# Chapter 4
# Morphological Evolution Repeatedly Caused by Mutations in Signaling Ligand Genes

**Arnaud Martin and Virginie Courtier-Orgogozo**

**Abstract** What types of genetic changes underlie evolution? Secreted signaling molecules (*syn*. ligands) can induce cells to switch states and thus largely contribute to the emergence of complex forms in multicellular organisms. It has been proposed that morphological evolution should preferentially involve changes in developmental toolkit genes such as signaling pathway components or transcription factors. However, this hypothesis has never been formally confronted to the bulk of accumulated experimental evidence. Here we examine the importance of ligand-coding genes for morphological evolution in animals. We use Gephebase (http://www.gephebase.org), a database of genotype-phenotype relationships for evolutionary changes, and survey the genetic studies that mapped signaling genes as causative loci of morphological variation. To date, 19 signaling genes represent 20% of the cases where an animal morphological change has been mapped to a gene (80/391). This includes the signaling gene *Agouti*, which harbors multiple cis-regulatory alleles linked to color variation in vertebrates, contrasting with the effects of coding variation in its target, the melanocortin receptor MC1R. In sticklebacks, genetic mapping approaches have identified 4 signaling genes out of 14 loci associated with lake adaptations. Finally, in butterflies, a total of 18 allelic variants of the *WntA* Wnt-family ligand cause color pattern adaptations related to wing mimicry, both within and between species. We discuss possible hypotheses explaining these cases of natural replication (genetic parallelism) and conclude that signaling ligand loci are an important source of sequence variation underlying morphological change in nature.

**Keywords** Signaling ligands • Genotype-phenotype relationships • Mutational target • cis-Regulatory alleles • Gephebase

A. Martin (✉)
Department of Biological Sciences, The George Washington University,
Washington, DC, USA
e-mail: arnaud@gwu.edu

V. Courtier-Orgogozo
Institut Jacques Monod, CNRS, UMR 7592, Université Paris Diderot, Paris, France

A key aim of developmental biology is to describe the molecular mechanisms underlying pattern formation, i.e., how gene expression patterns are established and how cell differentiation is orchestrated over time. Since the discovery of embryonic induction, which revealed that secreted molecules are capable of instructing and organizing cells in surrounding tissues (Waddington 1940; Spemann and Mangold 2001), cell-cell signaling has become a sine qua non mechanism of pattern formation in many (if not most) developmental systems (Meyerowitz 2002; Rogers and Schier 2011; Urdy 2012; Kicheva and Briscoe 2015). Experimental manipulations of extracellular signals can impact tissue patterning at a distance (Salazar-Ciudad 2006; Nahmad Bensusan 2011; Perrimon et al. 2012; Urdy et al. 2016). It follows that to understand how spatial information is deployed in differentiating tissues, it is critical to characterize the signals that mediate intercellular communication. A handful of genes coding extracellular proteins that act as signaling molecules between neighboring cells have been identified in animals (Nichols et al. 2006; Rokas 2008a; Perrimon et al. 2012): Wnt, TGF-beta, Hedgehog, Notch, EGF, RTK ligands, and TNFs, among other families. These signaling ligands are widely conserved and show highly regulated expression patterns (Salvador-Martínez and Salazar-Ciudad 2015).

In the 2000s it was proposed that the construction of multicellular organisms relies on a small set of conserved genes, referred to as the developmental genetic toolkit (DGT), which comprises a few hundred genes from a few dozen gene families involved in two major processes: cell differentiation and cell-cell communication (Carroll et al. 2005; Floyd and Bowman 2007; Rokas 2008b; Erwin 2009). On the other side, genes that are not part of the DGT were attached to vital routine functions such as metabolism, protein synthesis, or cell division. According to the DGT view, spatial information emerges from an interplay between genetic factors involved in signal transduction and transcriptional control. An inevitable consequence is that morphological evolution should be based, to a large extent, on reusing these toolkit components, and it follows that mutations in the DGT genes themselves should cause evolution of form (Carroll et al. 2005; Carroll 2008). Such proposition was formulated at the beginning of the twenty-first century, while few genes underlying morphological evolution had been identified – less than 50 cases in 2001 (Martin and Orgogozo 2013). As of today, the hypothesis that animal morphological evolution is mainly caused by mutations in DGT genes can now be tested further based on micro-evo-devo studies (Nunes et al. 2013) and the analyses of genotype-phenotype variation in nature (Orgogozo et al. 2015; Stern 2011). Here we investigate one aspect of the DGT view, the importance of genes encoding secreted signaling proteins in driving morphological evolution. We examine whether ligand-coding genes are preferential targets for the generation of morphological evolution. In addition, we confront existing data to predictions that the corresponding allelic variation should be (1) potentially adaptive (Barrett and Hoekstra 2011; Pardo-Diaz et al. 2015), (2) replicated over various phylogenetic levels (Gompel and Prud'homme 2009; Kopp 2009; Martin and Orgogozo 2013), and (3) cis-regulatory rather than coding (Prud'homme et al. 2007; Carroll 2008; Stern and Orgogozo 2008; Liao et al. 2010).

## 4.1  Gephebase: The Database of Genotype-Phenotype Variations

Experimental studies based on the manipulation of gene function in the laboratory – for instance, based on reverse genetics or on a mutant screen followed by forward genetics mapping – describe the overall architecture of the genotype-phenotype map in a given organism. However, the genetic causes of evolutionary change in nature do not necessarily equate to the mutations studied in the laboratory: evolutionary-relevant mutations may represent a particular subset of all possible mutations. To identify the genetic causes of natural differences between individuals, populations, and species, one can perform forward genetics studies that compare two naturally occurring phenotypic states – in general, using linkage mapping of quantitative trait loci or Mendelian genes or association mapping (Stern 2000). The so-called "loci of evolution" or "quantitative trait gene (QTG)" studies identify pairs of alleles linked to a specific phenotypic difference (Orgogozo et al. 2015), for instance between an ancestral and a derived state. These loci are typically genomic targets of selection when the variation is of adaptive or domesticating potential. Due to experimental limitations, the dataset is biased toward large-effect loci and thus misses a large fraction of what constitutes the total genetic template of evolution (Rockman 2012). Nevertheless, we think that it is crucial to gather the findings of this research program under the banner of a resource that would integrate, for comparative and meta-analytical purposes, our growing knowledge of genotype-phenotype relationships. To facilitate the curation and analysis of the relevant literature [see (Stern and Orgogozo 2008; Streisfeld and Rausher 2011; Martin and Orgogozo 2013) for previous examples], we have created Gephebase (http://www.gephebase.org), a database of genotype-phenotype relationships underlying natural and domesticated variation across Eukaryotes. Here, we use Gephebase to reflect on the importance of signaling ligand genes for morphological evolution in animals.

## 4.2  Method: Construction of Gephebase and Identification of Signaling Genes

Gephebase is a quality-controlled, manually curated database of published associations between genes and phenotypes in Eukaryotes – containing a total of 1400 entries as of December 31, 2016. For now, genes responsible for human disease and for aberrant mutant phenotypes in laboratory model organisms are excluded and can be found in other databases (OMIM, OMIA, FlyBase, etc.). QTL mapping studies whose resolution did not reach the level of the nucleotide or of the transcriptional unit are also excluded. In Gephebase, each genotype-phenotype association is attributed to only one type of experimental evidence among three possibilities: "association mapping," "linkage mapping," or "candidate gene." This

choice is made by Gephebase curators based on the best evidence available for a given genotype-phenotype relationship. Gene-to-phenotype associations identified by linkage mapping with resolutions below 500 kb have priority in the dataset (see Supplementary Materials in Martin and Orgogozo 2013). Association mapping studies are included based on individual judgment, with a strong bias toward SNP-to-phenotype associations that have been confirmed in reverse genetic studies. In other words, Gephebase intends to be more stringent than a compilation of statistically significant SNPs, and attempts to select studies where a given genotype-phenotype association is relatively well supported or understood.

Gephebase presents itself as a collection of entries, where each entry corresponds to an allelic difference at a given gene, either between two closely related species or between two individuals, its associated phenotypic change, and the relevant publications. As of today, the database contains a total of 391 entries related to animal morphological changes: 174 for domesticated or artificially selected traits, 172 for intraspecific trait variations, and 45 for interspecific changes (Table S1, available at http://virginiecourtier.wordpress.com/publications/. We identify 80 cases of natural morphological evolution and domestication in animals (out of 391) that involve 21 different ligand genes (Table 4.1; Table S2, available at http://virginiecourtier.wordpress.com/publications/).

To estimate the proportion of genes encoding signaling ligands in genomes (Table 4.2), we used the BioMart portal from Ensembl (Smedley et al. 2015). All the genes, which have both the following Gene Ontology (GO) annotations, "receptor binding" (Molecular Function, GO:0005102) and "extracellular region" (Cellular Component, GO:0005576), were considered as ligand genes. To count the number of genes with two GO annotations, we used BioMart to extract text files containing Ensembl Gene ID for each GO and each species. We then counted the number of genes having both GO in each species with the following Linux command: *comm -1 -2 <(sort human-GO0005102.txt) <(sort human-GO0005576.txt) | head -n -1 | wc -l* (note that the title line had to be excluded from the count).

---

**Box 4.1: Definitions**

*Admixture Mapping*: a method capitalizing on the current gene flow between two or more previously isolated populations to associate genetic loci to phenotypic traits. Admixture mapping is a form of association mapping.

*Association Mapping*: a forward genetics method for gephe identification based on a genome-wide statistical association between genetic variants and phenotypic traits, generally in a large cohort of unrelated individuals.

*Candidate Gene Approach*: a reverse genetics method that tests if a locus defined a priori, based on our current biological knowledge, underlies variation in a phenotype of interest. *Example*: opsin photoreceptor genes are typical candidate genes for differences in color vision.

(continued)

**Box 4.1** (continued)

*Forward genetics*: set of methods used to identify the genetic cause(s) of a given phenotypic trait ("from the phenotype to genes").

*Genetic hotspot*: a group of orthologous loci that have been associated multiple times to phenotypic variation due to independent mutational events in each lineage (Martin and Orgogozo 2013).

*Gephe* (neologism for genotype-phenotype relationship; pronounced *jay-fee*): an abstract entity composed of three elements: a variation at a genetic locus (two alleles), its associated phenotypic change (two distinct phenotypic states, e.g., an ancestral and a derived state), and their relationship (Orgogozo et al. 2015). A gephe is usually defined for a given genetic background and environment.

*Haplotype*: a set of closely linked alleles found on the same chromosome, which is inherited as a single piece.

*Heterotopy*: change that occurred during evolution in the location of a particular molecular event within the developing organism.

*Linkage Mapping*: a forward genetics method for gephe identification based on chromosome shuffling and crossing-overs, using the progeny of a hybrid cross. This includes the mapping of quantitative trait loci (QTL) and Mendelian loci.

*Mendelian Gene*: a segregating genetic unit which is detected through phenotypic differences associated with different alleles at the same locus (Orgogozo et al. 2016).

*Morphospace*: an abstract representation of all possible morphologies and shapes of an organism.

*Orthologous Loci*: pieces of DNA that share ancestry because of a speciation event and that are thus found in different species.

*Parallel Evolution*: here defined as independent repeated sequence variation at a same locus, underlying variation in a similar phenotypic trait (Stern 2013). For other definitions, see (Scotland 2011).

*Phenologue*: a similar phenotype caused by a conserved genetic mechanism in distant lineages (McGary et al. 2010; Lehner 2013). Used here as the phenotypic counterpart of a gephe involving several cases of parallel evolution.

*Quantitative Trait Locus*: a portion of DNA (the locus) that is associated with variation in a quantitative phenotypic trait.

*Reverse Genetics*: set of methods used to alter a given gene in order to characterize its function ("from genes to phenotypes").

**Table 4.1** Twenty-one ligand genes with known gephe variations identified in Gephebase – 19 of which related to cases of morphological evolution (accessed December 31, 2016)

| Ligand gene | Trait variation | Nb. of natural gephes (intra-/interspecific) in | Nb. of artificially selected gephes (domesticated) in | Comments | Reference PubMed ID |
|---|---|---|---|---|---|
| Agouti (ASIP) | Pigmentation | Mammals/birds (11) | Mammals/birds (14) | See Fig. 4.1 and Gephebase | See Gephebase |
| BMP2 | Fertility (females) + comb morphology (males) | – | Chicken (1) | Pleiotropic and joint effect with HAO1 | 22956912 24655072 |
| BMP3 | Craniofacial skeleton | – | Dog (1) | Uncertain (possible role of PRKG2) | 22876193 |
| BMP6 | Tooth number | Stickleback fish (1) | – | cis-Regulatory allele | 25205810 26062935 25732776 |
| BMP15 | Fertility (not a morphological trait) | – | Sheep (4) | Coding alleles | 10888873 23637641 |
| CBD103 | Pigmentation | Wolf/coyote (1) | Dog (same allele) | Coding allele | 19197024 17947548 |
| EDA | Armor plates + schooling behavior | Stickleback fish (1) | – | cis-Regulatory allele | 15790847 22481358 25629660 |
| EDN3 | Pigmentation | – | Chicken (1) | Large duplication | 22216010 25344733 25741364 |
| FGF5 | Hair length | – | Cat (1), dog (1), donkey (2) | Coding alleles | See Gephebase |
| GDF5 | Body size | Human (1) | – | Allele unknown (SNP association) | 18193045 |
| GDF6 | Skeletal traits | Stickleback fish (1); human (1) | – | cis-Regulatory alleles | 26774823 |

| Gene | Trait | | | Coding alleles | |
|---|---|---|---|---|---|
| GDF9 | Fertility (not a morphological trait) | – | Sheep (**2**) | Coding alleles | 19713444 20528846 |
| IGF1 | Body size | – | Dog (**1**) | cis-Regulatory allele | 17412960 |
| IGF2 | Muscle and fat content | – | Pig (**1**) | cis-Regulatory allele | 14574411 |
| KITLG | Pigmentation | Stickleback fish (**1**); human (**2**) | Cattle (**1**) | cis-Regulatory alleles | See text |
| Myostatin (GDF8) | Muscular growth | – | Mammals (**14**) | Coding (12) + cis-reg. (2) alleles | See Gephebase |
| Rspo2 | Hair length | – | Dog (**1**) | cis-Regulatory allele | 19713490 |
| scabrous | Bristle number | Fruit fly (**2**) | – | Distinct alleles affect thorax and abdomen | 7992053 |
| upd-like | Wing size | Jewel wasp (**1**) | – | cis-Regulatory allele QTL fractionation | 22363002 |
| wingless (Wnt1) | Pigment patterns (wing, larva) | Fruit fly (**1**) | Silkworm (**1**) | cis-Regulatory complex alleles | 23673642 26034272 |
| WntA | Pigment patterns (wing) | Butterflies (10) | – | cis-Regulatory alleles | See text |

**Table 4.2** Proportion of signaling ligand-encoding genes in the genome of several model species

| | Homo sapiens (GRCh38.p7) | Mus musculus (GRCm38.p4) | G. aculeatus (BROADS1) | D. melanogaster (BDGP6) | C. elegans (WBcel235) |
|---|---|---|---|---|---|
| Nb. of protein-coding genes | 22,285 | 22,222 | 20,787 | 18 | 20,362 |
| Nb. of genes with GO (molecular function) = "receptor binding" | 1592 | 1435 | 247 | 200 | 198 |
| Nb. of genes with GO (molecular function) = "extracellular region" | 4814 | 4225 | 334 | 1016 | 562 |
| Nb. of genes with GO (molecular function) = "extracellular region" and "receptor binding" | 930 | 771 | 115 | 105 | 86 |
| Proportion of signaling ligand genes | 4.17% | 3.47% | 0.55% | 0.75% | 0.42% |

Ligand-encoding genes are defined as the protein-coding genes associated with both "receptor binding" and "extracellular region" Gene Ontology terms

## 4.3 A Few Select Genes for Body-Wide Switches in Melanin Production in Tetrapods

Among 294 Gephebase morphology entries for tetrapods (Gephebase search term "Tetrapoda," including mammals and reptiles sensu *largo*), 206 genotype-phenotype relationships relate to pigment variation, including 193 entries identifying components of the melanocyte differentiation pathway. Both sampling and ascertainment biases explain this unusual enrichment. First, pigmentation shows a bulk of variation accessible to breeders and natural selection altogether (Protas and Patel 2008; Linderholm and Larson 2013). In combination with the fact that coloration variation often involves few genes, these features have made pigmentation a favorite target for exploring genotype-phenotype relationships (sampling bias). Second, there is predictability in the genetic basis of melanin pigment variation, as illustrated by the fact that the melanocortin 1 receptor (MC1R), a major regulator of melanocyte activation, is the most represented gene in Gephebase with 84 entries (6% of all 1400 entries). Interestingly, 80% of *MC1R* gephes (67/84) were identified by a candidate gene approach. This pattern illustrates well a latent ascertainment bias in the study of vertebrate pigment variation: when interested in the genetic basis of a color variation involving shifts in melanin types (mammalian coat, bird plumage, etc.), it has become a knee-jerk reflex for biologists to look for amino acid changes in MC1R, in particular in domains that had been functionally characterized. As a matter of fact, all of the 67 MC1R gephes based on a candidate gene approach involve mutations affecting the gene-coding region. Thus, both the phenotypic diversity of vertebrate pigmentation traits and their simple genetic basis explain the overrepresentation of MC1R to a large extent. This said, the fact that the remaining 20% of *MC1R* entries were identified by linkage or association mapping validates the idea that *MC1R* is a bona fide driver of color variation in vertebrates. As an explanation for this trend, it is likely that the MC1R protein hosts tuning sites that can modulate pigmentation without affecting other traits and that its mutations can show a dominant effect prone to a rapid adaptive spread (Mundy 2005; Kopp 2009; Kronforst et al. 2012; Reissmann and Ludwig 2013; Wolf Horrell et al. 2016). Other components of the melanocyte activation cascade also form gephes involved in natural and artificial selection of coloration traits (Fig. 4.1). This includes downstream targets of MC1R signal transduction such as the transcription factor gene *MITF* and the melanogenic genes *TYR*, *TYRP1*, and *Pmel17*, all involved in the biogenesis of eumelanosomes.

Upstream of MC1R, two signaling molecules that interact with receptor function are known as allelic sources of color variation in vertebrates. In particular, the antagonist ligand *Agouti/ASIP* is a genetic hotspot for pigment variation with a total of 28 entries in Gephebase. This includes numerous cases where this gene was identified by linkage or association mapping, both in natural and domesticated contexts (Fig. 4.1a–c), making *Agouti* one of the most commonly mapped genes in our dataset. Coding alleles of *Agouti* are recessive loss-of-function mutations
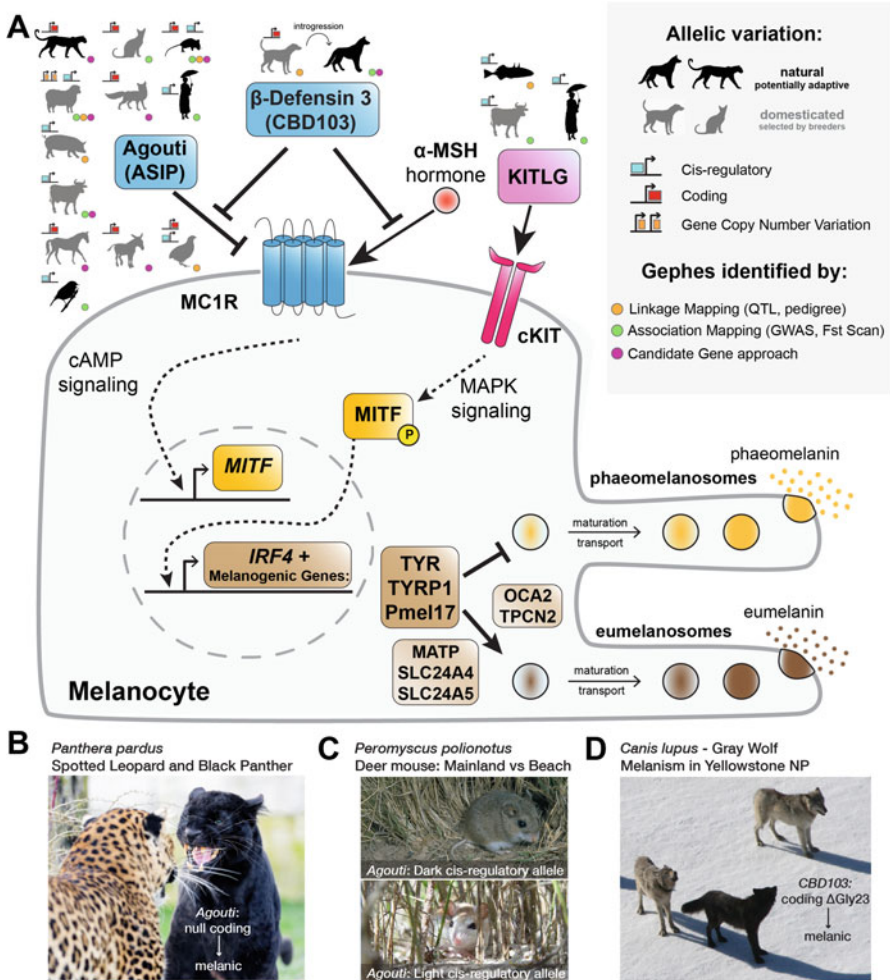
**Fig. 4.1** Alleles of secreted ligands associated to pigment variation in vertebrates. (**a**) The MC1R and cKIT signaling pathways each activate a signal transduction regulatory cascade converging on the MITF transcription factor that modulates the expression of melanogenic genes and ultimately activates the maturation and transport of dark eumelanin in melanosomes. Agouti and β-defensin3 are secreted extracellular modulators of MC1R, and KITLG is the agonist ligand of cKIT. Allelic variation at these three genes is associated to pigment variation in vertebrates. (**b**) Black panthers are leopards that carry a null mutation in *Agouti*. (**c**) Adaptive pigment variation in deer mice (*Peromyscus* spp.) has repeatedly involved sequence modifications at the *Agouti* locus. For instance, distinct populations of *P. polionotus* adapted to dark (mainland Florida; *top panel*) and light (coastal Florida; *bottom panel*) color backgrounds via cis-regulatory variants that modulate *Agouti* skin expression. (**d**) Black wolves can be seen at increasing frequencies in packs of the Yellowstone National Park (USA). The melanic allele corresponds to a single amino acid deletion, which was originally selected in domestic dogs and later introgressed in wild in North American wolves and coyotes by hybridization. *Photo credits* – (**b**) Emmanuel Keller (License CC BY-ND 2.0), (**c**) Roger Barbour (License CC BY-ND 2.0), (**d**) Doug Smith (Public Domain)

resulting in melanic phenotypes. This contrasts with the melanic gain-of-function coding alleles of MC1R which are dominant, a difference in allelic effects that is used to infer the genetic basis of melanism (Eizirik et al. 2003). The *Agouti* ligand inhibits the basal activation of the MC1R pathway. In an *Agouti*-null context, MC1R is hyper-activated by its active ligand, the pituitary melanocortin hormone α-MSH, which triggers a melanocyte regulatory cascade that culminates with eumelanin production. It has been proposed that wild-type Agouti can become an agonist of MC1R melanic variants (McRobie et al. 2014), suggesting that certain gain-of-function MC1R alleles reverse the responsiveness of the receptor to the Agouti ligand itself. In addition to Agouti, the β-defensin 3/CBD103 peptide is secreted by skin epithelia, strongly binds to MC1R, and was shown to be responsible for melanism in dogs (Candille et al. 2007). In certain melanic dog breeds, one amino acid deletion in β-defensin 3/CBD103 results in dominant melanism, possibly by blocking the inhibitory activity of Agouti or by losing its blocking of α-MSH stimulatory binding (Nix et al. 2013). Of note, the CBD103$^{\Delta G23}$ melanic allele is revealing a complex history that blurs the boundary between wild and domesticated. First, based on ancient DNA studies, it probably originated through domestication from a possible wolflike gene pool as early as 10,000 years ago (Ollivier et al. 2013), introgressing into modern dog breeds. Second, it propagated back in the wild, resulting in relatively recent segregation of melanic phenotypes in North American gray wolves, North American coyotes, and Italian gray wolves (Anderson et al. 2009). The melanic allele shows signatures of positive selection, but it remains unclear if this is due to a fitness effect of the melanic coat or, alternatively, to the antimicrobial properties of β-defensin 3. A few other cases of organism-wide color changes have been found to be positively selected (Vignieri et al. 2010; Barrett and Hoekstra 2011; Laurent et al. 2016).

In conclusion, mutations in MC1R and Agouti account for 54% (112/206) of the gephes dealing with tetrapod pigmentation variation in our current dataset. Such an overrepresentation cannot be explained by experimental bias alone and suggests that MC1R and Agouti are preferential targets for pigmentation evolution in tetrapods.

## 4.4  cis-Regulatory Evolution Drives Regional Specific Color Shifts

While ligand- and receptor-coding changes likely modulate the strength of signaling, and, thus, pigment synthesis in melanocytes, such changes are likely to affect all the body regions where these genes are expressed. In contrast, region-specific changes in coat, skin, or plumage coloration are more likely to involve cis-regulatory mutations. In a previous meta-analysis of the gephe literature, it was established empirically that localized morphological changes almost always involve cis-regulatory rather than coding variation (Stern and Orgogozo 2008).

*Agouti* is a hotspot of cis-regulatory evolution for pigment pattern modification and provides one of the most spectacular examples of QTL fractionation. Deer mice display extensive pigment variation matching the color of their environment (Manceau et al. 2010). Fine mapping of this variation revealed that not only the *Agouti* locus is the major driver of pigment variation (Manceau et al. 2011) but also this genetic region decomposes itself into multiple noncoding sub-loci, each tightly associated with parts of the total phenotype (Linnen et al. 2013). Various regulatory elements are involved in directing the expression of three alternative isoforms into different body regions (Mallarino et al. 2016). Each adaptive allele is a complex haplotype that is inherited as a package that underwent multiple local changes. This is of major importance to understand how small leaps in the morphospace occur, as it illustrates the principle that genetic hotspots, in addition to providing a somewhat predictable basis for phenotypic evolution between species, can also accumulate mutations that collectively result in large-effect variation within a single lineage (Stam and Laurie 1996; McGregor et al. 2007; Rebeiz et al. 2011; Martin and Orgogozo 2013; Linnen et al. 2013; Noon et al. 2016).

Thus, the studies of vertebrate pigment variation suggest that a receptor (*MC1R*) and its inverse agonist (*Agouti*/*ASIP*) are key regulators of melanocyte differentiation, driving adaptive variation in natural contexts as well as novel color features available to farmers and breeders. Coding evolution in either component results in body-wide color shifts, while cis-regulatory evolution of *Agouti*, by tuning the spatial deployment of an MC1R switch-off, permits subtle changes in morphology. The *Agouti*/*MC1R* axis is not a typical developmental pathway and plays little role during ontogenesis (e.g., see Gene Ontology annotations in Gephebase). In contrast, the *endothelin-3* ligand/*endothelin-receptor B* (EDN3/EDNRB) signaling axis has pleiotropic roles in the differentiation and migration of neural crest cells, and mutations in both EDN3 and EDNRB have been found to cause pigmentation changes in domesticated chicken, cattle, and horse (Santschi et al. 1998; Dorshorst et al. 2011; Qanbari et al. 2014). So far, only domesticated alleles of EDN3/EDNRB that may be under unrealistic selective regimes have been mapped. Thus, while it represents perhaps a genuine DGT component, it remains ambiguous if endothelin pathway genes can be a mutational target of evolution in a natural context. To truly assess the role of signaling ligand genes in morphological evolution, it is useful to focus on radiating lineages that allow a trait-by-trait dissection by forward genetics (i.e., taking advantage of natural variation between closely related lineages – populations and sister species) and, sometimes, natural experiments of replicated evolution (Kopp 2009; Powell and Mariscal 2015). In the next sections, we focus primarily on stickleback fishes and *Heliconius* butterflies, for which numerous linkage mapping efforts have been uncovering the genetic basis of several morphological adaptations.

## 4.5  Recent Stickleback Fish Adaptations Repeatedly Recruited Ligand Alleles

Three-spined sticklebacks (*Gasterosteus aculeatus*) are a species of marine fishes that repeatedly colonized freshwater environments following the retreat of the Pleistocene glaciers. Adapting to these novel niches involved numerous morphological, physiological, and behavioral modifications all available to genetic dissection by QTL mapping and population scans. Among the 14 gephes that have been mapped in sticklebacks (*Pitx1*, *TSHBeta2*, *KCNH4*, *KITLG*, *EDA*, *GDF6*, *BMP6*, *PRKCD*, *SOD3*, *KCNH4*, *ATP6V0A1*, *ATP1A1*, *Mucin*, *IGK*), 4 involve a secreted ligand gene. Analysis of well-annotated genomes indicates that secreted ligand genes represent less than 5% of the total number of genes within an animal genome (Table 4.2). The proportion of ligand gephes in sticklebacks (28%) is thus higher than expected with the null hypothesis that mutations responsible for phenotypic evolution occur randomly at any gene within a genome (chi$^2$ test: chi$^2 > 20$; $p < 10^{-5}$).

A single large-effect locus was identified as driving melanin pigment reduction in freshwater populations (Fig. 4.2a). Contrary to expectations, this trait mapped neither to the MC1R pathway nor at its downstream targets, but at the *Kit-ligand* (*KITLG*) locus (Miller et al. 2007; Jones et al. 2012), which encode the secreted signaling component of a parallel pathway (Fig. 4.1). KITLG is the ligand of the KIT receptor, which triggers a MAPK tyrosine kinase transduction cascade that modulates the differentiation and activity of melanocytes (Wehrle-Haller 2003). While the *KIT* receptor has been identified in a total of 17 color-related gephes, it is only linked to domesticated alleles in the cattle, pig, horse, donkey, domesticated fox, and domestic cat (see Advanced Search "Gene name and synonyms" = "KIT" at www.gephebase.org for a complete list). In contrast, cis-regulatory alleles of *KITLG* have been shown to underlie natural pigment variation not only in stickleback fishes but also in humans (Miller et al. 2007; Guenther et al. 2014). An Ala193Asp mutation in *KITLG* has also been shown to cause piebald coat color phenotypes in cattle breeds (Seitz et al. 1999; Qanbari et al. 2014). Of note, cis-regulatory *KITLG* variation may provide tissue-specific effects that limit its potential deleterious pleiotropic effects on cancer risks, as observed in other variant forms of this locus in humans (Karyadi et al. 2013; Litchfield et al. 2016).

Another locus, encoding the bone morphogenetic protein 6 (*BMP6*) ligand, was found to cause tooth gain in freshwater stickleback population (Cleves et al. 2014; Erickson et al. 2015) (Fig. 4.2b). The causal change is cis-regulatory and downregulates *BMP6* expression, late during oral development (see Cleves et al. 2014 correction). Surprisingly, genetic mapping of a second freshwater population revealed that another genomic locus has driven a similar phenotypic output (Ellis
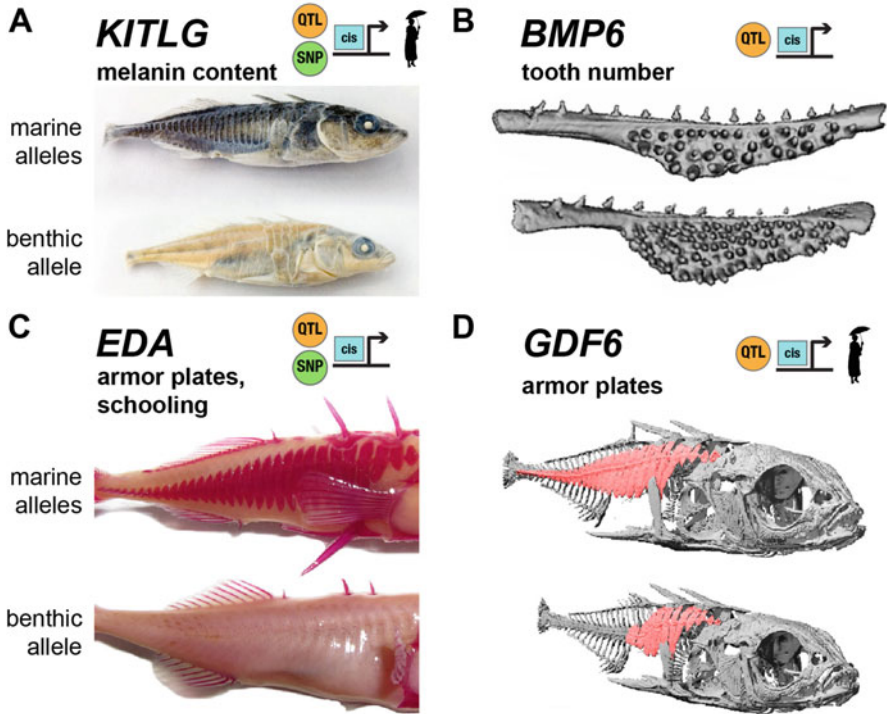
**Fig. 4.2** Secreted ligand loci involved in marine-to-lake adaptations in sticklebacks. (**a**) A *KITLG* cis-regulatory variant causes reduced melanization in lake populations (*bottom*) compared to marine alleles (*top*). (**b**) MicroCT images of the tooth plates of a marine vs. a lake-adapted ecotype. The freshwater cis-regulatory *BMP6*-derived allele causes increased tooth area and density. (**c**–**d**) Armor plates are lateral bony structures, here stained by Alizarin Red (**c**) and false-colored in MicroCT rendering (**d**, *pink*), which were repeatedly reduced or lost in freshwater populations. cis-Regulatory alleles of *EDA* and GDF6 cause distinct effects on plate distribution, number, and size (*Photo credits* – (**a**) Frank Chan and David Kingsley, (**b**) Craig Miller and David Kingsley, (**c**) Nicholas Ellis and Craig Miller, (**d**) Catherine Guenther, Vahan Indjeian, and David Kingsley)

et al. 2015). BMP ligands belong to the TGF-β family, are shared by all bilaterian animals, and play important roles for the regulation of development (De Robertis 2008). Compilation of current data suggests that mutations in TGF-β family genes are often involved in the tinkering of reproductive and skeletal traits during evolution and domestication. Several BMP alleles have been associated to increased fertility in domestic sheeps (*BMP15* and its paralog *GDF9*) (Monestier et al. 2014) and to fecundity and bone allocation in chicken (*BMP2*) (Johnsson et al. 2012). Genetic studies of craniofacial diversity mapped a QTL interval containing the *BMP4* gene in cichlid fishes (Albertson et al. 2005) and found a strong association between a single amino acid change in BMP3 and brachycephalic (short-skulled) dog breeds (Schoenebeck et al. 2012).

Body armor loss, via the reduction of lateral bony plates, has been a recurring adaptation to freshwater in sticklebacks. Two major loci have been characterized. The tumor necrosis factor superfamily gene *Ectodysplasin A* (*EDA*) harbors cis-regulatory variation existing at low frequency in the marine population that has been repeatedly recruited in continental populations to drive plate number reduction (Colosimo et al. 2005; Jones et al. 2012; O'Brown et al. 2015). The same locus also triggers a change in schooling behavior, as fishes from lake habitats have lost the ability to precisely align their body axis when swimming in a group, an effect that is reversed by transgenic overexpression of EDA (Greenwood et al. 2016). In addition, a combination of QTL mapping and genome scan has identified a freshwater-specific allele at the *growth/differentiation factor 6* (*GDF6*) locus, which results in a gain of expression of that gene in the developing epithelium and, ultimately, in a reduction of lateral plate size (Indjeian et al. 2016). Like for *KITLG*, this case also opened a window into human evolution as it was found that a *GDF6* hindlimb-specific enhancer was lost in the human lineage, with skeletal modifications obtained in mice that suggest a potential role in the evolution of bipedalism (Indjeian et al. 2016).

Forward genetics efforts in sticklebacks thus show that ligand genes belonging to classical developmental pathways are an important source of morphological variation of adaptive relevance. Noticeably, all the stickleback gephes described here are cis-regulatory, in accordance with the prediction that tinkering of developmental genes is more likely to involve cis-regulatory changes than coding mutations (Carroll 2008; Stern and Orgogozo 2008). Next, we focus on how accumulated changes in signaling ligand loci have enlarged the landscape of possible morphologies in insect wings.

## 4.6   The Wnt Beneath My Wings

There are few case studies that characterize adaptive variation for a same set of traits both within and between species. Butterflies of the *Heliconius* genus provide a rich phylogenetic template for such micro-evo-devo studies (Papa et al. 2008; Supple et al. 2014; Kronforst and Papa 2015; Merrill et al. 2015). They display a range of highly variable wing color pattern phenotypes involved in Müllerian mimicry (the collaborative display of similar morphologies to predators from multiple unpalatable species) and sexual selection that are amenable to hybrid crosses followed by linkage mapping. In addition, their natural hybrid zones form a system of choice for high-resolution admixture mapping, looking for SNP-phenotype associations and the smoking guns of selection that are the handful of Mendelian loci that keep adjacent populations phenotypically distinct in the face of constant gene flow and recombination. The Wnt-family signaling ligand WntA has emerged as a key genetic driver of wing pattern evolution in butterflies. Originally discovered as a Mendelian locus responsible for discrete shifts in pattern shapes in the *Heliconius erato* mimetic radiation, this gene shows striking
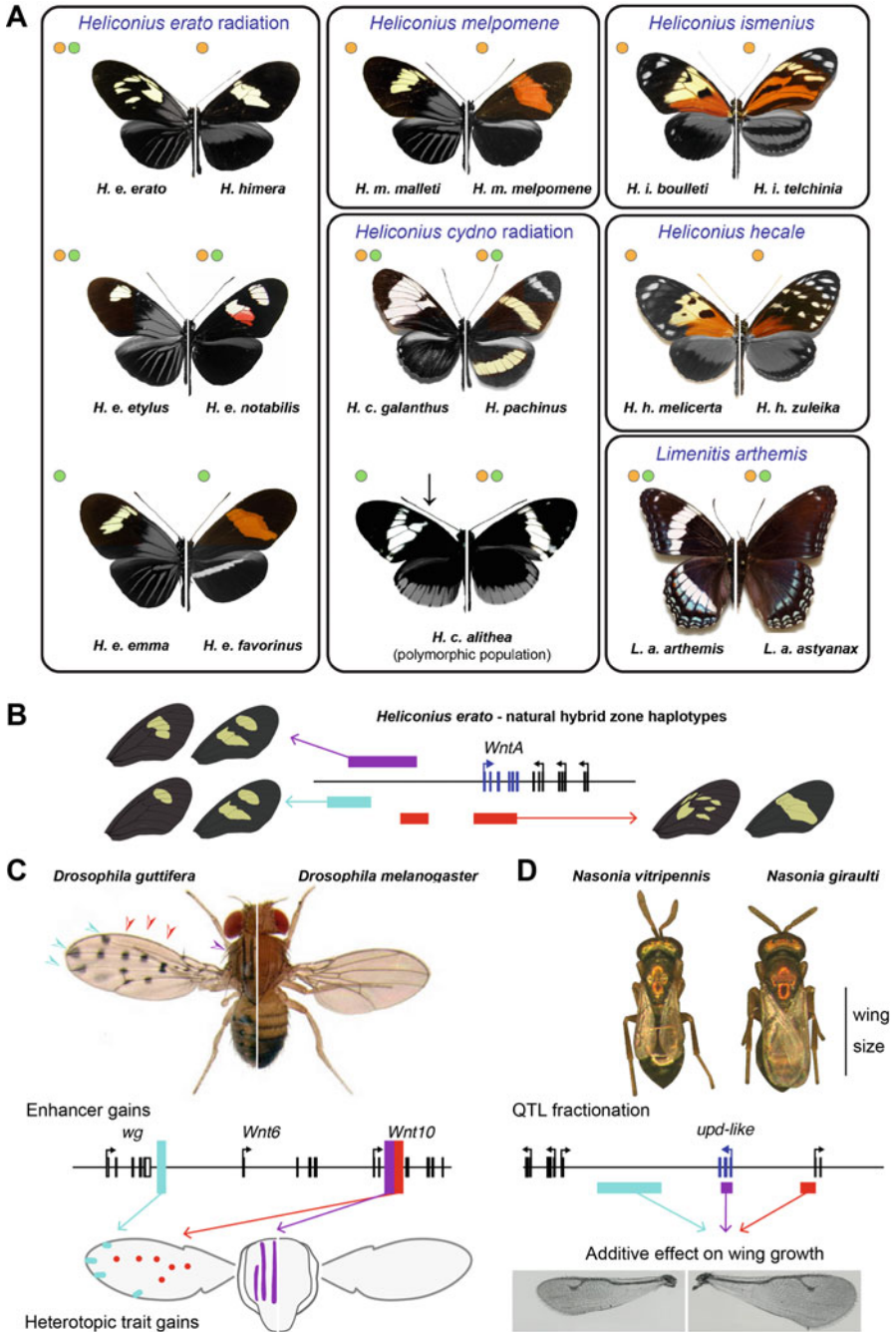
**Fig. 4.3** Mapped cis-regulatory alleles of *WntA*, a genetic hotspot of wing pattern shape variation. (**a**) A total of 18 *WntA* cis-regulatory variants have been identified by linkage mapping (*orange dots*) and admixture mapping in natural hybrid zones (*green dots*). Each allele is associated with spatial shifts in *WntA* expression that drive pattern shape variations, in particular, in the median

expression differences in larval wing disks that correlate tightly with the position of presumptive color elements and defines the black contours of forewing color patterns (Martin et al. 2012). Both linkage and admixture mapping approaches have revealed that a versatile pool of *WntA* alleles underlie marked phenotypic differences in at least six geographic races of *H. erato* (Fig. 4.3a, b). Following this discovery, additional mapping efforts discovered that *WntA* variants control pattern variation in four other *Heliconius* species, as well as in *Limenitis arthemis*, a species which diverged from the *Heliconius* genus 65 million years ago (Fig. 4.3a) (Gallant et al. 2014; Huber et al. 2015). All the mapped *WntA* alleles not only underlie phenotypic divergence within species but also convergence between sympatric morphs that evolved in distinct species, thus providing clear examples of adaptive tinkering and repeated evolution of similar patterns. As expected, the causative changes are not found in the *WntA* coding exons, which show little variation in amino acid sequence, but in the adjoining regulatory loci that control *WntA* expression during wing development. The role of *WntA* cis-regulatory mutations may very well extend to much broader phylogenetic levels, as *WntA* expression, which shows spectacular shifts in expression in all the butterfly species assessed so far, always correlates with color pattern features (Martin and Reed 2014). With a total of 18 alleles in 7 species, all associated with wing color pattern variation, *WntA* can be seen as a genuine genetic hotspot of adaptation (Martin and Orgogozo 2013) and a case model for linking regulatory sequence variation, pattern formation, and morphological evolution at multiple time scales.

## 4.7 Ligand Gene Modularity Allows Interspecific Differences

The current data suggest that the *WntA* locus contains multiple control regions and haplotypes, each being able to reconfigurate part of *WntA* expression and the overall organization of wing patterns. Association mapping reveals at least three adjacent haplotype regions with distinct patterning effects in *H. erato* (Fig. 4.3b) and a single 1.8 kb indel perfectly associated to a polymorphic variant in a sympatric *H. cydno alithea* population (Gallant et al. 2014; Van Belleghem et al. 2017). This said, the
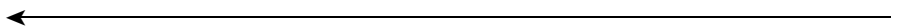
**Fig. 4.3** (continued) region of butterfly forewings. Each half-butterfly corresponds to a natural morph. *WntA*-independent color patterns were manually masked and shaded in *gray* to better highlight the wing pattern areas influenced by *WntA*. (**b**) Fractionation of the *H. erato WntA* locus at several haplotypic blocks, each perfectly associated with pattern shape variation across three natural hybrid zones (Van Belleghem et al. 2017). (**c**) Three novel cis-regulatory regions underlie the evolution of novel pigmentation traits in *D. guttifera*. (**d**) Fine QTL mapping of wing size variation in male *Nasonia* wasps identifies three intervals responsible for the differential spatio-temporal recruitment of the *upd-like* growth factor (*Photo credit (use with permission)* – (**c**) Nicolas Gompel and Shigeyuki Koshikawa and (**d**) David Loehlin)

functional dissection of these genetic elements is reaching a technical limitation at this moment due to the inability to test for the function of each cis-regulatory region in butterflies, and we must gain insight into the evolution of ligand gene expression in analog models to explore the logic of cis-regulatory control. Interestingly, detailed analyses of the cis-regulatory region of another *Wnt* locus, this time encompassing *wingless* (syn. *Wnt1*; *wg*) and its tandem paralogs *Wnt6* and *Wnt10* (Fig. 4.3c), show that three novel, tissue-specific cis-regulatory elements drive *wingless* expression and underlie novel color patterns on the wings and thorax of *Drosophila guttifera* fruit flies (Werner et al. 2010; Koshikawa et al. 2015). While these studies lack the phylogenetic resolution and replication observed in butterflies, they provide one of the most detailed mechanistic accounts of truly novel traits, where the deployment of Wnt expression in three different body regions is driven by independent cis-regulatory changes. Of note, *wg* is also associated to color patterns and wing contours in both flies and butterflies (Macdonald et al. 2010; Martin and Reed 2010; Koshikawa et al. 2015), and a redeployment of this gene to new body regions is likely to drive the evolution of new patterns as well, as it seemed to have occurred during the evolution of larval cuticle patterns in Lepidoptera (Yamaguchi et al. 2013). We note that while Koshikawa et al. did not detect any pattern-related *Wnt6* and *Wnt10* expression in *D. guttifera* developing wings (Koshikawa et al. 2015; S. Koshikawa, personal communication), these two paralogs are co-deployed with *wg* in butterflies where they may underlie a more complex architecture, with partially redundant ligand activities (Martin and Reed 2014). Beyond their obvious parallels (wing pigmentation traits; *Wnt* loci), the butterfly and *D. guttifera* data collectively depict a modular landscape of pattern evolution where acquisitions and modifications of cis-regulatory elements allow a fine-tuning of color patterns (Koshikawa 2015).

Another case study provides further support for linking gene regulatory region modularity at a ligand locus and interspecific variation (Loehlin and Werren 2012). Using two *Nasonia* wasp sister species, Loehlin and Werren mapped a male wing size variation QTL to the JAK/STAT pathway ligand gene *unpaired-like* (*upd-like*) and, by a genetic tour de force, were able to genetically break down this locus into three regulatory intervals, each with complementary effects on wing size. In fact, each mapped interval affects various complementary spatiotemporal expression patterns of *upd-like*, ultimately affecting wing growth. Thus, whether the phenotypic output is a growth trait (the *upd-like* case) or a color pattern (the *WntA* and *wg* cases), we have empirical evidence that morphological evolvability depends in these cases on the capacity to modify an expression pattern. In a nutshell, the different case studies linking insect wing variation and ligand genes highlight the importance of modular cis-regulatory architecture in the tinkering of anatomy.

## 4.8  How, When, and Why Ligand Genes Are Likely Drivers of Pattern Variation, or Not

Our cumulative knowledge of evolutionary genetics foreshadows a relative predictability in the genetic mechanisms that drive phenotypic change (Stern and Orgogozo 2009; Martin and Orgogozo 2013; Orgogozo 2015): by laying out what seems to be common mechanisms or trends in the generation of novelty, we can formulate post hoc expectations that can be generalized over broad taxonomic ranges. The cases of Wnt-based color pattern variation discussed above, *WntA* in nymphalid butterflies and *wg* in *D. guttifera*, both provide a useful model framework for understanding the molecular logic of pattern evolution due to their relative simplicity, as they take place in the two-dimensional canvas of the insect wing epithelium. To the best of our knowledge, these patterning systems are uncoupled from tissue growth, which prevents the complex dynamics found in many other morphological contexts (Salazar-Ciudad 2006; Salazar-Ciudad 2009; Urdy et al. 2016). As simplified spatial output of cellular differentiation, color patterns can be used as a proxy for more complex morphologies, providing fundamental insights that can be applied across all animals. A simple ascertainment emerges from the fly and butterfly data: cis-regulatory evolution of pattern-inducing signaling genes has repeatedly driven the evolution of new patterns and derived pattern shapes. We can elaborate upon a simple gradient model of positional information (Wolpert 1969) generating threshold-dependent pattern boundaries (Fig. 4.4a), to derive five types of ligand gene signaling that can produce morphological outcomes (Fig. 4.4b–f). Since cis-regulatory variation modulates gene expression in time and space, it can affect tissue patterning in multiple ways, and its effect on a ligand gene can be sufficient to induce a new pattern (Fig. 4.4b) or simply change its shape (Fig. 4.4c). In addition, cis-regulatory acquisition of localized repressors can dislocate a pattern and thus affect both pattern number and shape (Fig. 4.4d). Pattern size can also be affected by quantitative or temporal changes in the expression of a secreted factor, without requiring a change in the number of source cells, or, alternatively, by *trans*-interactions upstream of the ligand that would affect its secretion and transport (Fig. 4.4e). Finally, modification in the tissue responsiveness to the signal or its concentration or time-dependent interpretation may modulate the pattern thresholds (e.g., color composition) without affecting the overall size and shape of the pattern (Fig. 4.4f).

These distinct dimensions of pattern variation can be used to generate hypotheses on the molecular targets underlying a given phenotypic state. Below we illustrate this principle, building upon a set of observations made on the variable checkerspot (*Euphydryas chalcedona*). *E. chalcedona* checkerspots display a set of orange patterns outlined by black scales that are each expressing *WntA* or *wg/Wnt6/Wnt10* (Martin and Reed 2014). Each of these patterns can be contracted or expanded by an injection of dextran sulfate or heparin, respectively (Fig. 4.4g). These two sulfated polysaccharide compounds possess a high molecular weight, which restrict them to the extracellular space, and injections are only effective when
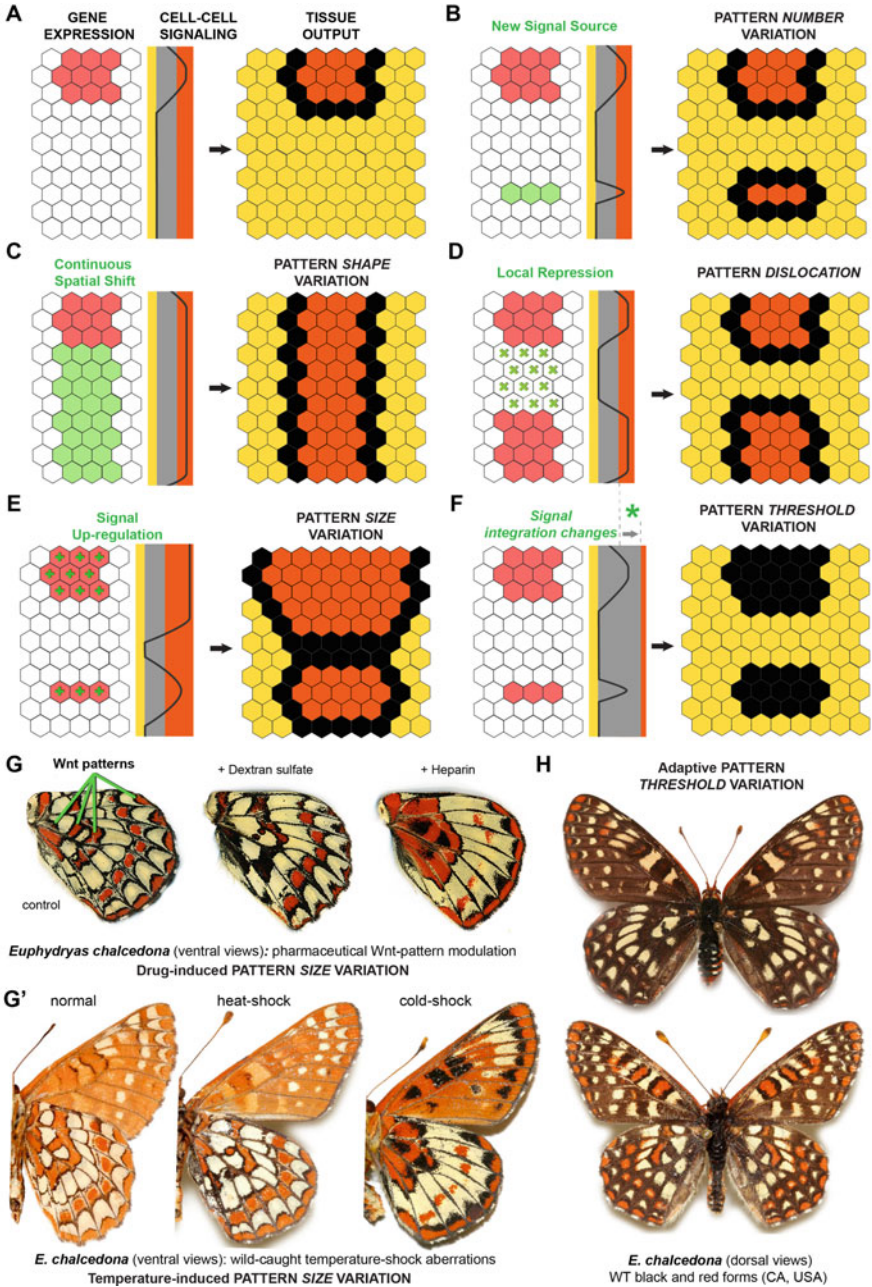
**Fig. 4.4** Distinct aspects of pattern variation may rely on different modes of ligand gene modification. (**a**) A three-step model of pattern formation. Ligand-expressing cells (*red hexagons*) deploy a signal that is interpreted by neighboring cells in a concentration-dependent manner, resulting in a three-state output (*yellow*, low signal; *black*, intermediate; *orange*, high). (**b**) Discrete gain of a novel ligand gene expression domain can generate novel pattern elements. (**c**) Continuous spatial modulation of ligand expression can generate new pattern shapes. (**d**) Local

performed within 24 h after pupation, revealing a short time window for pattern formation (Serfas and Carroll 2005; Martin and Reed 2014). Finally, both heparin and endogenous, heparin-like heparan sulfate proteoglycans (HSPGs) are known to bind Wnt ligands in the extracellular space, where they are of critical importance for signal secretion, stability, and transport (Lin 2004). These observations provide a simple alternative mechanism for modifying pattern size: rather than affecting signal strength directly, variation at genes involved in HSPG synthesis could also modulate the spread of Wnt ligands. Similarly, temperature shocks experienced during early pupal life create analogous pattern aberrations (Fig. 4.4g'), suggesting that specific physiological conditions are critical for normal patterning and that, here again, a broad range of molecular mechanisms taking place during cell-cell signaling (e.g., signal secretion, transport, reception, and degradation) could affect pattern size. The variable checkerspot takes its name from the extensive color pattern variations (Bowers et al. 1985; Long et al. 2014b) that can be observed between populations (Fig. 4.4h). Can we predict whether a ligand locus is involved in driving the difference between these Wnt-positive black vs. red/black patterns? Based on the framework developed above, we believe this is in fact an unlikely scenario. Indeed, the variation involves little differences in pattern shape or number and instead consists in color composition differences. A difference in signal sensitivity rather than signal strength between the two forms is more likely to explain the phenomenon, resulting in a threshold trait variation (see Allen et al. 2008 for a discussion of pattern size vs. color composition). We thus predict that this polymorphism could map to a Wnt-pathway gene or to a gene that can modify the output of the Wnt signaling pathway and that this gene should be active during the extracellular signaling phase or shortly thereafter. Alternatively, the threshold traits could also depend on signal temporal dynamics (Sorre et al. 2014). To be formally tested, these competing hypotheses will require linkage or association mapping between natural morphs and illustrate how our current knowledge can guide a different set of predictions, based on the type of observed trait variation.

**Fig. 4.4** (continued) loss of ligand expression can result in pattern dislocation. (**e**) Upregulation of a ligand gene can generate enlarged patterns. (**f**) Pattern composition may vary based on modifications of the signal interpretation process, downstream of the ligand gene itself (without affecting its expression or protein). (**g**) Sulfated polysaccharide injections in the variable checkerspot butterfly, performed within 24 h after pupation, affect the size of Wnt-positive patterns. Dextran sulfate results in Wnt pattern contractions, while heparin results in Wnt gain-of-function effects that expand the same patterns. Both compounds illustrate how genetic modulations of the extracellular environment can modulate pattern size. (**g'**) Temperature shocks during early pupal life result in pattern distortions (similar to G panel), indicating a sensitivity of the signaling step to physiological conditions. (**h**) The variable checkerspot is named after its color pattern polymorphism, involved in adaptive mimicry (Bowers et al. 1985; Long et al. 2014b). Differences in *red patterns* may be due to changes in genes modulating Wnt signal, rather than at a Wnt gene locus itself (see **f**)

## 4.9 Synthesis: Variations of Morphological Relevance in Ligand-Coding Genes Are cis-Regulatory, Complex, and Multiallelic

We have seen in this review that cis-regulatory alleles in ligand genes can drive morphological evolution in nature. Four cases stand out by the level of scrutiny at which they have been examined, as their experimental dissection shines by exceptional levels of phylogenetic replication or genetic resolution: *Agouti* (*Peromyscus maniculatus* – Nebraska Sandhills: light and dark alleles), *WntA* (*Heliconius* spp. and *Limenitis arthemis* butterflies: wing pattern shape variation), *wg* (*Drosophila guttifera*: acquisition of novel pigmented patterns), and *upd-like* (*Nasonia* spp.: wing size differences). Based on the data at hand, we propose a set of hypotheses that can now be confronted to future experimentation:

1. *Ligand cis-regulatory variation underlies heterotopies*. The four loci above provide clear illustrations of the principle that a local modification of morphology (heterotopy) is likely to be based on cis-regulatory variation. Due to their direct role in cell fate induction, ligand genes can be expressed in new places to influence developmental patterning and eventually anatomical phenotypes.

2. *Gene expression shifts require the accumulation of multiple changes clustered into complex alleles*. Fine mapping of the *Agouti* and *upd-like* loci reveals multiple sub-genic regions which independently contribute to the total phenotype (Loehlin and Werren 2012; Linnen et al. 2013). The same is true for *WntA* in a recent hybrid zone study (Fig. 4.3b), where three noncoding regions were each associated to pattern variation in distinct subareas of the butterfly wing, with their combination constituting the complete phenotype (Van Belleghem et al. 2017). Finally, the *wg* study reveals the modular evolution of three tissue-specific enhancers that collectively explain the pigmentation features of *D. guttifera* (Koshikawa et al. 2015; Koshikawa 2015). These four cases are conceptually similar and show that cis-regulatory evolution relies on the accumulation of multiple changes to generate large effects on ligand expression and final morphology.

3. *Parallel evolution is pervasive, even across distant lineages*. The repeated finding of the same orthologous gene causing similar visible trait changes across distinct lineages may be expected under the candidate gene approach, as a result of ascertainment bias. The replicated identification of coding alleles of *MC1R* and *Agouti* is of that order. However, when independent experiments happen to pinpoint the same locus by taking a linkage or association mapping approach, then we can firmly infer that gene reuse underlies a phenomenon of evolutionary repetition (Martin and Orgogozo 2013; Orgogozo et al. 2015). We have seen that cis-regulatory alleles of *Agouti* have been repeatedly mapped in several populations and species of *Peromyscus* deer mice as well as in humans. The stickleback *KITLG* cis-regulatory changes were mirrored by other cis-regulatory variants driving both skin and hair color variation in human populations (Miller et al. 2007; Guenther et al. 2014). Finally, the *WntA* locus was mapped as a

hotspot of wing pattern evolution in five *Heliconius* species as well as in a clade distant by about 65MY (Gallant et al. 2014). This implies that for a given phenotypic trait, the genetic basis of phenotypic variation may be relatively predictable in a post hoc fashion.

4. *Multiallelism could precede the aggregation of complex alleles.* The identification of multiallelism (syn. polyallelism, genetic heterogeneity) by forward genetic approaches is difficult in spite of their suspected importance in human disease (McClellan and King 2010). Indeed, detecting multiallelism requires a multiple-parent QTL scheme, and this has only been recently implemented in a handful of model organisms (Huang et al. 2011; Long et al. 2014a). Furthermore, GWAS studies typically underestimate the contributions of mixed alleles (Thornton et al. 2013). Several studies have nonetheless found that the pool of cis-regulatory variation influencing gene expression levels is multiallelic, to an overwhelming extent (Gruber and Long 2009; Zhang et al. 2011; King et al. 2014). Does this observation hold up for the spatial shift alleles considered here? As it turns out, replicated mapping within the *H. erato* and *H. cydno* radiations has identified six and four noncoding *WntA* alleles underlying ten distinct wing color shapes in these two species groups, respectively (Martin et al. 2012; Papa et al. 2013; Gallant et al. 2014). *WntA* thus exemplifies how repeated cis-regulatory modification of a ligand gene can replicate both within and between species, spanning a phylogenetic spectrum ranging from recently evolved populations (Van Belleghem et al. 2017) to distant lineages (Gallant et al. 2014). Importantly, this multiallelism probably acts as a prerequisite for the formation of complex alleles, as it is likely that adjacent regulatory regions evolve by recombination between blocks that exist as standing variation, rather than solely by cumulative de novo mutations on the same DNA molecule (Rebeiz et al. 2011; Martin and Orgogozo 2013). A chimeric, polyallelic origin can explain the cis-regulatory evolution of *optix* (Wallbank et al. 2015) (see also chapter by CD Jiggins in this volume), a transcription factor locus that, like *WntA*, shows extensive parallelism and multiallelism in the *Heliconius* genus (Reed et al. 2011; Papa et al. 2013; Martin et al. 2014; Kronforst and Papa 2015; Zhang et al. 2016). We expect that further examples of phenotypic radiations will uncover a multiallelic basis, as recently proposed in cichlid fishes (Roberts et al. 2016). The fact that we observe genetic heterogeneity shows that multiple variants can swarm in a gene pool and may thus provide the bricks of change to build novel cis-regulatory activities. We suggest that the large *Agouti*, *upd-like*, *WntA*, and *wg* haplotypes were agglomerated by recombination between multiple alleles segregating in ancestral populations (Martin and Orgogozo 2013).

## 4.10  Conclusion

In less than a decade, the DGT hypothesis has found validation in the forward genetics literature, where investigations that focused on a morphological difference (without a strong initial bias on the underlying genetics) eventually identified

genetic toolkit loci. This is particularly true for signaling genes: four out of seven morphological gephes in sticklebacks involve secreted signaling ligands, and 18 *WntA* alleles have been associated to wing pattern variation in butterflies. We hope that the continuous compilation of the genetic basis of phenotypic evolution into Gephebase will facilitate similarly minded questions of broad interest and perhaps yield to broader insights and meta-analytical thinking in evolutionary genetics.

# References

Albertson RC, Streelman JT, Kocher TD, Yelick PC (2005) Integration and evolution of the cichlid mandible: the molecular basis of alternate feeding strategies. Proc Natl Acad Sci U S A 102:16287–16292

Allen CE, Beldade P, Zwaan BJ, Brakefield PM (2008) Differences in the selection response of serially repeated color pattern characters: standing variation, development, and evolution. BMC Evol Biol 8:94

Anderson TM, Candille SI, Musiani M et al (2009) Molecular and evolutionary history of melanism in North American gray wolves. Science 323:1339–1343

Barrett RD, Hoekstra HE (2011) Molecular spandrels: tests of adaptation at the genetic level. Nat Rev Genet 12:767–780

Bowers MD, Brown IL, Wheye D (1985) Bird predation as a selective agent in a butterfly population. Evolution:93–103

Candille SI, Kaelin CB, Cattanach BM et al (2007) A β-defensin mutation causes black coat color in domestic dogs. Science 318:1418–1423

Carroll SB (2008) Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell 134:25–36

Carroll SB, Grenier JK, Weatherbee SD (2005) From DNA to diversity: molecular genetics and the evolution of animal design. Wiley, Somerset

Cleves PA, Ellis NA, Jimenez MT et al (2014) Evolved tooth gain in sticklebacks is associated with a cis-regulatory allele of Bmp6. Proc Natl Acad Sci 111:13912–13917

Colosimo PF, Hosemann KE, Balabhadra S et al (2005) Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. Science 307:1928–1933

De Robertis E (2008) Evo-devo: variations on ancestral themes. Cell 132:185–195

Dorshorst B, Molin A-M, Rubin C-J et al (2011) A complex genomic rearrangement involving the endothelin 3 locus causes dermal hyperpigmentation in the chicken. PLoS Genet 7:e1002412

Eizirik E, Yuhki N, Johnson WE et al (2003) Molecular genetics and evolution of melanism in the cat family. Curr Biol 13:448–453

Ellis NA, Glazer AM, Donde NN et al (2015) Distinct developmental genetic mechanisms underlie convergently evolved tooth gain in sticklebacks. Development 142:2442–2451

Erickson PA, Cleves PA, Ellis NA et al (2015) A 190 base pair, TGF-β responsive tooth and fin enhancer is required for stickleback Bmp6 expression. Dev Biol 401:310–323

Erwin DH (2009) Early origin of the bilaterian developmental toolkit. Philos Trans R Soc Lond Ser B Biol Sci 364:2253–2261

Floyd SK, Bowman JL (2007) The ancestral developmental tool kit of land plants. Int J Plant Sci 168:1–35

Gallant JR, Imhoff VE, Martin A et al (2014) Ancient homology underlies adaptive mimetic diversity across butterflies. Nat Commun 5:4817

Gompel N, Prud'homme B (2009) The causes of repeated genetic evolution. Dev Biol 332:36–47

Greenwood AK, Mills MG, Wark AR et al (2016) Evolution of schooling behavior in threespine sticklebacks is shaped by the eda gene. Genetics 203:677–681

Gruber JD, Long AD (2009) Cis-regulatory variation is typically polyallelic in Drosophila. Genetics 181:661–670

Guenther CA, Tasic B, Luo L et al (2014) A molecular basis for classic blond hair color in Europeans. Nat Genet 46:748–752

Huang X, Paulo M-J, Boer M et al (2011) Analysis of natural allelic variation in Arabidopsis using a multiparent recombinant inbred line population. Proc Natl Acad Sci 108:4488–4493

Huber B, Whibley A, Poul Y et al (2015) Conservatism and novelty in the genetic architecture of adaptation in Heliconius butterflies. Heredity 114:515–524

Indjeian VB, Kingman GA, Jones FC et al (2016) Evolving new skeletal traits by cis-regulatory changes in bone morphogenetic proteins. Cell 164:45–56

Johnsson M, Gustafson I, Rubin C-J et al (2012) A sexual ornament in chickens is affected by pleiotropic alleles at HAO1 and BMP2, selected during domestication. PLoS Genet 8: e1002914

Jones FC, Chan YF, Schmutz J et al (2012) A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. Curr Biol 22:83–90

Karyadi DM, Karlins E, Decker B et al (2013) A copy number variant at the KITLG locus likely confers risk for canine squamous cell carcinoma of the digit. PLoS Genet 9:e1003409

Kicheva A, Briscoe J (2015) Developmental pattern formation in phases. Trends Cell Biol 25:579–591

King EG, Sanderson BJ, McNeil CL et al (2014) Genetic dissection of the Drosophila melanogaster female head transcriptome reveals widespread allelic heterogeneity. PLoS Genet 10:e1004322

Kopp A (2009) Metamodels and phylogenetic replication: a systematic approach to the evolution of developmental pathways. Evolution 63:2771–2789

Koshikawa S (2015) Enhancer modularity and the evolution of new traits. Fly (Austin) 9:155–159

Koshikawa S, Giorgianni MW, Vaccaro K et al (2015) Gain of cis-regulatory activities underlies novel domains of wingless gene expression in Drosophila. Proc Natl Acad Sci 112:7524–7529

Kronforst MR, Papa R (2015) The functional basis of wing patterning in Heliconius butterflies: the molecules behind mimicry. Genetics 200:1–19

Kronforst MR, Barsh GS, Kopp A et al (2012) Unraveling the thread of nature's tapestry: the genetics of diversity and convergence in animal pigmentation. Pigment Cell Melanoma Res 25:411–433

Laurent S, Pfeifer SP, Settles ML et al (2016) The population genomics of rapid adaptation: disentangling signatures of selection and demography in white sands lizards. Mol Ecol 25:306–323

Lehner B (2013) Genotype to phenotype: lessons from model organisms for human genetics. Nat Rev Genet 14:168–178

Liao B-Y, Weng M-P, Zhang J (2010) Contrasting genetic paths to morphological and physiological evolution. Proc Natl Acad Sci 107:7353–7358

Lin X (2004) Functions of heparan sulfate proteoglycans in cell signaling during development. Development 131:6009–6021

Linderholm A, Larson G (2013) The role of humans in facilitating and sustaining coat colour variation in domestic animals. Semin Cell Dev Biol 24:587–593

Linnen CR, Poh Y-P, Peterson BK et al (2013) Adaptive evolution of multiple traits through multiple mutations at a single gene. Science 339:1312–1316

Litchfield K, Levy M, Huddart RA et al (2016) The genomic landscape of testicular germ cell tumours: from susceptibility to treatment. Nat Rev Urol 13:409–419

Loehlin DW, Werren JH (2012) Evolution of shape by multiple regulatory changes to a growth gene. Science 335:943–947

Long AD, Macdonald SJ, King EG (2014a) Dissecting complex traits using the Drosophila synthetic population resource. Trends Genet 30:488–495

Long EC, Hahn TP, Shapiro AM (2014b) Variation in wing pattern and palatability in a female-limited polymorphic mimicry system. Ecol Evol 4:4543–4552

Macdonald WP, Martin A, Reed RD (2010) Butterfly wings shaped by a molecular cookie cutter: evolutionary radiation of lepidopteran wing shapes associated with a derived cut/wingless wing margin boundary system. Evol Dev 12:296–304

Mallarino R, Linden TA, Linnen CR, Hoekstra HE (2016) The role of isoforms in the evolution of cryptic coloration in Peromyscus mice. Molecular Ecology 26(1):245–258

Manceau M, Domingues VS, Linnen CR et al (2010) Convergence in pigmentation at multiple levels: mutations, genes and function. Philos Trans R Soc Lond Ser B Biol Sci 365:2439–2450

Manceau M, Domingues VS, Mallarino R, Hoekstra HE (2011) The developmental role of Agouti in color pattern evolution. Science 331:1062–1065

Martin A, Orgogozo V (2013) The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. Evolution 67:1235–1250. doi:10.1111/evo.12081

Martin A, Reed RD (2010) Wingless and aristaless2 define a developmental ground plan for moth and butterfly wing pattern evolution. Mol Biol Evol 27:2864–2878

Martin A, Reed RD (2014) Wnt signaling underlies evolution and development of the butterfly wing pattern symmetry systems. Dev Biol 395:367–378. doi:10.1016/j.ydbio.2014.08.031

Martin A, Papa R, Nadeau NJ et al (2012) Diversification of complex butterfly wing patterns by repeated regulatory evolution of a Wnt ligand. Proc Natl Acad Sci U S A 109:12632–12637. doi:10.1073/pnas.1204800109

Martin A, McCulloch KJ, Patel NH et al (2014) Multiple recent co-options of Optix associated with novel traits in adaptive butterfly wing radiations. EvoDevo 5:1–14. doi:10.1186/2041-9139-5-7

McClellan J, King M-C (2010) Genetic heterogeneity in human disease. Cell 141:210–217

McGary KL, Park TJ, Woods JO et al (2010) Systematic discovery of nonobvious human disease models through orthologous phenotypes. Proc Natl Acad Sci 107:6544–6549

McGregor AP, Orgogozo V, Delon I et al (2007) Morphological evolution through multiple cis-regulatory mutations at a single gene. Nature 448:587–590

McRobie HR, King LM, Fanutti C et al (2014) Agouti signalling protein is an inverse agonist to the wildtype and agonist to the melanic variant of the melanocortin-1 receptor in the grey squirrel (*Sciurus carolinensis*). FEBS Lett 588:2335–2343

Merrill R, Dasmahapatra K, Davey J et al (2015) The diversification of Heliconius butterflies: what have we learned in 150 years? J Evol Biol 28:1417–1438

Meyerowitz EM (2002) Plants compared to animals: the broadest comparative study of development. Science 295:1482–1485

Miller CT, Beleza S, Pollen AA et al (2007) Cis-regulatory changes in kit ligand expression and parallel evolution of pigmentation in sticklebacks and humans. Cell 131:1179–1189

Monestier O, Servin B, Auclair S et al (2014) Evolutionary origin of bone morphogenetic protein 15 and growth and differentiation factor 9 and differential selective pressure between mono- and polyovulating species. Biol Reprod 91:83

Mundy NI (2005) A window on the genetics of evolution: MC1R and plumage colouration in birds. Proc R Soc Lond B Biol Sci 272:1633–1640

Nahmad Bensusan M (2011) Interpretation and scaling of positional information during development. Dissertation (Ph.D.), California Institute of Technology

Nichols SA, Dirks W, Pearse JS, King N (2006) Early evolution of animal cell signaling and adhesion genes. Proc Natl Acad Sci 103:12451–12456

Nix MA, Kaelin CB, Ta T et al (2013) Molecular and functional analysis of human β-defensin 3 action at melanocortin receptors. Chem Biol 20:784–795

Noon EP-B, Davis FP, Stern DL (2016) Evolved repression overcomes enhancer robustness. Dev Cell 39(5):572–584

Nunes MD, Arif S, Schlötterer C, McGregor AP (2013) A perspective on micro-evo-devo: progress and potential. Genetics 195:625–634.

O'Brown NM, Summers BR, Jones FC et al (2015) A recurrent regulatory change underlying altered expression and wnt response of the stickleback armor plates gene EDA. elife 4:e05290

Ollivier M, Tresset A, Hitte C et al (2013) Evidence of coat color variation sheds new light on ancient canids. PLoS One 8:e75110

Orgogozo V (2015) Replaying the tape of life in the twenty-first century. Interface Focus 5:20150057

Orgogozo V, Morizot B, Martin A (2015) The differential view of genotype–phenotype relationships. Front Genet 6:179

Orgogozo V, Peluffo A, Morizot B (2016) Chapter one-the "Mendelian Gene" and the "Molecular Gene": two relevant concepts of genetic units. Curr Top Dev Biol 119:1–26

Papa R, Martin A, Reed RD (2008) Genomic hotspots of adaptation in butterfly wing pattern evolution. Curr Opin Genet Dev 18:559–564

Papa R, Kapan DD, Counterman BA et al (2013) Multi-allelic major effect genes interact with minor effect QTLs to control adaptive color pattern variation in *Heliconius erato*. PLoS One 8: e57033

Pardo-Diaz C, Salazar C, Jiggins CD (2015) Towards the identification of the loci of adaptive evolution. Methods Ecol Evol 6:445–464

Perrimon N, Pitsouli C, Shilo B-Z (2012) Signaling mechanisms controlling cell fate and embryonic patterning. Cold Spring Harb Perspect Biol 4:a005975

Powell R, Mariscal C (2015) Convergent evolution as natural experiment: the tape of life reconsidered. Interface Focus 5:20150040

Protas ME, Patel NH (2008) Evolution of coloration patterns. Annu Rev Cell Dev Biol 24:425–446

Prud'homme B, Gompel N, Carroll SB (2007) Emerging principles of regulatory evolution. Proc Natl Acad Sci 104:8605–8612

Qanbari S, Pausch H, Jansen S et al (2014) Classic selective sweeps revealed by massive sequencing in cattle. PLoS Genet 10:e1004148

Rebeiz M, Jikomes N, Kassner VA, Carroll SB (2011) Evolutionary origin of a novel gene expression pattern through co-option of the latent activities of existing regulatory sequences. Proc Natl Acad Sci 108:10036–10043

Reed RD, Papa R, Martin A et al (2011) Optix drives the repeated convergent evolution of butterfly wing pattern mimicry. Science 333:1137–1141

Reissmann M, Ludwig A (2013) Pleiotropic effects of coat colour-associated mutations in humans, mice and other mammals. Semin Cell Dev Biol 24(6–7):576–586

Roberts RB, Moore EC, Kocher TD (2016) An allelic series at pax7a is associated with color polymorphism diversity in Lake Malawi cichlid fish. Mol Ecol 26(10):2615–2639

Rockman MV (2012) The QTN program and the alleles that matter for evolution: all that's gold does not glitter. Evolution 66:1–17

Rogers KW, Schier AF (2011) Morphogen gradients: from generation to interpretation. Annu Rev Cell Dev Biol 27:377–407

Rokas A (2008a) The molecular origins of multicellular transitions. Curr Opin Genet Dev 18:472–478

Rokas A (2008b) The origins of multicellularity and the early history of the genetic toolkit for animal development. Annu Rev Genet 42:235–251

Salazar Ciudad I (2006) On the origins of morphological disparity and its diverse developmental bases. BioEssays 28:1112–1122

Salazar-Ciudad I (2009) Looking at the origin of phenotypic variation from pattern formation gene networks. J Biosci 34:573–587

Salvador-Martínez I, Salazar-Ciudad I (2015) How complexity increases in development: an analysis of the spatial–temporal dynamics of 1218 genes in *Drosophila melanogaster*. Dev Biol 405:328–339

Santschi EM, Purdy AK, Valberg SJ et al (1998) Endothelin receptor B polymorphism associated with lethal white foal syndrome in horses. Mamm Genome 9:306–309

Schoenebeck JJ, Hutchinson SA, Byers A et al (2012) Variation of BMP3 contributes to dog breed skull diversity. PLoS Genet 8:e1002849

Scotland RW (2011) What is parallelism? Evol Dev 13:214–227

Seitz JJ, Schmutz SM, Thue TD, Buchanan FC (1999) A missense mutation in the bovine MGF gene is associated with the roan phenotype in belgian blue and shorthorn cattle. Mamm Genome 10:710–712

Serfas MS, Carroll SB (2005) Pharmacologic approaches to butterfly wing patterning: sulfated polysaccharides mimic or antagonize cold shock and alter the interpretation of gradients of positional information. Dev Biol 287:416–424

Smedley D, Haider S, Durinck S et al (2015) The BioMart community portal: an innovative alternative to large, centralized data repositories. Nucleic Acids Res 43:gkv350

Sorre B, Warmflash A, Brivanlou AH, Siggia ED (2014) Encoding of temporal signals by the TGF-β pathway and implications for embryonic patterning. Dev Cell 30:334–342

Spemann H, Mangold H (2001) Über induktion von embryoanlagen durch implantation artfremder organisatoren. Roux'Arch. Entwicklungsmech 1924; 100: 599–638 (translated and reprinted). Int J Dev Biol 45:13–38

Stam LF, Laurie CC (1996) Molecular dissection of a major gene effect on a quantitative trait: the level of alcohol dehydrogenase expression in *Drosophila melanogaster*. Genetics 144:1559–1564

Stern DL (2000) Perspective: evolutionary developmental biology and the problem of variation. Evolution 54:1079–1091

Stern DL (2011) Evolution, development, & the predictable genome. Roberts and Co. Publishers, Greenwood Village

Stern DL (2013) The genetic causes of convergent evolution. Nat Rev Genet 14:751–764

Stern DL, Orgogozo V (2008) The loci of evolution: how predictable is genetic evolution? Evolution 62:2155–2177

Stern DL, Orgogozo V (2009) Is genetic evolution predictable? Science 323:746–751

Streisfeld MA, Rausher MD (2011) Population genetics, pleiotropy, and the preferential fixation of mutations during adaptive evolution. Evolution 65:629–642

Supple M, Papa R, Counterman B, McMillan WO (2014) The genomics of an adaptive radiation: insights across the Heliconius speciation continuum. In: Ecological Genomics. Springer, pp 249–271

Thornton KR, Foran AJ, Long AD (2013) Properties and modeling of GWAS when complex disease risk is due to non-complementing, deleterious mutations in genes of large effect. PLoS Genet 9:e1003258

Urdy S (2012) On the evolution of morphogenetic models: mechano-chemical interactions and an integrated view of cell differentiation, growth, pattern formation and morphogenesis. Biol Rev 87:786–803

Urdy S, Goudemand N, Pantalacci S (2016) Chapter seven-looking beyond the genes: the interplay between signaling pathways and mechanics in the shaping and diversification of epithelial tissues. Curr Top Dev Biol 119:227–290

Van Belleghem SM, Rastas P, Papanicolaou A et al (2017) Complex modular architecture around a simple toolkit of wing pattern genes. Nat Ecol Evol 1:0052. doi:10.1038/s41559-016-0052

Vignieri SN, Larson JG, Hoekstra HE (2010) The selective advantage of crypsis in mice. Evolution 64:2153–2158

Waddington CH (1940) Organisers and genes. Cambridge University Press, Cambridge

Wallbank RW, Baxter SW, Pardo-Díaz C et al (2015) Evolutionary novelty in a butterfly wing pattern through enhancer shuffling. PLoS Biol 14(1):e1002353

Wehrle-Haller B (2003) The role of Kit-ligand in melanocyte development and epidermal homeostasis. Pigment Cell Res 16:287–296

Werner T, Koshikawa S, Williams TM, Carroll SB (2010) Generation of a novel wing colour pattern by the Wingless morphogen. Nature 464:1143–1148

Wolf Horrell EM, Boulanger MC, D'orazio JA (2016) Melanocortin 1 receptor: structure, function and regulation. Front Genet 7:95

Wolpert L (1969) Positional information and the spatial pattern of cellular differentiation. J Theor Biol 25:1–47

Yamaguchi J, Banno Y, Mita K et al (2013) Periodic Wnt1 expression in response to ecdysteroid generates twin-spot markings on caterpillars. Nat Commun 4:1857

Zhang X, Cal AJ, Borevitz JO (2011) Genetic architecture of regulatory variation in *Arabidopsis thaliana*. Genome Res 21:725–733

Zhang W, Dasmahapatra KK, Mallet J et al (2016) Genome-wide introgression among distantly related Heliconius butterfly species. Genome Biol 17:1