LETTERS

Repeated morphological evolution through *cis*-regulatory changes in a pleiotropic gene

Benjamin Prud'homme[1]*, Nicolas Gompel[1,2]*, Antonis Rokas[1]†, Victoria A. Kassner[1], Thomas M. Williams[1], Shu-Dan Yeh[3], John R. True[3] & Sean B. Carroll[1]

The independent evolution of morphological similarities is widespread[1,2]. For simple traits, such as overall body colour, repeated transitions by means of mutations in the same gene may be common[3–5]. However, for more complex traits, the possible genetic paths may be more numerous; the molecular mechanisms underlying their independent origins and the extent to which they are constrained to follow certain genetic paths are largely unknown. Here we show that a male wing pigmentation pattern involved in courtship display has been gained and lost multiple times in a *Drosophila* clade. Each of the cases we have analysed (two gains and two losses) involved regulatory changes at the pleiotropic pigmentation gene *yellow*. Losses involved the parallel inactivation of the same *cis*-regulatory element (CRE), with changes at a few nucleotides sufficient to account for the functional divergence of one element between two sibling species. Surprisingly, two independent gains of wing spots resulted from the co-option of distinct ancestral CREs. These results demonstrate how the functional diversification of the modular CREs of pleiotropic genes contributes to evolutionary novelty and the independent evolution of morphological similarities.

To address how complex traits are repeatedly gained and lost, we focused on the formation of spots of dark pigment at the tip of male wings in certain fruitfly (*Drosophila*) species of the *melanogaster*[6] and *obscura*[7] groups. These spots form under the control of multiple genes[8]. To determine how often male wing pigment spots have been gained or lost we established the phylogeny of these groups and reconstructed the evolution of the trait by means of bayesian phylogenetic inference (BI), which provides a statistical framework that explicitly accommodates phylogenetic and character mapping uncertainties in its ancestral character reconstruction estimates[9–11], and thus offers a more rigorous picture of character evolution.

The BI analysis provided evidence that the common ancestor of the clade was unspotted (Fig. 1, node 1) and that the wing spot was gained once within the *melanogaster* group (Fig. 1, node 2) and at least once more within the *obscura* group (Fig. 1, node 3). The analysis also indicated that the wing spot has been lost independently at least five times in the *melanogaster* group (Fig. 1, dashed branches). The availability of closely related species that differ with regard to wing spots enabled us to investigate the molecular mechanisms by which the spots were lost and gained.

D. elegans and *D. gunungcola* (Fig. 1, node 4) are interfertile sibling species that diverged 2–2.8 Myr ago and are characterized by differences in male-specific wing pigmentation[12] (Fig. 2b, c). The *yellow* (*y*) gene is a strong candidate for contributing to the difference in pigmentation because of its role in the production of black pigment in many *Drosophila* species[8,13,14] including *D. elegans* (S.-D.Y and J.R.T., unpublished observations). We found that in *D. elegans* pupal wings Yellow accumulates at high levels in the region where the adult pigmentation spot will form (Fig. 2b, d), whereas in *D. gunungcola* the Yellow protein is only expressed at low levels throughout the wing, correlating with the grey shading of the adult wing (Fig. 2c, e).

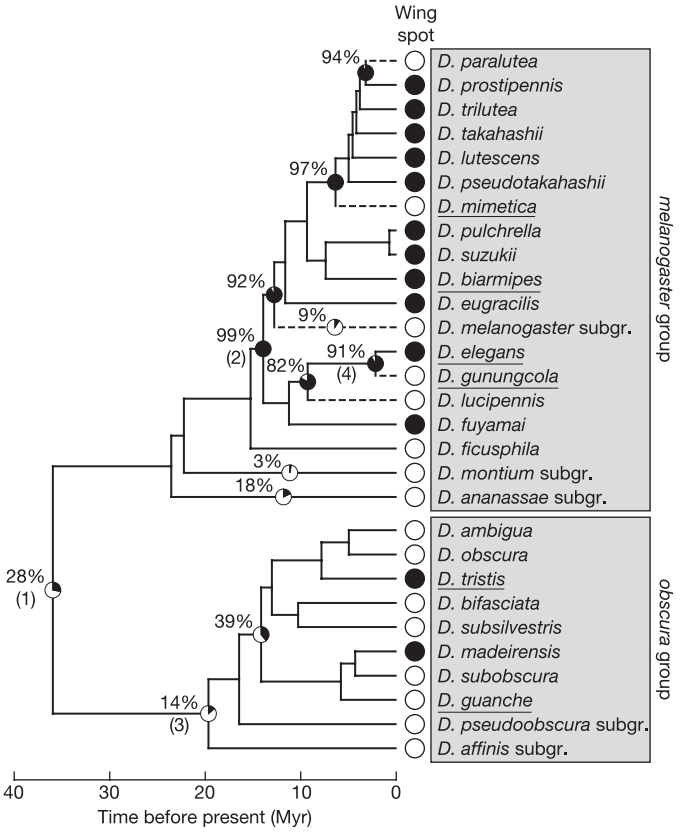

Figure 1 | Two independent gains and five losses of wing pigmentation spots in a *Drosophila* clade. For each species, the presence or absence of a wing spot is indicated with a solid black or a solid white circle, respectively. For key nodes, the black portion of each pie chart and the percentages shown next to them indicate the posterior probability that the ancestor was spotted. Dotted branches indicate inferred losses of the wing spot. Branch lengths correspond to absolute time estimates. Numbers in parenthesis identify nodes discussed in the text. Species examined in this study are underlined.

[1]University of Wisconsin and Howard Hughes Medical Institute, Bock Laboratories, 1525 Linden Drive, Madison, Wisconsin 53706, USA. [2]Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK. [3]Department of Ecology and Evolution, 650 Life Sciences Building, Stony Brook University, Stony Brook, New York 11794, USA. †Present address: Broad Institute of MIT and Harvard, 320 Charles Street, Cambridge, Massachusetts 02141-2023, USA.
*These authors contributed equally to this work.

The divergence with regard to *yellow* function between the two species is therefore regulatory in nature.

As a first step to determining whether regulatory changes have occurred at *y* in *D. gunungcola*, we sought to confirm that the formation of the wing spot of *D. elegans* and *D. biarmipes* share a common genetic basis. In *D. biarmipes*, the expression of *y* in the future wing spot region is controlled by the $spot^{bia}$ CRE, which evolved in the context of an ancestral CRE (*wing large*, Fig. 2a) driving uniform expression of *y* in the wings of unspotted species[8]. We isolated from *D. elegans* a fragment orthologous to the $spot^{bia}$ CRE, $spot^{ele}$ (Fig. 2a), and assayed its transcriptional regulatory activity in *D. melanogaster*. This fragment drives reporter expression in a pattern specifically confined to the spot region (Fig. 2f). These results indicate that the wing pigmentation spots of *D. elegans* and *D. biarmipes* (and, by phylogenetic inference, all of the wing-spotted species of the *melanogaster* group descended from the ancestor at node 2 in Fig. 1) are homologous.

We then isolated from the *D. gunungcola y* locus the 5′ non-coding DNA and the fragment orthologous to the *spot* CRE ($spot^{gun}$). When transferred into *D. melanogaster* the 5′ non-coding DNA drives uniform reporter expression in the wing, recapitulating the native *D. gunungcola y* expression pattern (not shown). In contrast, the $spot^{gun}$ CRE is largely inactive and drives barely detectable traces of reporter expression in the spot region (compare Fig. 2g with Fig. 2f). These results show that the major changes responsible for the loss of expression of *yellow* in *D. gunungcola* occurred in the $spot^{gun}$ CRE.

The close kinship of *D. elegans* and *D. gunungcola* allowed us to pursue the molecular characterization of the functional divergence of the *spot* CRE. The $spot^{ele}$ (775 base pairs (bp)) and $spot^{gun}$ (724 bp) CREs share 92% sequence identity (Fig. 3a). Fifty nucleotide substitutions and five indels (four of less than 4 bp and one of 47 bp) constitute all of the differences between these two functionally divergent sequences. To map the sequence changes responsible for the loss of activity of the element in *D. gunungcola* we constructed a series of $spot^{ele}$ chimaeric transgenes (*I* to *VI*, Fig. 3b) in which we systematically replaced subsets of *D. elegans*-specific nucleotides with their divergent *D. gunungcola* counterparts and assayed their transcriptional regulatory activity in *D. melanogaster*. One obvious candidate for the loss of activity was the 47-bp indel present in the $spot^{ele}$ sequence but absent from $spot^{gun}$. However, when this 47-bp stretch was selectively removed from the $spot^{ele}$ element (construct *Δ*), the transgene drove reporter expression similar to the native $spot^{ele}$ sequence (Fig. 3c). In fact, we found that most chimaeric constructs drove reporter levels and patterns in a similar manner to the $spot^{ele}$ element (Figs 2f and 3d, and not shown), thus ruling out significant functional contribution of these divergent nucleotides to the loss of activity of $spot^{gun}$. However, one chimaeric transgene, *III*, yielded a reporter expression similar to that of the $spot^{gun}$ element (Figs 2g and 3e). This construct included ten divergent nucleotides, indicating that the major functional differences between the two species' *spot* elements might be due to changes among these sites. To further delimit the number of nucleotides responsible for the divergence in gene expression, we made three sub-constructs (*III.1*, *III.2* and *III.3*) containing just three, four and three nucleotide differences, respectively. Whereas *III.3* yielded a normal spot pattern, lines carrying *III.1* or *III.2* showed a strong reduction in reporter expression (not shown). These results indicate that divergent nucleotides within both *III.1* and *III.2* are necessary for the function of $spot^{ele}$ and were involved in the functional change between the two species' CREs.

To test whether these divergent nucleotides contributed to the loss of Yellow spot expression in *D. gunungcola*, we made a reciprocal construct in which we replaced the ten divergent nucleotides of the $spot^{gun}$ element with their *D. elegans* counterparts. This construct, $spot^{gun}$ *rescue*, drives a spot expression pattern similar to that of $spot^{ele}$, thus restoring the ancestral function of the $spot^{gun}$ regulatory element (Fig. 3b, f). This result demonstrates the critical contribution of these ten nucleotides to the divergence in transcriptional regulatory activity between the two species' CREs and indicates that the $spot^{gun}$ CRE, although inactive, still contains some of the regulatory information required for generating the spot pattern. We deduce that at least two and no more than seven point mutations (presumably affecting two or more transcription-factor-binding sites) were sufficient to alter the function of the $spot^{gun}$ element and fully account for the divergence in *y* expression in the wing between *D. elegans* and *D. gunungcola*. We suggest that functional inactivation of CREs might require only a small number of mutational steps and might be the most likely path to the evolutionary loss of gene expression.

To test the generality of CRE inactivation in the loss of wing spots, we examined the expression and regulation of *y* in *D. mimetica* (Fig. 1, and Supplementary Fig. 2). As in *D. gunungcola*, the spot loss in *D. mimetica* is associated with changes in Yellow expression in the wing. Furthermore, the *D. mimetica spot* CRE, which shares 69% sequence identity with $spot^{ele}$, drives a very faint reporter expression in the spot region, similar in intensity to that of the $spot^{gun}$ CRE (Supplementary Fig. 2c). Taken together, our results show that the independent inactivation of the same CRE has contributed to the repeated loss of a wing pigmentation pattern in at least two species.

To determine whether the repeated gains of wing spots also had a common mechanistic basis, we examined Yellow expression in and CREs from *D. tristis*, which gained a wing spot independently from the *melanogaster* group species (Figs 1 and 4a). Remarkably, the trait has apparently evolved under sexual selection in both cases, along with a novel wing display step during the male courtship (Supplementary movies). Yellow expression in *D. tristis* male pupal wings, just as in spotted *melanogaster* group species, prefigures the wing pigmentation spot (Fig. 4a, b). We tested whether in *D. tristis* this expression was also controlled by the *wing large* CRE we had identified in *D. melanogaster* group species[8]. However, when

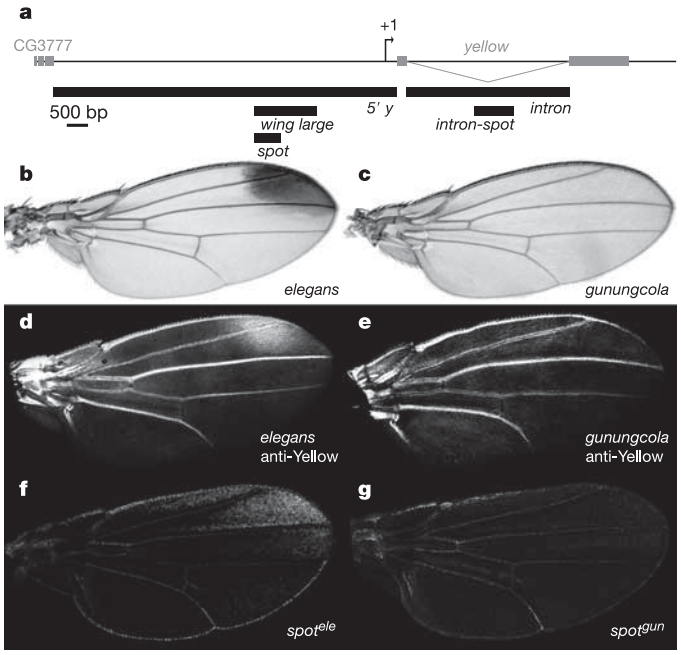

Figure 2 | Changes in the *yellow spot cis*-regulatory DNA underlie the loss of the pigmentation spot in *D. gunungcola*. **a**, Diagram of the *yellow* locus showing the position of the DNA fragments tested in this study (black bars). **b**–**e**, The adult male wings of *D. elegans* (**b**) and *D. gunungcola* (**c**) differ in their pigmentation patterns, which reflect the expression of the Yellow protein in developing pupal wings (**d** and **e**, respectively). **f**, **g**, The *D. elegans* CRE $spot^{ele}$, nested in the 5′ region, drives high levels of reporter expression in the spot region (**f**), whereas its *D. gunungcola* orthologue, $spot^{gun}$, has lost this activity (**g**).

transferred into *D. melanogaster* this CRE drove only uniform reporter expression in the wing (Fig. 4c), as expected for the ancestral *wing large* element, but no elevated expression in the spot region was observed.

We then considered whether a distinct CRE controlling Yellow expression in the spot had evolved at another position of the locus. We found that none of the sequences extending over the 5′ non-coding region of the locus drove any wing spot pattern (data not shown). However, the intron of the *D. tristis yellow* gene drove reporter expression in a spot pattern, as well as strong expression in the wing veins (Fig. 4d). Further dissection of the *D. tristis* intron revealed that expression in both the wing veins and the spot are controlled by a smaller fragment, *intron-spot* (927 bp; Fig. 2a, and data not shown). Because *D. tristis* contains a functional element orthologous to the *wing large* element of other species, and because we did not detect any sequence similarity between the *spot* and the *intron-spot* elements, we conclude that the evolution of *y* expression in the wing spot of *D. tristis* involved an entirely different *cis*-regulatory region. To determine whether the *D. tristis intron-spot* CRE evolved *de novo* or through the modification of a pre-existing element, we examined the transcriptional regulatory activity of the *y* intron from *D. guanche*, an unspotted species of the *obscura* group (Supplementary Fig. 3a). This sequence drove a wing vein expression pattern similar to that driven by the *D. tristis y intron-spot* CRE (Supplementary Fig. 3b). Thus, the proximity or overlap of an ancestral *wing vein* CRE and the derived *spot* CRE in the *D. tristis y* intron indicates that the spot activity might have evolved through the co-option of an ancestral *wing vein* CRE.

The molecular bases of spot evolution offer some general insights into how the course of evolution is determined by which genetic paths are possible, which are most probable, and which are permissible under natural selection. The gains of *yellow* expression in the wing spots of the *melanogaster* group species and in *D. tristis* through the co-option and modification of distinct, ancestral CREs of the *yellow* locus show that multiple molecular paths were possible for the evolution of spot formation (Fig. 4e). Yet both ancestral elements drove *y* expression in the pupal wing. We have suggested that novel CREs active in a particular tissue are most likely to evolve in the vicinity of pre-existing elements active in that tissue[8]. Indeed, rather than evolving the constellation of binding sites required for element function *de novo*, it is more likely that a new activity would evolve in the context of a functional element that already contained information (for example, particular binding sites) for the control of a gene. The co-option of available transcription factors and the co-option and modification of CREs to generate novel patterns illustrate, at the level of regulatory circuits, the sort of 'evolutionary tinkering' envisaged by François Jacob three decades ago[15].

Such tinkering is also a matter of what is permissible. Regulatory changes in one CRE, by selectively affecting one aspect of the spatio-temporal pattern of expression, circumvent the fitness reduction potentially associated with the global effects of changes in protein function. The *yellow* gene is highly pleiotropic[16,17] and the control of the different Yellow expression domains is governed by distinct CREs. Other pleiotropic genes also seem to have contributed to repeated evolutionary transitions through CREs[18–20]. Evolutionary changes in protein coding sequences have been observed in pigmentation genes in vertebrates, such as *MC1R*[4,5] and *Oca*[3], but these encode specialized, minimally pleiotropic proteins required for pigment production in melanocytes. These findings indicate that pleiotropy might be an important genetic constraint on the potential

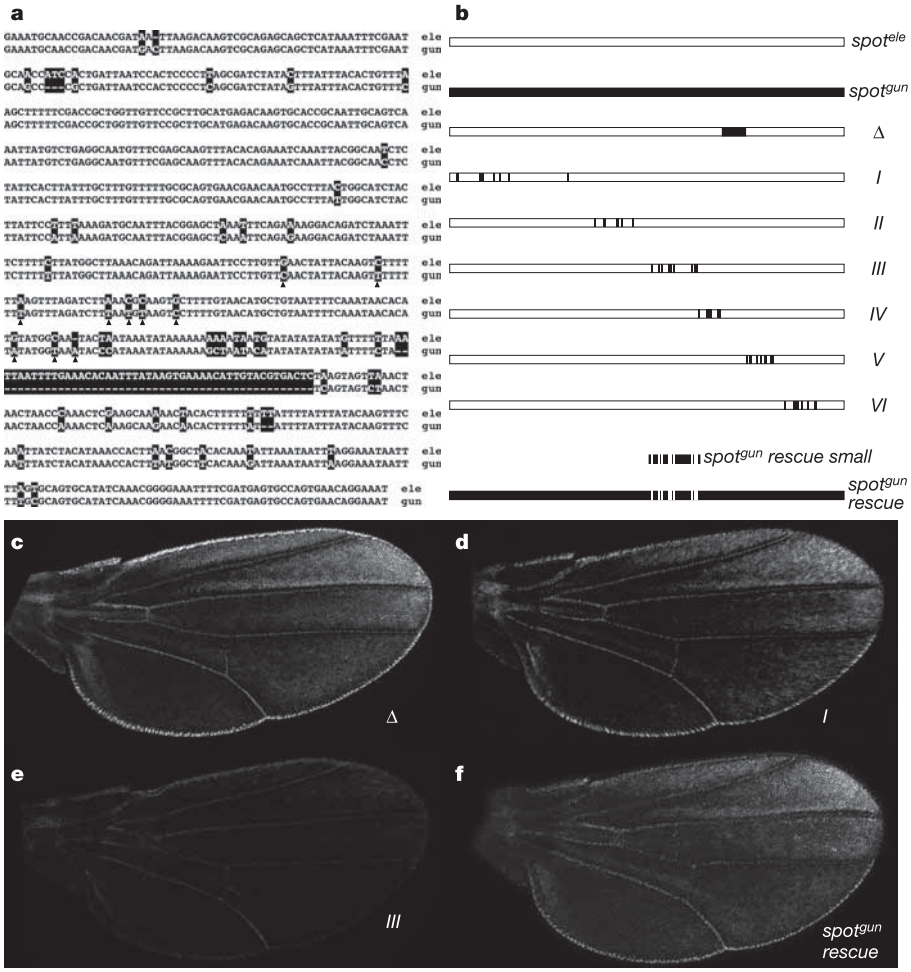

Figure 3 | A few divergent nucleotides account for the functional inactivation of the *spot*^{gun} element. **a**, *spot*^{ele} and *spot*^{gun} alignment; arrowheads indicate divergent nucleotides of construct *III*. **b**, Diagram of chimaeric constructs to map functional changes. **c**–**f**, Reporter expression in transformed pupal wings. Constructs *Δ* (**c**) and *I* (**d**) do not affect spot expression. However, construct *III* (**e**) drives a dramatically reduced reporter expression, similar to that of *spot*^{gun} (Fig. 2g). The reciprocal construct to *III*, *spot*^{gun} *rescue* (**f**), restores a spot pattern. However, the 90-bp subfragment, *spot*^{gun} *rescue small*, containing the ten key divergent nucleotides, is not active (not shown).

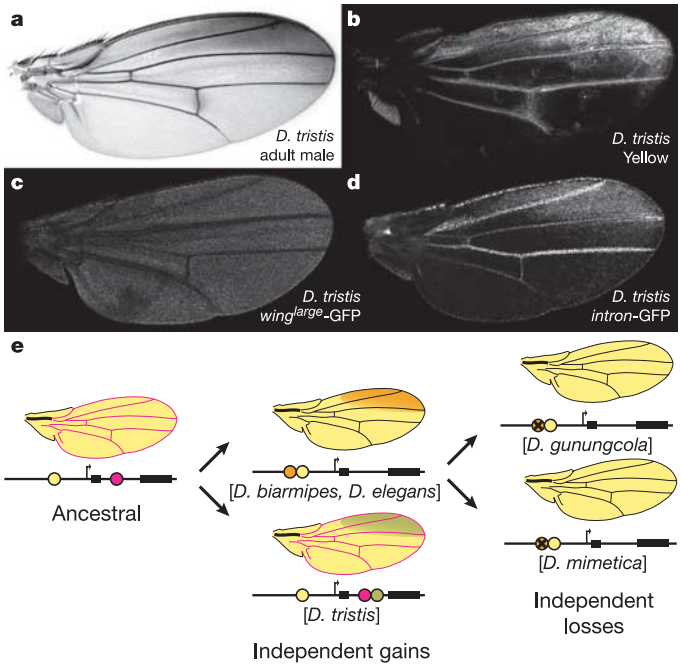

Figure 4 | Independent co-option of CREs of the *yellow* gene in wing spot evolution. a, **b**, In the *D. tristis* male pupal wing (**a**), Yellow expression prefigures the pigmentation pattern (**b**). **c**, **d**, The *D. tristis y wing large* orthologue (**c**) drives a uniform low level of reporter expression in the pupal wing, whereas the *D. tristis y intron* (**d**) drives the novel expression pattern in the spot region as well as the wing veins. **e**, *y* spot expression patterns evolved twice by the co-option and modification of two different pre-existing *y* CREs (symbolized by yellow → orange circles and pink → green circles) and were lost by the repeated inactivation of one CRE.

contribution of the different parts of genes (coding versus non-coding sequences) to morphological evolution. In highly pleiotropic genes, *cis*-regulatory changes are more likely to be accommodated than coding changes and should be expected to be major contributors to morphological evolution.

METHODS

Phylogenetics. We assembled a data matrix of 11 genes from 77 species (including new sequences and previously available data; Supplementary Tables 1 and 2) with an average of six genes per species. Selection of species was random in respect to their state of the wing spot trait and balanced across clades. Phylogenies were estimated under three optimality criteria: bayesian inference, maximum parsimony and maximum likelihood, as implemented in MrBayes[21], PAUP*[22] and PHYML[23], respectively. The resulting phylogeny is generally consistent with previous studies with smaller numbers of genes and taxa[24–28]. Ancestral character reconstruction was performed with BayesMultistate[9]. Divergence date estimates for the whole clade were obtained with the penalized likelihood method as implemented in r8s, version 1.7 (ref. 29). Details of all phylogenetic analyses are given in Supplementary Information, and the data matrix and analyses are available from the authors on request.

Immunochemistry, cloning and imaging. These were performed as described previously[8] (specific details are provided in Supplementary Information).

Received 4 November 2005; accepted 25 January 2006.

1. Conway Morris, S. *Life's Solution. Inevitable Humans in a Lonely Universe* (Cambridge Univ. Press, Cambridge, 2005).
2. West-Eberhard, M. J. *Developmental Plasticity and Evolution* (Oxford Univ. Press, Oxford, 2003).
3. Protas, M. E. *et al.* Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nature Genet.* **38,** 107–111 (2006).
4. Nachman, M. W., Hoekstra, H. E. & D'Agostino, S. L. The genetic basis of adaptive melanism in pocket mice. *Proc. Natl Acad. Sci. USA* **100,** 5268–5273 (2003).
5. Theron, E., Hawkins, K., Bermingham, E., Ricklefs, R. E. & Mundy, N. I. The molecular basis of an avian plumage polymorphism in the wild: a melanocortin-1-receptor point mutation is perfectly associated with the melanic plumage morph of the bananaquit, *Coereba flaveola*. *Curr. Biol.* **11,** 550–557 (2001).
6. Bock, I. R. & Wheeler, M. R. in *Studies in Genetics* (ed. Wheeler, M. R.) 1–102 (University of Texas, Austin, 1972).
7. Lakovaara, S. & Saura, A. in *The Genetics and Biology of Drosophila* Vol. 3b (eds Ashburner, M., Carson, H. L. & Thompson, J. N.) 1–59 (Academic, London, 1982).
8. Gompel, N., Prud'homme, B., Wittkopp, P. J., Kassner, V. A. & Carroll, S. B. Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in *Drosophila*. *Nature* **433,** 481–487 (2005).
9. Pagel, M., Meade, A. & Barker, D. Bayesian estimation of ancestral character states on phylogenies. *Syst. Biol.* **53,** 673–684 (2004).
10. Huelsenbeck, J. P., Rannala, B. & Masly, J. P. Accommodating phylogenetic uncertainty in evolutionary studies. *Science* **288,** 2349–2350 (2000).
11. Ronquist, F. Bayesian inference of character evolution. *Trends Ecol. Evol.* **19,** 475–481 (2004).
12. Yeh, S.-D., Liou, S.-R. & True, J. R. Genetics of divergence in male wing pigmentation and courtship behavior between *Drosophila elegans* and *D. gunungcola*. *Heredity* (in the press).
13. Wittkopp, P. J., Vaccaro, K. & Carroll, S. B. Evolution of *yellow* gene regulation and pigmentation in *Drosophila*. *Curr. Biol.* **12,** 1547–1556 (2002).
14. Wittkopp, P. J., True, J. R. & Carroll, S. B. Reciprocal functions of the *Drosophila yellow* and *ebony* proteins in the development and evolution of pigment patterns. *Development* **129,** 1849–1858 (2002).
15. Jacob, F. Evolution and tinkering. *Science* **196,** 1161–1166 (1977).
16. Drapeau, M. D., Cyran, S. A., Viering, M. M., Geyer, P. K. & Long, A. D. A *cis*-regulatory sequence within the *yellow* locus of *Drosophila melanogaster* required for normal male mating success. *Genetics* **172,** 1009–1030 (2006).
17. Walter, M. F. *et al.* Temporal and spatial expression of the *yellow* gene in correlation with cuticle formation and dopa decarboxylase activity in *Drosophila* development. *Dev. Biol.* **147,** 32–45 (1991).
18. Shapiro, M. D. *et al.* Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* **428,** 717–723 (2004).
19. Sucena, E., Delon, I., Jones, I., Payre, F. & Stern, D. L. Regulatory evolution of *shavenbaby/ovo* underlies multiple cases of morphological parallelism. *Nature* **424,** 935–938 (2003).
20. Sucena, E. & Stern, D. L. Divergence of larval morphology between *Drosophila sechellia* and its sibling species caused by *cis*-regulatory evolution of ovo/shaven-baby. *Proc. Natl Acad. Sci. USA* **97,** 4530–4534 (2000).
21. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17,** 754–755 (2001).
22. Swofford, D. L. PAUP*: *Phylogenetic Analysis Using Parsimony (*and Other Methods) 4.0 Beta* (Sinauer Associates, Sunderland, Massachusetts, 2002).
23. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52,** 696–704 (2003).
24. O'Grady, P. M. & Kidwell, M. G. Phylogeny of the subgenus Sophophora (Diptera: Drosophilidae) based on combined analysis of nuclear and mitochondrial sequences. *Mol. Phylogenet. Evol.* **22,** 442–453 (2002).
25. O'Grady, P. M. Reevaluation of phylogeny in the *Drosophila obscura* species group based on combined analysis of nucleotide sequences. *Mol. Phylogenet. Evol.* **12,** 124–139 (1999).
26. Goto, S. G. & Kimura, M. T. Phylogenetic utility of mitochondrial COI and nuclear Gpdh genes in *Drosophila*. *Mol. Phylogenet. Evol.* **18,** 404–422 (2001).
27. Kopp, A. & True, J. R. Phylogeny of the oriental *Drosophila melanogaster* species group: a multilocus reconstruction. *Syst. Biol.* **51,** 786–805 (2002).
28. Schawaroch, V. Phylogeny of a paradigm lineage: the *Drosophila melanogaster* species group (Diptera: Drosophilidae). *Biol. J. Linn. Soc.* **76,** 21–37 (2002).
29. Sanderson, M. J. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19,** 301–302 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Kopp for sharing unpublished data; J. J. Gao, S. Hagemann, M. T. Kimura, A. Kopp, V. Stamataki and the *Drosophila* Tucson Stock Center for providing fly stocks; M. Toda for fly identification; M. Bate for help with behavioural observations; A. Meade and M. Pagel for advice on bayesian ancestral character reconstruction; and B. Williams and J. Yoder for comments on the manuscript. N.G. thanks P. Simpson for hosting him in her laboratory. N.G. was an EMBO and a Human Frontier long-term fellow. The project was supported by the Howard Hughes Medical Institute (S.B.C.).

Author Information The sequences described in this paper have been deposited at the EMBL nucleotide database under accession numbers AM181668 to AM181680. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.B.C. (sbcarrol@wisc.edu).