in nature and discontinuities in the fossil record [4,8••].

## Conflict of interest statement

Nothing declared.

## Acknowledgements

This work was supported by funding from the United States National Science Foundation (DEB-1638997) and the United States Department of Agriculture (NIFA 2015-67013-23289).

## References

1. Nosil P: *Ecological Speciation*. Oxford University Press; 2012.

2. Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, Hohenlohe PA, Peichel CL, Saetre G-P, Bank C, Brännström A *et al.*: **Genomics and the origin of species**. *Nat Rev Genet* 2014, **15**:176-192.

3. Darwin C: *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. John Murray; 1859.

4. Gould SJ: *The Structure of Evolutionary Theory*. Belknap Press; 2002.

5. Mayr E: *Animal Species and Evolution*. Harvard University Press; 1963.

6. •• Flaxman SM, Wacholder AC, Feder JL, Nosil P: **Theoretical models of the influence of genomic architecture on the dynamics of speciation**. *Mol Ecol* 2014, **23**:4074-4088.

This theoretical study establishes the conditions under which GWC is likely and lays the foundation for further work on tipping points in transitions between host races and species. They showed that, in the face of strong gene flow, the accumulation of LD among weakly selected loci could couple to drive rapid transitions from low to high RI genome-wide.

7. Feder JL, Nosil P, Wacholder AC, Egan SP, Berlocher SH, Flaxman SM: **Genome-wide congealing and rapid transitions across the speciation continuum during speciation with gene flow**. *J Hered* 2014, **105**:810-820.

8. •• Nosil P, Feder JL, Flaxman SM, Gompert Z: **Tipping points in the dynamics of speciation**. *Nat Ecol Evol* 2017, **1**:0001-0008.

Building on the idea of GWC, this review of potential tipping points in speciation, links speciation to systems biology work on emergent properties and alternate stable states.

9. •• Schilling M, Mullen S, Kronforst M, Safran R, Nosil P, Feder J, Gompert Z, Flaxman S: **Transitions from single- to multi-locus processes during speciation with gene flow**. *Genes (Basel)* 2018, **9**:274.

This paper extends the original GWC framework to include predicted patterns of divergence for neutral loci and an examination of the temporal dynamics of within deme LD, as well as incorporating a classical measure of coupling among both selected and neutral loci. The work also includes an explicit test of GWC model predictions, utilizing whole-genome sequence data from *Heliconius* butterflies.

10. Berlocher SH, Feder JL: **Sympatric speciation in phytophagous insects: moving beyond controversy?** *Annu Rev Entomol* 2002, **47**:773-815.

11. Drès M, Mallet J: **Host races in plant-feeding insects and their importance in sympatric speciation**. *Philos Trans R Soc Lond B Biol Sci* 2002, **357**:471-492.

12. Forbes AA, Devine SN, Hippee AC, Tvedte ES, Ward AKG, Widmayer HA, Wilson CJ: **Revisiting the particular role of host shifts in initiating insect speciation**. *Evolution (N Y)* 2017, **71**:1126-1137.

13. Vertacnik KL, Linnen CR: **Evolutionary genetics of host shifts in herbivorous insects: insights from the age of genomics**. *Ann N Y Acad Sci* 2017, **1389**:186-212.

14. Feder JL, Egan SP, Nosil P: **The genomics of speciation-with-gene-flow**. *Trends Genet* 2012, **28**:342-350.

15. Yeaman S, Whitlock MC: **The genetic architecture of adaptation under migration-selection balance**. *Evolution* 2011, **65**:1897-1911.

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

17. • Ragland GJ, Doellman MM, Meyers PJ, Hood GR, Egan SP, Powell THQ, Hahn DA, Nosil P, Feder JL: **A test of genomic modularity among life-history adaptations promoting speciation with gene flow**. *Mol Ecol* 2017, **26**:3926-3942.

This study demonstrates that adult eclosion timing, a critical trait underlying differentiation between *R. pomonella* host races, is a polygenic trait and genetically independent from prewinter diapause depth [30]. Together with Egan *et al.* [30], this study lays the groundwork for testing for GWC in the *R. pomonella* complex by identifying loci putatively under differential selection.

18. •• Riesch R, Muschick M, Lindtke D, Villoutreix R, Comeault AA, Farkas TE, Lucek K, Hellen E, Soria-Carrasco V, Dennis SR *et al.*: **Transitions between phases of genomic differentiation during stick-insect speciation**. *Nat Ecol Evol* 2017, **1**:0082-0094.

This study of genomic differentiation across a spectrum of divergence in *Timema* stick insects demonstrates patterns of genic divergence between sympatric races experiencing ongoing gene flow and genomic divergence between sympatric reproductively isolated species, with a gap in levels of divergence between.

19. Nosil P, Villoutreix R, de Carvalho CF, Farkas TE, Soria-Carrasco V, Feder JL, Crespi BJ, Gompert Z: **Natural selection and the predictability of evolution in Timema stick insects**. *Science* 2018, **359**:765-770.

20. Gompert Z, Jahner JP, Scholl CF, Wilson JS, Lucas LK, Soria-Carrasco V, Fordyce JA, Nice CC, Buerkle CA, Forister ML: **The evolution of novel host use is unlikely to be constrained by trade-offs or a lack of genetic variation**. *Mol Ecol* 2015, **24**:2777-2793.

21. Gompert Z, Messina FJ: **Genomic evidence that resource-based trade-offs limit host-range expansion in a seed beetle**. *Evolution (N Y)* 2016, **70**:1249-1264.

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

23. Charlesworth B, Barton NH: **The spread of an inversion with migration and selection**. *Genetics* 2018, **208**:377-382.

24. Feder JL, Nosil P: **Chromosomal inversions and species differences: when are genes affecting adaptive divergence and reproductive isolation expected to reside within inversions?** *Evolution* 2009, **63**:3061-3075.

25. Feder JL, Nosil P, Flaxman SM: **Assessing when chromosomal rearrangements affect the dynamics of speciation: implications from computer simulations**. *Front Genet* 2014, **5**:295.

26. Yeaman S: **Local adaptation by alleles of small effect**. *Am Nat* 2015, **186**:S74-S89.

27. Orr HA: **The population genetics of speciation: the evolution of hybrid incompatibilities**. *Genetics* 1995, **139**:1805-1813.

28. Orr HA: **The genetic basis of reproductive isolation: insights from *Drosophila***. *Proc Natl Acad Sci U S A* 2005, **102**:6522-6526.

29. Coyne JA, Orr HA: *Speciation*. Sinauer Associates Inc.; 2004.

30. • Egan SP, Ragland GJ, Assour L, Powell THQ, Hood GR, Emrich S, Nosil P, Feder JL: **Experimental evidence of genome-wide impact of ecological selection during early stages of speciation-with-gene-flow**. *Ecol Lett* 2015, **18**:817-825.

This study experimentally demonstrates that selection on pre-winter diapause depth, a critical axis of divergence between apple and hawthorn host races of *R. pomonella*, produces genome-wide allele frequency differences that mirrored natural host race divergence.

31. Doellman MM, Egan SP, Ragland GJ, Meyers PJ, Hood GR, Powell THQ, Lazorchak P, Hahn DA, Berlocher SH, Nosil P *et al.*: **Divergent natural selection on eclosion time affects genome wide patterns of geographic and host related differentiation for apple and hawthorn races of Rhagoletis pomonella flies**. *Ecol Evol* 2018 http://dx.doi.org/10.1002/ece3.4758.

32. Doellman MM, Ragland GJ, Hood GR, Meyers PJ, Egan SP, Powell THQ, Lazorchak P, Glover MM, Tait C, Schuler H *et al.*: **Genomic differentiation during speciation-with-gene-flow: comparing geographic and host-related variation in divergent life history adaptation in Rhagoletis pomonella**. *Genes (Basel)* 2018, **9**:262.

33. Soria-Carrasco V, Gompert Z, Comeault AA, Farkas TE, Parchman TL, Johnston JS, Buerkle CA, Feder JL, Bast J, Schwander T *et al.*: **Stick insect genomes reveal natural selection's role in parallel speciation**. *Science* 2014, **344**:738-742.

34. Peccoud J, Ollivier A, Plantegenest M, Simon J-C: **A continuum of genetic divergence from sympatric host races to species in the pea aphid complex**. *Proc Natl Acad Sci U S A* 2009, **106**:7495-7500.

35. Peccoud J, Mahéo F, de la Huerta M, Laurence C, J-C Simon: **Genetic characterisation of new host-specialised biotypes and novel associations with bacterial symbionts in the pea aphid complex**. *Insect Conserv Divers* 2015, **8**:484-492.

36. • Nouhaud P, Gautier M, Gouin A, Jaquiéry J, Peccoud J, Legeai F, Mieuzet L, Smadja CM, Lemaitre C, Vitalis R *et al.*: **Identifying genomic hotspots of differentiation and candidate genes involved in the adaptive divergence of pea aphid host races**. *Mol Ecol* 2018, **27**:3287-3300.

This study uses a pooled sequencing genome scan approach to identify hotspots of differentiation harboring potential candidate genes in closely related pea aphid host races. These hot spots are distributed throughout the genome and contain many candidate chemosensory genes, while background genomic divergence remains low.

37. Nouhaud P, Peccoud J, Mahéo F, Mieuzet L, Jaquiéry J, Simon J-C: **Genomic regions repeatedly involved in divergence among plant-specialized pea aphid biotypes**. *J Evol Biol* 2014, **27**:2013-2020.

38. Barton NH: **Multilocus clines**. *Evolution* 1983, **37**:454-471.

39. • Butlin RK, Smadja CM: **Coupling, reinforcement, and speciation**. *Am Nat* 2018, **191**:155-172.

A comprehensive review of coupling research in speciation and reinforcement. This work includes both a thorough survey of classic work on coupling of barriers underlying RI and a call for speciation work to broaden its perspective on the origins and coupling of barrier loci.

40. Feder JL, Flaxman SM, Egan SP, Comeault AA, Nosil P: **Geographic mode of speciation and genomic divergence**. *Annu Rev Ecol Evol Syst* 2013, **44**:73-97.

41. • Eyres I, Duvaux L, Gharbi K, Tucker R, Hopkins D, Simon J-C, Ferrari J, Smadja CM, Butlin RK: **Targeted re-sequencing confirms the importance of chemosensory genes in aphid host race differentiation**. *Mol Ecol* 2017, **26**:43-58.

Targeted sequencing is used to track patterns of differentiation in chemosensory genes, putative targets of differential selection among closely related host races, across the speciation continuum in pea aphids (*A. pisum*). They demonstrate that chemosensory genes show greater divergence than putative control genes across many host race comparisons, while *A. pisum* associated with *L. pratensis* show genome-wide divergence among both chemosensory and control genes.

42. • Southcott L, Kronforst MR: **A neutral view of the evolving genomic architecture of speciation**. *Ecol Evol* 2017, **7**:6358-6366.

Theoretical results from this study demonstrate that, under-specific circumstances, neutral processes can produce genomic patterns similar to those predicted under GWC. Their results highlight the importance of understanding the demographic histories of study systems.

43. Gompert Z, Mandeville EG, Buerkle CA: **Analysis of population genomic data from hybrid zones**. *Annu Rev Ecol Evol Syst* 2017, **48**:207-229.

44. Wolf JBW, Ellegren H: **Making sense of genomic islands of differentiation in light of speciation**. *Nat Rev Genet* 2017, **18**:87-100.

45. Fontaine MC, Pease JB, Steele A, Waterhouse RM, Neafsey DE, Sharakhov IV, Jiang X, Hall AB, Catteruccia F, Kakani E *et al.*: **Extensive introgression in a malaria vector species complex revealed by phylogenomics**. *Science* 2015, **347**.

46. Nadeau NJ, Martin SH, Kozak KM, Salazar C, Dasmahapatra KK, Davey JW, Baxter SW, Blaxter ML, Mallet J, Jiggins CD: **Genome-wide patterns of divergence and gene flow across a butterfly radiation**. *Mol Ecol* 2013, **22**:814-826.

47. Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, Simpson F, Blaxter M, Manica A, Mallet J, Jiggins CD: **Genome-wide evidence for speciation with gene flow in Heliconius butterflies**. *Genome Res* 2013, **23**:1817-1828.

48. Noor MAF, Bennett SM: **Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species**. *Heredity (Edinb)* 2009, **103**:439-444.

49. Cruickshank TE, Hahn MW: **Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow**. *Mol Ecol* 2014, **23**.

50. Hermisson J, Pennings PS: **Soft sweeps and beyond: understanding the patterns and probabilities of selection footprints under rapid adaptation**. *Methods Ecol Evol* 2017, **8**:700-716.

51. Comeault AA, Flaxman SM, Riesch R, Curran E, Soria-Carrasco V, Gompert Z, Farkas TE, Muschick M, Parchman TL, Schwander T *et al.*: **Selection on a genetic polymorphism counteracts ecological speciation in a stick insect**. *Curr Biol* 2015, **25**:1975-1981.

52. Via S, Conte G, Mason-Foley C, Mills K: **Localizing F(ST) outliers on a QTL map reveals evidence for large genomic regions of reduced gene exchange during speciation-with-gene-flow**. *Mol Ecol* 2012, **21**:5546-5560.

53. Gompert Z, Comeault AA, Farkas TE, Feder JL, Parchman TL, Buerkle CA, Nosil P: **Experimental evidence for ecological selection on genome variation in the wild**. *Ecol Lett* 2014, **17**:369-379.

54. • Gompert Z, Egan SP, Barrett RDH, Feder JL, Nosil P: **Multilocus approaches for the measurement of selection on correlated genetic loci**. *Mol Ecol* 2017, **26**:365-382.

This work discusses the extension of estimates of direct selection on loci to the scale of genome-wide data, where the number of parameters to estimate far outweighs the number of sampled individuals. They show that as the genetic architecture of traits become more diffuse, selection is more difficult to estimate and it may not be possible to genotype a large enough sample size for accurate estimates of selection. They provide key insights into estimating selection in cases of highly polygenic traits likely to underlie GWC conditions.

55. Chaturvedi S, Lucas LK, Nice CC, Fordyce JA, Forister ML, Gompert Z: **The predictability of genomic changes underlying a recent host shift in Melissa blue butterflies**. *Mol Ecol* 2018, **27**:2651-2666.

56. Sousa V, Hey J: **Understanding the origin of species with genome-scale data: modelling gene flow**. *Nat Rev Genet* 2013, **14**:404-414.

57. Ravinet M, Faria R, Butlin RK, Galindo J, Bierne N, Rafajlović M, Noor MAF, Mehlig B, Westram AM: **Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow**. *J Evol Biol* 2017, **30**:1450-1477.

58. Raj A, Stephens M, Pritchard JK: **fastSTRUCTURE: variational inference of population structure in large SNP data sets**. *Genetics* 2014, **197**.

59. Pritchard JK, Stephens M, Donnelly P: **Inference of population structure using multilocus genotype data**. *Genetics* 2000, **155**:945-959.

60. Gompert Z, Lucas LK, Buerkle CA, Forister ML, Fordyce JA, Nice CC: **Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants**. *Mol Ecol* 2014, **23**:4555-4573.

61. Jombart T, Devillard S, Balloux F: **Discriminant analysis of principal components: a new method for the analysis of genetically structured populations**. *BMC Genet* 2010, **11**:94.

62. Berlocher SH: **Radiation and divergence in the Rhagoletis pomonella species group: inferences from allozymes**. *Evolution* 2000, **54**:543-557.

63. Xie X, Rull J, Michel AP, Velez S, Forbes AA, Lobo NF, Aluja M, Feder JL: **Hawthorn-infesting populations of *Rhagoletis pomonella* in Mexico and speciation mode plurality**. *Evolution* 2007, **61**:1091-1105.

64. Xie X, Michel AP, Schwarz D, Rull J, Velez S, Forbes AA, Aluja M, Feder JL: **Radiation and divergence in the *Rhagoletis pomonella* species complex: inferences from DNA sequence data**. *J Evol Biol* 2008, **21**:900-913.

65. Rull J, Aluja M, Feder JL: **Evolution of intrinsic reproductive isolation among four North American populations of Rhagoletis pomonella (Diptera: Tephritidae)**. *Biol J Linn Soc* 2010, **100**:213-223.

66. Powell THQ, Hood GR, Murphy MO, Heilveil JS, Berlocher SH, Nosil P, Feder JL: **Genetic divergence along the speciation continuum: the transition from host race to species in *Rhagoletis* (Diptera: tephritidae)**. *Evolution* 2013, **67**:2561-2576.

67. Powell THQ, Forbes AA, Hood GR, Feder JL: **Ecological adaptation and reproductive isolation in sympatry: genetic and phenotypic evidence for native host races of *Rhagoletis pomonella***. *Mol Ecol* 2014, **23**:688-704.

68. Bierne N, Welch J, Loire E, Bonhomme F, David P: **The coupling hypothesis: why genome scans may fail to map local adaptation genes**. *Mol Ecol* 2011, **20**:2044-2072.

69. Hoban S, Kelley JL, Lotterhos KE, Antolin MF, Bradburd G, Lowry DB, Poss ML, Reed LK, Storfer A, Whitlock MC: **Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions**. *Am Nat* 2016, **188**:379-397.

70. Lindtke D, Yeaman S: **Identifying the loci of speciation: the challenge beyond genome scans**. *J Evol Biol* 2017, **30**:1478-1481.

71. Feder JL, Nosil P, Gompert Z, Flaxman SM, Schilling MP: **Barnacles, barrier loci and the systematic building of species**. *J Evol Biol* 2017, **30**:1494-1497.

72. Davey JW, Barker SL, Rastas PM, Pinharanda A, Martin SH, Durbin R, McMillan WO, Merrill RM, Jiggins CD: **No evidence for maintenance of a sympatric *Heliconius* species barrier by chromosomal inversions**. *Evol Lett* 2017, **1**:138-154.

73. Kronforst MR, Hansen MEB, Crawford NG, Gallant JR, Zhang W, Kulathinal RJ, Kapan DD, Mullen SP: **Hybridization reveals the evolving genomic architecture of speciation**. *Cell Rep* 2013, **5**:666-677.

74. Merrill RM, Dasmahapatra KK, Davey JW, Dell'Aglio DD, Hanly JJ, Huber B, Jiggins CD, Joron M, Kozak KM, Llaurens V *et al.*: **The diversification of *Heliconius* butterflies: what have we learned in 150 years?** *J Evol Biol* 2015, **28**:1417-1438.

75. The Heliconius Genome Consortium: **Butterfly genome reveals promiscuous exchange of mimicry adaptations among species**. *Nature* 2012, **487**:94.