CHAPTER SIX

Identifying Coopted Networks and Causative Mutations in the Origin of Novel Complex Traits

A. Monteiro[*,†,1], **M.D. Gupta**[‡]
[*]National University of Singapore, Singapore, Singapore
[†]Yale-NUS College, Singapore, Singapore
[‡]Max Planck Institute for Plant Breeding Research, Köln, Germany
[1]Corresponding author: e-mail address: antonia.monteiro@nus.edu.sg

Contents

Abstract

One of the central goals of the field of evo-devo is to understand how novel complex traits originate. Novel complex traits are often old, and this makes understanding the genetic basis of their origin difficult. The traditional genetics approach for identifying the causative mutations for trait origin, of crossing species with and without the trait, is often impossible when the species are too distantly related. Alternatively, if the species are closely related, the genetic basis of their differences is often the recent loss, rather than the gain, of the trait in one of them, and mutations resulting in trait loss are not always equivalent to those that led to trait gain. Here, we reexamine an evo-devo study of the origin of melanic spots in the wings of flies, which is presented in more than one mainstream undergraduate textbook on Evolution, as an example of molecular evolution leading to the origin of a novel trait. We put forth an alternative to the previously

Current Topics in Developmental Biology, Volume 119
ISSN 0070-2153
http://dx.doi.org/10.1016/bs.ctdb.2016.03.003

proposed scenario and, in our view, a more likely evolutionary framework that explains the data, the CRE–DDC model, and then review other case studies and avenues of research that should help identify where new complex traits come from, as well as the actual causative mutations underlying their origin.

1. INTRODUCTION

One of the goals of the field of evo-devo is to understand how gene regulatory networks (GRNs) evolve to give rise to novel complex traits (Wagner & Lynch, 2010). In order to properly examine this goal, we will first clarify what we mean by "novel" and by "complex." A novel trait here follows a character—rather than a functional-based definition (Peterson & Mueller, 2013) and is a qualitatively novel feature of an organism, that arises at some point in a lineage so that it is absent from the sister lineage and from the common ancestor of both. A complex trait is a trait whose development depends on a large number of interacting genes, even though it is unclear at this point what an appropriate threshold for this large number should be. The pursuit of the evolution of novel complex traits is, thus, a pursuit of the mechanisms that lead to the origin of novel discrete traits in organisms, rather than the mechanisms that modify preexistent traits and networks in a more quantitative fashion.

The recent evo-devo literature has been making exciting headways into the evolution of novel traits but several papers appear to confound the evolution of gene expression with the evolution of novel traits. This current confusion in the field is what prompted us to write this review. Later, we review some of the proposed theory about the evolution of novel GRNs; then we highlight an example that has made its way into mainstream Evolution and Evo-devo undergraduate level textbooks (Bergstrom & Dugatkin, 2012; Futuyma, 2009; Gilbert & Epel, 2009) whose underlying genetic mechanism may not be as simple as previously proposed; we then highlight other work that takes alternative and complementary approaches that provide unique perspectives to this complex problem, and we end by proposing novel avenues for progress to be made in this field.

2. THE DIFFICULTY OF RESEARCH ON THE EVOLUTION OF NOVEL TRAITS

One of the central paradigms in the field of evo-devo is the understanding that when a new complex trait arises in an organism, let us say a new morphology, then, something had to change during development

for the trait to appear. Inevitably, this involves changes in the regulation of one or more genes involved in building that trait, directly or indirectly, as a result of DNA mutations. These mutations can either occur in *cis*-regulatory elements (CREs) or in other parts of the DNA, such as in protein or microRNA-coding sequences.

One of the main difficulties of performing research on novel traits lies in the identification of the root causes of the origin of the novel trait. Crossing species with and without the trait, and following the inheritance pattern of the trait and of closely linked genetic markers helps to identify the genomic location of mutations underlying trait origin. However, this type of experiment is often impossible to conduct for most novel traits, either because the species with and without the trait are too distantly related and cannot be crossed, or because, if closely related, the genetic basis of their differences is often the recent loss, rather than the gain, of the trait in one of them, and mutations resulting in trait loss may not be equivalent to those that lead to trait gain. When species can be crossed, however, as in the case of the multiple races of a single species of butterfly, this type of research is very powerful and has led to the identification of mutations in regulatory regions of genes that caused the evolution of novel wing color phenotypes (Reed et al., 2011; Wallbank et al., 2016).

Because identifying mutations leading to the origin of the novel traits in more distantly related species is difficult to do, researchers have opted to examine whether mutations in genes known to be required for trait development, and whose changes in expression correlate with changes in trait expression, hold the key mutations that led to trait origins. Sometimes, mutations that are not causative of the novel trait will accumulate in CREs of these genes, and these mutations can be mistakenly inferred to be important contributors to the appearance or loss of the trait. Instead, it is possible that many of these mutations are neutral, ie, resulting from relaxed selection, or merely aid in strengthening or weakening the expression of the novel trait, after its origin.

We will begin by reviewing some novel complex traits and the mechanisms proposed to underlie their origin. Then, we will discuss some of the claims regarding the proposed molecular basis for their origin.

3. EXAMPLES OF NOVEL COMPLEX TRAITS AND THEIR LIKELY ORIGIN VIA COOPTION OF PREEXISTENT NETWORKS

Many complex traits appear to evolve from preexisting networks of interconnected genes that are rewired to perform new functions. The

cooption of preexisting networks is often loosely inferred based on similarity of gene expression patterns that occur during the development of the ancestral trait and the novel trait. These novel traits include the insect wing, which was proposed to have originated via the merging of two preexisting gene networks, one involved in lateral body margin development, and the other involved in leg lobe development (Medved et al., 2015; Niwa et al., 2010). Other examples are the evolution of limbs in early bilaterians from a coopted anterior–posterior head axis patterning system (Lemons, Fritzenwanker, Gerhart, Lowe, & McGinnis, 2010), the evolution of beetle horns from a coopted limb network (Moczek & Rose, 2009); the evolution of a novel embryonic skeleton of sea urchins via the reuse of the adult skeleton network (Gao & Davidson, 2008); the evolution of butterfly eyespots from a variety of alternative preexisting gene networks, including those for limb development, wound healing, and wing margin development (reviewed in Monteiro, 2015); the evolution of paired fins from a preexistent median fin network (Freitas, Zhang, & Cohn, 2006); and the evolution of a novel genital lobe in *Drosophila melanogaster* from a preexistent GRN underlying the development of larval breading spiracles (Glassford et al., 2015).

For all of the traits described earlier, there is no actual knowledge of the causative mutations that led to the appearance of the novel trait. However, all these traits appear to originate via the cooption of one or more preexisting GRNs that are redeployed in novel contexts. GRNs can be modular and context insensitive and can work in different locations in the body by the simple activation of a few top regulatory gene(s) at that novel location (Halder, Callaerts, & Gehring, 1995; Monteiro, 2012; Schlosser, 2004). In addition, these core GRNs can become activated via novel inputs, or become rewired to activate a different set of downstream genes (Davidson & Erwin, 2006) to give rise to truly novel traits.

The simplest molecular mechanism by which a modular GRN can be coopted to a novel developmental context involves the evolution of a novel CRE in a top regulatory gene of the network that coopts this gene to the novel context. Because this top regulatory gene is prewired to a series of downstream targets, these other genes become expressed in the novel context as well, without suffering any direct molecular evolution in their CREs, at least immediately following the cooption event (Gao & Davidson, 2008; Monteiro, 2012; Monteiro & Podlaha, 2009). In other words, the CREs of internal or terminal network genes become pleiotropic and multifunctional following a cooption event (Monteiro & Podlaha, 2009). An elegant demonstration of this pleiotropy was recently found in *D. melanogaster*, where

CREs of multiple genes involved in building a novel genital lobe were found to also drive the expression of the same genes in a different developmental context, in the development of the larval spiracle (Glassford et al., 2015). The authors initially assumed that two distinct CREs, one for each developmental context, had to lay adjacent or intermixed with each other at each locus; but this was not the case, as fine dissection of these elements showed that their function was indivisible.

4. MISCONCEPTIONS AROUND THE COOPTION OF PREEXISTING CREs REQUIRED FOR THE EVOLUTION OF NOVEL TRAITS

The initial assumption of Glassford and colleagues that the pleiotropic CREs they isolated in the genes belonging to the anal lobe/larval spiracle GRNs in *Drosophila* should be divisible for function lies at the heart of a likely misconception that we want to address directly in this review. This misconception involves the idea that evolution in the sequence of a preexisting CRE of a gene, regardless of the gene's position in a GRN, can coopt the gene to perform a new function, and therefore contribute to the appearance of novel traits. This idea largely derives from studies that look at the evolution of gene expression (Gompel, Prud'homme, Wittkopp, Kassner, & Carroll, 2005; Jeong et al., 2008; Rebeiz, Jikomes, Kassner, & Carroll, 2011), rather than studies that look at the evolution of novel traits. Furthermore, this idea is potentially valuable in investigations of trait loss but has not been validated in the investigations of trait origin. However, the idea has been used to illustrate the origin of certain novel traits and translated into general undergraduate textbooks (Bergstrom & Dugatkin, 2012; Futuyma, 2009; Gilbert & Epel, 2009).

The cooption and modification of a preexisting CRE idea is generally pitched like this: a gene that already has a functional CRE that leads to its expression in developmental context X, acquires new mutations inside it or in its immediate vicinity, that lead to the gene now being expressed in developmental context Y, in addition to X, and acquiring a new function during development (Prud'homme et al., 2006; Rebeiz et al., 2011).

The CRE cooption idea largely derives from studies that have focused on candidate genes known to be involved in building the trait of interest rather than on genes known to regulate the entire network of genes required to build the trait. Furthermore, the genes examined are often terminal effector genes, at the ends of GRNs, rather than developmental genes with the

potential to coordinate the expression of multiple effector genes in a GRN. An example of the former type of gene, is *yellow*, an enzyme involved in melanin synthesis, whose CREs have been dissected in extreme detail across *Drosophila* species differing in melanization patterns (Prud'homme et al., 2006). Fig. 1 describes the CRE cooption model where the independent tweaking of two preexisting CREs for *yellow* were interpreted to contribute to the independent gain of a novel *yellow* expression domain and a novel melanized spot in the wing tip of each species (Fig. 1).

These studies tend to correlate evolution of phenotype, eg, gain or loss of black pigmentation, with evolution of terminal gene regulation, eg, *yellow* gene expression, but they fail to show causation. They first identify the

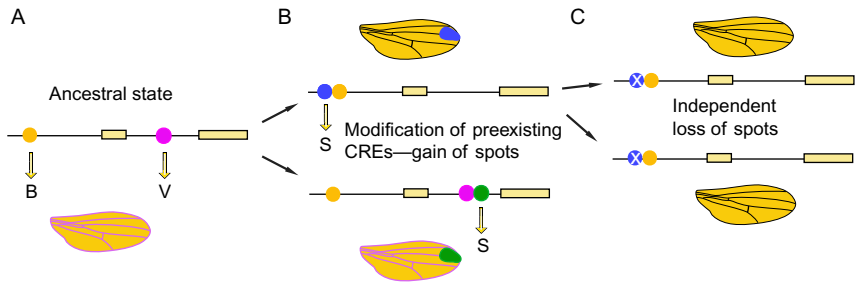

Fig. 1 A popular study on the origin of a novel trait, wing spots, discovers the possible molecular basis of spot loss, but perhaps not how spots were actually gained. Prud'homme and collaborators (Gompel et al., 2005; Prud'homme et al., 2006) have studied the evolution of gains and losses of wing spots in several species of *Drosophila* by comparing the genetic changes that took place in the candidate pigmentation gene *yellow*. (A) They studied two sets of closely related lineages that separately evolved wing spots (S), arguably from the modification of two preexisting CREs, one driving *yellow* in the wing blade (B), and one along the wing veins (V). Both of these areas contain melanin pigments. (B) Each lineage modified a separate CRE to drive *yellow* expression in the tip of the wing (*green* and *blue* areas). (C) In two descendent lineages from one of the lineages with spots, spots were lost via changes in the same CRE (*blue*). The authors claim that evolution in preexisting CREs of *yellow* led to the gain of the spots in each of the lineages. However, they do not show that these changes are sufficient to create spots. In fact, when the two spot CREs of *yellow* were expressed in *D. melanogaster* (a species without wing spots) they were insufficient to produce pigmentation spots. Other factors need to be jointly coregulated as well, pointing to the cooption of *yellow* as part of a larger modular pigmentation network. More recent work (Arnoult et al., 2013) showed that evolution at the transcription factor *Distal-less* is necessary and sufficient to regulate expression of *yellow* in the wing spot region, as well as regulate other enzymes that need to be present or absent simultaneously for black melanin to be produced on the wing (see Fig. 2).

CREs that drive the precise spatial expression of the terminal gene in the different species, and then show that evolution in those sequences often matches the evolution of gain or loss of the phenotype in the adult. What is often overlooked in the translation of this research into general evolution textbooks, despite the mention of caveats in the discussion of these papers, is that evolution in these CREs while perhaps sufficient to cause loss of the novel trait, or to cause gain of a novel gene expression pattern, is not sufficient for gain of the novel trait, eg, black pigmentation. For instance, it remains unclear whether regulatory evolution at the *yellow* locus is implicated in the gain of a novel pigmentation pattern in *Drosophila* across any of the myriad studies where evolution of CREs of *yellow* have been investigated. Evolution of *yellow* expression in novel locations of the body is not sufficient to drive evolution of pigmentation because other enzymes are needed, and others need to be suppressed, in the same cells for melanin synthesis to take place (Gompel et al., 2005; True, Edwards, Yamamoto, & Carroll, 1999; Wittkopp & Beldade, 2009).

A piecemeal recruitment of individual genes to a novel location in the body that, on their own, cannot produce a visible phenotype, is unlikely to be at the origin of novel traits that require multiple genes to work in concert. This is because regulatory mutations driving expression of each gene at the novel location would not create a phenotype visible to selection and, thus, would be unlikely to persist until the other required genes are recruited to the same location and the final phenotype would come under purifying selection.

The most likely way that organisms are evolving novel traits is via changes in the regulation of developmental genes that are able to coordinate the expression of complete modular GRNs that lead to a phenotype immediately visible to selection. An example of this type of gene is *shavenbaby*, a gene both necessary and sufficient to initiate trichome development in flies, and whose collection of distinct CREs add distinct patches of trichomes to the body of a larvae (McGregor et al., 2007). In the case of melanin spots in flies, a recent candidate with a similar function to that of *shavenbaby* appears to be *Distal-less* (*Dll*) (Arnoult et al., 2013; Monteiro et al., 2013). The evolution of novel *Dll* expression domains appears to be sufficient to explain the origin of spots of pigmentation in some fly wings. A novel *Dll* expression domain in the tip of the pupal wing in a spotted *Drosophila* species, drives *yellow* expression in the same exact domain, and induces melanin synthesis (Arnoult et al., 2013) (Fig. 2). Overexpression of *Dll* in this species drives *yellow* expression and wing pigmentation across the whole wing, whereas

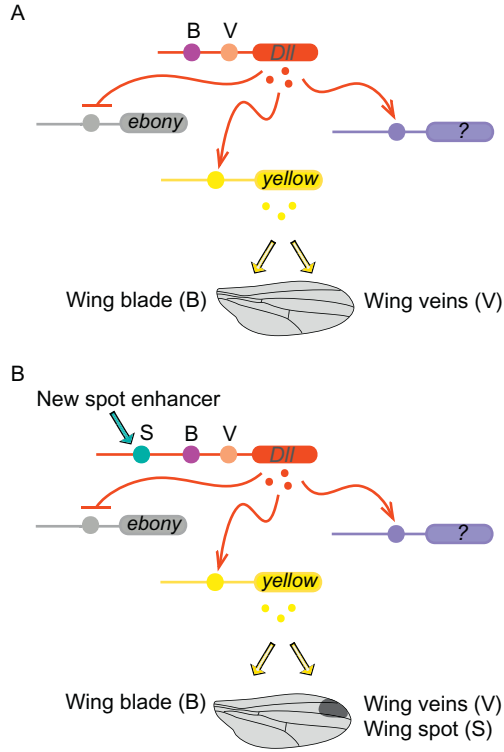

Fig. 2 A model of wing spot evolution based on Arnoult et al. (2013). (A) *Distal-less* (*Dll*) is prewired to melanin synthesis genes, such as *ebony*, *yellow*, and others, before the origin of a distal wing spot. This network is likely responsible for the slight melanization across the whole wing blade (B) and along the wing veins (V) presumably via distinct *cis*-regulatory elements (CREs) at *Dll*, B, and V. (B) *Dll* acquires a novel CRE (S) that leads to a novel gene expression domain in the wing tip of a fly wing late during pupal wing development. This leads to the expression of *yellow* in this area of the wing, as well as other gene expression changes, causing the gain of a melanized wing spot (S). *yellow* is responding to Dll inputs, and Dll binding sites at *yellow* predate the cooption of Dll to the wing tip.

downregulation of *Dll* leads to loss of melanin spots, and loss of *yellow* expression in those spots (Arnoult et al., 2013). In *Bicyclus anynana* butterflies, *Dll* appears to have the same function—it is sufficient to drive black pigmentation on the wing when ectopically expressed during early pupal development, and black areas of pigmentation become reduced when *Dll* is downregulated at that same time in development (Monteiro et al., 2013). What this means is that novel pigmentation patterns can appear on the wing without evolution happening in CREs of terminal genes at ends

of networks such as *yellow*. These terminal genes become active in the novel wing domains because top regulatory genes of a pigmentation GRN, in this case *Dll*, get recruited (or expressed at higher levels) in these novel locations, and not because the individual members of the network do so on their own. It is important to note that we are not arguing that evolution of the novel *Dll* expression domain alone in the flies with black wing spots contributed to the appearance of the spots in all their contrast and intensity. We are arguing that the novel *Dll* expression domain enabled the recruitment of a functioning melanin synthesis GRN and a visible phenotype. We believe that later genetic variants, for instance those identified in the *yellow* locus, contributed to accentuate the final phenotype, but not to its origin. The data that support this idea comes from testing the evolved *yellow* CRE of *Drosophila biarmipes* attached to *yellow* in a *D. melanogaster ebony* mutant line, and this leading to a slight darkening in the anterior area of the wing (Gompel et al., 2005).

Regulatory connections between genes like *Dll* and terminal genes like *yellow* had to form at some point for the modular pigmentation GRN to come to existence. Arnoult et al. (2013) argue that the *Dll-yellow* connection formed within the *Drosophila* clade, as *Dll* overexpression in one of the basal-branching unspotted members of the group, *Drosophila ananassae*, did not lead to *yellow* expression nor wing melanization. However, it is also possible that the *Dll-yellow* connection formed earlier and was later lost in *D. ananassae*. While no direct evidence exists that ectopic *Dll* leads to ectopic *yellow* expression in *B. anynana*, the experiments in this species suggest that such connection (direct or indirect) exists, and may conceivably predate the divergence of flies and butterflies and be part of an ancestral "melanin synthesis regulatory network" (Monteiro et al., 2013). Only future comparative work with additional *Drosophila* and/or butterfly species will be able to address this question. For instance, testing whether *Dll* also regulates black spot development in species of the *obscura* group, or repeating the ectopic expression of *Dll* in a few additional basal lineages of the *melanogaster* group, as was done for *D. ananassae*, should help lend support for either the ancestral melanin GRN hypothesis or the independent and convergent melanin GRN evolution hypothesis in flies and butterflies.

Earlier, we have argued that there are few molecular routes underlying the gain of a complex trait. The loss of a complex trait, however, can take more than one genetic route. A loss does not necessarily require the loss of expression of the top regulatory gene in the developmental context that led to trait gain, although this is one way a loss can happen (see example from

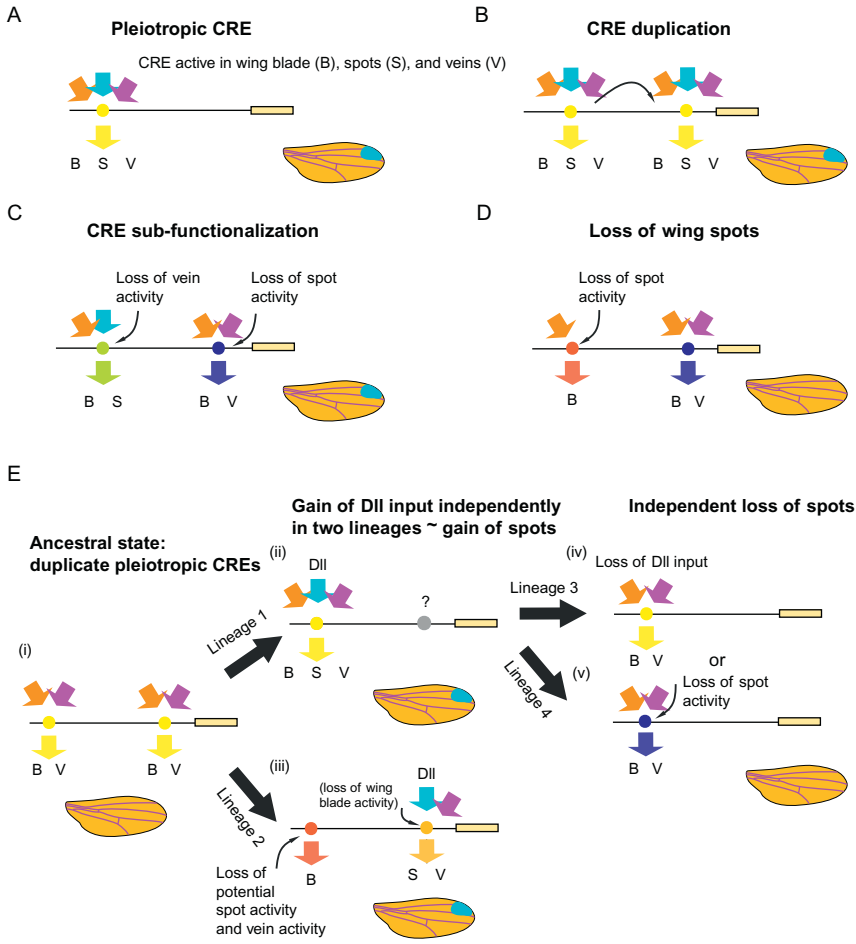

Fig. 3 An alternative model, which we call the CRE–DDC model, to that proposed by Prud'homme et al. (2006) regarding origin and loss of wing tip spots. Origin of wing spots does not necessarily involve molecular evolution at the *yellow* locus, whereas spot loss may. (A) If the gene *yellow* is part of a modular gene regulatory network (see Fig. 2), when the network is recruited to a novel developmental context the *cis*-regulatory element (CRE) of *yellow*, without undertaking any evolution, becomes pleiotropic, and drives *yellow* in novel expression domains. In this example, the network has been reused three times and the CRE drives gene expression in the wing blade, B, wing spots S, and wing veins, V. (B) The CRE duplicates and both CREs can now be used to drive *yellow* expression in all the three wing areas. (C) The redundant CREs begin to subfunctionalize via neutral process of evolution (see text), separating some of their ancestral functions. Here the left CRE (*green*) loses the ability to drive *yellow* in the vein tissue (loss of *pink arrow* above CRE) perhaps via the evolution of binding sites for repressor molecules present only in the vein tissue. The right CRE loses the ability to drive *yellow* in the wing spots (loss of *light blue*

(Continued)

Arabidopsis later). A loss of phenotype can also happen via loss of expression of the individual genes that are part of the coopted network, via mutations in their CRE. The loss of expression of these internal genes, however, can potentially affect the development of more than one trait, because these genes are originally pleiotropic. However, this pleiotropy can be circumvented if the CREs of these genes duplicate and subfunctionalize (Fig. 3). Many of the detailed examples of trait evolution (trait loss) in the evo-devo literature are probably documenting this type of scenario.

Force et al. (1999) proposed that one of the most likely explanations for the maintenance of duplicated genes in genomes, ie, the protein-coding sequence plus associated CREs controlling the gene's expression in different developmental contexts—was the subfunctionalization of the gene's original functions. This would happen merely due to relaxed selection, that is, mutations accumulating in redundant duplicated CREs. Their model is called the duplication, degeneration, and complementation (DDC) model. For instance, if a locus containing multiple CREs duplicates into copies A and B, mutations that accumulate in one of the CREs in copy A, leads to purifying selection on the ortholog CRE in copy B, and vice versa for mutations that accumulate first in a CRE in copy B, leading to the maintenance of its ortholog CRE in copy A. The remaining single copy CRE at each locus, now requires the presence of a functional protein-coding sequence in *cis* (next door) for its function to persist, and the gene duplicates are, thus, maintained in the genome.

Fig. 3—Cont'd *arrow*). (D) Further mutations in the left CREs, eliminates *yellow* expression and melanin spots on the wing. (E) Alternative scenario to that proposed by Prud'homme et al. (2006), depicted in Fig. 1. (i) The gene *yellow* has undergone duplication of an ancestral pleiotropic CRE. (ii–iii) Two separate lineages of flies (lineages 1 and 2) gain pigmentation spots because of molecular evolution upstream of *yellow*, leading to a novel *Distal-less* (*Dll*) input taking place in the wing tip (*light blue arrows*). The gray CRE has not been investigated experimentally. In lineage 2, the *red* CRE has lost its potential activity in the distal wing spot region (as well as vein activity), so reuse of a preexisting pleiotropic CRE can only take place at the other CRE (iii). In addition, the functional spot CRE of lineage 2 has also lost wing blade activity. (iv–v) In subsequent descendants of lineage 1 (lineages 3 and 4) there are two independent losses of spots that may have occurred due to loss of *Dll* activity in the wing tip (loss of *light blue arrow*) (iv), or due to molecular evolution at the yellow CRE (v), preventing CRE activity specifically in the wing tip.

5. THE CRE–DDC MODEL

The DDC model described earlier starts with the premise that most genes have complex modular regulation, ie, multiple CREs, but does not explain how this complex regulation may originate. An extension of this model, however, can be applied to the creation and maintenance of novel CREs in genes, starting with single pleiotropic CREs that duplicate and subfunctionalize. Pleiotropic CREs arise automatically in genes in the middle or terminus of GRNs due to network cooption events (Fig. 2). However, once a pleiotropic CRE duplicates, it becomes free to subfunctionalize. This subfunctionalization can arise merely due to nonadaptive processes, as in the DDC model. For instance, mutations may accumulate in one of the CRE's copies that destroy its function in one of the developmental contexts but not in the other. This can happen via the accumulation of mutations in the CRE that create binding sites for repressors that are present in one developmental context but not in the other (Fig. 3). Once one of the CRE's pleiotropic functions is eliminated, that same function comes under purifying selection in the second CRE and that second CRE is retained in the genome. If mutations that eliminate a different function in the second CRE occur, then that function needs to be maintained by the first CRE, and both CREs come under purifying selection and are retained in the genome (Fig. 3). This is a likely mechanism by which multiple CREs in the gene *yellow* as well as in many other genes in the middle or terminus of GRNs came to be.

6. HOW TO DIFFERENTIATE *CIS*-ELEMENT EVOLUTION VS *CIS*-ELEMENT REUSE VIA GENE NETWORK COOPTION IN THE ORIGIN OF NOVEL TRAITS

The subfunctionalization of duplicated CREs of internal and terminal network genes essentially makes these genes resemble top network regulatory genes in their modular regulatory architecture (McKay & Lieb, 2013). By this we mean, each type of gene, regardless of their position in a network, may end up evolving a series of modular CREs that regulate its expression in only a single or a few different developmental contexts. However, mutations that lead to trait gain can generally only occur in top regulators of the network, not in genes that lie in the middle or terminus of such networks. So, when we see evolution of CREs of genes in closely related species, as previously documented for those of the gene *yellow*, associated with the gain of a

novel trait (Prud'homme et al., 2006), it is tempting to conclude that these mutations contributed to the origin of the trait. However, this may not necessarily be so. In these cases, it is important to test whether the observed expression changes are actually responsible for the gain of the trait on their own. In many cases, these mutations may instead be involved in the process of CRE subfunctionalization via purely neutral processes, or alternatively, involved in quantitative enhancements or modifications of the novel trait but not its origin.

The hypothesis of CRE evolution leading to the origin of novel traits proposes that molecular evolution takes place at the CRE, either next door or inside the element, for the element to acquire the novel expression pattern associated with the origin of the novel trait. In the CRE reuse hypothesis, no new molecular evolution is required for this CRE to acquire the novel expression domain seen in the lineage with the novel trait. This is because molecular evolution took place in a top regulatory gene further up in the network, coopting it to the novel developmental context. This important distinction on where causative mutations are leading to the origin of novel traits can be used in tests that support or refute the CRE evolution vs the CRE reuse hypotheses.

One way of testing whether molecular evolution at a preexisting CRE is actually responsible for coopting the CRE and the associated gene to perform a novel function (see *Arabidopsis* and *Cardamine* example later), is to take the putatively novel CRE attached to the gene it is supposed to regulate, and express it in a close related host species without the trait. If the CRE-gene construct leads to the novel trait, then one can conclude that evolution in its sequence was causal. If not, then either evolution at the CRE is not causal or the host species has other mutations that prevent the proper function of the entire network. A different host can be tried in these instances. A reverse approach is to try and identify an orthologous CRE in a sister lineage that does not display the novel trait and test the expression of this orthologous element in the species with the novel trait. If this CRE leads to a novel gene expression domain in the *trans*-regulatory environment of the species with the novel trait, then evolution of its sequence is not required to lead to the origin of the novel trait. This is because the CRE likely belongs to a gene in the middle or terminus of a regulatory network and will automatically be able to drive gene expression in a novel developmental context as long as the activating *trans*-factors are expressed in the novel context as well (see the transition from Fig. 3Ei to ii).

7. NOVEL APPROACHES TO IDENTIFY TOP REGULATORS OF NETWORKS AND CAUSATIVE MUTATIONS IN THE ORIGIN OF NOVEL COMPLEX TRAITS: THE POWER OF FORWARD GENETIC SCREENS

Although today we have an extraordinary array of cutting-edge genomic tools to study the regulation of developmental processes, the traditional method of forward genetic screens still continues to be one of the most powerful techniques available to uncover the location of causative mutations that lead to the origin of novel traits.

The main idea behind a traditional forward genetic screen is to randomly mutate places in the genome, both protein-coding and regulatory regions, which alter the development of traits, and by doing so, implicate these regions in the regulatory network that builds the trait. Because most genes in networks are internal genes, the majority of these screens discover internal and terminal genes in networks, but a recent screen discovered the actual locus that caused the evolution of a morphological novelty in plants.

Early members of the mustard family of plants (eg, *Aethionema arabicum*) have simple leaves, whereas later derived members (eg, *Cardamine hirsuta*) acquired dissected or compound leaves, the novel trait. Some derived members within the dissected leaf lineage (eg, *Arabidopsis thaliana*), however, reverted back to simple leaves and to the ancestral trait (Fig. 4). Therefore, *C. hirsuta* and *A. thaliana* represent a pair of closely related species where a novel trait, dissected leaves, was gained and then lost, respectively.

Using a forward genetic screen in *C. hirsuta*, Vlad and colleagues were able to identify a mutant with simplified leaves where much of the dissected character was lost (Vlad et al., 2014). Genetic fine mapping showed that the simplified leaf phenotype in *C. hirsuta* was caused by the reduced function of a homeobox gene (*Reduced Complexity*, *RCO*), which exists as a duplicate of another gene, *LMI1*-type (*Late Meristem Identity-1*) (Fig. 4). The simplification of the leaf in the *C. hirsuta* mutant was due to the loss of *RCO*. This mimicked the process of leaf simplification in *Arabidopsis*, which also lost *RCO* during its evolution. Interestingly, the coding region of *RCO* is functionally equivalent to *LMI1*, but the CREs of each of these copies are different. In fact, a novel CRE in *RCO*, leading to its novel expression near the base of the developing leaflets, caused the origin of the dissected and lobed leaves in the Brassicaceae family. Expressing *RCO* and its flanking genomic sequences (containing the novel CRE) from *C. hirsuta* in *A. thaliana* led to

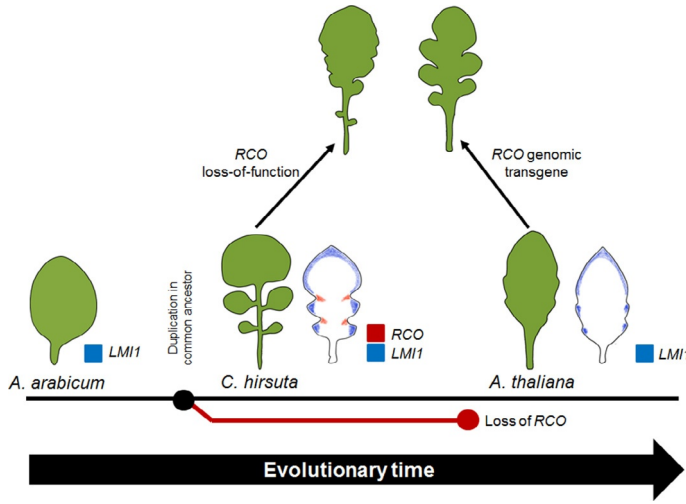

Fig. 4 Duplication and regulatory divergence leading to a morphological novelty, dissected leaves, in Brassicaceae. *Aethionema arabicum*, a simple leaved species, has a single copy of *LMI1*. *RCO* arose by the duplication of *LMI1* leading to dissected leaves in species like *C. hirsuta*. The loss of *RCO* in species like *A. thaliana* led to leaf simplification. While the loss of *RCO* function in *C. hirsuta* leads to formation of simpler leaves, the expression of the genomic copy of *RCO* from *C. hirsuta* in *A. thaliana* leads to formation of deeply lobed leaves. The *LMI1* coding sequence can complement the *RCO* function in *C. hirsuta* only when placed under the *RCO* regulatory sequence. The schematics of the expression patterns of *RCO* and *LMI1* in developing *C. hirsuta* and *A. thaliana* leaves are shown alongside the adult leaf outlines. The novel expression domain of *RCO* in *C. hirsuta*, at the base of developing leaflets, suppresses growth leading to dissections in the adult leaf. *Adapted from Vlad, D., Kierzkowski, D., Rast, M. I., Vuolo, F., Dello Ioio, R., Galinha, C., et al. (2014). Leaf shape evolution through duplication, regulatory diversification, and loss of a homeobox gene.* Science, 343, *780–783.*

formation of deeply lobed leaves resembling the dissected leaves of *C. hirsuta*. This conceptually simple experiment showed that a novel CRE in *RCO* created a novel phenotype, and its loss led to trait loss. The discovery of the causative mutation in this case was aided by the fact that the *RCO* gene was solely functioning in leaf dissection, so the mutation that led to its loss had no pleiotropic effects and could be easily studied. However, it must be pointed out that in this study the novel CRE from *C. hirsuta* was transplanted into *A. thaliana*, which had secondarily lost the lobed leaf character to represent the ancestral state. A more rigorous test to definitively prove that the novel CRE in *C. hirsuta* caused the gain of dissected leaf trait during evolution would be to express the genomic region of *RCO* in a more basally divergent

simple leaf crucifer species (eg, *A. arabicum*) and verify whether it can drive *RCO* expression in the novel domain and confer the novel trait. We should also point out that although this study is an excellent example to show trait loss due to the loss of a top regulator, it does not lend proof to the CRE–DDC model specifically because the novel CRE was experimentally moved into a species where the trait was secondarily lost.

Forward genetic screens were also used to identify the direct regulator of *yellow* in the CRE that drove this gene in the tip of the *D. biarmipes* wing (Arnoult et al., 2013). Several hundred wing transcription factors were individually knocked-down in *D. melanogaster* RNAi transgenic lines, and the lines were screened for their ability to disrupt the activity of the *D. biarmipes* yellow spot CRE in a *D. melanogaster* host. This approach eventually led to the identification of *Dll* as the coordinator of the set of enzymes needed for melanin synthesis.

8. HOW TO IDENTIFY COOPTED GRNs USING NOVEL GENOMIC TOOLS

The examples above illustrate that forward genetic screens can be used to identify top regulators of GRNs that, when coopted to novel developmental contexts, create novel traits. However, while identifying causative mutations that lead to novel traits is certainly an important consideration, an equally important question concerns identifying the GRNs that are being controlled by these top regulators for a better understanding of the evolutionary process. In essence, we also want to learn the identity of the preexisting GRNs that get coopted and modified to produce novel traits.

The identity of a coopted GRN can, in principle, be guessed via the identification of a pleiotropic CRE that drives gene expression in the old and in the novel developmental contexts (Glassford et al., 2015; Monteiro, 2012) (Fig. 2A). For instance, when CREs of genes involved in the development of a new genital lobe in *Drosophila* were attached to a reporter gene, they drove reporter gene expression in the lobe and in the larval spiracles, showing that a gene network involved in building a spiracle was coopted to build the novel genital lobe (Glassford et al., 2015; Monteiro, 2012). As we have outlined earlier, it is likely that over time even these CREs may lose some of their pleiotropic functions and subfunctionalize, preventing the identification of the preexisting developmental context wherein they were active. However, given the predicted large number of internal genes in coopted networks, many, if not most, of these genes should

still contain some pleiotropic function, and should contain the identity of the GRN they originally functioned in.

The identification of pleiotropic CREs, however, is not straightforward. This is especially the case in nonmodel systems where blindly testing candidate regulatory sequences for function in transgenics can be very time consuming, if not entirely prohibitive. New genomic tools, however, are currently available to facilitate the identification of CREs that do not require the use of transgenics. These tools involve methods that identify open chromatin regions surrounding the genes of interest in the specific tissues of interest at the specific times of development when the genes are being expressed (Fig. 5). Techniques like ChIP-seq and DNase-seq have been successfully used to identify regulatory sequences (Meyer & Liu, 2014). However, ChIP-seq relies on the availability of specific antibodies and identifies regions of DNA bound specifically by certain proteins thereby biasing

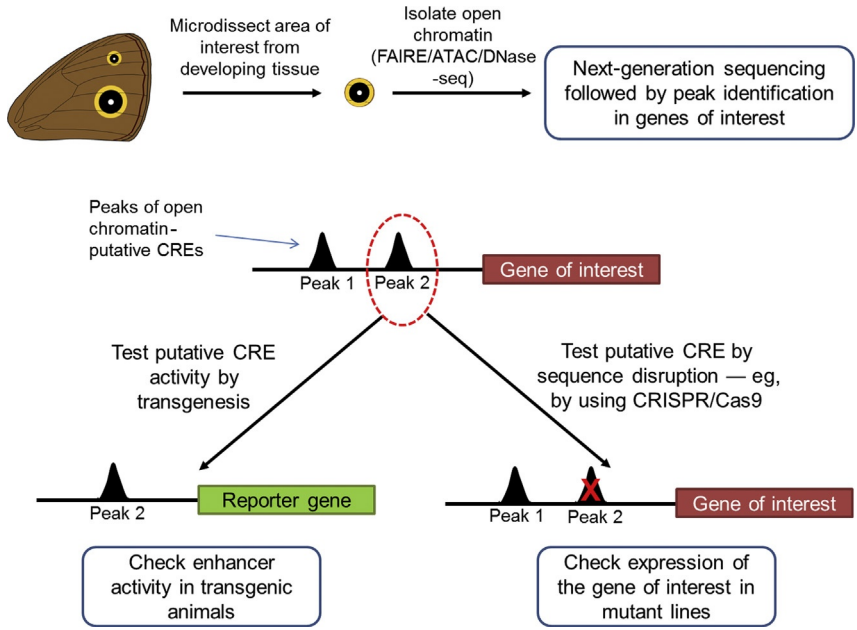

Fig. 5 A conceptual framework showing how new genomic tools that are used to isolate open chromatin regions may be employed to identify pleiotropic CREs and thus, used to identify GRNs that are coopted to give rise to novel traits. In this particular example, isolation of open chromatin from developing eyespots of butterfly wings may aid in the identification of pleiotropic CREs that drive gene expression in multiple developmental contexts. The multiple expression domains may highlight the type of ancestral GRN that may have given rise to eyespots.

the results (Meyer & Liu, 2014). DNase-seq also serves as a good technique to identify regulatory sequences; but the additional step of enzymatic digest can introduce variability across samples (Simon, Giresi, Davis, & Lieb, 2012). Some of the more straightforward and simpler techniques that have recently been used to successfully identify CREs are FAIRE (formaldehyde-assisted isolation of regulatory elements) and ATAC (assay for transposase accessible chromatin) followed by next-generation sequencing (Buenrostro, Giresi, Zaba, Chang, & Greenleaf, 2013; Giresi, Kim, McDaniell, Iyer, & Lieb, 2007). Both techniques enrich the samples for genome-wide open chromatin region, whose reads can be mapped back to the genomic regions to identify "peaks" of open chromatin. The underlying assumption behind the use of open chromatin as a proxy for a regulatory region is that the complexes of transcription factors regulating gene expression displace the nucleosomes creating a transient "open" state in the chromatin (Giresi & Lieb, 2009).

FAIRE relies on the crosslinking of genomic DNA to bound proteins using formaldehyde and a phase separation step where the DNA–protein complexes are extracted into an organic phase and the unbound DNA stays in the aqueous phase (Simon et al., 2012). Since open chromatin is relatively free of nucleosomes, the regulatory elements are enriched in the aqueous phase. Several recent studies have shown that FAIRE-seq results are consistent with DNase-seq and aggregate ChIP-seq results in human cell lines and *Drosophila* (Giresi & Lieb, 2009; McKay & Lieb, 2013; Song et al., 2011). Another advantage of using this technique is that it requires much smaller amounts of starting material compared to techniques like ChIP-seq, thereby simplifying the technical aspects where amounts of the starting material can be limiting (Simon et al., 2012). In a recent study, McKay and Lieb used FAIRE-seq to identify peaks of open chromatin along the genomic tracks of actively transcribed genes in *Drosophila* at different developmental time points (McKay & Lieb, 2013). The putative CREs (500–1 kb genomic fragments) were cross-validated by placing them upstream of a reporter gene, and inserting the construct back into the fly's genome. The reporter gene showed sharp spatiotemporal expression patterns depending on the developmental time points at which the open chromatin regions were identified.

The recently developed ATAC-seq relies on the ability of a modified Tn3 transposase (Illumina) to bind selectively to nucleosome-depleted regions (Buenrostro et al., 2013). Once bound, the transposase attaches adaptor oligos to the end of the fragmented DNA in a single step (Buenrostro et al., 2013). This technique requires even smaller amounts

of starting tissue compared to FAIRE and greatly simplifies the library preparation procedure since the genomic DNA is sheared and tagged with sequencing adaptors in a single step. These features make ATAC-seq a suitable technique to isolate regulatory sequences specifically from developing tissues that have relatively few cells and require microdissection.

Therefore, in principle, these new genomic tools can be used to identify CREs regulating the spatiotemporal expression of genes that are candidate internal genes in a suspected coopted network that is building a novel trait. Once specific CREs are identified, they can be further investigated for their pleiotropic function. Either by placing then upstream of a reporter gene and observing whether they drive the reporter in multiple developmental contexts, or by eliminating these CREs using genome editing tools (eg, CRISPR-Cas9) and testing whether pleiotropic effects of the deletion are observed. The identity of the different developmental contexts encoded by the same CRE will highlight the likely nature of the ancestral GRN reused to produce the novel trait. Some caveats of the CRE knockout experiments, however, include the presence of duplicate enhancers, also known as shadow enhancers (Cannavo et al., 2016; Hong, Hendrix, & Levine, 2008), which, if not also targeted by the same guide RNA, may prevent the detection of a phenotype. Furthermore, if the targeted CREs are indeed pleiotropic, then their loss might in certain cases impairs embryonic survival. This latter caveat, however, may be overcome by the typical mosaicism of a first generation CRISPR knockout.

9. SUMMARY AND FUTURE PROSPECTS

It is difficult to identify the actual mutations that cause the cooption of preexisting gene networks to novel locations in the body to function in the differentiation of novel complex traits. These mutations are expected to involve the cooption of top regulators of modular networks to novel developmental contexts. These top regulators are often difficult to identify because crosses between individuals with and without the trait often cannot be performed when the two species are distantly related. Internal and terminal network genes are more easily identified simply because they constitute the bulk of the network and can be picked up via RNA-seq approaches or candidate gene approaches. Because of their relative higher abundance in networks, however, researchers have been examining the evolution of the expression of these internal network genes as a proxy for trait evolution. In particular, they have found that evolution in the CREs of these genes

leads to novel gene expression patterns that correlate with the presence of the novel trait. This has led to the idea that evolution in the sequence of a preexisting CRE of a gene, regardless of the gene's position in a GRN, can coopt the gene to perform a new function. The argument we tried to make earlier is that novel complex traits can indeed appear via modification of preexisting CREs, but not via the modification of CREs of genes in the terminus of complex regulatory networks. Complex traits only appear when modifications are done to genes in higher regulatory positions of complex networks. Modifications to CREs of genes in the middle or terminus of GRNs may lead to trait loss, but not to trait origin.

Forward genetic screens can be used to identify mutations that cause the origin of novel traits, and novel genomic tools can be used to discover pleiotropic CREs belonging to genes in the middle of coopted GRNs. The discovery of these pleiotropic CREs can provide one of the most important insights about the origin of novel traits: the identity of the preexisting network or networks that were coopted to give rise to the novel trait. This information is a very illuminating piece of the puzzle of how novel complex traits originated and came to be. The identity of top regulators of these pleiotropic CREs, and the identity of the mutations that move these genes and associated networks to new locations can then be pursued as a separate and complementary endeavor.

ACKNOWLEDGMENTS

We thank Kathy Su, Virginie Orgogozo, and another anonymous reviewer for their comments and suggestions. This research was funded by the Ministry of Education of Singapore (MOE R-154-000-602-112 and MOE2014-T2-1-146) and by the National University of Singapore and Yale-NUS College.

REFERENCES

Arnoult, L., Su, K. F. Y., Manoel, D., Minervino, C., Magrina, J., Gompel, N., et al. (2013). Emergence and diversification of fly pigmentation through evolution of a gene regulatory module. *Science*, *339*, 1423–1426.

Bergstrom, C. T., & Dugatkin, L. A. (2012). *Evolution*. New York, NY: W.W. Norton & Company.

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., & Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods*, *10*, 1213–1218.

Cannavo, E., Khoueiry, P., Garfield, D. A., Geeleher, P., Zichner, T., Gustafson, E. H., et al. (2016). Shadow enhancers are pervasive features of developmental regulatory networks. *Current Biology*, *26*, 38–51.

Davidson, E. H., & Erwin, D. H. (2006). Gene regulatory networks and the evolution of animal body plans. *Science*, *311*, 796–800.

Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L., & Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*, *151*, 1531–1545.

Freitas, R., Zhang, G. J., & Cohn, M. J. (2006). Evidence that mechanisms of fin development evolved in the midline of early vertebrates. *Nature*, *442*, 1033–1037.

Futuyma, D. J. (2009). *Evolution* (2nd ed.). Sunderland, MA: Sinauer Associates.

Gao, F., & Davidson, E. H. (2008). Transfer of a large gene regulatory apparatus to a new developmental address in echinoid evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 6091–6096.

Gilbert, S. F., & Epel, D. (2009). *Ecological developmental biology: Integrating epigenetics, medicine, and evolution*. Sunderland, MA: Sinauer Associates.

Giresi, P. G., Kim, J., McDaniell, R. M., Iyer, V. R., & Lieb, J. D. (2007). FAIRE (formaldehyde-assisted isolation of regulatory elements) isolates active regulatory elements from human chromatin. *Genome Research*, *17*, 877–885.

Giresi, P. G., & Lieb, J. D. (2009). Isolation of active regulatory elements from eukaryotic chromatin using FAIRE (formaldehyde assisted isolation of regulatory elements). *Methods*, *48*, 233–239.

Glassford, W. J., Johnson, W. C., Dall, N. R., Smith, S. J., Liu, Y., Boll, W., et al. (2015). Co-option of an ancestral Hox-regulated network underlies a recently evolved morphological novelty. *Developmental Cell*, *34*, 520–531.

Gompel, N., Prud'homme, B., Wittkopp, P. J., Kassner, V. A., & Carroll, S. B. (2005). Chance caught on the wing: Cis-regulatory evolution and the origin of pigment patterns in Drosophila. *Nature*, *433*, 481–487.

Halder, G., Callaerts, P., & Gehring, W. J. (1995). Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila. *Science*, *267*, 1788–1792.

Hong, J. W., Hendrix, D. A., & Levine, M. S. (2008). Shadow enhancers as a source of evolutionary novelty. *Science*, *321*, 1314.

Jeong, S., Rebeiz, M., Andolfatto, P., Werner, T., True, J., & Carroll, S. B. (2008). The evolution of gene regulation underlies a morphological difference between two Drosophila sister species. *Cell*, *132*, 783–793.

Lemons, D., Fritzenwanker, J. H., Gerhart, J., Lowe, C. J., & McGinnis, W. (2010). Co-option of an anteroposterior head axis patterning system for proximodistal patterning of appendages in early bilaterian evolution. *Developmental Biology*, *344*, 358–362.

McGregor, A. P., Orgogozo, V., Delon, I., Zanet, J., Srinivasan, D. G., Payre, F., et al. (2007). Morphological evolution through multiple cis-regulatory mutations at a single gene. *Nature*, *448*, 587–590.

McKay, D. J., & Lieb, J. D. (2013). A common set of DNA regulatory elements shapes Drosophila appendages. *Developmental Cell*, *27*, 306–318.

Medved, V., Marden, J. H., Fescemyer, H. W., Der, J. P., Liu, J., Mahfooz, N., et al. (2015). Origin and diversification of wings: Insights from a neopteran insect. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 15946–15951.

Meyer, C. A., & Liu, X. S. (2014). Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. *Nature Reviews. Genetics*, *15*, 709–721.

Moczek, A. P., & Rose, D. J. (2009). Differential recruitment of limb patterning genes during development and diversification of beetle horns. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 8992–8997.

Monteiro, A. (2012). Gene regulatory networks reused to build novel traits. *Bioessays*, *34*, 181–186.

Monteiro, A. (2015). Origin, development, and evolution of butterfly eyespots. *Annual Reviews of Entomology*, *60*, 253–271.

Monteiro, A., Chen, B., Ramos, D. M., Oliver, J. C., Tong, X., Guo, M., et al. (2013). Distal-less regulates eyespot patterns and melanization in Bicyclus butterflies. *Journal of Experimental Zoology Part B*, *320*, 321–331.

Monteiro, A., & Podlaha, O. (2009). Wings, horns, and butterfly eyespots: How do complex traits evolve? *PLoS Biology*, *7*(2), e1000037.

Niwa, N., Akimoto-Kato, A., Niimi, T., Tojo, K., Machida, R., & Hayashi, S. (2010). Evolutionary origin of the insect wing via integration of two developmental modules. *Evolution & Development*, *12*, 168–176.

Peterson, T., & Mueller, G. B. (2013). What is evolutionary novelty? Process versus character based definitions. *Journal of Experimental Zoology Part B*, *320*, 345–350.

Prud'homme, B., Gompel, N., Rokas, A., Kassner, V. A., Williams, T. M., Yeh, S. D., et al. (2006). Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. *Nature*, *440*, 1050–1053.

Rebeiz, M., Jikomes, N., Kassner, V. A., & Carroll, S. B. (2011). Evolutionary origin of a novel gene expression pattern through co-option of the latent activities of existing regulatory sequences. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 10036–10043.

Reed, R. D., Papa, R., Martin, A., Hines, H. M., Counterman, B. A., Pardo-Diaz, C., et al. (2011). optix drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science*, *333*, 1137–1141.

Schlosser, G. (2004). The role of modules in development and evolution. In G. Schlosser & G. P. Wagner (Eds.), *Modularity in development and evolution* (pp. 519–582). Chicago: University of Chicago Press.

Simon, J. M., Giresi, P. G., Davis, I. J., & Lieb, J. D. (2012). Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA. *Nature Protocols*, *7*, 256–267.

Song, L., Zhang, Z., Grasfeder, L. L., Boyle, A. P., Giresi, P. G., Lee, B. K., et al. (2011). Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Research*, *21*, 1757–1767.

True, J. R., Edwards, K. A., Yamamoto, D., & Carroll, S. B. (1999). Drosophila wing melanin patterns form by vein-dependent elaboration of enzymatic prepatterns. *Current Biology*, *9*, 1382–1391.

Vlad, D., Kierzkowski, D., Rast, M. I., Vuolo, F., Dello Ioio, R., Galinha, C., et al. (2014). Leaf shape evolution through duplication, regulatory diversification, and loss of a homeobox gene. *Science*, *343*, 780–783.

Wagner, G. P., & Lynch, V. J. (2010). Evolutionary novelties. *Current Biology*, *20*, R48–R52.

Wallbank, R. W. R., Baxter, S. W., Pardo-Diaz, C., Hanly, J. J., Martin, S. H., Mallet, J., et al. (2016). Evolutionary novelty in a butterfly wing pattern through enhancer shuffling. *PLoS Biology*, *14*, e1002353.

Wittkopp, P. J., & Beldade, P. (2009). Development and evolution of insect pigmentation: Genetic mechanisms and the potential consequences of pleiotropy. *Seminars in Cell & Developmental Biology*, *20*, 65–71.