

Are we there yet? Tracking the development of new model systems

Arhat Abzhanov¹, Cassandra G. Extavour¹, Andrew Groover^{2,3}, Scott A. Hodges⁴, Hopi E. Hoekstra¹, Elena M. Kramer¹ and Antonia Monteiro⁵

¹ Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA

² Department of Plant Biology, University of California, Davis, CA 95616, USA

³ USDA Forest Service, Institute of Forest Genetics, Davis, CA 95616, USA

⁴ Department of Ecology, Evolution and Marine Biology, University of California, Santa Barbara, CA 93106, USA

⁵ Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520, USA

It is increasingly clear that additional ‘model’ systems are needed to elucidate the genetic and developmental basis of organismal diversity. Whereas model system development previously required enormous investment, recent advances including the decreasing cost of DNA sequencing and the power of reverse genetics to study gene function are greatly facilitating the process. In this review, we consider two aspects of the development of new genetic model systems: first, the types of questions being advanced using these new models; and second, the essential characteristics and molecular tools for new models, depending on the research focus. We hope that researchers will be inspired to explore this array of emerging models and even consider developing new molecular tools for their own favorite organism.

The need for new genetic model systems

Determining how and why the diversity of complex life forms that surround us originated is a major question in biology. Given the millions of species on our planet, understanding the evolution of organismal form, physiology or behavior will require a sustained effort to expand the currently small set of experimental model organisms to include many others at key branches of the tree of life.

Investigators, however, are interested in addressing many different questions about organismal diversity. Some of us want to understand the evolutionary process at the twigs of the tree of life – in the last few million years of evolution – and focus on the genetic, developmental and ecological changes that underlie differentiation of closely related species. Others are more interested in the changes that evolved hundreds of millions of years ago and now are only found in representatives of surviving crown groups (see Glossary). There are advantages and challenges to working on both evolutionary time scales. When sampling the twigs on a single branch, we are often able to identify the genetic basis of phenotypic variation because we can take advantage of the ability to cross closely related species; however, these recent transitions often involve relatively small changes in phenotype. By contrast, when sampling single twigs in different branches, we can study major phenotypic transitions (e.g. changes in body plan) that represent major evolutionary changes.

However, because there are often few surviving intermediate species and genetic crosses are not feasible, the precise molecular mechanisms responsible for these phenotypic transitions are usually more difficult to pinpoint.

Regardless of our individual preferences in sampling the tree of life, a clearer picture of organismal evolution will emerge only as we study more taxa. In addition, many of us want to understand evolutionary processes at the molecular and/or developmental level, processes that are at work in natural populations. Although traditional models can be studied in the wild (see below), some researchers are moving away from existing laboratory systems (e.g. *Drosophila melanogaster*, *Caenorhabditis elegans*, *Mus musculus*, *Arabidopsis thaliana*) and are developing new genomic resources and tools for organisms at diverse branches of the tree of life [1,2].

Glossary

Adaptation: the process by which natural selection favors those individuals with heritable morphological, physiological or behavioral traits that increase fitness in a particular environment.

Crown group: a group or clade of organisms defined by only extant species. It includes the last common ancestor of a given clade as well as all living descendants of that ancestor.

Developmental evolution (Evo-Devo): a field of study that integrates traditional research on organismal evolutionary biology (systematics, paleontology, and comparative anatomy) with molecular embryology, genetics and genomics with the goal of understanding how changes in developmental genetic programs produce morphological diversity.

Expressed sequence tag (EST): a short sequence of transcribed RNA that is produced by random sequencing of cloned cDNA pools.

Genotype × environment interactions: differential developmental or physiological responses of certain genotypes in different environments; these might reflect varying degrees of phenotypic plasticity.

Hybridization: the interbreeding of distinct species.

Macroevolution: evolutionary changes that are studied above the level of species, often associated with differences between families or phyla.

Microevolution: the process of allele frequency change over many generations, usually measured at the species level or below. Microevolutionary processes are also often studied among closely related, often interfertile, species.

Near isogenic lines (NILs): a genotype, generally derived by repeated backcrossing, which differs from another genotype by only one genetic region.

Phenotypic plasticity: the ability of an organism of a given genotype to alter its phenotype in response to environmental conditions.

Quantitative trait loci mapping: quantitative traits are aspects of a phenotype that are controlled by more than one locus and often vary continuously. The genomic localization of these loci is conducted by statistical analysis of the segregation of multiple genetic markers and variation in the phenotypic trait(s) of interest.

Speciation: the evolutionary process by which new species are formed.

Corresponding author: Kramer, E.M. (ekramer@oeb.harvard.edu).

The development of new model systems, however, is not trivial and requires a significant investment of money and time; thus, it is crucial to make careful choices about the most appropriate organism to study. As most research today is hypothesis driven, it is also important to evaluate whether the specific questions cannot be better addressed in traditional systems. Notably, some of the most successful recently developed systems are not actually 'new' but have been the subjects of ecological and evolutionary study for decades (e.g. cichlids, sticklebacks, deer mice, butterflies, Darwin's finches and monkey flowers). What is new is our ability to apply advanced molecular tools to these ecologically well-characterized species.

The goals of this review are twofold. First, we provide an overview of the diverse array of questions that newly emerging model systems can help to elucidate, including the genetic basis of speciation and adaptation, the evolution of morphological and ecological novelty and the nature of the metazoan genetic toolkit. Second, we outline the fundamental characteristics and tools that are essential for the development of new model systems. Whereas the former serves to highlight the important contributions already being made by new model systems, the latter might guide researchers who are considering the development of new tools for their own study organisms.

Research questions for new model systems

The growing number of species for which significant genetic resources are available is sparking a new era of study in which fundamental genetic questions underlying phenotypic evolution, adaptation and speciation can be addressed with rigor. One area of progress is the merging of fields (e.g. population genetics and development) and a concomitant expansion of the questions that can be answered. This type of synergy is exemplified by studies in developmental evolution (termed 'evo-devo'), which are increasingly intersecting with the genomic scale study of natural variation underlying adaptive traits within species. Although far from complete, the studies described below, all drawn from emerging models, are beginning to answer some long-standing questions about the genetics of phenotypic evolution at both the macro- and micro-levels (Boxes 1 and 2).

Macroevolutionary changes

In our quest to understand the molecular basis of phenotypic traits, novel traits are of special interest because they represent major discontinuities in the diversity of life. Given that many genes have retained the same function over millions of years, how do novel functions and traits evolve? Generally, two interrelated processes have been proposed to account for such evolution. One process is gene duplication followed by divergence – one gene copy retains its original function, whereas the other copy is free to diverge and take on a new role or expression pattern. A second process is to increase the complexity of regulatory control of genes whereby additional spatial or temporal patterns of gene expression occur (i.e. co-option). Understanding how these processes have contributed to the evolution of novelty, and thus biodiversity, has been studied at many time scales. For example, in columbines

Box 1. Models and tools for macroevolutionary studies

Phylogenetic position is an important consideration in choosing models for macroevolutionary studies. New models, which represent undersampled lineages, are useful in investigating the complement of genetic pathways present in the last common ancestor of a given clade. This type of question can be asked at many phylogenetic levels, and systems such as the Crustacean *Parhyale* are helping us to understand how deeply conserved particular genetic programs are. Crucial components of this research include established expression protocols and reverse genetics, which allow candidate genes to be functionally assessed at many different levels [12,65]. Another important goal in macroevolution is to understand the mechanisms by which morphological novelties arise. Morphological innovation can be recognized throughout the tree of life but, in some cases, these evolutionary events have occurred recently enough to offer the chance to fully tease apart their evolution. In the butterfly *Bicyclus*, as well as in other Lepidopterans, researchers are drawing on a wealth of natural variation, combined with comparative gene expression and elegant transgenic techniques, to understand how the genetic pathways controlling eyespots evolved [52,66,67]. In this example we see an important new trend: the bridging of micro- and macroevolutionary scales to inform one another. Choosing macroevolutionary models with this information in mind (e.g. taxa with tractable genome sizes and natural variation) will greatly improve the overall utility of the system.

(*Aquilegia*), a novel fifth floral organ, the staminodium, evolved within the last 12–15 million years [3]. One type of floral organ identity gene, *APETALA3* (*AP3*), has three paralogs, one of which is specific to the novel staminodium, suggesting that duplications of the *AP3* locus facilitated the evolution of new floral organ types [4]. By contrast, in several butterflies, eyespot pattern evolution seems to be associated with the co-option of pre-existent small circuits

Box 2. Models and tools for microevolutionary studies

Choosing a system that harbors significant recent phenotypic variation is an absolute necessity for microevolutionary studies. Several new models that meet this requirement are now used to address a series of long standing questions in evolutionary biology. Do convergent phenotypes result from changes in the same genes or even the same mutations? Does adaptation arise from standing genetic variation or from new mutations? Do adaptive mutations occur in protein-coding regions or noncoding regions? For example, new molecular tools, including a large database of interspecific genetic variation, are helping to elucidate the basis of recent adaptive radiation in the genus *Aquilegia*. This database was generated through deep express sequence tag sequencing of interspecies hybrids, producing both an important resource for candidate gene characterization and for mapping major quantitative traits (<http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gud-b=aquilegia>). There is perhaps no better known example of adaptive radiation than Darwin's finches, the genus *Geospiza*. By combining a candidate gene approach with creative microarray analyses, the developmental pathways responsible for the evolution of diverse, complex, and, in some cases, convergent beak morphologies are finally being elucidated [34,35]. The ability to conduct genetic crosses (or alternatively, to collect detailed pedigree information) allows for not only the identification of genes, but in some cases also of the precise mutations contributing to adaptation. The recent development of a molecular genetic map in deer mice (*Peromyscus*) highlights how different types of mutations (those in both protein-coding regions and putative *cis*-regulatory regions) can contribute to phenotypic evolution [17,53]. Thus, the integration of modern molecular genetic tools is rapidly advancing the development of long-standing evolutionary models into fully realized genetic systems.

of interacting genes that also function in patterning the anterior-posterior wing compartment [5], the insect ventral appendages and/or genes that enable wound healing [6]. At some level, both of these examples involve co-option of pre-existing genetic components. The key difference is whether gene duplication facilitates the co-option process, as is seen in columbine, or if the genetic module can be directly recruited, as observed for butterfly eyespots.

Major evolutionary transitions can also be studied by comparing genome sequences at a much deeper phylogenetic level. When did the evolutionary ‘toolkits’ necessary for major new features evolve? For example, genome comparisons among early branching animals have identified homologs for most synapse components in the demosponge, *Amphimedon queenslandica*, which might have facilitated the later evolution of a nervous system in metazoans [7]. In plants, whole genome duplications followed by diploidization have been a recurrent theme. This process could provide the building blocks for large-scale diversification by duplicating entire genetic pathways, which in turn enables greater phenotypic plasticity and can create novel genetic interactions [8]. However, additional genome sequences at key phylogenetic positions are needed to date the timing of ancestral duplication events and determine whether they are correlated with major plant diversification events [9].

Another important contribution of such sequencing efforts is to elucidate the evolutionary dynamics of microRNA (miRNA)-based regulatory mechanisms across plants and animals [10,11]. The fact that many genetic modules, including miRNAs, are deeply conserved, has offered numerous opportunities for convergence at the molecular level. Similarly, new model systems for macroevolutionary questions allow us to measure degrees of conservation and diversification at both molecular and developmental levels [12–14].

Microevolutionary changes

A major question in evolutionary biology concerns the number, location and the effect size of loci underlying phenotypic variation. Darwin laid the groundwork for a ‘micromutational’ view of evolution in which numerous genetic changes of small effect underlie differences in phenotypic and adaptive traits [15]. Variation in some phenotypic traits described using quantitative trait locus (QTL) mapping is consistent with this view (e.g. floral characters associated with self-pollination in *Mimulus nasutus* × *M. guttatus* crosses) [16]. However, QTL studies from a variety of species have identified traits for which a modest number of genes can explain large amounts of observed phenotypic variation. For instance, coat pigmentation differences in beach mice (*Peromyscus polionotus*) are largely explained by variation at two genes [17], whereas many distinct wing patterns in *Heliconius* butterflies seem to be produced by variation at the same locus [18]. Moreover, major QTLs have been identified in Black Cottonwood (*Populus trichocarpa*) for both stem growth and an adaptive trait, spring bud burst [19]. Although the micromutational view could still hold, it has become clear that evolutionary genetic changes can also be concentrated in few genomic regions (i.e. many mutations of small effect could occur at the same locus).

A challenge in research using established and emerging model species is to increase the power and resolution of genome-scale strategies to identify the precise sequence changes that lead to phenotypic evolution. In addition, the role of epistasis and other higher-order interactions (e.g. among duplicated genes or chromosomes) are poorly defined in the regulation and evolution of phenotypic traits. Similarly, in only a few cases has the potential genetic basis of crucial genotype × environment (G×E) interactions been identified. Plants hold a special potential to assist in this process because they respond strongly to environmental cues. In the major model system *Arabidopsis*, analysis of natural variation has provided considerable insight into the role of genes such as those that encode the Phytochrome family in responses to light and temperature [20,21]. This work highlights how existing major model systems can be used to address evolutionary and ecological questions when examined in nature [22,23]. Of course, more rigorous approaches, including studies of network biology, are probably required to fully understand G×E and other complex interactions.

A related question concerning the nucleotide basis for phenotypic evolution is whether mutations are primarily found in *cis*-regulatory regions or in structural, protein-coding regions and whether these locations differ among traits or developmental timing. Examples for both have been found in plants and animals, although the total number of causative mutations discovered and functionally verified remains modest (for recent reviews, see Refs [24,25]). One example of a *cis*-regulatory change is the evolution of light pigmentation in threespine stickleback (*Gasterosteus aculeatus*) populations, which result from upstream changes at the *Kit Ligand* locus [26]. In plants, variation in the promoter regions of two *KNOX* (*Knotted1-like homeobox*) genes is associated with the diversification of leaf morphology in the *A. thaliana* relative *Cardamine* [27]. (This latter case underscores the utility of developing new systems that are closely related to existing models.) Conversely, numerous examples can be found for structural changes underlying differences in morphological phenotypes (e.g. Refs. [24,28,29]). Interestingly, adaptive differences in beach mouse coat color are attributed to the interaction of both a structural mutation (*Melanocortin-1 receptor*) and a *cis*-regulatory element that affects *Agouti* expression [17]. More data are needed to have a complete view of the relative contribution of *cis*-regulatory versus structural differences to phenotypic evolution. Most notably, the ability to identify *cis*-regulatory regions is challenging, because they can exist far from the coding sequence of their target genes, are variable in length and are not always well conserved [30].

Another route to understanding how phenotypic evolution proceeds is to study independent origins of the same adaptive traits and ask whether the same pathways, genes and mutations have been used (i.e. whether convergence has occurred at the molecular level; see Ref. [31] for a recent review). At one extreme is the finding that independent origins of melanism have evolved through different genetic mechanisms in different populations of rock pocket mice (*Chaetodipus intermedius*) [32]. By contrast, in threespine sticklebacks, the evolution of low-plated

phenotypes in freshwater locations has occurred through the repeated selection of *Ectodyplasin* (*Eda*) alleles derived from the same ancestral haplotype [33]. Similarly, adaptive evolution of beak morphologies in Darwin's finches (*Geospiza*) might have tweaked only two genetic pathways to generate considerable morphological variation [34,35]. In the plant columbine (*Aquilegia*), a comparison of the expression patterns of the anthocyanin pathway genes across multiple species revealed similar gene-specific downregulation in lineages that had independently lost the production of these floral pigments, suggesting similar mutations in *trans*-regulatory factors [36].

The ability to identify the degree of molecular convergence for particular traits depends on the genetic resources available. For instance, QTL mapping can indicate that similar genomic regions are responsible for a trait but alone will not determine whether the same genes and mutations are involved. Similarly, expression studies might show that different levels of a particular gene correlate with trait variation, but either *cis*-regulatory or *trans*-regulatory mutations could be responsible for the expression differences. An important goal will be to identify particular constraints or biases that promote molecular convergence and the circumstances under which they occur [37].

What characteristics, resources, and tools are important for new model systems?

The recent explosion in new model systems has resulted in large part from the increasing accessibility of medium- and large-scale sequencing, which together with improvements to bioinformatics tools, produced many genome and expressed sequence tag (EST) resources (Figure 1). New model system development has also gained momentum from demonstrability: in sticklebacks, for example, the identification of the *Eda* gene discussed above was conducted via a mapping approach alone, despite the fact that the locus was not an *a priori* candidate gene for armor plating [38]. However, the tools and characteristics that should be available in a model system depend to a large extent on the particular questions being asked. Although we expect major model systems to possess the full complement of molecular tools, model systems that are especially developed to address evolutionary and ecological questions might not possess this entire spectrum of resources (Figure 2). Although a full complement of tools allows greater precision and confidence in identifying and functionally verifying the causative genetic changes responsible for phenotypic evolution, the same level of precision is not always necessary and, depending on the questions being asked, some tools are more relevant than others.

Macroevolutionary tools

Models for answering macroevolutionary questions must be selected with careful attention to their phylogenetic position. The sampling of new lineages that are not well represented by existing models, such as hemichordates or lower land plants, is a priority. The experimental tools best suited for these studies place an emphasis on candidate gene approaches, gene expression and functional analyses (Box 1).

Obtaining candidate gene sequences. A common approach in evo-devo research is to document changes in the expression of candidate developmental genes that have previously been shown to play an important role in homologous traits in other taxa. These expression changes in the underlying developmental program are then correlated with phenotypic change. Species-specific probes for such genes are crucial for this approach. One starting point is to clone the gene of interest using degenerate primers, given the sequence from a known homolog(s). Another strategy is to produce a large EST database from a cDNA library (ideally, a normalized library). EST sequencing can be an important resource for any new model system and will facilitate future lines of research, ranging from candidate gene discovery to phylogenomics to eventual genome sequencing [39,40].

Reliable expression protocols Established *in situ* hybridization protocols are also an essential tool for any system, but even more so for evo-devo in which studies are often rooted in comparative gene expression. This technique allows expression patterns to be visualized in greater detail, both spatially and temporally, relative to coarser approaches such as RT-PCR. Species-specific or cross-reactive antibodies are also useful because antibody localization techniques are more technically forgiving than *in situ* hybridization. Cross-reactive antibodies have been a boon to many broad comparative studies for plant and animal systems and could negate the need for gene cloning (e.g. [41,42]).

Large numbers of organisms at many developmental stages. Given that *in situ* hybridization can be a challenging technique to optimize, its success depends on the availability of a copious supply of organisms at different developmental stages to ensure that the correct time points and tissues are sampled. Ideally, these individuals can be raised in large numbers in a controlled laboratory environment, although this is not an absolute necessity. It is possible to use organisms collected from the wild, as long as they are available in large numbers at appropriate stages. Admittedly, this approach has limitations, but several studies have shown that broad expression surveys and even functional experiments can be conducted on wild-collected embryos [13,35,43].

At least one functional tool (but the more, the better). Inferring gene function from gene expression is not always a straightforward process. Some developmental genes are expressed in very precise patterns and yet have no known function(s) (e.g. Ref. [44]), whereas others have conserved expression patterns but new function(s) (either developmental, biochemical or both; e.g. Ref. [28,29]). Therefore, it is important to develop reverse genetic tools to test the function of the gene of interest, for example, by the knockdown of endogenous gene function. With the advent of RNA interference (RNAi) and morpholino technology, these reverse genetic approaches are becoming easily accessible [45–47]. As overexpression techniques are also developed, including direct injection of mRNA molecules or proteins into tissues, viral vectors or transgenic expression assays,

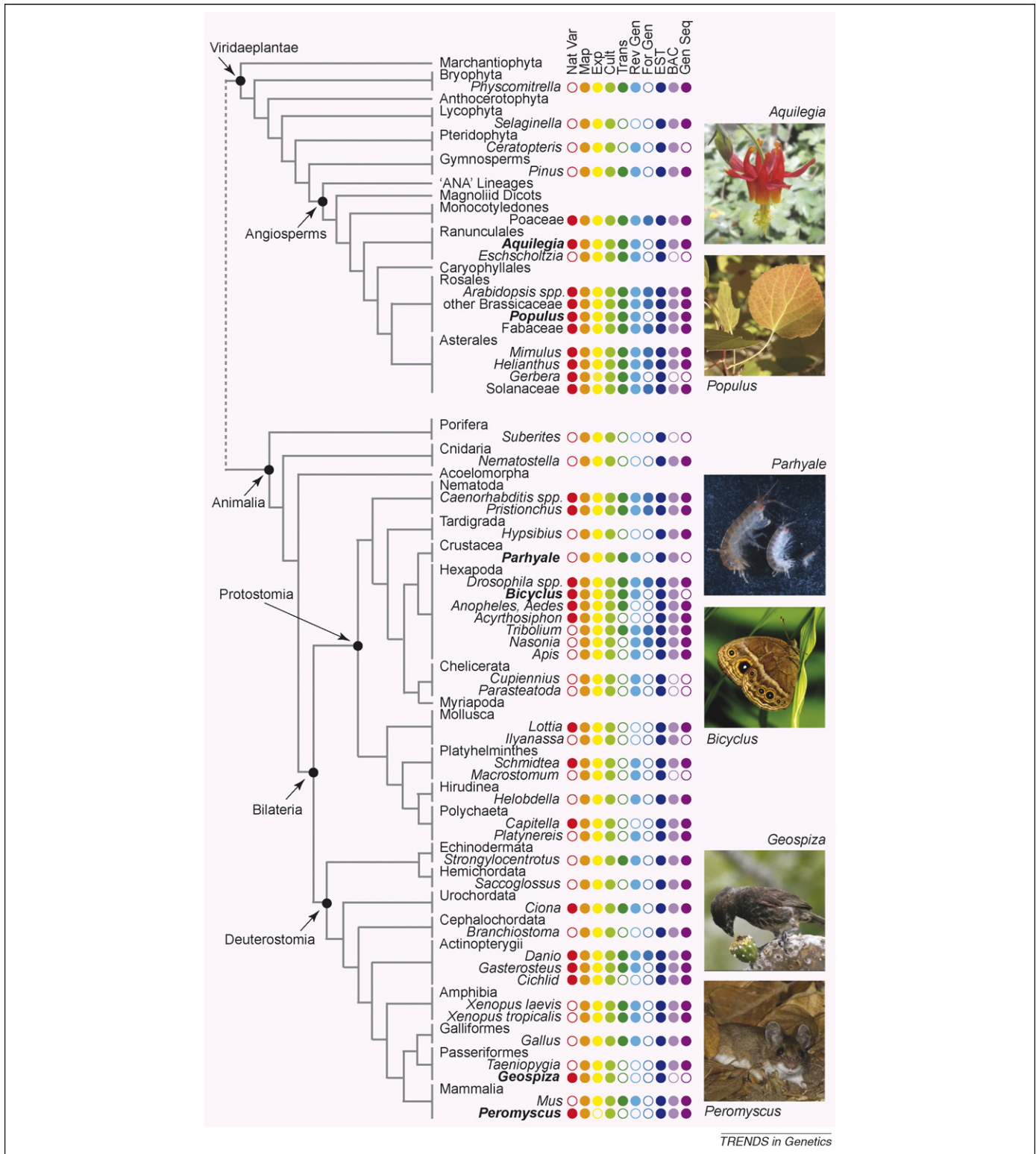


Figure 1. Established and emerging model systems across the metazoan and plant kingdoms. Simplified phylogenies of the metazoan and plant kingdoms show the relationships among both well-developed and nascent model systems. The colored circles to the right of each genus indicate the spectrum of available tools and resources, as determined by search of PubMed, Web of Knowledge and public websites, as well as personal communications. The availability of ten different classes of tools are considered for each model: Nat Var is the presence of natural genetic variation; Map, the availability of a genetic map, which is often, but not necessarily associated, with the ability to perform classic genetic crosses; Exp, established gene expression techniques; Cult, the ability to culture the organism in the laboratory or easily obtain it in the field; Trans, established transgenic techniques; Rev Gen, established reverse genetic tools; For Gen, the ability to conduct forward genetic screens; EST, the availability of an express sequence tag (EST) database; BAC, the availability of bacterial artificial chromosome (BAC) library resources and associated physical genome maps; and Gen Seq, having a full genome sequence. Filled circles indicate that the tool is currently available or actively under development; empty circles, that it is not yet available. In some cases (e.g. *Arabidopsis* spp., *Caenorhabditis* spp.), the genus name represents several developed model species. Images for models that are shown in bold are displayed at right.

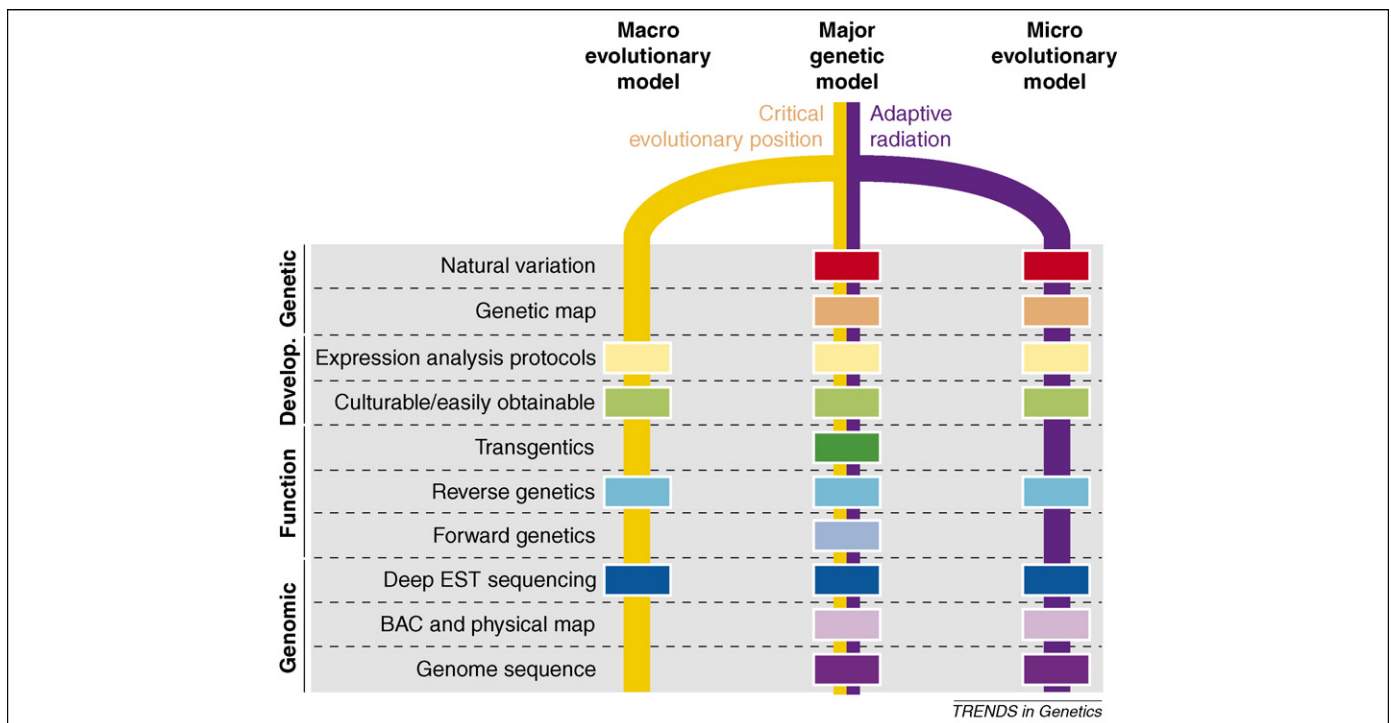


Figure 2. Essential tools for new model systems. Depending on the research question and the evolutionary timescale, different tools are important. For macroevolutionary questions, taxa are chosen to sample deep genetics and phenotypic divergence, whereas for microevolutionary questions, recent evolutionary history, such as adaptive radiation, is needed. Major genetic models can contribute to either of these types of questions. Although major genetic models are expected to have the full range of genetic tools and resources, it is not necessary for all model species to have all tools. The major tools we highlighted comprise four classes: genetic, developmental, functional and genomic. The tools are color coded to correspond to the broader survey shown in Figure 1. The most crucial tools for all models are expression analysis protocols, reverse genetic tools, expressed sequence tag (EST) databases and either the ability to be cultured in the laboratory or easily obtained from the wild. Microevolutionary models also require natural variation within the species, classical genetic approaches that depend on genetic maps and a fully sequenced genome or, at least, the availability of BAC library resources. Transgenic approaches are useful for any model type, but are not always essential.

phenotypes of interest can be further tested and are expected to mirror the results of the downregulation experiments. Such complementary tools can establish both requirement and sufficiency (e.g. when ectopic expression leads to the activation of a modular developmental cascade) for a gene's role in the development of a particular trait.

Microevolutionary tools

To address microevolutionary questions, which typically involve closely related species, populations of a species or phenotypic plasticity within a species, it is a good idea to choose models with an interesting and well-documented recent diversification (e.g. adaptive radiation, hybridization, adaptive phenotypic plasticity). The tools and resources required to uncover the molecular basis of trait evolution at this time scale have some overlap with macroevolutionary tools but also have some important differences (Box 2).

A large EST database from a normalized cDNA library.

In the context of microevolutionary studies, an EST database can be useful for identifying novel candidate genes in specific tissues associated with the trait under investigation. Alternatively, a comprehensive library can serve as the foundation for microarray design, which can be used to identify differentially expressed genes across seasonal forms (phenotypic plasticity), populations of one species or closely related species (e.g. Ref. [34]). In addition, EST sequencing that uses tissue from a F1 or F2 hybrid enables the rapid identification of polymorphisms

for genetic mapping (e.g. Ref. [48], see below). This mapping can itself be conducted using microarrays, thereby facilitating QTL identification [49].

A genetic map with reliable markers. A well-resolved genetic map is an essential tool for any study that aims to understand how many loci are involved in phenotypic variation, along with their relative effect size and genomic location. Although marker development and the subsequent construction of a genetic map was once an extremely onerous task, many array-based and new sequencing approaches have revolutionized this process [50,51].

The ability to do traditional genetic crosses. The ability to cross organisms in the laboratory, for at least two generations, is required to map the alleles causing the divergent phenotypes. When variation at candidate genes is suspected, an association study with single-pair crosses bearing distinct markers placed at the candidate loci can help to further implicate these loci in trait evolution (see Refs. [52,53]). Another option, however, is to use existing natural variation (e.g. population genomics), which can allow mapping approaches without controlled crosses (see [54] for a review). Although the use of controlled crosses is ideal, because it can also control for environmental effects (e.g. Box 3), the use of large pedigrees is often the only option when studying large or long-lived individuals (e.g. reed deer, sheep, birds; reviewed in Refs. [55,56]).

Box 3. Models for global climate change

Along with the evolutionary questions outlined here, an increasingly important and practical issue is how to best study and quantify the impact of abiotic change, such as global warming, on biodiversity and ecology. As knowledge accumulates about the genetic variation underlying adaptive traits within populations, a new question emerges: can this new information be used to make better decisions about conserving vital genetic variation, or making informed decisions about determining appropriate genotypes for restoration? These and related challenges are addressed by the applied fields of conservation genetics and genomics, which require basic information from studies of evolution, population genetics and genomics, as well as the dissection of adaptive traits. In addition, many emerging model systems (e.g. *Nematostella*, *Saccoglossus*, *Populus* and *Aquilegia*, among others) have the potential to serve as 'sentinel species' – canaries in the coal mine so to speak – because their sensitive biology can provide feedback on changing habitats. For example, like many plants, *Populus* ties its life cycle to environmental cues, such as temperature [68], making it a useful model to study responses to climate change. In addition, *Populus* might be useful for such applications as carbon sequestration, bioremediation and even biofuels; thus, the recent completion of the *Populus* genome sequence represents an important milestone [69].

A sequenced genome or bacterial artificial chromosome library. Many QTL studies identify candidate genomic areas containing hundreds of genes. A genome sequence allows additional targeted markers and/or the best candidate genes to be identified for further examination. Thus, the genome sequence of a new model species (or a close relative with conserved linkage) is particularly important for microevolutionary questions. Bacterial artificial chromosome (BAC) libraries, on the other hand, can substitute for full genome sequencing as they allow relevant BAC clones to be identified and sequenced on a more limited scale.

At least one functional tool. Once candidate loci are identified, functional assays are required to test whether changes in that gene(s) is causative. To determine whether an amino acid change contributes to the trait, one option is to produce *in vivo* or *in vitro* assays to test alternative proteins. If transcriptional regulatory changes are suspected, gene overexpression or knock-down (e.g. RNAi, see above) can determine whether these manipulations affect the trait, and specific regulatory elements can be tested by *in vivo* or *in vitro* reporter assays. In plants, near isogenic lines (NILs) are also valuable to assess the specific contribution of a locus to a particular trait as they allow QTL regions to be studied in isolation (e.g. Ref. [57]). Transgenics, however, are often the most informative assays, and also are the most difficult to perform, because they can determine the specific contribution of genes or mutations to a phenotype.

For microevolutionary systems, recent phenotypic diversification, deep EST sequencing and ease of culture or availability are essential (Figure 2). Many other techniques including expression protocols and reverse genetics are highly valuable, whereas the importance of transgenics and classical genetics depends on the questions being asked.

Concluding remarks and future perspectives

Recently developed model species already have made a significant contribution to our understanding of ecological

and evolutionary processes. In addition to the major questions discussed above, the concept of developmental modularity has been refined with data using emerging model organisms including non-*melanogaster* Drosophilids and finches [34,35,58]. Also, the persistently intractable issue of homology is being re-examined using comparative gene expression and functional tests in models such as ferns, butterflies and sea anemones [6,14,59]. Novel systems can also offer information about phylogenetic relationships by providing new informative markers either as sequence data [40] or morphological and developmental characters [60,61]. Finally, if current models prove inefficient or unsatisfactory, new model systems can and should be developed to address questions of practical and clinical significance. For example, new insights into stem cell biology and regeneration are promised by flatworms [62], and transgenic advances in chickens, including the development of economically important traits, could have important applications [63].

Establishing a new model system is not as laborious or expensive as it once was. This discussion of important characteristics to answer specific types of biological questions should help pinpoint candidate species for new model development. As long as samples can be easily obtained, even if they cannot be cultured in the laboratory, deep express sequence tag (EST) and genome sequencing is possible. However, developing a new model system can present challenges: a pre-existing research community might not exist, and specific tools might have to be developed, possibly requiring significant investments of time and/or money. Moreover, we acknowledge that every study organism is not suited for development as a model system, whether it is because of factors such as large genome size, extremely long generation time or limited research questions. Ultimately, increased resources from funding agencies will have to stretch to support the development of new community resources (e.g. transgenics or culture facilities) and the creation of preliminary tools (e.g. EST collections or bacterial artificial chromosome libraries) for emerging model systems, making it even more important that choices be made with care. If given the proper support, August Krogh's comment that 'For many problems there is an animal [or plant] on which it can be most conveniently studied' [64] can be realized.

References

- Jenner, R.A. and Wills, M.A. (2007) The choice of model organisms in evo-devo. *Nat. Rev. Genet.* 8, 311–319
- Mandoli, D.F. and Olmstead, R. (2000) The importance of emerging model systems in plant biology. *J. Plant Growth Regul.* 19, 249–252
- Tucker, S.C. and Hodges, S.A. (2005) Floral ontogeny of *Aquilegia*, *Semiquilegia*, and *Isopyrum* (Ranunculaceae). *Int. J. Plant Sci.* 166, 557–574
- Kramer, E.M. *et al.* (2007) Elaboration of B gene function to include the identity of novel floral organs in the lower eudicot *Aquilegia* (Ranunculaceae). *Plant Cell* 19, 750–766
- Keys, D.N. *et al.* (1999) Recruitment of a *hedghog* regulatory circuit in butterfly eyespot evolution. *Science* 283, 532–534
- Monteiro, A. *et al.* (2006) Comparative insights into questions of lepidopteran wing pattern homology. *BMC Dev. Biol.* 6, 52
- Sakarya, O. *et al.* (2007) A post-synaptic scaffold at the origin of the animal kingdom. *PLoS ONE* 2, e506
- Ohno, S. (1970) *Evolution by Gene Duplication*, Springer-Verlag
- Floyd, S.F. and Bowman, J.L. (2007) The ancestral developmental tool kit of land plants. *Int. J. Plant Sci.* 1, 1–35

- 10 Barakat, A. *et al.* (2007) Conservation and divergence of microRNAs in *Populus*. *BMC Genomics* 8, 481
- 11 Stark, A. *et al.* (2007) Systematic discovery and characterization of fly microRNAs using 12 *Drosophila* genomes. *Genome Res.* 17, 1865–1879
- 12 Extavour, C.G. (2005) The fate of isolated blastomeres with respect to germ cell formation in the amphipod crustacean *Parhyale hawaiiensis*. *Dev. Biol.* 277, 387–402
- 13 Lowe, C.J. *et al.* (2006) Dorsoventral patterning in hemichordates: insights into early chordate evolution. *PLoS Biol.* 4, e291
- 14 Finnerty, J.R. *et al.* (2004) Origins of bilateral symmetry: Hox and dpp expression in a sea anemone. *Science* 304, 1335–1337
- 15 Orr, H.A. (2005) The genetic theory of adaptation: a brief history. *Nat. Rev. Genet.* 6, 119–127
- 16 Fishman, L. *et al.* (2002) Minor quantitative trait loci underlie floral traits associated with mating system divergence in *Mimulus*. *Evolution Int. J. Org. Evolution* 56, 2138–2155
- 17 Steiner, C.C. *et al.* (2007) Adaptive variation in beach mice produced by two interacting pigmentation genes. *PLoS Biol.* 5, e219
- 18 Joron, M. *et al.* (2006) A conserved supergene locus controls colour pattern diversity in *Heliconius* butterflies. *PLoS Biol.* 4, e303
- 19 Bradshaw, H.D. and Stettler, R.F. (1995) Molecular genetics of growth and development in *Populus*: mapping QTLs with large effects on growth, form and phenology traits in a forest tree. *Genetics* 139, 963–973
- 20 Blazquez, M.A. *et al.* (2003) A thermosensory pathway controlling flowering time in *Arabidopsis thaliana*. *Nat. Genet.* 33, 168–171
- 21 Maloof, J.N. *et al.* (2001) Natural variation in light sensitivity of *Arabidopsis*. *Nat. Genet.* 29, 441–446
- 22 Gutteling, E.W. *et al.* (2007) Mapping phenotypic plasticity and genotype-environment interactions affecting life-history traits in *Caenorhabditis elegans*. *Heredity* 98, 28–37
- 23 Mitchell-Olds, T. and Schmitt, J. (2006) Genetic mechanisms and evolutionary significance of natural variation in *Arabidopsis*. *Nature* 441, 947–952
- 24 Hoekstra, H.E. and Coyne, J.A. (2007) The locus of evolution: Evo devo and the genetics of adaptation. *Evolution Int. J. Org. Evolution* 61, 995–1016
- 25 Wray, G.A. (2007) The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* 8, 206–216
- 26 Miller, C.T. *et al.* (2007) cis-regulatory changes in kit ligand expression and parallel evolution of pigmentation in sticklebacks and humans. *Cell* 131, 1179–1189
- 27 Hay, A. and Tsiantis, M. (2006) The genetic basis for differences in leaf form between *Arabidopsis thaliana* and its wild relative *Cardamine hirsuta*. *Nat. Genet.* 38, 942–947
- 28 Galant, R. and Carroll, S.B. (2002) Evolution of a transcriptional repression domain in an insect Hox protein. *Nature* 415, 910–913
- 29 Ronshaugen, M. *et al.* (2002) Hox protein mutation and macroevolution of the insect body plan. *Nature* 415, 914–917
- 30 Wray, G.A. *et al.* (2003) The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* 20, 1377–1419
- 31 Arendt, J. and Reznick, D. (2008) Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends Ecol. Evol.* 23, 26–32
- 32 Hoekstra, H.E. and Nachman, M.W. (2003) Different genes underlie adaptive melanism in different populations of rock pocket mice. *Mol. Ecol.* 12, 1185–1194
- 33 Colosimo, P.F. *et al.* (2005) Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science* 307, 1928–1933
- 34 Abzhanov, A. *et al.* (2006) The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. *Nature* 442, 563–567
- 35 Abzhanov, A. *et al.* (2004) Bmp4 and morphological variation of beaks in Darwin's finches. *Science* 305, 1462–1465
- 36 Whittall, J.B. *et al.* (2006) Convergence, constraint and the role of gene expression during adaptive radiation: floral anthocyanins in *Aquilegia*. *Mol. Ecol.* 15, 4645–4657
- 37 Orr, H.A. (2005) The probability of parallel evolution. *Evolution Int. J. Org. Evolution* 59, 216–220
- 38 Colosimo, P.F. *et al.* (2004) The genetic architecture of parallel armor plate reduction in threespine sticklebacks. *PLoS Biol.* 2, E109
- 39 Delsuc, F. *et al.* (2005) Phylogenomics and the reconstruction of the tree of life. *Nat. Rev. Genet.* 6, 361–375
- 40 Dunn, C.W. *et al.* (2008) Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* 452, 745–749
- 41 Panganiban, G. *et al.* (1997) The origin and evolution of animal appendages. *Proc. Natl. Acad. Sci. U. S. A.* 94, 5162–5166
- 42 Bharathan, G. *et al.* (2002) Homologies in leaf form inferred from KNOX gene expression during development. *Science* 296, 1858–1860
- 43 Lowe, C.J. *et al.* (2002) Gene expression and larval evolution: changing roles of distal-less and orthodenticle in echinoderm larvae. *Evol. Dev.* 4, 111–123
- 44 Choe, C.P. *et al.* (2006) A pair-rule gene circuit defines segments sequentially in the short-germ insect *Tribolium castaneum*. *Proc. Natl. Acad. Sci. U. S. A.* 103, 6560–6564
- 45 Gunsalus, K.C. and Piano, F. (2005) RNAi as a tool to study cell biology: building genome-phenome bridge. *Curr. Opin. Cell Biol.* 17, 3–8
- 46 Watson, J.M. *et al.* (2005) RNA silencing platforms in plants. *FEBS Lett.* 579, 5962–5987
- 47 Ekker, S.C. and Larson, J.D. (2001) Morphant technology in model developmental systems. *Genesis* 30, 89–93
- 48 Beldade, P. *et al.* (2006) A wing expressed sequence tag resource for *Bicyclus anynana* butterflies, an evo-devo model. *BMC Genomics* 7, 130
- 49 Gilad, Y. and Borevitz, J. (2006) Using DNA microarrays to study natural variation. *Curr. Opin. Genet. Dev.* 16, 553–558
- 50 Miller, M.R. *et al.* (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res.* 17, 240–248
- 51 Wittenberg, A.H. *et al.* (2005) Validation of the high-throughput marker technology DArT using the model plant *Arabidopsis thaliana*. *Mol. Genet. Genomics* 274, 30–39
- 52 Beldade, P. and Brakefield, P.M. (2002) The genetics and evo-devo of butterfly wing patterns. *Nat. Rev. Genet.* 3, 442–452
- 53 Hoekstra, H.E. (2006) Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity* 97, 222–234
- 54 Stinchcombe, J.R. and Hoekstra, H.E. (2007) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* 100, 158–170
- 55 Slate, J. (2005) Quantitative trait locus mapping in natural populations: progress, caveats and future directions. *Mol. Ecol.* 14, 363–379
- 56 Ellegren, H. and Sheldon, B.C. (2008) Genetic basis of fitness differences in natural populations. *Nature* 452, 169–175
- 57 Bradshaw, H.D. and Schemske, D.W. (2003) Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature* 426, 176–178
- 58 Ludwig, M.Z. *et al.* (1998) Functional analysis of eve stripe 2 enhancer evolution in *Drosophila*: rules governing conservation and change. *Development* 125, 949–958
- 59 Sano, R. *et al.* (2005) KNOX homeobox genes potentially have similar function in both diploid unicellular and multicellular meristems, but not in haploid meristems. *Evol. Dev.* 7, 69–78
- 60 Kadner, D. and Stollewerk, A. (2004) Neurogenesis in the chilopod *Lithobius forficatus* suggests more similarities to chelicerates than to insects. *Dev. Genes Evol.* 214, 367–379
- 61 Stollewerk, A. *et al.* (2001) Neurogenesis in the spider *Cupiennius salei*. *Development* 128, 2673–2688
- 62 Sánchez Alvarado, A. and Kang, H. (2005) Multicellularity, stem cells, and the neoblasts of the planarian *Schmidtea mediterranea*. *Exp. Cell Res.* 306, 299–308
- 63 van de Lavoie, M.C. *et al.* (2006) Germline transmission of genetically modified primordial germ cells. *Nature* 441, 766–769
- 64 Krebs, H.A. (1975) The August Krogh Principle: “for many problems there is an animal on which it can be most conveniently studied”. *J. Exp. Zool.* 194, 221–226
- 65 Browne, W.E. *et al.* (2006) Expression of otd orthologs in the amphipod crustacean, *Parhyale hawaiiensis*. *Dev. Genes Evol.* 216, 581–595
- 66 Monteiro, A. *et al.* (2007) The combined effect of two mutations that alter serially homologous color pattern elements on the fore and hindwings of a butterfly. *BMC Genet.* 8, 22
- 67 Ramos, D.M. and Monteiro, A. (2007) Transgenic approaches to study wing color pattern development in Lepidoptera. *Mol. Biosyst.* 3, 530–535
- 68 Bohlenius, H. *et al.* (2006) CO/FT regulatory module controls timing of flowering an seasonal growth cessation in trees. *Science* 312, 1040–1043
- 69 Groover, A.T. (2007) Will genomics guide a greener forest biotech? *Trends Plant Sci.* 12, 234–238