Phylogenies and the Evolution of Development

51

Ehab Abouheif

SUMMARY   The new field of evolutionary developmental biology (EDB) promises to resolve the black box that lies between evolution of the phenotype and evolution of the genotype. Phylogenies have played a central role in helping EDB achieve this goal. This chapter highlights several examples in the field to show how phylogenies have been important for drawing inferences on ancestral states of development and evolutionary modes of developmental change, and for distinguishing between homology and the convergence of developmental processes. Finally, the importance of phylogenies in selecting potential model systems for future evolutionary developmental studies is discussed.

As a discipline, evolutionary developmental biology (EDB) is both old and new, with its origins stretching as far back as Aristotle. The relatively recent discovery that all animals share a conserved set of developmental **regulatory genes** (transcription factors and signaling molecules) has induced a renaissance in EDB over the last few decades. This has placed EDB in a good position to fulfill its main promise of elucidating the black box between evolution of the genotype and evolution of the phenotype, as well as understanding how developmental genes and processes evolve. Phylogenies have always played a central role in EDB and, for the most part, this role has been a positive one. In some historical cases, however, phylogenies have had a negative impact on the progress of EDB. The most famous case is that of the renowned evolutionary developmental biologist, Ernst Haeckel, who actually coined the term *phylogeny*. Haeckel pioneered the use of ancestral adult features, which appear during the embryonic or juvenile stages of their descendants, to infer the evolutionary history of countless organisms. He claimed, for example, that the appearance of gill arches during the early ontogeny of a mammalian embryo was evidence that fish are the ancestors of mammals. This view was encapsulated in the evolutionary phrase (biogenic law) "ontogeny recapitulates phylogeny." Although this approach can be useful for reconstructing phylogenies, and is in fact still used today, it was Haeckel's assertion that phylogeny was the mechanical cause of ontogeny that most likely led to the eventual divorce between evolutionary and developmental biology for more than seventy years. Only

recently has phylogeny regained its central and positive role in EDB, although this time not as a mechanical cause of development, but as a necessary tool for drawing important interpretations about the history and evolution of developmental programs. The goal of this chapter is to briefly outline several ways in which phylogeny has and still can positively impact modern research in EDB. It will discuss how phylogenies have been critical for inferring ancestral states and evolutionary modes of development, as well as for distinguishing between **homology** and **convergence** in developmental processes. Finally, the chapter ends with a brief discussion of phylogenetic sampling and the selecting of appropriate model systems in EDB.

Phylogenies Are Important for Inferring Ancestral States of Development during Evolution

Inferring ancestral states of developmental processes or genes is critical to understanding how complexity in plants and animals has evolved. Phylogenetic analyses often focus on an important class of regulatory genes known as **homeobox genes,** which predate the split between plants, fungi, and animals. Evolutionary developmental biologists have expended much time and effort into understanding how the ancient origin of homeobox genes has affected the morphological diversity we currently observe in extant plants and animals. One general consequence of the ancient origin of homeobox genes is that they have been evolution-

arily conserved and reused in different combinations to give rise to similar phenotypes in closely and distantly related groups. The developmental genetic basis underlying the independent evolution of petals in different groups of flowering plants (angiosperms) is a case in point. Morphological and phylogenetic criteria show that petals have evolved independently at least six different times within the angiosperms. The independent evolution of these petals most likely occurred through the repeated recruitment of a suite of developmental control genes, including *APETALA3* (*AP3*) and *PISTILLATA1* (*P1*), which in angiosperms are a class of homeobox genes known as **MADS-box genes**. The different types of petals are thought to have evolved by recruiting *AP3* and *P1* in slightly different ways in each independent invention of the petal (see Chapter 14, Eudicotyledons). For example, both *AP3* and *P1* are involved in the development of petals in eudicotyledons, such as mouse-ear cress (*Arabidopsis thaliana*), as well as in monocotyledons such as rice. However, it is thought that the differences in these petal types have arisen through the evolution of differences in the downstream targets regulated by *AP3* and *P1* or through evolution of their interactions with their protein-binding partners.

Reconstructing the ancestral states of homeobox genes is also critical for understanding whether their ancient origin has resulted in a gradual increase in their number, and whether an increase in their number has in turn been associated with a corresponding increase in the complexity of plant and animal morphologies. Initial work on an important class of homeobox genes, called the **Hox genes** (see Box *Hox genes*), implied that this may have been the case. Given the remarkable ability of *Hox* genes to transform the body plans of animals (**Figure 51.1**), it was hypothesized that the number of *Hox* clusters is correlated to morphological complexity within taxonomic groups like the chordates. This hypothesis was based on the striking observation that amphioxus, the most basal extant chordate, possesses only one *Hox*

cluster in its genome, while the agnathan (lamprey and hagfish) genomes possess two, and mammals four. This correlation generally implies that the ancestors of all metazoan animals possess few and relatively simple complements of *Hox* and other developmental genes. Since this initial discovery, however, increased phylogenetic sampling of *Hox* clusters in different animals has indicated that the correlation between the number of developmental regulatory genes and morphological complexity is at best loose and inconclusive. The horseshoe crab, for example, considered a living chelicerate fossil because it shows many ancestral features, has four *Hox* clusters—more than expected relative to the more complex chordate lineages. Another example is the important model system, the zebrafish (*Danio rerio*), which possesses more *Hox* genes than mammals.

These conclusions regarding *Hox* genes can be generally extended to other important developmental genes, such as the *wingless* gene family, which encodes secreted signaling molecules that control cell fate in animal development. The sea anemone, *Nematostella vectensis*, is an ancestral cnidarian species that sits at the base of the metazoan tree. The *Nematostella* genome contains almost as many copies of the *wingless* gene family as are found in vertebrates. A phylogenetic analysis of all gene copies known from several animal phyla, such as chordates, arthropods, and nematodes, showed that each copy of the *wingless* gene family in *Nematostella* had an equivalent counterpart in humans. In fact more counterparts were found in *Nematostella* than in other important model systems, such as *Drosophila melanogaster* and *Caenorhabditis elegans*. This example clearly illustrates that basal metazoans possess a complex inventory of this important gene family, and has shattered the notion of a correlation between the increase in number of developmental genes and morphological complexity. In fact, it appears that the loss of developmental genes may be just as significant for the evolution of metazoan body plans as the gain of novel developmental genes through gene duplication.

Hox genes

In most animals *Hox* genes are expressed in the same relative order along the main body axis as they are positioned along the chromosome. Evolutionary changes in the number or expression of *Hox* genes have a profound influence on adult morphology in organisms. Changes in the expression of *Hox* genes along the antero-posterior axis can lead to changes in the identity of body regions, and are called homeotic transformations. In one of the most famous examples of a homeotic transformation, E. B. Lewis demonstrated that mutations in one of the *Hox*

genes, *Ultrabithorax*, switches the identity or fate of the third thoracic segment in *Drosophila* into the second. The third thoracic segment possesses a pair of small balancing organs called halteres, whereas the second possesses a pair of fully functional wings. Because the third thoracic segment acquires the identity of the second, the mutant adult fly develops a pair of wings instead of halteres, and the transformed mutant has two pairs of wings instead of one (see Figure 51.1).

Basic terms

Convergence: A trait that has evolved independently in two distantly related species, i.e., the trait was not present in the most recent common ancestor of the two species. Traditionally, convergent traits are thought not to share a common developmental basis.

Homeobox genes: Genes that contain a conserved stretch of 60 amino acid residues that are important for regulating development through transcription factor binding.

Homology: A trait that is present in the most recent common ancestor between two species.

Hox genes: Genes encoding transcription factors that control the identity or fate of different regions along the antero-posterior axis of the embryo, and are generally clustered (each cluster typically contains nine Hox genes) within the genome.

MADS-Box genes: Plant and animal genes encoding transcription factors that contain a highly conserved DNA-binding domain called the *MADS domain*. These genes play important roles in controlling the identity or fate of roots, flowers, and angiosperms.

Parallelism: A trait that has evolved independently in two closely related species, i.e., the trait was not present in the most recent common ancestor of the two species. Traditionally, parallel traits are thought to share a common developmental basis.

Regulatory gene: Genes that encode transcription factors or signaling molecules, which play important roles in regulating the expression of other genes during development.

Phylogenies Are Important for Inferring Evolutionary Modes of Development

Phylogenies are also critical for inferring modes of evolutionary change in development. Although the analytical tools to estimate evolutionary modes of change have been developed and applied, the use of phylogenies for inferring modes of developmental evolution has been underutilized. Little is known about whether developmental processes evolve in a gradual, neu-tral, or punctuated manner or whether development can evolve independent of adult morphology. In some cases, selection can act directly on developmental processes independent of adult morphology. For instance, in the sea urchin genus *Heliocidaris*, there are profound and rapid changes in the form of the larva in closely related species, even though the embryonic and adult forms are relatively conserved. Combining the fossil record and phylogenetic analyses, it has been argued that these larvae may have evolved according to a punctuated model of evolutionary change.

Another recent example comes from work on the nematode vulva. This is an essential sexual organ in adult nematode roundworms, and it is a highly conserved and homologous feature across species. Despite its conserved nature in adults, the developmental processes that build the nematode vulva vary substantially across species. For example, the inductive signals (received from surrounding cells) and the size of the equivalence group (cells that are competent to receive an inductive signal) that will give rise to the adult vulva have evolved between two closely related nematode species, *Caenorhabditis elegans* and *Pristionchus pacificus*. This phenomenon has been called developmental system drift (DSD). It was hypothesized that DSD underlying the adult nematode vulva evolves neutrally. However, phylogenetic analyses of 40 characteristics of vulval development in more than 52 nematode species showed that the majority do not evolve neutrally, but are instead biased. This bias is most likely a consequence of selection on these developmental characters. This finding is consistent with simulations performed on the gene network that underlies the development of wings in the winged and
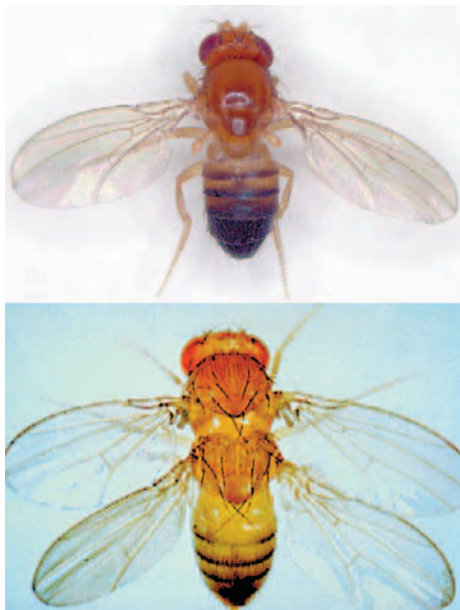

Figure 51.1   The *Ultrabithorax* mutation. Wild type (top) versus four-winged fly (bottom).
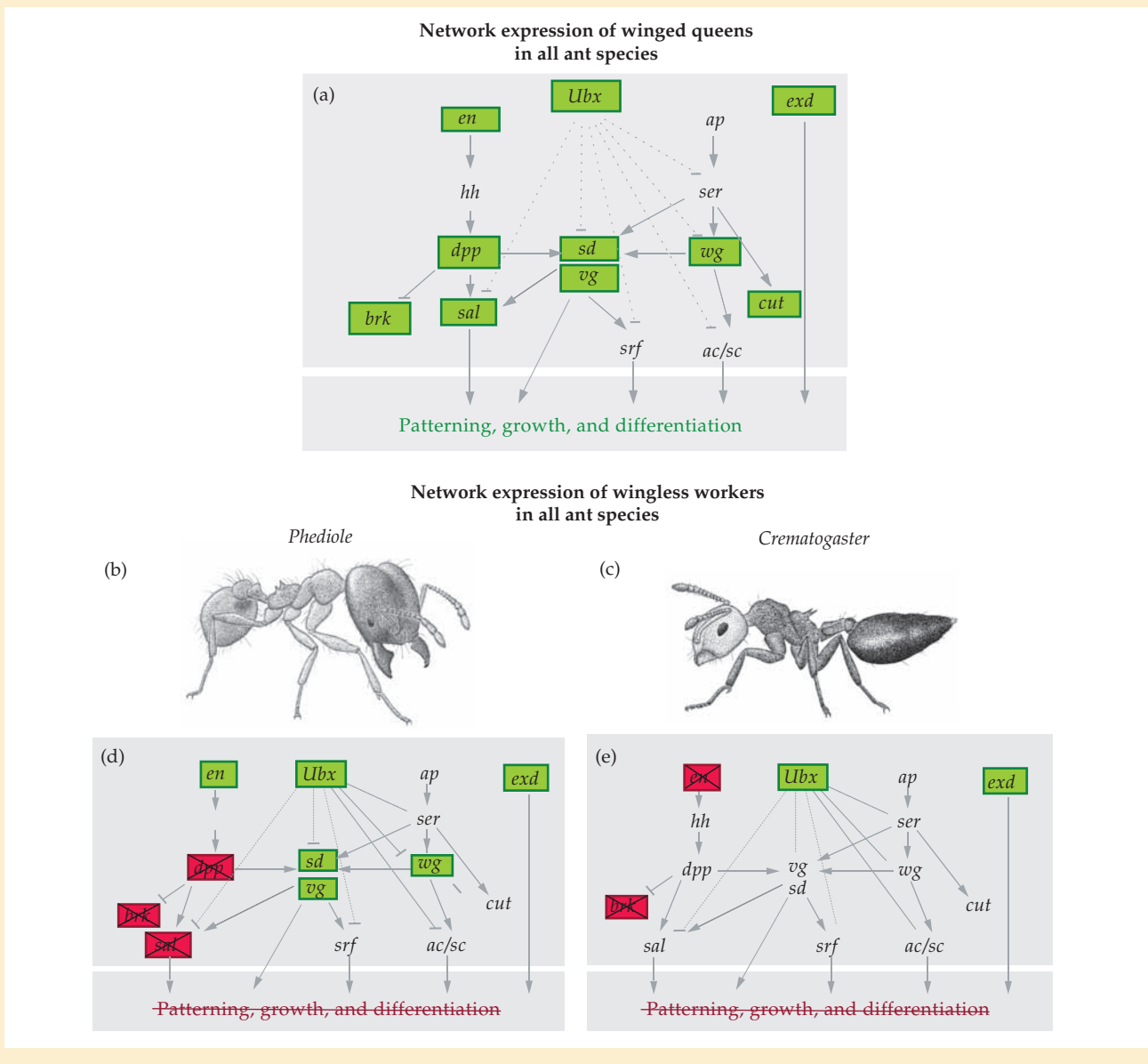
wingless castes in ants (see Box *Evolution of the gene network underlying wing polyphenism in ants*). They suggest that DSD in gene networks may generally be a consequence of a combination of selection and neutral processes.

Finally, phylogenies can specifically be used as a tool to quantitatively infer the mode of evolution in gene expression. For example, a maximum-likelihood approach was used to infer that gene expression in yeast evolves according to a nonphylogenetic model;

Evolution of the gene network underlying wing polyphenism in ants

Wing polyphenism is the ability of a single genome to produce winged queens that develop fully functional wings to fly up in the air for mating and dispersal, and workers that are completely wingless so they can live underground in response to environmental cues. Wing polyphenism in ants evolved just once, approximately 125 million years ago; as a consequence, the worker caste in all ant species is wingless. Expression of the gene network that controls wing development in *Dro-*

sophila and other holometabolous insects is conserved in the winged queens and males in all ant species examined to date (as indicated by green boxes and color in (A). In contrast, in wingless workers, expression of specific genes in the network is interrupted (as indicated by the red boxes and color) in different places in the workers of different ant species, like those from the genus *Pheidole* (B) and (D), and from the genus *Crematogaster* (C) and (E).

Network expression of winged queens in all ant species

(a)

Network expression of wingless workers in all ant species

Phediole

(b)

Crematogaster

(c)

(d)

(e)

in other words, that closely related genes are no more likely than distantly related genes to share similar expression patterns. This means that gene expression in yeast evolves so rapidly that any historical signal in gene expression has been erased. It is also possible to identify life-history or ecological factors that drive evolutionary changes in gene expression. In different species of echinoderms, there is wide variation in the timing of expression of *msp130*, a gene that encodes a protein that is part of the mesodermally derived envelope surrounding the echinoderm endoskeleton. The concentrated changes test, which is a parsimony-based approach, shows that large delays in the timing of expression of *msp130* (10–20 hours) are significantly associated with species that have independently evolved a switch from feeding to nonfeeding larvae. Conversely, small delays in the timing of *msp130* showed no significant association with any life history, larval, or adult features, and thus, are more likely to evolve neutrally.

Phylogenies Are Important for Distinguishing between Homology and Convergence of Developmental Genes and Processes

Distinguishing between homology and convergence is fundamental to any field of biology that either builds or uses phylogeny as a tool for inference. For the systematist, the primary motivation for distinguishing between homology and convergence is the simple fact that homology forms the basis on which phylogenies are built. Thus, when a systematist attempts to reconstruct the evolutionary history of a group of species, homologous traits are considered as *signal*, whereas convergent traits are considered *noise*. Conversely, for the comparative biologist, the primary motivation for distinguishing between homology and convergence is to understand how convergent traits have arisen in each of the species. His emphasis on convergence is based on the belief that convergent traits provide independent replicates or natural experiments for the evolution of a trait by natural selection. The finding that large delays in the onset of *msp130* expression are correlated with three species that independently evolved a switch from feeding to nonfeeding larvae is a good example of this emphasis and search for evidence of natural selection as the basis for convergent traits.

For the evolutionary developmental biologist, however, the primary motivation for distinguishing between homology and convergence is not only to identify the selective factors that may have driven the evolution of developmental genes and processes, but also to understand how developmental genes evolve relative to other levels of biological organization. Thus, the evolutionary developmental biologist must search for evolutionary associations or dissociations among the different levels of biological organization, such as embryology, genes, gene expression, and gene networks that underlie the development of morphological traits. Many striking examples of associations and dissociations exist and continue to emerge in the literature, such as the aforementioned example of the nematode vulva, in which it is homologous across species, but underlying developmental processes are not. Assessing associations and dissociations is best accomplished by simultaneously mapping traits from different levels of biological organization onto a phylogenetic tree. **Figure 51.2** outlines three possible scenarios. Scenario 1 (see Figure 51.2A) shows that a homologous morphological trait is encoded by homologous embryological and developmental genetic origins. A good example of this scenario is the evolution of locomotion and feeding appendages in 13 species of "crustaceans." Changes in the expression of *Ubx* are tightly associated with modification of anterior thoracic limbs into feeding appendages called *maxillipeds*.

Scenario 2 (see Figure 51.2B) shows that a non-homologous (convergent) morphological trait and embryological origin in two distantly related species develop from homologous developmental gene networks and genes. One of the best examples of this scenario comes from studies on the gene *distal-less* in different animal appendages. Distal-less protein is a transcription factor that plays an important role in organizing the growth and patterning of the proximodistal axes of limbs in *D. melanogaster*, and is expressed in the distal regions of the developing limbs. Distal-less is expressed in the appendages of animals from four additional phyla: (1) vertebrate fins and limbs, and tunicate ampullae within Chordata, (2) annelid parapodia, (3) onychophoran lobopodia, and (4) echinoderm tube feet. Historical evidence supports the independent (convergent) evolution of these different limb types. The fossil record of vertebrates, as well as extant hagfishes and lampreys for example, clearly indicates that the earliest members of this group lack limbs entirely. In contrast, comparisons of sequence and expression data indicate that the *distal-less* gene, and possibly its role in patterning proximodistal axes, is homologous in all of these phyla. This example highlights the possibility that the evolution of novel structures may involve recruiting existing machinery for generic developmental tasks, such as patterning or cell signaling, rather than inventing them completely de novo.

Scenario 3 (see Figure 51.2C) is similar to Scenario 2 (see Figure 51.2B) in that a nonhomologous morphological trait is being patterned by a homologous developmental gene or network. The main difference between the two scenarios is that independently
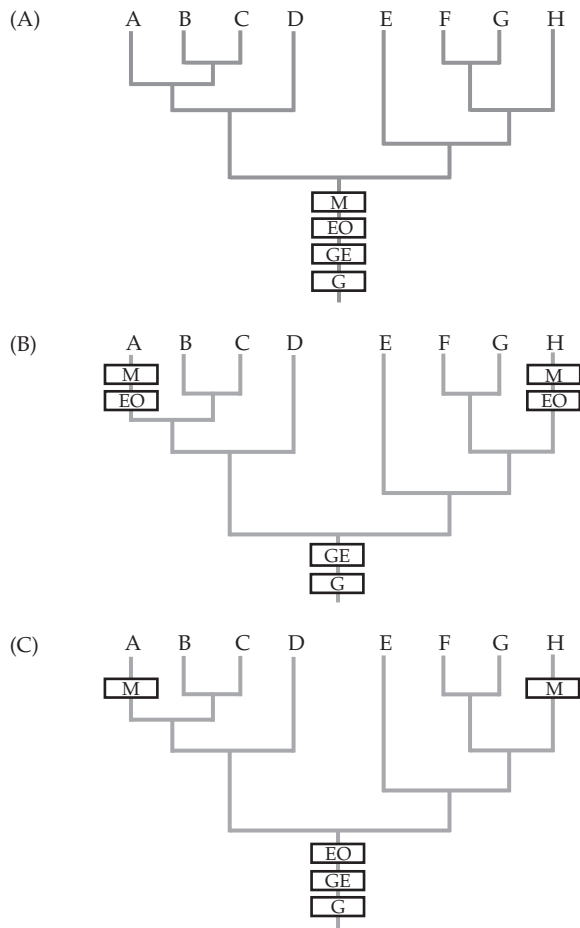
Figure 51.2 Three hypothetical scenarios. The boxes represent traits derived from several levels of the biological hierarchy: genes (G), gene expression patterns (GE), embryonic origins (EO), and morphological structure (M). The hierarchical traits are mapped onto a hypothetical eight-taxon phylogenetic tree (taxa labeled A–H). (A) Scenario 1 represents a scenario where all levels are homologous, and an evolutionary change at one level is reflected at others. (B) Scenario 2 represents a scenario of convergent evolution, because the morphological structure and its embryological origins evolved independently, while the underlying regulatory genes and their expression in a primitive or unrelated structure are homologous across distantly related taxa. (C) Scenario 3 represents a scenario of parallel evolution; the morphological structure evolved independently, while its underlying embryological origins, genes, and gene expression are homologous across closely related taxa.

evolved morphologies are being derived from homologous embryological origins, indicating that the species being compared are most likely closely related. A good example of this scenario is a study showing that expression of the gene *yellow* is associated with both the independent gain and loss of pigmentation spots on the wings of male drosophilid flies. In two drosophilid species, which have gained the spot independently, *yellow* expression has been activated to give rise to the spot through two different regulatory

binding sites. Conversely, in two drosophilid species that have lost the spot independently, *yellow* expression has been inactivated through the same regulatory binding site.

This comparison between Scenarios 2 and 3 has important implications for evolutionary theory in general. It questions the classical distinction between convergence and **parallelism**, two forms of independent evolution. Traditionally, convergence was defined as the independent evolution of traits between distantly related species that do not share a common developmental basis, whereas parallelism was defined as the independent evolution of traits between closely related species that do share a developmental basis. In this context, Scenario 3 differs from Scenario 2 only in a matter of degree and evolutionary time. Scenario 2 occurs between distantly related species whose morphologies and the embryological origins are truly convergent, whereas genes and the network they are embedded in are homologous (although they have often diverged considerably in their sequence and upstream regulators and downstream targets). Scenario 3, on the other hand, occurs largely between closely related species, where the morphologies and embryological origins are strikingly similar and the homologous genes encoding for them have diverged only in their *cis*-regulation. At the level of morphology, the types of changes inherent in Scenario 2, such as the independent evolution of limbs and eyes in flies and vertebrates, would have been clearly considered to be truly convergent prior to the discovery that they are also encoded by homologous genes. Therefore, the classical definition of convergence and parallelism is challenged, because both convergent and parallel traits can share a homologous developmental genetic basis, though to different degrees.

This ambiguity has led to a general disagreement about how to define or classify parallelism. Some authors have argued that the term parallelism should be replaced by the term convergence, whereas others suggest that the term convergence should be replaced by the term parallelism. Finally, parallelism has even been classified as a form of homology. Instead of trying to classify away this gray zone and ambiguity between these terms, it is important to acknowledge this gray zone as a crucial and missing phase in the evolutionary process. More recently, parallelism (Scenario 3) has been conceived as a mesoevolutionary phase that links micro- and macroevolution. Although it is beyond the scope of this chapter to fully explore the importance of this gray zone for evolutionary theory, this discussion clearly highlights that we still have much to learn from this hierarchical approach, and that there is immense potential in simultaneously mapping traits at different levels of biological organization onto a phylogenetic tree.

Phylogenies Are Important for Selecting Study Taxa in EDB

One final issue to consider is the role that phylogenies play in the selection of taxa to study in EDB studies. There are two important aspects to examine in this discussion. The first is how to choose model systems for study in EDB, and the second is whether to sample broadly across species or select a few species for in-depth analysis. Selecting a particular species or a group of species as model systems for study requires a large investment of resources, such as building up genetic tools and obtaining a genome sequence for the species of interest. How, then, shall we select our model systems—should this choice be based on the phylogenetic position of a species in the larger taxonomic group, or should it be based on ease of experimental and genetic manipulation? Clearly, there needs to be a balance between both, although it is arguable that phylogenetic considerations for the selection of model organisms should be at least as important as ease for lab studies, if not more so. Within the angiosperms (flowering plants), one model system for developmental genetic studies is the tobacco plant *Nicotiana tabacum* (**Figure 51.3A**). Although *Nicotiana* has revealed fundamental genetic architectures, it is a core eudicotyledon, and is considered to be a derived angiosperm. To understand the genetic basis underlying the wide variation in floral morphologies, other phylogenetically more basal species must be studied—for instance, the opium poppy plant, *Papaver somniferum* (**Figure 51.3B**), which belongs to the order Ranunculales. This order is closely related to, but is not included in, the core eudicots, and thus occupies a key phylogenetic position with respect to the genetic pathways that control petal specification. It is therefore useful for elucidating whether these genetic pathways are conserved in more distant angiosperm taxa or if they have diverged. In fact, MADS-box genes have undergone gene duplication and have diverged in function in the opium poppy plant relative to some of the core eudicots. From the perspective of experimental and genetic manipulation in the lab, establishing the opium poppy plant as a model system is challenging: it produces few flowers per plant, and stable transformation systems are just now being established. Furthermore, the gene duplication events make it difficult to assess gene function relative to the core eudicots. It is clear, however, that if the goal is to understand floral diversity, then given the phylogenetic position of the opium poppy plant, it must be made a priority for study.

In the above example, the challenge of balancing phylogenetic position and ease of experimental manipulation in selecting model systems for study goes far beyond trying to establish just one or two alternative model systems like the opium poppy plant. When we consider that there are around 300,000 species of
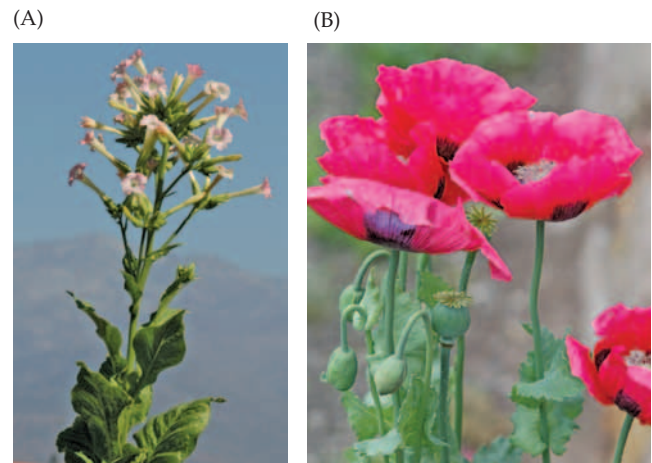
Figure 51.3 In angiosperms, *Nicotiana tabacum* (A) is a model system where functional analyses of MADS-box genes have been performed by means of heterologous ectopic overexpression in flowers. In contrast, *Papaver somniferum* (B) is a non-model species where virus-induced gene silencing has been shown to be an effective tool for gene knockouts. A strong case can be made that *Papaver somniferum* (B) would make a better Evo-Devo model system than *Nicotiana tabacum* (A) because of its key phylogenetic position.

flowering plants, a fundamental question arises—to what extent should we sample broadly across different species within a group, versus going in depth between groups? The field of phylogenetic systematics has dealt with an analogous problem. There has been much debate as to whether phylogenetic resolution and accuracy are improved by sequencing more genes from few taxa or sequencing few genes from many taxa. Clearly both would be desirable, but in a research climate of limited resources, we are forced to choose between the two strategies. While adding particular taxa can break up long branches in a phylogeny, thus avoiding incorrect topologies, sequencing more genes within fewer taxa, a phylogenomic approach, has been shown to increase phylogenetic resolution. Evolutionary developmental biologists face a similar conundrum. There are clearly not enough resources to develop every species in a model system, and at the same time, it is important to assess broad evolutionary trends in developmental processes across taxa. Although there is no easy solution, the suggestion of creating *anchor taxa* that both occupy key phylogenetic positions and are also easily manipulable in the lab seems sound.

In conclusion, phylogenies have had a fundamental impact on studies in EDB. The above examples show just a fraction of the potential for how phylogenies can be used to make inferences about evolution and development. As developmental data within and between species accumulate, and analytical tools become more sophisticated, phylogenies will continue to play a central and positive role in EDB.

Basic Bibliography

Abouheif, E. 1997. Developmental genetics and homology: A hierarchical approach. *Trends in Ecology and Evolution* 12: 405-408.

Carroll, S. B., Grenier, J. K., and Weatherbee, S. D. 2005. *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design.* Malden, MA: Blackwell Publishing.

Gould, S. J. 2002. *The Structure of Evolutionary Theory*. Cambridge, MA: Harvard University Press.

Hall, B. K. and Olson, W. M. (eds.). 2003. *Keywords and Concepts in Evolutionary Developmental Biology*. Cambridge, MA: Harvard University Press.

Wilkins, A. S. 2002. *The Evolution of Developmental Pathways*. Sunderland, MA: Sinauer Associates.