Inversion breakpoints and the evolution of supergenes

Romain Villoutreix, Diego Ayala, Mathieu Joron, Zachariah Gompert, Jeffrey L Feder, Patrik Nosil

Inversion breakpoints and the evolution of supergenes

Romain Villoutreix[1], Diego Ayala[2], Mathieu Joron[1], Zachariah Gompert[3], Jeffrey L. Feder[4], Patrik Nosil[1].

Molecular Ecology 30 (12): 2738-2755. https://doi.org/10.1111/mec.15907

[1] CEFE, CNRS, Univ. Montpellier. 1919 Route de Mende. 34293 Montpellier. France

[2] UMR MIVEGEC, IRD, CNRS, Univ. Montpellier, 911 avenue Agropolis, BP 64501, 34394 Montpellier, France.

[3] Department of Biology, Utah State University, Utah 84322, USA

[4] Department of Biological Sciences, University of Notre Dame, Notre Dame, Indiana 46556, USA

Abstract

The coexistence of discrete morphs that differ in multiple traits is common within natural populations of many taxa. Such morphs are often associated with chromosomal inversions, presumably because the recombination suppressing effects of inversions helps maintain adaptive combinations for traits that are controlled by different genes (supergene hypothesis). However, inversions can also harbor adaptive mutations at their breakpoints, leading to rise in inversion frequency in addition to (or independent from) a role for recombination suppression (breakpoint-mutation hypothesis). In this review, we first

describe the myriad of ways that breakpoints can create mutations. We then critically examine the evidence for both hypotheses in examples of putative supergenes. Despite decades of focused study, we find that evidence that inversions spread due to recombination suppression, as proposed by the supergene hypothesis, is often indirect. Breakpoints may harbor mutations that are adaptive in many cases, but again the evidence is indirect and characterization of inversion breakpoints at the sequence level is incomplete in most systems. Direct tests of the role of suppressed recombination and breakpoint mutation in inversion evolution are needed. Finally, we emphasize how the two hypotheses can act in conjunction, with implications for understanding the dynamics of supergene evolution. We conclude that breakpoint characterization could be essential for understanding the origin of complex, discrete phenotypic forms in nature.

Keywords: chromosomal inversion, linkage, mutation, recombination, genome evolution.

Supergenes and complex phenotypes

The origin and maintenance of discrete morphs within natural populations has long fascinated evolutionary biologists (Darwin, 1862; Fisher, 1930; Ford, 1971). Because of the discrete nature of the phenotypic differences observed, this variation was often assumed to have a genetic origin, making its study of primary interest for early geneticists (see Chapter 6 of (Ford, 1971) for a good overview of genetic programs of the time). Moreover, the diverse nature of the traits that usually differ between these morphs (color, morphology, sexual-compatibility, etc.), led to the hypothesis that the genetic basis of this discrete variation is a functional unit containing multiple loci, coined as a "Super-gene" (spelled supergene hereafter; see glossary) by Darlington and Mather in 1949 (Fisher, 1930; Ford, 1965; Mather, 1955). Fisher anticipated that genetic linkage between multiple selected loci could evolve under natural selection (Fisher, 1930) and Ford later suggested that linkage could be increased via translocation or chromosomal inversions (Ford, 1971). Chromosomal inversions can indeed strongly reduce recombination between the genomic segments they contain, a property leading to their discovery in 1921 by Sturtevant (Sturtevant, 1921).

Examples of discrete morphs associated with chromosomal inversions (*i.e.*, candidate supergenes; see glossary) are accumulating rapidly, leading to peaked interest in supergene and inversion evolution. There are currently two core hypotheses for how and why supergenes and inversions evolve. First and perhaps more accepted, inversions can be selected for their capacity to impede recombination between sets of epistatic or locally adapted genes (*i.e.,* supergene hypothesis; see glossary) (Kirkpatrick, 2010; Kirkpatrick & Barton, 2006). This hypothesis predicts that loci underlying trait variation will generally be distributed across the inversion. Second, inversions can create adaptive mutations at their

breakpoints (*i.e.* breakpoint-mutation hypothesis; see glossary; Fig. 1A) (Dobzhansky, 1947, 1970; Kirkpatrick, 2010), leading to their rise in frequency via selection on breakpoint variants. This hypothesis predicts that adaptive mutations will be localized at breakpoints, rather than more widely distributed throughout the inversion. Notably, a breakpoint mutation can, in principle, be pleiotropic and affect multiple traits. Therefore, even when multiple traits are associated with an inversion, a role for recombination suppression needs to be demonstrated rather than assumed. As we discuss in more detail below, these hypotheses are not mutually exclusive and might combine to drive supergene evolution (*i.e.* breakpoint-linkage hypothesis; see glossary; Fig. 1A). Indeed, theory implies that for new inversions to establish under the supergene hypotheses, they must capture all segregating variants at selected loci in a favorable combination (Kirkpatrick & Barton, 2006). Alternatively, an inversion with an adaptive breakpoint mutation may be able to initially rise in frequency and persist for a sufficient amount of time to allow for the sorting out of any initially captured maladapted alleles (e.g., via gene conversion and double-recombination events).

Distinguishing among these hypotheses (Fig. 1A) is important as it will help answer the following major questions concerning adaptation and genome rearrangement (Fig 1.B). Do divergent selected variants evolve first and subsequently become locked together via suppressed recombination within an inversion (Kirkpatrick & Barton, 2006) ? Do inversions establish first because of an adaptive breakpoint mutation such that they are only subsequently turned into supergenes by the emergence of new variants within them (Navarro & Barton, 2003) ? Or, in contrast, pre-existing divergently selected variants and adaptive breakpoint mutations tend to coincide when supergenes form via the emergence

of an inversion? The answers to these questions are important because they determine the roles for mutation, pre-standing genetic variation and recombination in adaptation.

In this review, we critically examine the evidence for the three hypotheses mentioned previously regarding inversion evolution. We begin with a discussion of the mechanisms by which breakpoints can create adaptive mutations, as this is often a less appreciated facet of inversions than is recombination suppression. We then empirically review well-known cases of candidate supergenes (Box 2 – Fig. 1). Our collective findings lead us to propose that characterization of breakpoints at the sequence level could be crucial for understanding the evolutionary dynamics of supergenes. Specifically, the most common molecular mechanism generating inversions also generates duplications at both end of the inverted segment (Ranz et al., 2007). This offers a unique opportunity to date inversions, independently from their content, and test if they emerged before or after the divergently selected loci they may contain. Thus, breakpoint characterization could be essential for understanding the origin and maintenance of complex, discrete phenotypic forms in nature.

Molecular mechanisms by which inversion breakpoints can create adaptive mutations

The idea that inversions might be selected because of an adaptive mutation at one of their breakpoints is not new (Dobzhansky, 1947). However, this hypothesis is yet to receive highly focused study in supergene evolution, despite known molecular mechanisms that are likely to create a breakpoint mutation whenever inversions form. Two main processes can generate chromosomal inversions, ectopic recombination and double strand staggered breaks. Ectopic recombination generates inversions via a recombination event between two homologous sequences oriented head to head in a sequence (Box 1 for further details; Box 1 – Fig. 1). Double strand staggered breaks can generate inversions via the complete

detachment of a DNA fragment and its subsequent reattachment in the opposite orientation (Box 1; Box 1 – Fig. 1). Most inversions are generated by this latter mechanism (Ranz et al., 2007).

Both the mechanisms mentioned above can generate ample phenotypic effects by changing the expression of a gene (or set of genes) at breakpoints (Fig. 2). Breakpoints can also create chimeric genes (see glossary), which might sometimes have a beneficial effect. Indeed, examples of the potential adaptive potential off chimeric genes are accumulating in a diverse set of taxa (Carvunis et al., 2012; Guillén & Ruiz, 2012; Jones et al., 2012; Stewart & Rogers, 2019). For example, in *Drosophila mojavensis* a chimeric gene at an inversion breakpoint is associated with the detoxification of toxic compounds in cacti (Guillén & Ruiz, 2012). Another example is found in the threespine stickleback *Gasterosteus aculeatus,* where a gene located at an inversion breakpoint on chromosome XI harbors different exon sequences depending on the inversion orientation, and is associated with differences in muscular development between lake and oceanic phenotypes (Jones et al., 2012). While speculative, an inversion breakpoint could create a new pleiotropic chimeric gene that affects multiple traits. Moreover, the most common inversion mechanism (double strand staggered breaks) often generate duplications at both breakpoints in the derived sequence (Box1 – Fig. 1), allowing for the creation of chimeric gene without loss of the donor genes. This property will also facilitate neo-functionalization of any gene duplicated by this inversion generation mechanism.

As we will see below, direct evidence that breakpoint mutations drive supergene evolution is scarce, but the mechanisms by which they might do so are established. With these

considerations in place we turn to evaluating the role of suppressed recombination and breakpoint mutation in the evolution of candidate examples of supergenes.

The supergene hypothesis: what is the evidence for recombination suppression among multiple loci controlling selected trait variation?

A core assumption of the supergene hypothesis (*i.e.*, that suppressed recombination is advantageous) is that multiple, linked genetic variants control trait variation that differs between inversion forms. In contrast, if variation is controlled by a single locus with pleiotropic effects on several traits, suppressed recombination is not necessarily favored or required to explain the existence of discrete morphs that differ in these traits.

Our review of the literature revealed that we still know very little about the number, location and identity of selected loci within candidate supergenes. This is not surprising given that determining the number of adaptive loci within a region of suppressed recombination is difficult, because suppressed recombination frustrates attempts at fine-scale genetic mapping. The presence of two or more selected loci in a supergene was historically inferred from genetic crosses aimed at detecting rare recombinants (*i.e.*, intermediate phenotypic forms) (Cain & Sheppard, 1954; Mather, 1950). For example, Mather described a genetic model where multiple selected loci locked within an inversion control heterostyly in *Primula sinensis*, a classic example of a putative supergene (Mather, 1950). However, recent work using modern genomic approaches suggests that Mather's simple genetic model is erroneous and needs modification (Li et al., 2016). Specifically, a 178 kilo base-pair indel containing 5 predicted genes is associated with the different floral types (Li et al., 2016), with one floral type being hemizygous for this segment and the other completely lacking it (Table 1) (Li et al., 2016).

A revision of the classic genetic model also occurred for the iconic land snail *Cepaea nemoralis* (Table 1), where cryptic shell color-pattern was long thought to be associated with five loci locked within an inversion (Cain & Sheppard, 1954). Recent work showed an absence of recombination in the putative supergene, casting doubt on the number and linkage relationships of loci controlling shell color (Gonzalez, Aramendia, & Davison, 2019). These recent developments urge for more rigorous tests on the true number of selected variants within supergenes and their precise genomic location(s).

Indeed, to our knowledge, the existence of multiple selected loci within a supergene associated with an inversion has been demonstrated in only two systems, both plants (Table 1; Box 2). In addition, indirect but fairly convincing evidence exists in two insect systems. Details of these example are contained in Box 2. Clearly, further work determining the number of loci within inversions and candidate supergenes that contribute to trait variation is warranted. In this regard, emerging techniques that allow functional genetic manipulation of structural variants (e.g., CRISPR-Cas9) hold promise for making such progress (Hopkins, Tyukmaeva, Gompert, Feder, & Nosil; Kraft et al., 2015).

The breakpoint hypothesis: what is the evidence that functional loci reside in or near inversion breakpoints?

A core assumption of the breakpoint hypothesis is that functional loci reside in or near inversion breakpoints. We found indications for this hypothesis in three systems (*Papilio polytes, Solenopsis invicta* and *Timema cristinae*), but expect this number to increase when breakpoint characterization is completed in more systems (Table 1).

In the mimetic butterfly *Papilio polytes*, an inversion breakpoint disrupts a transcriptional regulator and affects the expression in embryonic wings of three neighboring genes

potentially involved in wing color-pattern development (*UXT, U3T* and *prospero;* Table 1) (Nishikawa et al., 2015). In the fire ant *Solenopsis invicta,* a region of reduced recombination generated via three successive inversions (Yan et al., 2020) is associated with different social structures of colonies (Wang et al., 2013). A breakpoint of the second inversion is increasing the expression of a neighboring gene (*SI2.2.0_02248*), which is a strong candidate to explain differences in colony social structures (Huang, Dang, Chang, & Wang, 2018; Yan et al., 2020). In *Timema cristinae* a massive deletion occurring at one likely breakpoint of a putative inversion deleted multiple loci known to affect color in another species of the genus (Villoutreix et al., 2020).

Interestingly, with the exception of *Timema cristinae*, indications of putative selected loci within the inversion at locations other than the breakpoint exist for these systems. We discuss the implication of this observation in the following section.

The breakpoint-linkage hypothesis and implications for evolutionary dynamics

As mentioned previously, the supergene and breakpoint-mutation hypotheses are not mutually exclusive. The selective advantage of an inversion could come from its property of impeding recombination between at least one locus within the inversion and a favorable breakpoint mutation. We refer to this mixed hypothesis as the breakpoint-linkage hypothesis (see glossary), which predicts that loci controlling trait variation will be localized both at one (or both) breakpoint(s) and within the inversion.

In theory, this hypothesis could involve different processes and dynamics compared to the classic supergene hypothesis. A supergene composed of a selected locus within an inversion and an adaptive breakpoint mutation suggests two possible evolutionary scenarios. First, pre-existing genetic variants within the inversion could become associated with a new

196 adaptive breakpoint mutation. Although formal modeling is required, it is possible that such
197 a coupling could help relax some of the restrictions of a model based strictly on suppressed
198 recombination (Kirkpatrick & Barton, 2006). For example, such coupling might mean that
199 the initial inversion needs not to capture essentially all favorable segregating alleles in the
200 correct linkage phase with one another. Instead, such sorting could occur after the initial
201 establishment of the inversion, with the initial establishment of the inversion being
202 dependent on the sum of adaptive values of the breakpoint and other genetic variants.

203 Second, an inversion could establish first because of an adaptive breakpoint mutation
204 (Guerrero, Rousset, & Kirkpatrick, 2012), and subsequently evolve into a supergene by the
205 emergence of additional genetic variants within it. This latter scenarios could be initiated in
206 allopatry, where an inversion with an adaptive breakpoint sweeps to fixation and is selected
207 for its recombination suppressing properties only later upon secondary contact (or
208 introgression) and gene flow. In this latter case, allopatry allows for the evolution and
209 capture of alternative sets of favorable alleles in alternate arrangements, by passing
210 difficulties of the supergene emergence when gene flow occurs and adaptive alleles are
211 segregating within populations (Feder, Gejji, Powell, & Nosil, 2011).

212 In terms of empirical evidence, we found indications for the breakpoint-linkage hypothesis
213 in two systems (Table 1). In the mimetic butterfly *Papilio polytes*, an inversion contains a
214 locus shown to affect wing color-pattern (*doublesex*) and a breakpoint mutation affecting
215 the expression of three other genes potentially also affecting wing color-pattern (*UXT, U3T*
216 and *prospero*) (Kunte et al., 2014; Nishikawa et al., 2015). In the fire ant *Solenopsis invicta,* a
217 region of reduced recombination generated via three successive inversions contains a
218 strong candidate gene for odor generation/reception (*Gp-9*), a mechanisms at the root of

social structure differences in fire ants (Wang et al., 2013). A breakpoint of the second inversion increase the expression of another strong candidate gene for odor reception (*SI2.2.0_02248*) (Huang et al., 2018; Yan et al., 2020).

While functional work is needed to better elucidate the role of these different loci in *P. polytes* and *S. invicta*, these two systems highlight the possibility for breakpoints to be one of several selected loci of a supergene. The vast majority of candidate supergene systems still lack a complete characterization of inversion breakpoints (usually one breakpoint is characterized but not the other; Table 1). For this reason we expect more systems to support the breakpoint-linkage hypothesis after further investigation.

Future directions

Future work could usefully focus on two directions. First, better information on the number, location and identify of loci controlling trait variation within inversion is required. This can be achieved using crosses to create recombinant forms, as for the two plant examples discussed in Box 2. The regions identified with this technique are likely still large, and the number of loci identified here is likely to be an under-estimation. Moreover, the identification of causal variants at breakpoints will be challenging as many variants will be linked due to the lack of recombination at breakpoints (Hoffmann & Rieseberg, 2008; Matzkin, Merritt, Zhu, & Eanes, 2005). Alternatively to controlled crosses, genome wide association mapping can be applied in species that exhibit more continuous variation in the same traits as those of close relatives with discrete morphs and inversions, but that lack inversion polymorphism. This will yield more precise estimates, but need to be complemented with functional manipulation in the species with the inversion. Ultimately, functional transformations such as those offered by CRISPR-Cas9 may be required to

precisely and firmly establish the variants functionally contributing to trait variation, possibly by 'engineering' recombinant alleles to validate the independent effects of a given variant on trait variation (Hopkins et al.; Kraft et al., 2015).

Second, once the selected loci within the inversion are identified, further work can be implemented to characterize supergene origin, more specifically to test if the genetic variants locked in the inversion predated the inversion. To do so, we need to date the inversion independently from the selected variants it contains. An estimation of inversion age by averaging the divergence time over the whole inverted sequence is likely to underestimate inversion age, possibly resulting in the erroneous conclusion that the variants appearance predated the inversion. Indeed, recombination still occurs between inverted segments, and will only be prevented for selected loci and breakpoints. To circumvent this caveat, one can take advantage of a useful feature of most inversions: duplication at breakpoints (Ranz et al., 2007). Dating the duplication event will allow one to accurately date the inversion independently from its genomic content. Such dating could also be done for genes within the inversion and near breakpoints, but this will be less precise, depending on the amount a recombination they experience.

Such dating was shown to be effective in *Heliconius numata* butterflies (Jay et al., 2018). It requires sequence-level characterization of breakpoints, a difficult endeavor because breakpoints often reside in highly repetitive regions of the genomes (Russell B. Corbett-Detig et al., 2019; da Silva et al., 2019), which are difficult to assemble using short sequencing read technologies (Fig. 3A). Nonetheless, creative approaches such as those applied by Corbett-Detig and colleagues in *Drosophila* (Russell B. Corbett-Detig & Hartl, 2012) hold promise (Fig. 3B), as do long reads technologies such as Nanopore and PacBio.

Concluding remarks

Rapidly emerging candidate supergenes systems urge a careful assessment of the state of research in this field. We outlined how inversions can facilitate supergene evolution either via suppressed recombination or by inducing adaptive mutation at breakpoints. To our knowledge, definitive evidence for the existence of multiple selected loci within candidate supergenes has only been gathered in two systems. Thus direct evidence even for the classic supergene hypothesis is still scarce, although many examples are consistent with this idea. Little is known about the origin of supergenes, specifically if genetic variants or inversion appeared first. As an alternative or in addition to the supergene hypothesis, inversions may spread due to adaptive breakpoint mutations. Observations in multiple candidate supergenes suggest a joint role for suppressed recombination and breakpoint mutation in supergene evolution. Thus, further tests of the breakpoint-linkage hypothesis are warranted. Distinguishing and confirming the role of breakpoints in supergene evolution will likely require characterization of inversion breakpoints at the sequence level, which has yet to be completed in most systems. We propose that further study of inversion breakpoints will facilitate understanding the origin and maintenance of complex, discrete phenotypic forms in nature, and the dynamics by which supergenes evolve. Although we focused here on morphs, similar processes could apply to the evolution of other discrete units of diversity such as ecotypes, species, and sexes.

Glossary:

Breakpoint-linkage hypothesis: the selective advantage to an inversion stems from both a favorable effect of mutation at one of its breakpoint and the recombination suppression between this mutation and at least one other selected variant within the inversion.

Breakpoint-mutation hypothesis (also referred as 'position effect hypothesis' in the literature): the selective advantage to an inversion stems from a favorable effect(s) associated with a mutation at one or both of its breakpoints.

Candidate supergene: case where multiple co-varying and discrete phenotypes are associated with an inversion; the existence of multiple loci associated with these traits must be demonstrated to advance the example from a candidate to a bonafide supergene.

Chimeric gene: a new gene constituted of sequences not previously associated together, for example of parts of two genes, or parts of genes and non-coding DNA. Here, we consider cases where the different DNA segments are brought together by an inversion breakpoint (see Fig. 2B).

Chromosomal inversion: a chromosomal rearrangement in which a segment of a chromosome is reversed end to end. This causes the inverted segment to have an opposite sequence order from the ancestral sequence and reduces recombination in inversion heterozygotes. Recombination is especially reduced near chromosomal inversion breakpoints.

Chromosomal inversion breakpoint (shortened to 'inversion breakpoint' or simply 'breakpoint' in the main text and hereafter): the position(s) on a chromosome where DNA sequence orientation shifts after the occurrence of a chromosomal inversion (i.e., the end of

306 an inversion where the genome shifts from inverted back to collinear). An inversion
307 generally has two breakpoints.

308 (Chromosomal) breakpoint mutation: sequence modification(s) at breakpoints resulting
309 from a chromosomal inversion. Mutations may also occur near breakpoints after an
310 inversion has occurred, however, here, we refer here to changes associated with the
311 creation of the rearrangement itself.

312 Discrete morphs: individuals of the same species that exhibit discrete phenotypic
313 differences (e.g., in morphology, color, behavior, physiology, etc.). In this piece, we focus on
314 cases where morphs are differentiated by multiple co-varying phenotypic traits such that
315 multiple genes may be involved.

316 Supergene: a group of genes affecting different phenotypes so seldom separated by
317 crossing-over that in effect they operate as a single genetic entity (i.e., a non-recombining
318 locus). This end may be achieved by close genetic proximity on a chromosome (i.e., tight
319 physical linkage), by inclusion within an inversion, or by other molecular mechanisms
320 impeding crossing-over. This definition is modified from that of E. B. Ford (Ford, 1965).

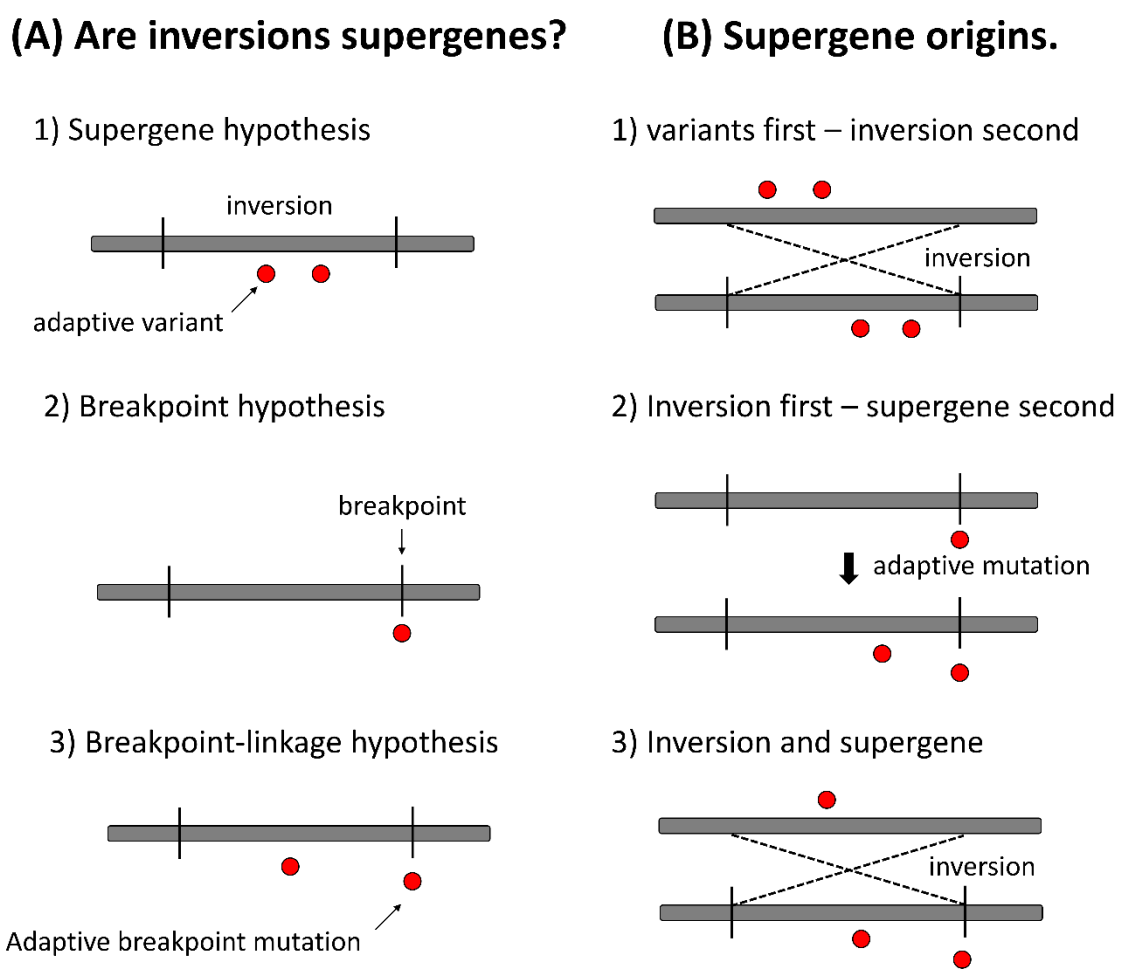
321 Supergene hypothesis: supergenes evolve because suppressed recombination within them
322 maintains favorable allelic combinations at multiple loci that control different traits.

323 Switch gene: A gene controlling the development of alternative discrete morphs.

324

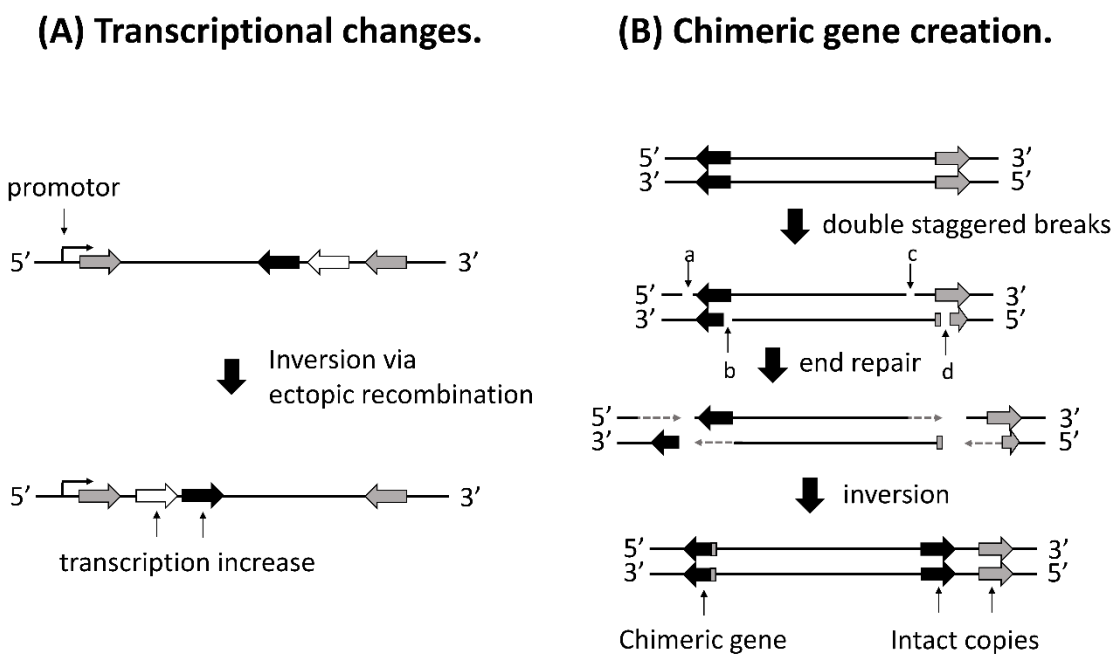
Acknowledgements

This manuscript greatly benefited from discussions with Pr. Russel Corbett-Detig, Pr. Angus Davison, Pr. Haruhiko Fujiwara, Dr. Clemens Küpper, Pr. Cheng-Ruei Lee, Pr. Thomas Mitchell-Olds and Dr. John Wang on their respective study systems. We are grateful to Pr. Cheng-Ruei Lee for providing us photographs of *Boechera stricta.* We are grateful to Dr. Venera Tyukmaeva for insights on CRISPR-Cas9 and functional genomics. R.V. and P.N. were supported by a grant from the European Research Council (EE-Dynamics 770826, https://erc.europa.eu/). D.A. was supported by an ANR grant (ANR-18-CE35-0002-01 – WILDING). M. J. was supported by an ANR grant (ANR-18-CE02-0019-02 supergène). Z. G. was supported by the NSF (DEB 1638768). J. F. was supported by grants from the USDA NIFA program and NSF. All authors contributed to writing the manuscript. The authors declare having no competing interests.



338
339
340

341 **<u>Figure 1 – Hypotheses and evolutionary scenarios:</u> A.** Hypotheses behind inversion
342 evolution. **B.** Evolutionary scenarios for supergene evolution. Grey squares represents a
343 chromosome; red dots are divergently selected loci; black bars are inversion breakpoints.



344

345 **<u>Figure 2 – Possible mutations at inversion breakpoints:</u> A.** An inversion generated through
346 ectopic recombination put two genes into a different transcription context, altering their
347 level of expression. **B.** An inversion generated through the staggered break mechanism
348 generates a new chimeric switch gene, while avoiding gene loss via duplication. Arrow
349 represent genes, with orientation indicated via arrow direction; a, b, c, d represent single
350 strand breaks, resulting here in two staggered breaks.

351

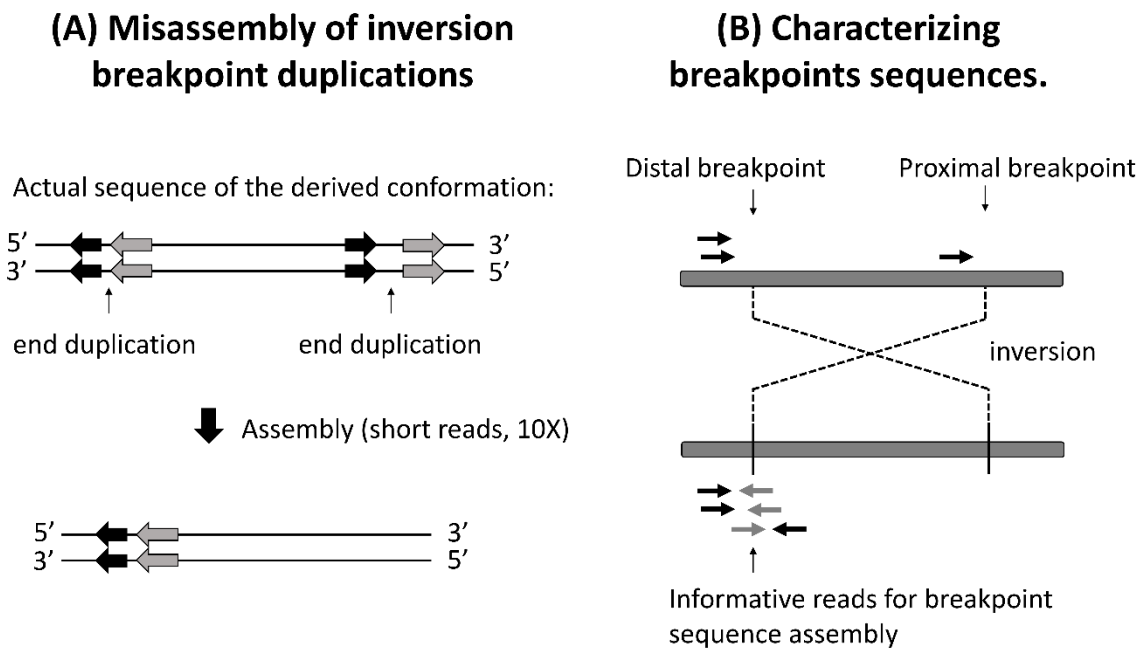


Figure 3 – Characterizing breakpoint at the sequence level: A. Using short reads technologies can lead to misassembly of inversion breakpoints, for example due to duplications generated through the staggered breaks mechanisms (Ranz et al., 2007). **B.** A method to characterize breakpoint sequences using a single reference genome and short read data, as developed by (Russell B. Corbett-Detig, Cardeno, & Langley, 2012). Black arrows symbolize aligned reads on reference genome, while grey arrows symbolize their unaligned paired reads. Breakpoint sequence characterization is possible by using these read pairs to generate and *de novo* assembly of the region. Figure redrawn and modified from (Russell B. Corbett-Detig et al., 2012) with the permission of the authors.

Box 1 - Inversion breakpoints can create a diversity of types of genetic variation via mutation.

We here discuss how inversions are formed and the associated effects of induced mutations on genetic variation at breakpoints. Two known molecular mechanisms can generate inversions: (1) ectopic recombination and, (2) double strand staggered breaks (Box 1. - Fig1) (Ling & Cordaux, 2010; Ranz et al., 2007).

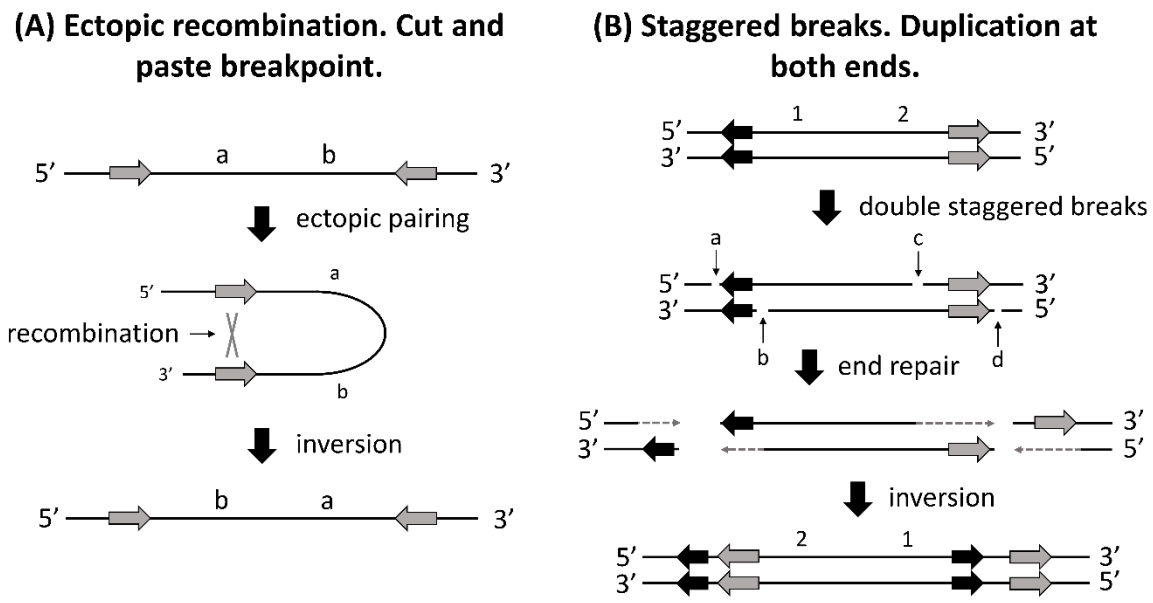
The first mechanism occurs from recombination between homologous sequences that are oriented head-to-head in the original sequence, subsequently resulting in inversion of the sequence located between these elements (Box 1 – Fig. 1A). This mechanism leaves little molecular traces and generates 'cut-and-paste' type breakpoints. The second mechanism involves double strand breaks occurring in two positions of the original sequences (Ranz et al., 2007). Such breaks are usually staggered, meaning that they will result in fragments with stretches of single stranded DNA at their extremities (Box 1. - Fig. 1B). The DNA repair mechanism that synthesize the reverse complement of these single stranded stretches then join the DNA sequences back together in a non-homologous way, sometimes reinserting the DNA segment in the opposite orientation, creating an inversion. When the breaks are heavily staggered (*i.e.,* with long single strand stretches) this results in inversions with duplicated sequences at both breakpoints in the derived sequence (Box 1. - Fig. 1B) (Ranz et al., 2007). In contrast, when the breaks are blunt or lightly staggered, 'cut-and-paste' type breakpoints without duplications are created in the derived sequence.

Despite these clear mechanisms of inversion formation, we still know very little about genetic variation at breakpoints and the mutations that breakpoint generates (Table 1). This stems largely from the fact that inversion breakpoints often lies in very repetitive regions

(Russell B. Corbett-Detig et al., 2019; da Silva et al., 2019) which are difficult to sequence and assemble and thus challenging to study (Fig. 3A). Nonetheless, some progress has been made in *Drosophila spp.* (R. B. Corbett-Detig, 2016; Russell B. Corbett-Detig & Hartl, 2012; McBroome, Liang, & Corbett-Detig, 2020; Ranz et al., 2007), where double strand staggered breaks appear to be the most common mechanism generating inversions, very often resulting in duplications at breakpoints (Ranz et al., 2007).

We note, however, that locating breakpoints is not sufficient to identify their phenotypic effects. Obtaining the derived sequence is necessary to try to understand possible phenotypic effects. This is not a straightforward task, as disruption of gene sequence by breakpoints can sometimes leads to increase in their expression as seen in fire ants (Yan et al., 2020). Nonetheless, phenotypic effects are expected, for example because both the aforementioned mechanisms can change the expression context of set of genes (Fig. 2A). In addition, the double strand staggered breaks mechanism can have another potential effect, in which duplications of sequences leads to the creation of a new chimeric gene. This occurs via merging the initial part of a gene with another coding sequence such as another gene or pseudo-gene (Fig. 2B). This phenomenon can happen without the loss of the originally disrupted genes (via their duplication), mitigating costs.

Thus, there are several possibilities by which breakpoints can create genetic variation with likely phenotypic effects. The observation that similar genomic regions are often re-used repeatedly as breakpoints for different inversions, including in different taxa, raised the possibility for a recurrent and even somewhat predictable role for breakpoints in evolution (Pevzner & Tesler, 2003).



Box 1 - Figure 1 - Known molecular mechanisms generating inversions: Black and grey arrows represent genes, with orientation indicated by the direction of the arrow **A.** Ectopic recombination between similar elements (transposable elements or duplicated genes) in opposite orientation in a DNA segment leads to the inversion of the interspacing sequence. **B.** Fours single strand breaks (a, b, c, and d) lead to two staggered breaks. Through end repair and reattachment in the opposite orientation an inversion in generated with sequence duplication at both breakpoints (black and grey genes).

Box 2 - Determining the number of loci within inversions that contribute to trait variation

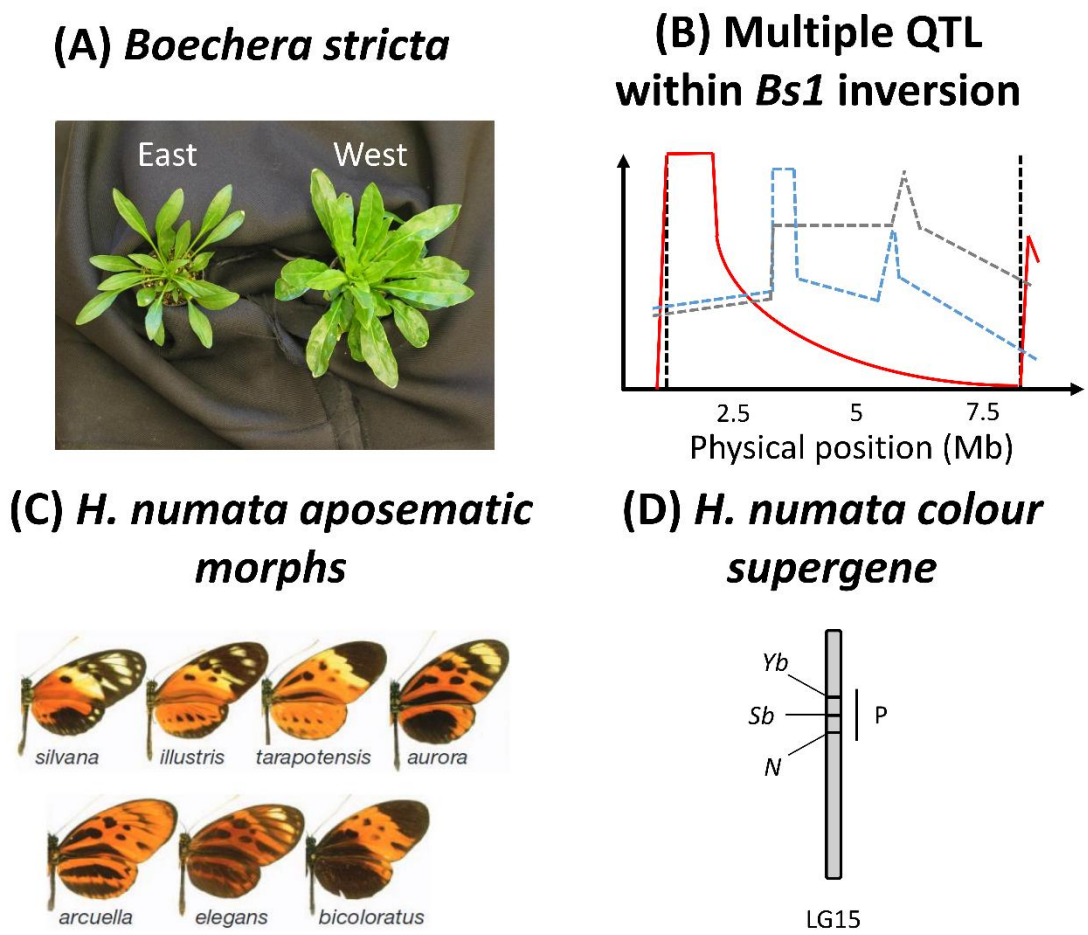
Despite seventy years of research on supergene, evidence for the existence of multiple selected loci within an inversion has only been gathered in a handful of systems.

Evidence for multiple loci controlling trait variation within an inversion has been demonstrated in *Boechera stricta* (Box 2 – Fig. 1A), a perennial plant from the Rocky Mountains. Here, two subspecies adapted to different climates have entered in secondary contact (Lee et al., 2017). The subspecies differ in multiple traits and through crosses it was demonstrated that three different QTL are associated with these divergent traits (Box 2 – Fig. 2B). Interestingly, the inversion itself has an effect on these traits, independent of the genotype at the three QTL, but the reason for this is unknown. This supergene has also had its evolutionary origins characterized such that it is known that the inversion is relatively young and likely evolved after secondary contact (Lee et al., 2017). The existence of multiple selected loci within an inversion has also been demonstrated, again via crosses, in the *Mimulus guttatus* species complex, where two life strategies (perennial *vs* annual) are maintained through divergent selection (Lowry & Willis, 2010). Specifically, through crosses between the *Mimulus guttatus* annual type and a closely related perennial species with a collinear gene order, *Mimulus tilingii*, Coughlan and Willis (Coughlan & Willis, 2018) identified two QTL within the inversion associated with the traits experiencing divergent selection.

Less direct, but reasonably strong, evidence for multiple loci has also been reported in two insect systems (Table 1). Here, the approach was to compare the genetic basis of traits in closely related taxa that exhibit versus lack inversion polymorphism. In the polymorphic butterfly *Heliconius numata* a region of reduced recombination generated via multiple

441 successive inversions control variation in color morphs. The same genomic region is known
442 to contain at least three recombining loci associated with color variation in other species of
443 the genus (Box 1 – Fig. 1D) (Huber et al., 2015; Mathieu Joron et al., 2011; Van Belleghem et
444 al., 2017). It is therefore likely that the inversion in *H. numata* harbor variants at these three
445 loci that recombine in other *Heliconius* species. A similar example stems from *Timema*, a
446 genus of stick insect that is endemic of the western USA and Mexico and that shows within-
447 population variation for cryptic coloration (i.e., green versus melanistic morphs). In several
448 species, morphs are highly discrete and controlled by a region of suppressed recombination
449 on the linkage group eight (named '*Mel-Stripe*') (Nosil et al., 2018; Villoutreix et al., 2020).
450 However, a recent study in a species displaying more quantitative color variation (*Timema*
451 *chumash*) revealed that multiple genetic variants, and likely multiple genes within *Mel-*
452 *Stripe*, control color variation (Villoutreix et al., 2020).

453



454

455 **Box 2 - Figure 1 – Flagship examples of supergenes: A.** *Boechera stricta* subspecies types.
456 The East subspecies has smaller leaves with more trichomes. Photo credit: Cheng-Ruei Lee.
457 **B.** Multiple QTL were detected in the *Bs1* inversion. The red, blues dashed and grey dashed
458 lines correspond to different composited traits obtained through discriminant function
459 analysis. See (Lee et al., 2017) for more details. Black dashed lines correspond to inversion
460 breakpoint positions. Breakpoint are of cut-and-paste type for the *Bs1* inversion. Figure
461 redrawn and modified with the permission of the authors. **C.** *Heliconius numata* aposematic
462 morphs, involved in Müllerian mimicry rings with other butterflies. **D.** Aposematic morphs in
463 *H. numata* are associated with an inversion (named P) on linkage group LG15, encompassing
464 at least three loci associated with aposematic color morph in *H. melpomene*: Yb, Sb and N.

465

Species	Multiple traits?	Selection on phenotypes?	Map to an inversion?	Multiple loci?	Selection on recomb. sup.?	Adaptive breakpoint mutation?	Breakpoints sequence?	Variants first?
Aves								
Ruff - *Philomachus pugnax*	Yes. (Kupper et al., 2016; Lamichhaney et al., 2016)	Possibly. (Kupper et al., 2016; Lamichhaney et al., 2016)	Yes. (Kupper et al., 2016; Lamichhaney et al., 2016)	Likely. (Kupper et al., 2016; Lamichhaney et al., 2016)	Not tested.	No. Counter-selected. (Kupper et al., 2016; Lamichhaney et al., 2016)	Yes. (Kupper et al., 2016; Lamichhaney et al., 2016)	Not known.
White-throated sparrow - *Zonotrichia albicollis*	Yes. (Tuttle et al., 2016)	Yes. (Tuttle et al., 2016)	Yes. (Thomas et al., 2008; Thorneycroft, 1966; Throneycroft, 1975; Tuttle et al., 2016)	Likely. (Tuttle et al., 2016)	Not tested.	Not known.	No.	Not known.
Invertebrates								

Anopheles – *Anopheles gambiae*	Possibly. (Cheng, Tan, Hahn, & Besansky, 2018)	Possibly. (Fouet, Gray, Besansky, & Costantini, 2012; Gray, Rocca, Costantini, & Besansky, 2009; Rocca, Gray, Costantini, & Besansky, 2009)	Yes. (Cheng et al., 2018; Coluzzi, Sabatini, della Torre, Di Deco, & Petrarca, 2002; Coluzzi, Sabatini, Petrarca, & Di Deco, 1979; Fouet et al., 2012; Gray et al., 2009; Rocca et al., 2009)	Possibly. (Cheng et al., 2012)	Not tested.	No.	Partly characterized. (Cheng et al., 2018; Sharakhov et al., 2006)	Not known.
Grove snail - *Cepaea nemoralis*	Yes. (Cain & Sheppard, 1950, 1952)	Possibly. (Cain & Sheppard, 1950, 1954; Cameron & Cook, 2012)	Not known. (Gonzalez et al., 2019)	Not known.	Not known.	Not known.	No.	Not known.
Numata longwing - *Heliconius numata*	Yes. (Mathieu Joron et al., 2011)	Yes. (Benson, 1972; Chouteau, Llaurens, Piron-Prunier, & Joron, 2017; M. Joron,	Yes. (Mathieu Joron et al., 2011).	Very likely. (Huber et al., 2015; Mathieu Joron et al., 2011; Van Belleghem	Not tested.	No.	Partly characterized. (Jay et al., 2018)	Not known.

		Wynne, Lamas, & Mallet, 1999; Kapan, 2001; Mallet & Barton, 1989; Merrill et al., 2012)		et al., 2017)				
Common Mormon - *Papilio polytes*	Yes. (Kunte et al., 2014; Nishikawa et al., 2015)	Yes. (Clarke & Sheppard, 1972)	Yes. (Kunte et al., 2014; Nishikawa et al., 2015)	Likely. (Nishikawa et al., 2015)	Not tested.	Possibly. (Nishikawa et al., 2015)	Yes. (Nishikawa et al., 2015)	Not known.
Fire ant - *Solenopsis invicta*	Yes. (Wang et al., 2013).	Possibly. (Ross & Keller, 1995)	Yes. (Wang et al., 2013)	Possibly. (Wang et al., 2013)	Not tested.	Possibly. (Huang et al., 2018; Yan et al., 2020)	Yes. (Yan et al., 2020)	Not known.
Timema walking sticks - *Timema cristinae*.	Yes. (Comeault et al., 2015)	Yes. (Cristina P. Sandoval, 1994; C. P. Sandoval & Nosil, 2005)	Likely. (Villoutreix et al., 2020)	Very likely. (Villoutreix et al., 2020)	Not tested.	Likely. (Villoutreix et al., 2020)	Partly characterized. (Nosil et al., 2018; Villoutreix et al., 2020)	Not known.

Mammals								
Humans – *Homo sapiens*	Yes. (Puig, Casillas, Villatoro, & Caceres, 2015; Stefansson et al., 2005)	Likely. (Stefansson et al., 2005)	Yes (Puig et al., 2015; Stefansson et al., 2005)	Possibly (Puig et al., 2015)	Not tested.	Possibly. (de Jong et al., 2012)	Yes. (Stefansson et al., 2005)	Not known.
Plants								
Drummond's rockcress - *Boechera stricta*	Yes. (Lee et al., 2017)	Yes. (Anderson, Lee, & Mitchell-Olds, 2011; Anderson, Lee, Rushworth, Colautti, & Mitchell-Olds, 2013; Lee et al., 2017)	Yes. (Lee et al., 2017)	Yes. (Lee et al., 2017)	Yes, but debated. (Charlesworth & Barton, 2018; Lee et al., 2017)	No. (Lee et al., 2017).	Yes. (Lee et al., 2017).	Yes.

Common yellow monkeyflower - *Mimulus guttatus* species complex	Yes. (Coughlan & Willis, 2018; Lowry & Willis, 2010)	Yes. (Lowry & Willis, 2010)	Yes. (Coughlan & Willis, 2018; Lowry & Willis, 2010)	Yes. (Coughlan & Willis, 2018)	Not tested.	Not known.	No.	Not known.
Primrose - *Primula sp.*	Yes. (Darwin, 1862)	Likely. (Piper & Charlesworth, 2008)	No. (Li et al., 2016)	Possibly. (Li et al., 2016)	Not tested.	N.A. Indel. (Li et al., 2016)	Yes. (Li et al., 2016)	N. A.
Teleosts								
Atlantic Cod – *Gadus morhua*.	Yes. (Berg et al., 2016; Sinclair-Waters et al., 2018)	Likely. (Berg et al., 2016; Sinclair-Waters et al., 2018)	Yes. (Berg et al., 2016; Kirubakaran et al., 2016; Sinclair-Waters et al., 2018)	Possibly. (Berg et al., 2016; Kirubakaran et al., 2016; Sinclair-Waters et al., 2018)	Not tested.	Not known.	No.	Not known.
Threespine sticklebacks - *Gasterosteus aculeatus*	Yes. (Jones et al., 2012)	Likely. (Jones et al., 2012)	Yes. (Jones et al., 2012)	Possibly. (Albert et al., 2007;	Not tested.	Yes. (Jones et	Yes. (Jones et al., 2012)	Not known.

				Colosimo et al., 2004)		al., 2012)		
Steelhead/rainbow trout – *Oncorhynchus mykiss*	Yes. (Nichols, Edo, Wheeler, & Thorgaard, 2008; Pearse, Miller, Abadia-Cardoso, & Garza, 2014)	Not tested.	Possibly. [(Nichols et al., 2008; O'Malley, Sakamoto, Danzmann, & Ferguson, 2003; Pearse et al., 2014; Robison, Wheeler, Sundin, Sikka, & Thorgaard, 2001)	Likely. (Lemopoulos, Uusi-Heikkilä, Huusko, Vasemägi, & Vainikka, 2018)	Not tested.	Not known.	No.	Not known.

Table 1 – Examples of candidate supergene. Species: common and latin naming. **Multiple traits:** Does morphs different by multiple traits? **Selection on phenotypes:** Has selection on these traits been shown? **Map to an inversion:** Are these traits associated with and inversion? **Multiple loci:** Has the existence of multiple loci controlling the aforementioned traits been demonstrated in this system? **Selection on recomb. sup.:** Has the selective advantage of an inversion been demonstrated to come from its property of impeding recombination between multiple selected loci? **Adaptive breakpoint mutation:** Is their evidence for this inversion to be selected for or against because of one of its breakpoints. **Breakpoint sequence:** Have the breakpoints been characterized at the sequence level. **Variants first?**: Do adaptive variants in the putative supergene predate the origins of the inversion?

Bibliography

Albert, A. Y. K., Sawaya, S., Vines, T. H., Knecht, A. K., Miller, C. T., Summers, B. R., . . . Schluter, D. (2007). The genetics of adaptative shape shift in stickleback: Pleiotropy and effect size. *Evolution*, *62*(1), 76-85.

Anderson, J. T., Lee, C. R., & Mitchell-Olds, T. (2011). Life-History Qtls and Natural Selection on Flowering Time in Boechera Stricta, a Perennial Relative of Arabidopsis. *Evolution*, *65*(3), 771-787. doi:10.1111/j.1558-5646.2010.01175.x

Anderson, J. T., Lee, C. R., Rushworth, C. A., Colautti, R. I., & Mitchell-Olds, T. (2013). Genetic trade-offs and conditional neutrality contribute to local adaptation. *Mol Ecol*, *22*(3), 699-708. doi:10.1111/j.1365-294X.2012.05522.x

Benson, W. W. (1972). Natural-Selection for Mullerian Mimicry in Heliconius-Erato in Costa-Rica. *Science*, *176*(4037), 936-939. doi:DOI 10.1126/science.176.4037.936

Berg, P. R., Star, B., Pampoulie, C., Sodeland, M., Barth, J. M. I., Knutsen, H., . . . Jentoft, S. (2016). Three chromosomal rearrangements promote genomic divergence between migratory and stationary ecotypes of Atlantic cod. *Scientific Reports*, *6*(23246). doi:10.1038/srep23246

Cain, A. J., & Sheppard, P. M. (1950). Selection in the polymorphic land snail Cepæa nemoralis. *Heredity*, *4*(3), 275-294. doi:10.1038/hdy.1950.22

Cain, A. J., & Sheppard, P. M. (1952). The effects of natural selection on body colour in the land snail Cepaea

nemoralis. *Heredity*, *6*(2), 217-231. doi:10.1038/hdy.1952.22

Cain, A. J., & Sheppard, P. M. (1954). NATURAL SELECTION IN CEPAEA. *Genetics*, *39*(1), 89-116.

Cameron, R., & Cook, L. (2012). CORRELATED PHENOTYPIC RESPONSES TO HABITAT DIFFERENCE IN CEPAEA NEMORALIS (L.). *FOLIA MALACOLOGICA*, *20*(4), 255-263.

Carvunis, A.-R., Rolland, T., Wapinski, I., Calderwood, M. A., Yildirim, M. A., Simonis, N., . . . Vidal, M. (2012). Proto-genes and de novo gene birth. *Nature*, *487*(7407), 370-374. doi:10.1038/nature11184

Charlesworth, B., & Barton, N. H. (2018). The Spread of an Inversion with Migration and Selection. *Genetics*, *208*(1), 377-382.

Cheng, C. D., Tan, J. C., Hahn, M. W., & Besansky, N. J. (2018). Systems genetic analysis of inversion polymorphisms in the malaria mosquito Anopheles gambiae. *Proc Natl Acad Sci U S A*, *115*(30), E7005-E7014. doi:10.1073/pnas.1806760115

Cheng, C. D., White, B. J., Kamdem, C., Mockaitis, K., Costantini, C., Hahn, M. W., & Besansky, N. J. (2012). Ecological Genomics of Anopheles gambiae Along a Latitudinal Cline: A Population-Resequencing Approach. *Genetics*, *190*(4), 1417-1432. doi:10.1534/genetics.111.137794

Chouteau, M., Llaurens, V., Piron-Prunier, F., & Joron, M. (2017). Polymorphism at a mimicry supergene maintained by opposing frequency-dependent selection pressures. *Proc Natl Acad Sci U S A*, *114*(31), 8325-8329. doi:10.1073/pnas.1702482114

Clarke, C. A., & Sheppard, F. R. S. (1972). The genetics of the mimetic butterfly Papilio polytes L. *Philosophical*

Transactions of the Royal Society of London. B, Biological Sciences, *263*(855), 431-458. doi:10.1098/rstb.1972.0006

Colosimo, P. F., Peichel, C. L., Nereng, K., Blackman, B. K., Shapiro, M. D., Schluter, D., & Kingsley, D. M. (2004). The Genetic Architecture of Parallel Armor Plate Reduction in Threespine Sticklebacks. *Plos Biology*, *2*(5), e109. doi:10.1371/journal.pbio.0020109

Coluzzi, M., Sabatini, A., della Torre, A., Di Deco, M. A., & Petrarca, V. (2002). A polytene chromosome analysis of the Anopheles gambiae species complex. *Science*, *298*(5597), 1415-1418. doi:10.1126/science.1077769

Coluzzi, M., Sabatini, A., Petrarca, V., & Di Deco, M. A. (1979). Chromosomal differentiation and adaptation to humn environments in the Anopheles gambiae complex. *Royal Society of Tropical Medicine and Hygiene, 73*(5), 483-497.

Comeault, Aaron A., Flaxman, Samuel M., Riesch, R., Curran, E., Soria-Carrasco, V., Gompert, Z., . . . Nosil, P. (2015). Selection on a Genetic Polymorphism Counteracts Ecological Speciation in a Stick Insect. *Current Biology*, *25*(15), 1975-1981. doi:10.1016/j.cub.2015.05.058

Corbett-Detig, R. B. (2016). Selection on Inversion Breakpoints Favors Proximity to Pairing Sensitive Sites in Drosophila melanogaster. *Genetics, 204*(1), 259-265. doi:10.1534/genetics.116.190389

Corbett-Detig, R. B., Cardeno, C., & Langley, C. H. (2012). Sequence-Based Detection and Breakpoint Assembly of Polymorphic Inversions. *Genetics, 192*(1), 131-137. doi:10.1534/genetics.112.141622

Corbett-Detig, R. B., & Hartl, D. L. (2012). Population Genomics of Inversion Polymorphisms in Drosophila

melanogaster. *PLoS Genet, 8*(12), e1003056. doi:10.1371/journal.pgen.1003056

Corbett-Detig, R. B., Said, I., Calzetta, M., Genetti, M., McBroome, J., Maurer, N. W., . . . Besansky, N. J. (2019). Fine-Mapping Complex Inversion Breakpoints and Investigating Somatic Pairing in the Anopheles gambiae Species Complex Using Proximity-Ligation Sequencing. *Genetics, 213*(4), 1495-1511. doi:10.1534/genetics.119.302385

Coughlan, J. M., & Willis, J. H. (2018). Dissecting the role of a large chromosomal inversion in life history divergence throughout the Mimulus guttatus species complex. *Mol Ecol, 28*(6), 1343-1357. doi:doi:10.1111/mec.14804

da Silva, V. H., Laine, V. N., Bosse, M., Spurgin, L. G., Derks, M. F. L., van Oers, K., . . . Groenen, M. A. M. (2019). The Genomic Complexity of a Large Inversion in Great Tits. *Genome Biology and Evolution, 11*(7), 1870-1881. doi:10.1093/gbe/evz106

Darwin, C. (1862). On the Two Forms, or Dimorphic Condition, in the Species of Primula, and on their remarkable Sexual Relations. *Journal of the Proceedings of the Linnean Society of London. Botany, 6*(22), 77-96. doi:10.1111/j.1095-8312.1862.tb01218.x

de Jong, S., Chepelev, I., Janson, E., Strengman, E., van den Berg, L. H., Veldink, J. H., & Ophoff, R. A. (2012). Common inversion polymorphism at 17q21.31 affects expression of multiple genes in tissue-specific manner. *BMC Genomics, 13*(458). doi:10.1186/1471-2164-13-458

Dobzhansky, T. G. (1947). Adaptive Changes Induced by Natural Selection in Wild Populations of Drosophila. *Evolution, 1*(1/2), 1-16. doi:10.2307/2405399

606 Dobzhansky, T. G. (1970). *Genetics of the Evolutionary*
607 *Process*: Columbia University Press.
608 Feder, J. L., Gejji, R., Powell, T. H. Q., & Nosil, P. (2011).
609 ADAPTIVE CHROMOSOMAL DIVERGENCE
610 DRIVEN BY MIXED GEOGRAPHIC MODE OF
611 EVOLUTION. *Evolution, 65*(8), 2157-2170.
612 doi:10.1111/j.1558-5646.2011.01321.x
613 Fisher, R. A. (1930). *The Genetical Theory of Natural*
614 *Selection*: Oxford.
615 Ford, E. B. (1965). *Genetic Polymorphism*. London: Studies
616 Faber & Faber.
617 Ford, E. B. (1971). *Ecological Genomics* (3rd ed.). London:
618 Chapman and Hall LTD.
619 Fouet, C., Gray, E., Besansky, N. J., & Costantini, C. (2012).
620 Adaptation to Aridity in the Malaria Mosquito
621 Anopheles gambiae: Chromosomal Inversion
622 Polymorphism and Body Size Influence Resistance to
623 Desiccation. *PLoS One, 7*(4), e34841.
624 doi:10.1371/journal.pone.0034841
625 Gonzalez, D. R., Aramendia, A. C., & Davison, A. (2019).
626 Recombination within the Cepaea nemoralis supergene is
627 confounded by incomplete penetrance and epistasis.
628 *Heredity, 123*(2), 153-161. doi:10.1038/s41437-019-
629 0190-6
630 Gray, E. M., Rocca, K. A. C., Costantini, C., & Besansky, N.
631 J. (2009). Inversion 2La is associated with enhanced
632 desiccation resistance in Anopheles gambiae. *Malaria*
633 *Journal, 8*. doi:10.1186/1475-2875-8-215
634 Guerrero, R. F., Rousset, F., & Kirkpatrick, M. (2012).
635 Coalescent patterns for chromosomal inversions in
636 divergent populations. *Philosophical Transactions of the*
637 *Royal Society B-Biological Sciences, 367*(1587), 430-
638 438.

Guillén, Y., & Ruiz, A. (2012). Gene alterations at Drosophila inversion breakpoints provide prima facie evidence for natural selection as an explanation for rapid chromosomal evolution. *BMC Genomics, 13*(1), 53. doi:10.1186/1471-2164-13-53

Hoffmann, A. A., & Rieseberg, L. H. (2008). Revisiting the Impact of Inversions in Evolution: From Population Genetic Markers to Drivers of Adaptive Shifts and Speciation? *Annual Review of Ecology, Evolution, and Systematics, 39*(1), 21-42. doi:10.1146/annurev.ecolsys.39.110707.173532

Hopkins, D. P., Tyukmaeva, V. I., Gompert, Z., Feder, J. L., & Nosil, P. Functional genomics offers new tests of speciation hypotheses. *in press*.

Huang, Y.-C., Dang, V. D., Chang, N.-C., & Wang, J. (2018). Multiple large inversions and breakpoint rewiring of gene expression in the evolution of the fire ant social supergene. *Proceedings of the Royal Society B: Biological Sciences, 285*(1878), 20180221. doi:10.1098/rspb.2018.0221

Huber, B., Whibley, A., Poul, Y. L., Navarro, N., Martin, A., Baxter, S., . . . Joron, M. (2015). Conservatism and novelty in the genetic architecture of adaptation in Heliconius butterflies. *Heredity, 114*(5), 515-524. doi:10.1038/hdy.2015.22

Jay, P., Whibley, A., Frézal, L., Rodríguez de Cara, M. Á., Nowell, R. W., Mallet, J., . . . Joron, M. (2018). Supergene Evolution Triggered by the Introgression of a Chromosomal Inversion. *Current Biology, 28*(11), 1839-1845.e1833. doi:10.1016/j.cub.2018.04.072

Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., . . . Kingsley, D. M. (2012). The genomic basis of adaptive evolution in threespine

sticklebacks. *Nature*, *484*, 55-61. doi:10.1038/nature10944

https://www.nature.com/articles/nature10944#supplementary-information

Joron, M., Frezal, L., Jones, R. T., Chamberlain, N. L., Lee, S. F., Haag, C. R., . . . ffrench-Constant, R. H. (2011). Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature*, *477*(7363), 203-206. doi:10.1038/nature10341

Joron, M., Wynne, I. R., Lamas, G., & Mallet, J. (1999). Variable selection and the coexistence of multiple mimetic forms of the butterfly Heliconius numata. *Evolutionary Ecology*, *13*(7-8), 721-754. doi:10.1023/A:1010875213123

Kapan, D. D. (2001). Three-butterfly system provides a field test of mullerian mimicry. *Nature*, *409*(6818), 338-340. doi:10.1038/35053066

Kirkpatrick, M. (2010). How and Why Chromosome Inversions Evolve. *Plos Biology*, *8*(9), e1000501. doi:10.1371/journal.pbio.1000501

Kirkpatrick, M., & Barton, N. (2006). Chromosome Inversions, Local Adaptation and Speciation. *Genetics*, *173*(1), 419-434. doi:10.1534/genetics.105.047985

Kirubakaran, T. G., Grove, H., Kent, M. P., Sandve, S. R., Baranski, M., Nome, T., . . . Andersen, O. (2016). Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of Atlantic cod. *Mol Ecol*, *25*(10), 2130-2143. doi:10.1111/mec.13592

Kraft, K., Geuer, S., Will, Anja J., Chan, Wing L., Paliou, C., Borschiwer, M., . . . Andrey, G. (2015). Deletions, Inversions, Duplications: Engineering of Structural

Variants using CRISPR/Cas in Mice. *Cell Reports, 10*(5), 833-839. doi:https://doi.org/10.1016/j.celrep.2015.01.016

Kunte, K., Zhang, W., Tenger-Trolander, A., Palmer, D. H., Martin, A., Reed, R. D., . . . Kronforst, M. R. (2014). doublesex is a mimicry supergene. *Nature, 507*(7491), 229-232. doi:10.1038/nature13112

Kupper, C., Stocks, M., Risse, J. E., dos Remedios, N., Farrell, L. L., Mcrae, S. B., . . . Burke, T. (2016). A supergene determines highly divergent male reproductive morphs in the ruff. *Nature Genetics, 48*(1), 79-83. doi:10.1038/ng.3443

Lamichhaney, S., Fan, G. Y., Widemo, F., Gunnarsson, U., Thalmann, D. S., Hoeppner, M. P., . . . Andersson, L. (2016). Structural genomic changes underlie alternative reproductive strategies in the ruff (Philomachus pugnax). *Nature Genetics, 48*(1), 84-88. doi:10.1038/ng.3430

Lee, C. R., Wang, B. S., Mojica, J. P., Mandakova, T., Prasad, K. Y. S. K., Goicoechea, J. L., . . . Mitchell-Olds, T. (2017). Young inversion with multiple linked QTLs under selection in a hybrid zone. *Nature Ecology & Evolution, 1*(5). doi:10.1038/s41559-017-0119

Lemopoulos, A., Uusi-Heikkilä, S., Huusko, A., Vasemägi, A., & Vainikka, A. (2018). Comparison of Migratory and Resident Populations of Brown Trout Reveals Candidate Genes for Migration Tendency. *Genome Biology and Evolution, 10*(6), 1493-1503. doi:10.1093/gbe/evy102

Li, J., Cocker, J. M., Wright, J., Webster, M. A., McMullan, M., Dyer, S., . . . Gilmartin, P. M. (2016). Genetic architecture and evolution of the S locus supergene in Primula vulgaris. *Nature Plants, 2*(12), 16188. doi:10.1038/nplants.2016.188

Ling, A., & Cordaux, R. (2010). Insertion Sequence Inversions Mediated by Ectopic Recombination between

Terminal Inverted Repeats. *PLoS One, 5*(12), e15654. doi:10.1371/journal.pone.0015654
Lowry, D. B., & Willis, J. H. (2010). A Widespread Chromosomal Inversion Polymorphism Contributes to a Major Life-History Transition, Local Adaptation, and Reproductive Isolation. *Plos Biology, 8*(9), e1000500.
Mallet, J., & Barton, N. H. (1989). Strong Natural-Selection in a Warning-Color Hybrid Zone. *Evolution, 43*(2), 421-431. doi:Doi 10.2307/2409217
Mather, K. (1950). The Genetical Architecture of Heterostyly in Primula sinensis. *Evolution, 4*(4), 340-352. doi:10.2307/2405601
Mather, K. (1955). Polymorphism as an Outcome of Disruptive Selection. *Evolution, 9*(1), 52-61. doi:10.2307/2405357
Matzkin, L. M., Merritt, T. J. S., Zhu, C.-T., & Eanes, W. F. (2005). The structure and population genetics of the breakpoints associated with the cosmopolitan chromosomal inversion In(3R)Payne in Drosophila melanogaster. *Genetics, 170*(3), 1143-1152. doi:10.1534/genetics.104.038810
McBroome, J., Liang, D., & Corbett-Detig, R. (2020). Fine-scale position effects shape the distribution of inversion breakpoints in Drosophila melanogaster. *Genome Biology and Evolution*. doi:10.1093/gbe/evaa103
Merrill, R. M., Wallbank, R. W. R., Bull, V., Salazar, P. C. A., Mallet, J., Stevens, M., & Jiggins, C. D. (2012). Disruptive ecological selection on a mating cue. *Proceedings of the Royal Society B-Biological Sciences, 279*(1749), 4907-4913. doi:10.1098/rspb.2012.1968
Navarro, A., & Barton, N. H. (2003). Chromosomal Speciation and Molecular Divergence--Accelerated

Evolution in Rearranged Chromosomes. *Science*, *300*(5617), 321-324. doi:10.1126/science.1080600

Nichols, K. M., Edo, A. F., Wheeler, P. A., & Thorgaard, G. H. (2008). The genetic basis of smoltification-related traits in Oncorhynchus mykiss. *Genetics*, *179*(3), 1559-1575. doi:10.1534/genetics.107.084251

Nishikawa, H., Iijima, T., Kajitani, R., Yamaguchi, J., Ando, T., Suzuki, Y., . . . Fujiwara, H. (2015). A genetic mechanism for female-limited Batesian mimicry in Papilio butterfly. *Nature Genetics*, *47*(4), 405-409. doi:10.1038/ng.3241

Nosil, P., Villoutreix, R., de Carvalho, C. F., Farkas, T. E., Soria-Carrasco, V., Feder, J. L., . . . Gompert, Z. (2018). Natural selection and the predictability of evolution in Timema stick insects. *Science*, *359*(6377), 765-770. doi:10.1126/science.aap9125

O'Malley, K. G., Sakamoto, T., Danzmann, R. G., & Ferguson, M. M. (2003). Quantitative Trait Loci for Spawning Date and Body Weight in Rainbow Trout: Testing for Conserved Effects Across Ancestrally Duplicated Chromosomes. *Journal of Heredity*, *94*(4), 273-284. doi:10.1093/jhered/esg067

Pearse, D. E., Miller, M. R., Abadia-Cardoso, A., & Garza, J. C. (2014). Rapid parallel evolution of standing variation in a single, complex, genomic region is associated with life history in steelhead/rainbow trout. *Proceedings of the Royal Society B-Biological Sciences*, *281*(1783). doi:10.1098/rspb.2014.0012

Pevzner, P., & Tesler, G. (2003). Human and mouse genomic sequences reveal extensive breakpoint reuse in mammalian evolution. *Proceedings of the National Academy of Sciences*, *100*(13), 7672. doi:10.1073/pnas.1330369100

Piper, J., & Charlesworth, B. (2008). The evolution of distyly in Primula vulgaris. *Biological Journal of the Linnean Society, 29*(2), 123-137. doi:10.1111/j.1095-8312.1986.tb01827.x

Puig, M., Casillas, S., Villatoro, S., & Caceres, M. (2015). Human inversions and their functional consequences. *Briefings in Functional Genomics, 14*(5), 369-379. doi:10.1093/bfgp/elv020

Ranz, J. M., Maurin, D., Chan, Y. S., von Grotthuss, M., Hillier, L. W., Roote, J., . . . Bergman, C. M. (2007). Principles of Genome Evolution in the Drosophila melanogaster Species Group. *Plos Biology, 5*(6), e152. doi:10.1371/journal.pbio.0050152

Robison, B. D., Wheeler, P. A., Sundin, K., Sikka, P., & Thorgaard, G. H. (2001). Composite Interval Mapping Reveals a Major Locus Influencing Embryonic Development Rate in Rainbow Trout (Oncorhynchus mykiss). *Journal of Heredity, 92*(1), 16-22. doi:10.1093/jhered/92.1.16

Rocca, K. A. C., Gray, E. M., Costantini, C., & Besansky, N. J. (2009). 2La chromosomal inversion enhances thermal tolerance of Anopheles gambiae larvae. *Malaria Journal, 8*. doi:10.1186/1475-2875-8-147

Ross, K. G., & Keller, L. (1995). Ecology and Evolution of Social-Organization - Insights from Fire Ants and Other Highly Eusocial Insects. *Annual Review of Ecology and Systematics, 26*, 631-656. doi:DOI 10.1146/annurev.es.26.110195.003215

Sandoval, C. P. (1994). Differential visual predation on morphs of Timema cristinae (Phasmatodeae:Timemidae) and its consequences for host range. *Biological Journal of the Linnean Society, 52*(4), 341-356. doi:doi:10.1111/j.1095-8312.1994.tb00996.x

Sandoval, C. P., & Nosil, P. (2005). Counteracting selective regimes and host preference evolution in ecotypes of two species of walking-sticks. *Evolution, 59*(11), 2405-2413.
Sharakhov, I. V., White, B. J., Sharakhova, M. V., Kayondo, J., Lobo, N. F., Santolamazza, F., . . . Besansky, N. J. (2006). Breakpoint structure reveals the unique origin of an interspecific chromosomal inversion 2La in the Anopheles gambiae complex. *Proceedings of the National Academy of Sciences, 103*(16), 6258-6262. doi:10.1073/pnas.0509683103
Sinclair-Waters, M., Bradbury, I. R., Morris, C. J., Lien, S., Kent, M. P., & Bentzen, P. (2018). Ancient chromosomal rearrangement associated with local adaptation of a postglacially colonized population of Atlantic Cod in the northwest Atlantic. *Mol Ecol, 27*(2), 339-351. doi:10.1111/mec.14442
Stefansson, H., Helgason, A., Thorleifsson, G., Steinthorsdottir, V., Masson, G., Barnard, J., . . . Stefansson, K. (2005). A common inversion under selection in Europeans. *Nature Genetics, 37*(2), 129-137. doi:10.1038/ng1508
Stewart, N. B., & Rogers, R. L. (2019). Chromosomal rearrangements as a source of new gene formation in Drosophila yakuba. *PLoS Genet, 15*(9), e1008314. doi:10.1371/journal.pgen.1008314
Sturtevant, A. H. (1921). A Case of Rearrangement of Genes in Drosophila. *Proc Natl Acad Sci U S A, 7*(8), 235-237. doi:10.1073/pnas.7.8.235
Thomas, J. W., Cáceres, M., Lowman, J. J., Morehouse, C. B., Short, M. E., Baldwin, E. L., . . . Martin, C. L. (2008). The Chromosomal Polymorphism Linked to Variation in Social Behavior in the White-Throated Sparrow (Zonotrichia albicollis) Is a Complex Rearrangement and

Suppressor of Recombination. *Genetics, 179*(3), 1455-1468. doi:10.1534/genetics.108.088229

Thorneycroft, H. B. (1966). Chromosomal Polymorphism in the White-Throated Sparrow, Zonotrichia albicollis (Gmelin). *Science, 154*(3756), 1571-1572.

Throneycroft, H. B. (1975). A Cytogenetic Study of the White-Throated Sparrow, Zonotrichia albicollis (Gmelin). *Evolution, 29*(4), 611-621. doi:10.2307/2407072

Tuttle, E. M., Bergland, A. O., Korody, M. L., Brewer, M. S., Newhouse, D. J., Minx, P., . . . Balakrishnan, C. N. (2016). Divergence and Functional Degradation of a Sex Chromosome-like Supergene. *Current Biology, 26*(3), 344-350.

Van Belleghem, S. M., Rastas, P., Papanicolaou, A., Martin, S. H., Arias, C. F., Supple, M. A., . . . Papa, R. (2017). Complex modular architecture around a simple toolkit of wing pattern genes. *Nature Ecology & Evolution, 1*, 0052. doi:10.1038/s41559-016-0052

http://www.nature.com/articles/s41559-016-0052#supplementary-information

Villoutreix, R., de Carvalho, C. F., Soria-Carrasco, V., Lindtke, D., De-la-Mora, M., Muschick, M., . . . Nosil, P. (2020). Large-scale mutation in the evolution of a gene complex for cryptic coloration. *Science, 369*(6502), 460-466. doi:10.1126/science.aaz4351

Wang, J., Wurm, Y., Nipitwattanaphon, M., Riba-Grognuz, O., Huang, Y. C., Shoemaker, D., & Keller, L. (2013). A Y-like social chromosome causes alternative colony organization in fire ants. *Nature, 493*(7434), 664-668. doi:10.1038/nature11832

Yan, Z., Martin, S. H., Gotzek, D., Arsenault, S. V., Duchen, P., Helleu, Q., . . . Keller, L. (2020). Evolution of a

supergene that regulates a trans-species social polymorphism. *Nature Ecology & Evolution*, *4*, 240-249.