Delicious poison: genetics of *Drosophila* host plant preference

Noah K. Whiteman and Naomi E. Pierce

Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA

Insects use chemical cues to identify host plants, which suggests that chemosensory perception could be a target of natural selection during host specialization. Five papers using data from the 12 recently sequenced *Drosophila* genomes examined chemosensory gene function and evolution across specialist and generalist species. A functional study identifies odorant binding proteins that mediate loss of toxin avoidance in a specialist, and targeted genomic studies indicate specialists and island endemics lose chemosensory genes more rapidly than generalist and mainland relatives. Together, these studies suggest a mode of chemoreceptor evolution dominated by birth/death dynamics, coupled with a low level of potential positive selection.

Evolution of host specificity in insects

"Now I must feed myself with most delicious poison," declares Cleopatra in Shakespeare's *Antony and Cleopatra*, with reference to the tonic of Mandrake root *Mandragora officinarum* she is about to imbibe to help her sleep through Antony's unbearable absence. The concept of a 'delicious poison' captures the essence of the evolution of host plant specialization by many herbivorous insects, which imbibe plant toxins to escape from less-than-ideal circumstances. In the coevolutionary process, compounds that once served to deter attacks by plant enemies are often co-opted as feeding or oviposition stimulants by specialists that have evolved the ability to detoxify them [1]. In insects, this transition likely involves mutations conferring toxin resistance or detoxification by immatures, coupled with an affinity for the new host plants by ovipositing adults. The process of specialization might also select for changes in the chemosensory repertoire, including gene loss and/or gain. Whereas detoxification mechanisms have been relatively well studied and characterized [1], whether selection has acted upon genes involved in preference and avoidance during and after specialization is largely unknown [2]. The molecular and genomic consequences of host shifts, specialization and ecologically driven divergence potentially leading to speciation are now beginning to be illuminated with the help of genetic tools and comparative analysis of the 12 newly available *Drosophila* genomes [3] (Figure 1).

We here highlight five studies on the evolution of host specificity of *Drosophila* species that set a foundation for future research in this area. A functional study by Matsuo

and colleagues [4] demonstrates how an odorant binding protein gene–gene interaction and alteration in function mediates loss of gustatory avoidance behavior in the specialist *D. sechellia*. In two genomic studies, McBride [5] and McBride and Arguello [6] compare chemosensory families (olfactory and gustatory receptors) of species in the *D. melanogaster* subgroup (Figure 1) and find evidence for more rapid loss of these loci in the specialist than in the generalist sister species. Both studies also report higher rates of amino acid changing substitutions in intact receptor genes in specialists versus generalists. A fourth study by Vieira and colleagues [7] focuses on odorant binding proteins that interact with olfactory and gustatory receptors, and likewise observes more gene losses and elevated rates of amino acid substitutions in specialist species. However, a fifth analysis of the evolution of olfactory and gustatory receptor gene families by Gardiner and colleagues [8] argues that specialization and demography are sufficiently confounded in these groups that purely adaptive interpretations are not warranted.

Chemosensory genes and host preference in *Drosophila* species

Because adult female *Drosophila* must identify and lay eggs on prospective host plants for their offspring that subsequently feed on the microbes living on rotting plant tissue, factors influencing host selection are essential in understanding host specialization [2]. Loci underlying the chemoreception pathways include olfactory receptors (*Or*) and gustatory receptors (*Gr*) [9] that interact in specific ways with odorant binding proteins (*Obp*) (Figure 2). Changes in both receptor [10] and binding protein [4] gene expression are hypothesized to be important targets of evolution during host specialization and ecologically driven divergence in insects. For example, overexpression of the *Or43a* antennal receptor in *D. melanogaster* resulted in decreased avoidance of the fruit volatile benzaldehyde, showing how simple changes in gene expression in a single olfactory receptor can yield ecologically interesting phenotypes [11]. Intriguingly, the detoxification and chemosensory pathways might be linked, because one potential role for *Obp*s is to remove toxic ligands from the lymph (Figure 2) around the olfactory or gustatory receptor neuron [12].

The genus *Drosophila* encompasses well-known examples of specialist and generalist taxa, as well as those that appear to be in transition between these extremes [13]. For example, *D. sechellia* and *D. erecta* are well-known specialists that are close relatives of the genetic model *D. melanogaster*, a generalist (Figure 1).

Corresponding author: Whiteman, N.K. (nwhiteman@oeb.harvard.edu).
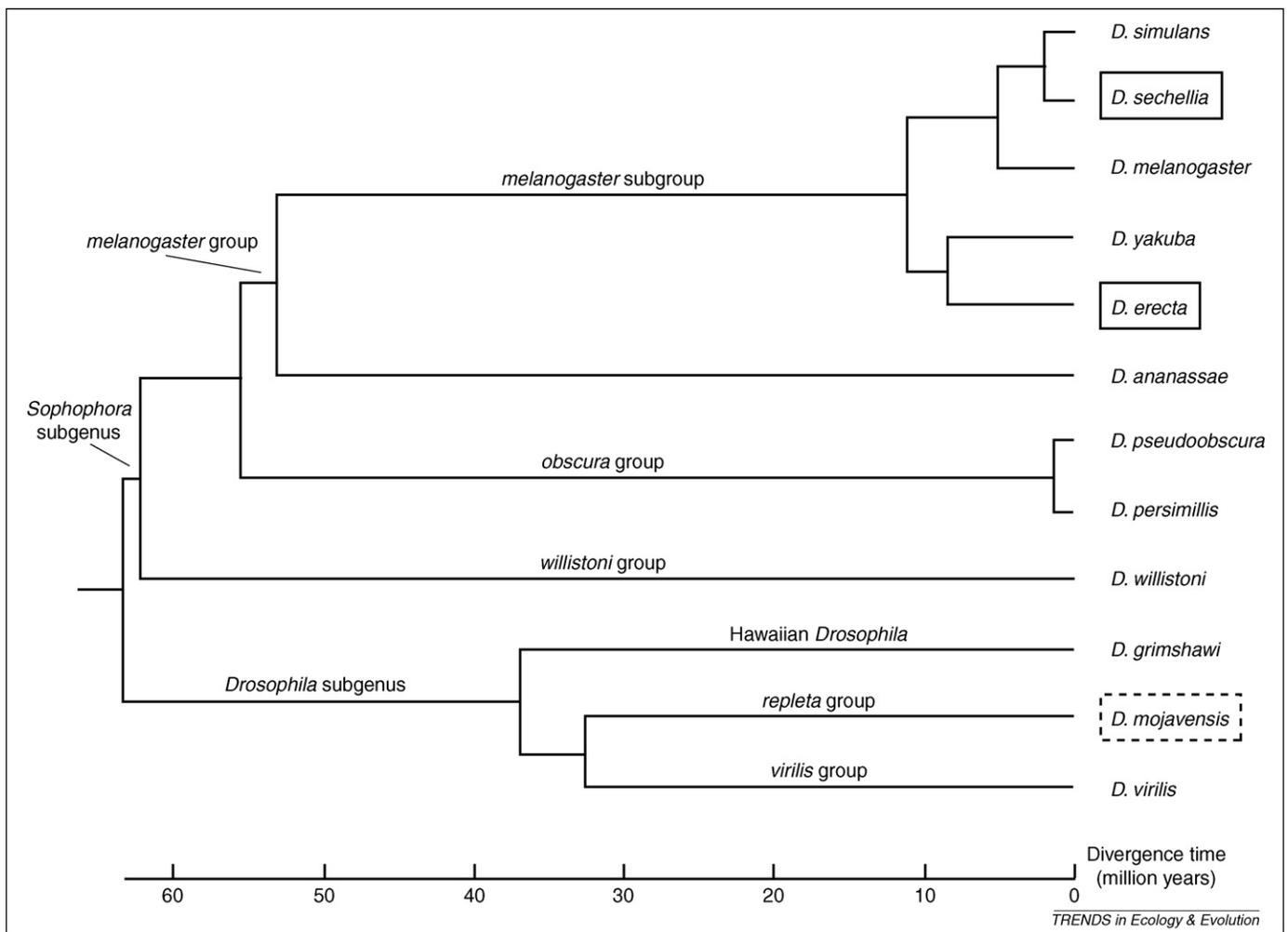
Figure 1. Whole-genome phylogeny of the 12 fully sequenced *Drosophila* species. Names of specialist species are enclosed in rectangles: the two *melanogaster* subgroup fruit specialists are in solid rectangles and the repleta-group cactus specialist is in a dashed rectangle. Adapted from Hahn [21]. © 2007 Hahn, Creative Commons Attribution License.

D. sechellia is a particularly interesting case. The species is endemic to the Seychelles Islands and lays eggs and feeds only on the ripe 'vomit fruit' of the Noni shrub *Morinda citrifolia* [14–17]. The fruit is toxic to all other *Drosophila* species, but *D. sechellia* is resistant to the toxins. Whereas female *D. sechellia* are attracted to the Noni, congeners are repelled by it, including the putative sister species *D. simulans*, a widely distributed fruit generalist [14–17]. Two main toxins in the Noni fruits, octanoic acid (OA) and hexanoic acid (HA), mediate interactions with *Drosophila* and are among the primary attractants for egg-laying females of *D. sechellia* [17]. The major quantitative trait loci responsible for conferring resistance to OA have been identified [17]. Key differences in both the number and the responses of particular chemosensilla to HA were also found between *D. sechellia* and *D. melanogaster* [18]. Important remaining pieces of the puzzle include identifying the basis of the avoidance of HA and OA by *D. simulans*, and the basis of tolerance or preference for these toxins in *D. sechellia*.

Matsuo and colleagues [4] created hybrids of *D. sechellia* and *D. melanogaster* to locate genes underlying host preference. The most promising gene localized was *Obp57e*, which is expressed in sensilla in the legs of *D. simulans*.

However, the translated region of this gene was essentially identical across *D. melanogaster*, *D. simulans* and *D. sechellia*, suggesting that regulatory and not structural differences were of key importance; indeed, a 4 base pair deletion in the region upstream of this gene in *D. sechellia* prevented *Obp57e* expression in transgenic *D. melanogaster* green fluorescent protein promoter assays driven by *Obp57e*.

To further explore the role of *Obp57e*, Matsuo and colleagues generated knockouts of *Obp57e* and of a second linked gene, *Obp57d*, in *D. melanogaster*. In an oviposition preference assay, *D. melanogaster* with the *D. simulans Obp57d/e* genes behaved like *D. simulans* and avoided HA and OA. The *D. melanogaster* flies with the *D. sechellia Obp57d/e* genes preferred high concentrations of OA, which partially recovered the *D. sechellia* phenotype. The authors interpret their findings in the context of the three factors hypothesized to underpin specialization in *D. sechellia*, including toxin resistance [14–17], olfactory preference [18] and gustatory avoidance by *D. simulans* [4]. They argue that the first mutation enabling the ancestor of *D. sechellia* to utilize the Noni was likely in a gene mediating the loss of gustatory avoidance. If this is correct, the question arises as to whether the loss of avoidance
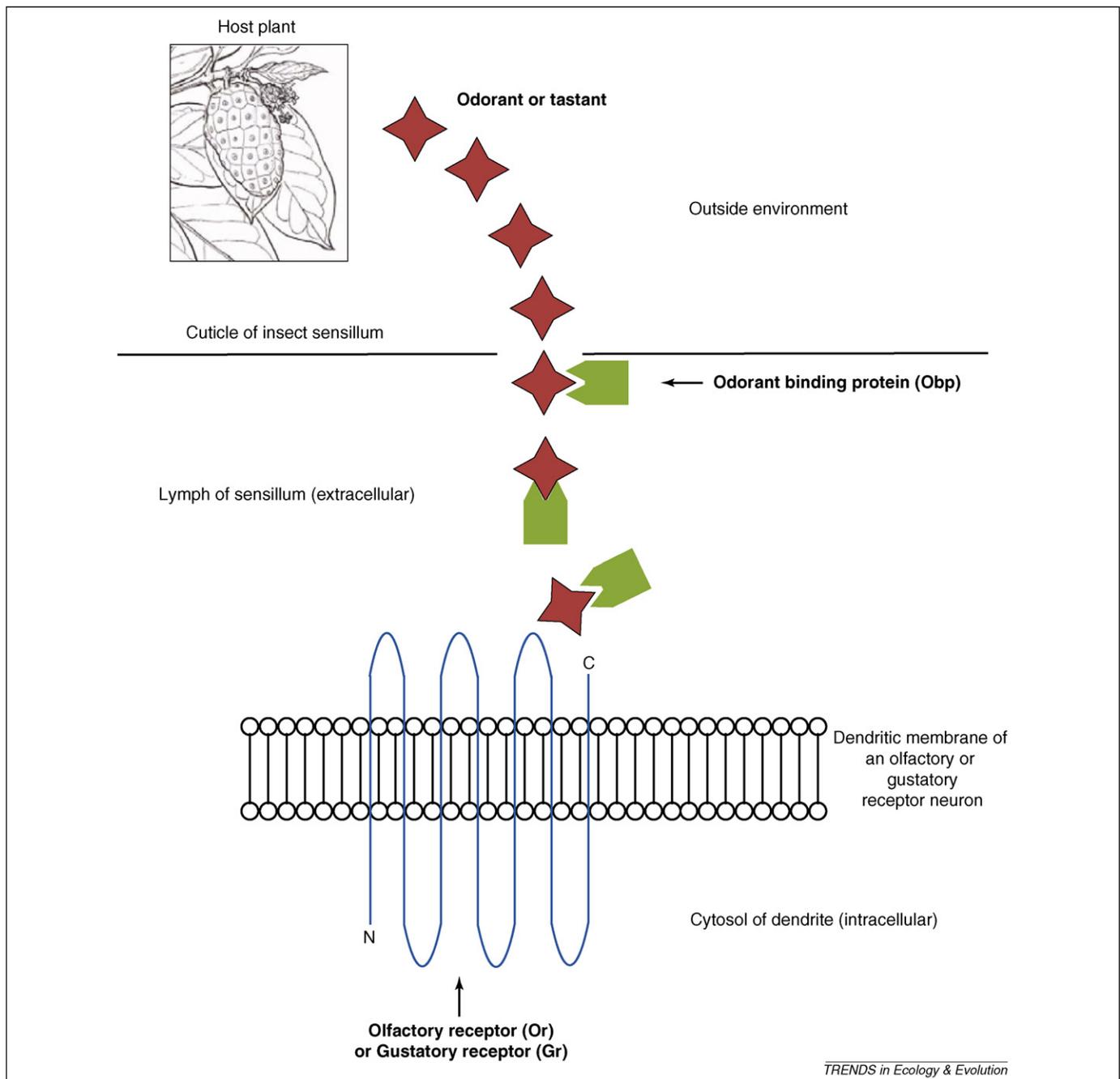
Figure 2. Simplified illustration of the first segment of the olfactory and gustatory signaling transduction pathway in an insect, depicting how olfactory receptors (*Or*) and gustatory receptors (*Gr*) might interact with odorant binding proteins (*Obp*). An odorant or tastant source produces a particular ligand that is recognized within an insect olfactory or gustatory sensillum. The ligand enters through a pore in the chemosensory sensillum, after which it encounters the aqueous lymph cavity housing the dendrite of the chemosensory receptor neuron. Crossing the lymph to the dendrite of the receptor neuron is an obstacle. Most odorants are hydrophobic, and a leading hypothesis explaining how the ligand crosses this hydrophilic fluid is that upon entering the lymph, an Obp, produced by accessory cells, binds to the ligand and transports it to the dendritic membrane, where it is presented to a transmembrane Or or Gr. The binding of a ligand to a receptor molecule results in depolarization of the dendrite and the production of an action potential [9]. Not depicted are the alternative hypotheses for the role of the odorant binding proteins [22]. The illustration of *M. citrifolia* was downloaded from the public domain website, http://www.fs.fed.us/global/iitf/pdf/shrubs/Morinda%20citrifolia.pdf.

behavior (and a change in gene function leading to specialization) is a special case or a general phenomenon.

A genomic view of the consequences of host specialization and niche reduction

The functional study of Matsuo and colleagues suggests that gene–gene interactions and changes in *Obp* gene expression, notably loss of chemosensory gene function, are key in mediating host preference in *D. sechellia*. Given these findings, one might expect that a shift from generalist to specialist diets is accompanied by a loss of chemosensory genes due to changes in selection pressure associated with specialization. Carolyn McBride [5] studied all members of the *Or* and *Gr* families along the *D. sechellia* and *D. simulans* lineages, and found that the specialist *D. sechellia* has lost *Or* and *Gr* genes ten times more rapidly than its generalist sister species *D. simulans*. Subsequently, McBride and Arguello [6]

Table 1. Characteristics of the three gene families considered, including a summary of the results of comparative genomic analyses within the *D. melanogaster* subgroup, which comprises five fully sequenced species, *D. erecta*, *D. melanogaster*, *D. sechellia*, *D. simulans* and *D. yakuba*, out of nine extant species

Gene family	Number of genes in *D. melanogaster*	Genomic location	Pattern of gene family loss and gain in *D. melanogaster* subgroup (five species)	Elevated rate of gene loss along lineages leading to *D. sechellia* and *D. erecta* versus closest generalist relatives in the *D. melanogaster* subgroup?	Higher rate of amino acid changing substitutions in *D. sechellia* and *D. erecta* versus generalists (relative to a random set of genes in each lineage)?	Refs
Olfactory receptor	61	Scattered throughout, clusters comprise only a few genes	64 present in most recent common ancestor, 4 duplications, 12 losses	Yes, 6 losses in *D. sechellia* and none in *D. simulans*; no in *D. erecta* or *D. yakuba*	Yes, but <1 (values >1 are typically suggestive of strong positive selection)	[5,6]
Gustatory receptor	68	Scattered throughout, clusters comprise only a few genes	74 present in most recent common ancestor, 0 duplications, 35 losses	Yes, 13 losses in *D. sechellia* and 2 losses *D. simulans*; yes, 11 losses in *D. erecta* and none in *D. yakuba*	Yes, but <1	[5,6]
Odorant binding protein	51	Highly clustered on a few chromosomes, ca. 80% on Muller elements C and E	51 present in most recent common ancestor, 5 duplications, 5 losses	Potentially in both; 1 loss in *D. sechellia* and none in *D. simulans*, 2 losses in *D. erecta* and 1 in *D. yakuba*	Yes, but <1	[7]

extended this analysis to include the other three sequenced members of the *D. melanogaster* subgroup, including another specialist, *D. erecta* (Figure 1). Overall, they found that: (i) the specialists *D. erecta* and *D. sechellia* showed a fivefold faster rate of gene loss in the *Gr* family; (ii) gene loss was particularly rapid in those genes important in detecting bitter compounds (Table 1); and (iii) the two specialists, *D. sechellia* and *D. erecta*, had higher ω ratios (Box 1) in intact *Gr* and *Or* genes versus their generalist relatives. The authors suggest that higher ω ratios could be explained by several non-mutually exclusive processes,

including genetic drift, purifying selection and diversifying selection (Box 1).

More recently, Gardiner and colleagues [8] compared gene gain and loss and ω ratios in the same *Gr* and *Or* gene families across all 12 sequenced *Drosophila* species. They were interested in disentangling the effects of restricted distribution (endemism) and specialization, because the remaining sequenced *Drosophila* species include the Hawaiian endemic *D. grimshawi*, a resource generalist, and *D. mojavensis*, a cactus specialist inhabiting the desert Southwest of North America. The most important finding from this study is that the proportion of pseudo-genized genes did not differ between the three specialist and nine generalist species, but the two endemic species had significantly more losses than the ten mainland species. The authors conclude that small population size might be more important than niche specialization in accounting for chemosensory gene loss, namely through genetic drift.

Vieira and colleagues [7] show that across the 12 *Drosophila* species with completely sequenced genomes, *Obp* genes in the two specialist lineages (*D. sechellia* and *D. erecta*) within the sequenced members of the *melanogaster* subgroup had higher ω ratios than the three generalist species (Table 1). This was the same pattern found for the *Gr* and *Or* genes [7,8]. Specifically, higher ω ratios were found for *Or*, *Gr* and *Obp* families as well as at the random set of loci along lineages leading to the two specialist species, suggesting that demography played an important role in producing the observed patterns [19]. Vieira and colleagues did not observe dramatic gene loss of *Obp* loci in specialist lineages, although some of the specialist lineages lost more genes than their closest generalist relatives. Interestingly, about the same number of genes are found in the *Or*, *Gr* and *Obp* gene families across the *melanogaster* group (Table 1), but there are major differences in copy number between the receptor genes and the *Obp*s in species outside of this clade [8].

Box 1. What are ω ratios?

The redundancy of the DNA code means that in protein-coding genes, nucleotide substitutions can be either synonymous (the amino acid coded by the codon remains the same) or non-synonymous (the corresponding amino acid changes). Comparing the rates of non-synonymous/synonymous substitutions (the ω rate ratio) of a gene or locus of interest between species can provide evidence as to whether that gene or locus is under selection [23]. A variety of maximum likelihood and Bayesian methods are available to estimate ω ratios for a given phylogenetic gene tree [23]. When $\omega < 1$, purifying selection is inferred because non-synonymous substitutions are deleterious with respect to fitness; when $\omega = 1$, neutral evolution is inferred because there is no difference in fitness between non-synonymous and synonymous substitutions; and when $\omega > 1$, positive selection is inferred because non-synonymous substitutions are favored by natural selection. In their most general form, ω ratios are averaged across all nucleotide sites, but because non-synonymous rates are often quite variable across a gene, ω values can also be estimated for individual loci.

Although it is possible to test for significant differences among ω values, the most conservative interpretation holds that adaptive evolution has occurred only when ω values are >1. However, even when ω values are >1, demographic forces can elevate ω ratios if there is an imbalance between genetic drift and purifying selection, such as in situations where the effective population size is low [5,6]. Because several non-mutually exclusive factors can affect ω ratios, comparisons using these data, which are always only correlative in nature, need to be interpreted with caution.

The research horizon

The five studies highlighted here lay an excellent foundation for a more informed exploration of chemosensory gene evolution and function in the Drosophilidae and other insects (Box 2). Clearly, factors associated with host specialization and demography (e.g. endemism) could be important in mediating the evolution of chemosensory gene families in *Drosophila*, but there is much left to learn. Correlative genomic studies cannot provide causal links; only careful functional studies will illuminate underlying mechanisms.

The functional study by Matsuo and colleagues illustrates the difficulty in dissecting the genetic basis of complex phenotypes even with sophisticated tools developed for *D. melanogaster*. If this is representative of the effort required to understand even partially the functional basis of candidate host preference genes in other species, the road ahead is likely to be challenging: *D. sechellia* and *D. simulans* are the only members of the 12 sequenced species that can be crossed with *D. melanogaster*. More generally, the study illustrates that it is not the detection of host cues *per se* that has changed during specialization, but rather the response to the signal (*D. melanogaster* and *D. simulans* were repelled by OA whereas *D. sechellia* was attracted to it).

The studies reviewed here address the general question of whether life-history specialization has resulted in gene loss. An underlying argument of most of these papers is that host plant specialization has resulted in a concomitant loss of genes, presumably because host plant specialists do not need to detect and respond to as wide a variety of toxic plant compounds as their generalist sister taxa. However, Gardiner and colleagues note that because host specialization and population size are confounded for many of the species, the observed gene loss might be due to demographic factors rather than host plant specialization. When demography is confounded with specialization, signatures of positive selection could also be explained by genetic drift (Box 1). Nonetheless, Gardiner and colleagues do identify several *Or* and *Gr* loci with ω ratios >1, which is the strongest evidence to date of positive selection acting on these loci.

Genomic studies of a greater number of taxa with different life histories might provide insights into alternative hypotheses for gene loss/gain associated with life-history evolution. For example, Gilad and colleagues [19] compare the evolution of *Or* genes in humans and chimpanzees, and find that only ~40% of *Or* genes in humans have intact coding regions compared with ~70% in great apes. Humans have a decreased reliance on smell and heightened visual acuity relative to great apes, and cook food to remove plant toxins, and so the authors argue that these life-history differences might have generated the striking pattern. However, as the authors admit, tests of selection in this study generally yielded ω values <1, and so although appealing, adaptationist interpretations might not be justified.

In conclusion, the five studies presented here help to set the course for understanding the role of chemosensory genes in host specialization. Overall, the comparative genomics studies show strong evidence of birth/death

Box 2. Future directions

Below we outline some future research directions for comparative genomic studies.

What is the right scale of comparison?
The comparative genomics studies [5–7] that focused on the sequenced members of the *melanogaster* subgroup considered a closely related set of species recently diverged from a common ancestor. However, although such a method promises to capture the immediate impact of host specialization on receptor evolution, it has low statistical power, and resource specialization and small population size can be confounded (as in *D. sechellia* and *D. erecta*). In particular, reduced effective population size in endemic species can result in higher ω ratios because substitutions that are slightly deleterious can become fixed at a faster rate in smaller populations than in larger ones [24]. Studies including a large number of species (e.g. all 12 sequenced drosophilids) have a larger sample size, but as phylogenetic distance between taxa increases, the ability to discern the effects of host specialization on genomic consequences might decrease. For example, the closest sequenced relative of the specialist *D. mojavensis* is the generalist *D. virilis*, and the two diverged from a common ancestor tens of millions of years ago [3].

Including ecology
Without full knowledge of the life histories of species, comparative conclusions are hard to draw. This was highlighted by the difficulty that Gardiner and colleagues had in categorizing different species in their comparative analysis. In accordance with the current view, they coded *D. grimshawi* as a generalist, but pointed out that some *D. grimshawi* strains are specialists, and so the group might best be characterized as polymorphic. Similarly, they coded *D. mojavensis* as a mainland species, but noted that the strain that was sequenced is a genetically distinct strain found on Santa Catalina Island, California, so it might be considered an endemic [25]. They coded *D. erecta* as a mainland species, but as its distribution is restricted to central Africa, it could be considered to be an endemic. *D. pseudoobscura* was coded as a generalist, even though little is known about the plant substrates it uses. These examples highlight how difficult it is to draw strong conclusions from comparative analyses using a limited sample size of species, especially without good ecological information.

Including other species
To better understand the consequences of host specialization and demography, it will be important to compare entire gene repertoires in suites of closely related lineages that vary in resource niche breadth and geographic range. Several immediate candidates in *Drosophila* that are closely related to members of the 12 completely sequenced species include: (i) a specialist lineage of *D. grimshawi* [9]; (ii) mainland *D. mojavensis* and its sister species *D. arizonae*, a cactus generalist; (iii) *D. mauritiana*, an island endemic and generalist whose genome is currently being sequenced; and (iv) *D. santomea*, an island endemic closely related to the generalist *D. yakuba* [13]. In addition, concerted effort should be made toward understanding the ecology and natural history of the *Drosophila* radiation.

Including other genes
Comparative analyses should be expanded to include other loci that might be expected to also experience relaxed selection pressure as a result of niche reduction, such as detoxification pathway genes. Intriguingly, specialist lineages also lose genes encoding cytochrome p450 at a faster rate than generalist species [3]. These genes are known to be important in toxin detoxification, but again, effects of demography versus selection would need to be teased apart.

dynamics for each gene family mediated by novel selective pressure and demographic flux during specialization and founding events, with a potential role for positive selection at some *Gr* and *Or* loci. The functional study has yielded

insight into how quantitative trait mapping of a complex behavioral trait can be used as a basis for a detailed candidate gene study. Another recent study [20] shows the promise of using whole-genome gene expression studies in this context, and demonstrates that, rather than structural differences, changes in the expression of *Or* and *Obp* genes might be important in the evolution of drosophilid host preference. Taken together, these studies show that the union of ecology, evolutionary genomics, physiology and neurobiology will illuminate how host plant toxins can become co-opted as 'delicious poison,' a question that lies at the heart of insect–plant coevolution and the evolution of specialization.

Acknowledgements

References

1 Wheat, C.W. *et al.* (2007) The genetic basis of a plant-insect coevolutionary key innovation. *Proc. Natl. Acad. Sci. U. S. A.* 104, 20427–20431
2 Bernays, E.A. and Chapman, R.F. (1994) *Host-Plant Selection by Phytophagous Insects,* Chapman & Hall
3 *Drosophila* 12 Genomes Consortium (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450, 203–218
4 Matsuo, T. *et al.* (2007) Odorant-binding proteins Obp57d and Obp57e affect taste perception and host-plant preference in *Drosophila sechellia*. *PLoS Biol.* 5, e118
5 McBride, C.S. (2007) Rapid evolution of smell and taste receptor genes during host-specialization in *Drosophila sechellia*. *Proc. Natl. Acad. Sci. U. S. A.* 104, 4996–5001
6 McBride, C.S. and Arguello, J.R. (2007) Five *Drosophila* genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. *Genetics* 177, 1395–1416

7 Vieira, F.G. *et al.* (2007) Comparative genomic analysis of the odorant-binding protein family in 12 *Drosophila* genomes: purifying selection and birth-and-death evolution. *Genome Biol.* 8, R235
8 Gardiner, A. *et al.* (2008) *Drosophila* chemoreceptor gene evolution: selection, specialization and genome size. *Mol. Ecol.* 17, 1648–1657
9 Hallem, E.A. *et al.* (2006) Insect odor and taste receptors. *Annu. Rev. Entomol.* 51, 113–135
10 Keller, A. and Vosshall, L.B. (2007) Influence of odorant receptor repertoire on odor perception in humans and fruit flies. *Proc. Natl. Acad. Sci. U. S. A.* 104, 5614–5619
11 Stortkuhl, K.F. *et al.* (2005) An increased receptive field of olfactory receptor Or43a in the antennal lobe of *Drosophila* reduces benzaldehyde-driven avoidance behavior. *Chem. Senses* 30, 81–87
12 Steinbrecht, R.A. (1998) Odorant-binding proteins: expression and function. *Ann. N. Y. Acad. Sci.* 855, 323–332
13 Markow, T.A. and O'Grady, P.M. (2005) Evolution genetics of reproductive behavior in *Drosophila*. *Annu. Rev. Genet.* 39, 263–291
14 Jones, C.D. (1998) The genetic basis of *Drosophila sechellia*'s resistance to a host plant toxin. *Genetics* 149, 1899–1908
15 Jones, C.D. (2001) The genetic basis of larval resistance to a host plant toxin in *Drosophila sechellia*. *Genet. Res.* 78, 225–233
16 Coyne, J.A. *et al.* (1991) Genetics of morphological differences and hybrid sterility between *Drosophila sechellia* and its relatives. *Genet. Res.* 57, 113–122
17 Jones, C.D. (2005) The genetics of adaptation in *Drosophila sechellia*. *Genetica* 123, 137–145
18 Dekker, T. *et al.* (2006) Olfactory shifts parallel superspecialism for toxic fruit in *Drosophila melanogaster* sibling, *D. sechellia*. *Curr. Biol.* 16, 101–109
19 Gilad, Y. *et al.* (2003) Natural selection on the olfactory receptor gene family in humans and chimpanzees. *Am. J. Hum. Genet.* 73, 489–501
20 Kopp, A. *et al.* (2008) Evolution of gene expression in the *Drosophila* olfactory system. *Mol. Biol. Evol.* 25, 1081–1092
21 Hahn, M.W. (2007) Bias in phylogenetic tree reconciliation methods: implications for vertebrate genome evolution. *Genome Biol.* 8, R141
22 Kaissling, K-E. (2001) Olfactory perireceptor events in moths: a kinetic model. *Chem. Senses* 26, 125–150
23 Yang, Z. (2003) Adaptive molecular evolution. In *Handbook of Statistical Genetics* (Balding, D.J. *et al.*, eds), pp. 229–254, John Wiley and Sons
24 Johnson, K.P. and Seger, J. (2001) Elevated rates of nonsynonymous substitution in island birds. *Mol. Biol. Evol.* 18, 874–881
25 Reed, L.K. *et al.* (2007) Evolutionary relationships of *Drosophila mojavensis* geographic host races and their sister species *Drosophila arizonae*. *Mol. Ecol.* 16, 1007–1022