# Genetic mapping with SNP markers in *Drosophila*

Jürg Berger*, Takashi Suzuki*, Kirsten-André Senti, Janine Stubbs, Gotthold Schaffner & Barry J. Dickson

*These authors contributed equally to this work.

Published online: 19 November 2001, DOI: 10.1038/ng773

**Map-based positional cloning of *Drosophila melanogaster* genes is hampered by both the time-consuming, error-prone nature of traditional methods for genetic mapping and the difficulties in aligning the genetic and cytological maps with the genome sequence. The identification of sequence polymorphisms in the *Drosophila* genome will make it possible to map mutations directly to the genome sequence with high accuracy and resolution. Here we report the identification of 7,223 single-nucleotide polymorphisms (SNPs) and 1,392 insertions/deletions (InDels) in common laboratory strains of *Drosophila*. These sequence polymorphisms define a map of 787 autosomal marker loci with a resolution of 114 kb. We have established PCR product–length polymorphism (PLP) or restriction fragment–length polymorphism (RFLP) assays for 215 of these markers. We demonstrate the use of this map by delimiting two mutations to intervals of 169 kb and 307 kb, respectively. Using a local high-density SNP map, we also mapped a third mutation to a resolution of approximately 2 kb, sufficient to localize the mutation within a single gene. These methods should accelerate the rate of positional cloning in *Drosophila*.**

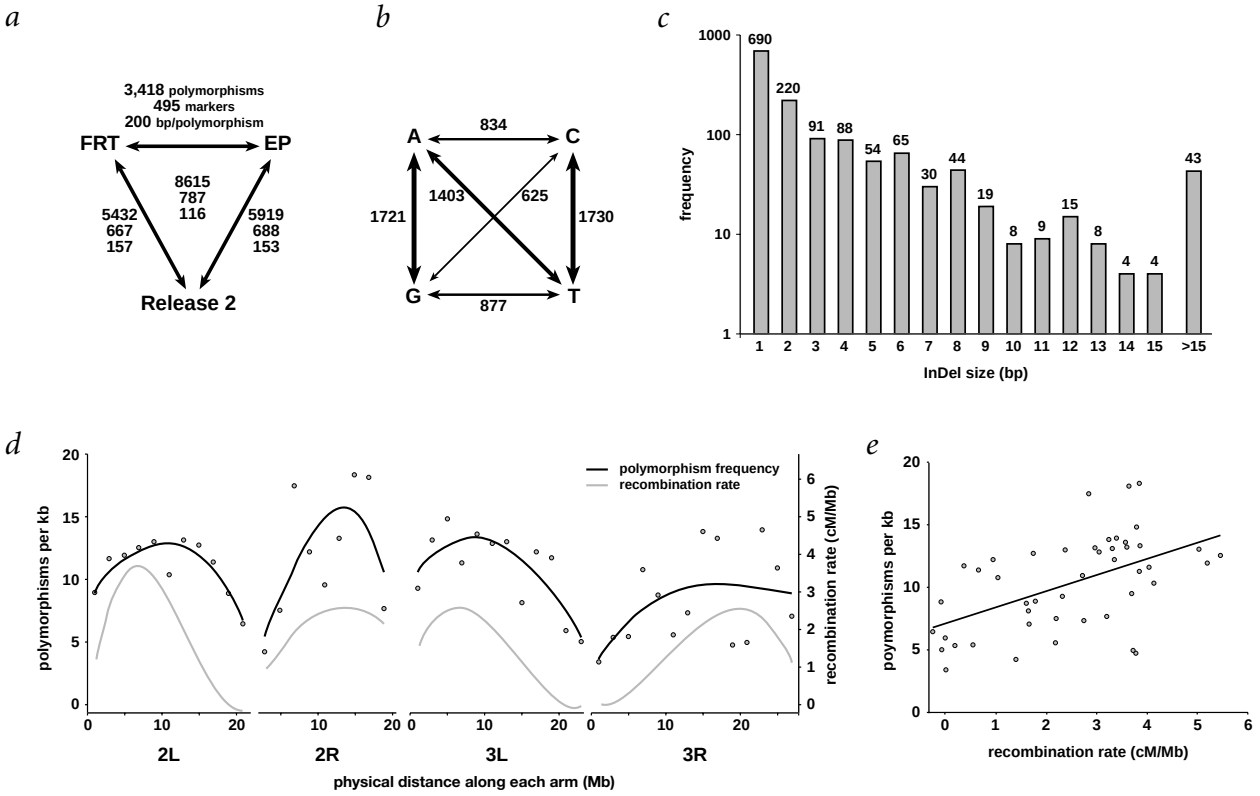


**Fig. 1** Classes and distribution of sequence polymorphisms in the *Drosophila* genome. ***a***, Number of sequence polymorphisms identified in pairwise comparisons of FRT and EP strains and the Release 2 genome sequence. The total number of polymorphisms identified, the number of markers these polymorphisms define and the average number of base pairs per polymorphism are indicated for each (arrows) and any (center) pair of strains. ***b***, Class distribution of 7,190 biallelic SNPs. In our sample, transversions slightly (1.08×) but significantly ($P$=0.0007) outnumbered transitions. Other surveys of sequence polymorphisms in *Drosophila* have reported a slight bias in favor of transitions[5,8,23]. Among transversions, AT polymorphisms are markedly overrepresented (1.50×, $P$<0.00001) and CG polymorphisms underrepresented (0.69×, $P$<0.00001). POLYBAYES reported three different nucleotides for another 33 SNPs. ***c***, Size distribution of 1,392 InDels. The largest InDel identified was 169 bp. ***d***, Polymorphism density and recombination rates across the major autosomes. Polymorphism densities (circles) were calculated in 2-Mb intervals across each arm, and fitted with a third-order polynomial (black line). We calculated recombination rates by fitting a third-order polynomial (gray line) to a plot of genetic distance versus physical distance between adjacent cytological letter divisions across each arm. This plot was in turn derived from tables provided by FlyBase that link each cytological division to the genetic (flybase.bio.indiana.edu/maps/lk/cytotable.txt) and physical (flybase.bio.indiana.edu/maps/lk/genome-cyto-seq-map/) maps. ***e***, Scatterplot of polymorphism density versus recombination rate in 2-Mb intervals across the four major autosomal arms. Polymorphism density increases with the rate of recombination[17]. The null hypothesis (that the slope of the regression line is zero) is rejected with high probability ($P$=0.00008).

*Research Institute of Molecular Pathology, Dr. Bohr-Gasse 7 A-1030, Vienna, Austria. Correspondence should addressed to B.J.D. (e-mail: dickson@nt.imp.univie.ac.at).*

A dense map of sequence polymorphisms is a valuable resource for mapping human disease loci[1] and experimentally induced mutations in model organisms[2–6]. For genetic mapping in model organisms, SNPs and InDels offer several key advantages over conventional genetic markers: (i) they are highly abundant (ii) they exist in only two codominant variants and (iii) in most cases, they are phenotypically neutral.

To enable genetic mapping with SNP markers, it is first necessary to establish a map of sequence polymorphisms between the strain used for mutagenesis and a divergent strain that can be used to generate recombinants for mapping. If laboratory strains are largely isogenic, as is the case in *Caenorhabditis elegans*, a sufficiently divergent natural isolate must be used for mapping purposes. A polymorphism map can then be constructed simply by carrying out light shotgun sequencing of this strain and aligning these sequences to the genome sequence of the laboratory strain[4]. The difficulty in this case is that the genetic background of the new isolate is poorly defined and does not include any visible genetic markers to assist in the recovery of recombinants[4]. Ideally, well-defined laboratory strains would themselves have sufficient nucleotide diversity for mapping purposes, as we show here to be the case for *Drosophila*. This poses the challenge of constructing a map of polymorphisms between two strains, neither of which may correspond to the reference genome.

Two general approaches to constructing such a map may be considered. One strategy would be to carry out light shotgun sequencing of one strain, align this to the reference genome and then genotype the second strain for any polymorphisms thus identified. An alternative approach is to directly amplify and resequence specific loci from each of the two strains of interest. A mathematical analysis of this problem (see Methods and Web Note A) indicates that

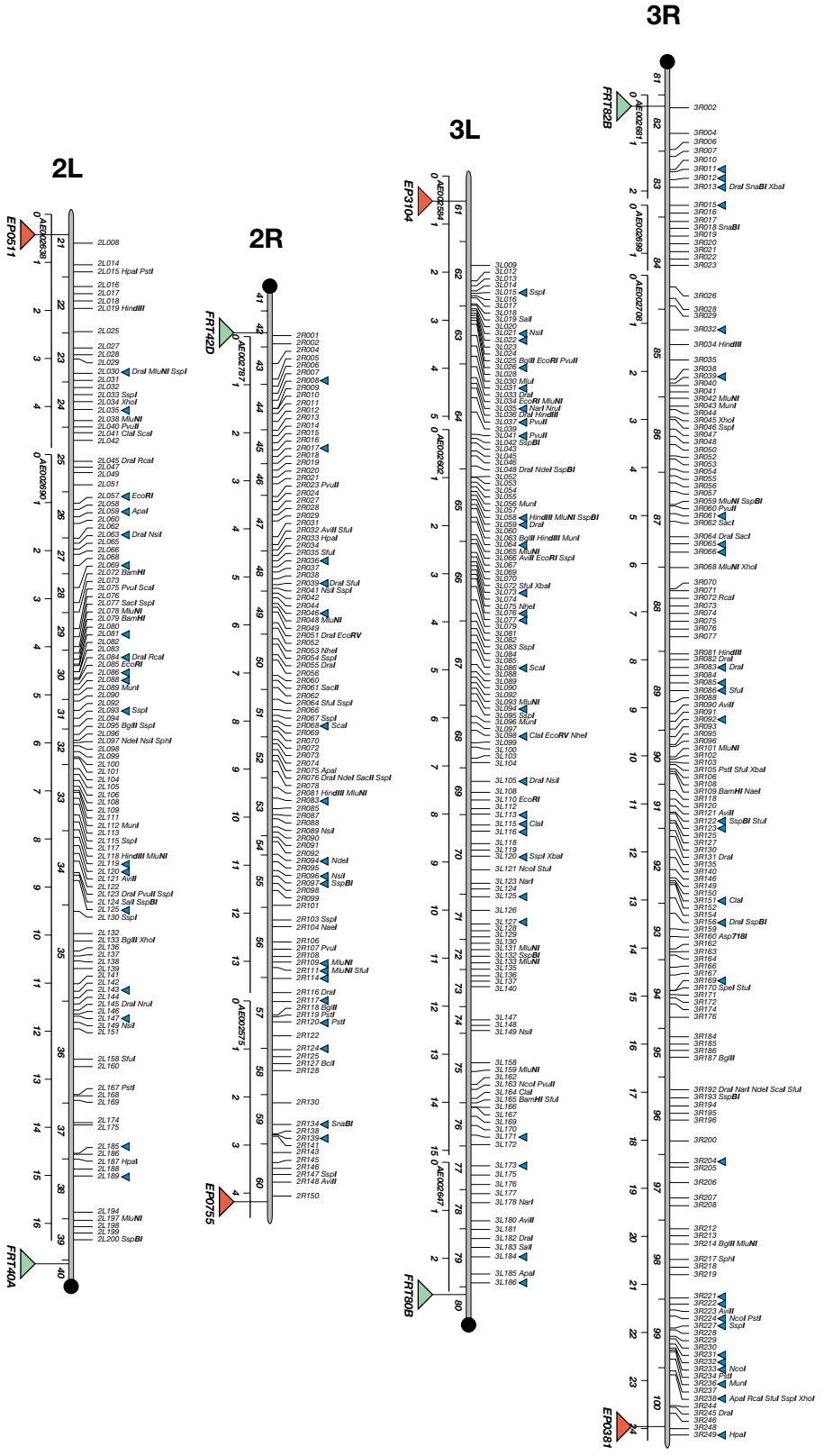


**Fig. 2** Polymorphism map of FRT versus EP strains. The locations of 495 SNP marker loci that discriminate between FRT and EP strains are shown. For each arm, the right scale shows the approximate extent of each cytological division, whereas the left scale indicates the precise location (in Mb) of each marker on one of the ten contigs of at least 1 Mb in Release 2 of the *Drosophila* genome sequence. The locations of the proximal FRT and distal EP insertions are shown. Arrowheads indicate markers for which PLP assays have been established. Restriction endonucleases for verified RFLP-SNPs are also indicated. The exact location and allelic variants for each polymorphism, including additional polymorphisms between these strains and the Release 2 genome sequence, are listed in Web Tables A–D.

the directed approach is more likely to be preferable if polymorphisms are abundant; the shotgun approach, if they are rare.

In *Drosophila*, polymorphisms are sufficiently abundant to favor the directed approach. Indeed, sequence-tagged sites generated during the *Drosophila* genome project[7] have recently been resequenced from several different wildtype strains and natural isolates to construct a map of 474 SNP marker loci[5,8]. We aimed to construct a similar map for useful laboratory strains, also using the directed approach. One additional advantage of this strategy is that it allows different strains to be selected for different genomic regions. We therefore selected a set of paired strains for each chromosome arm: one strain carrying an FRT insertion[9] at the base of the arm, and a corresponding strain carrying an insertion of an EP element[10] at the tip. Our choice was based on the following criteria. First, because of their usefulness in generating genetic mosaics, the FRT strains are frequently used as parental strains in large scale mutagenesis screens[11–15]. Second, traditional methods are particularly inadequate for mapping mutations that can be detected only in genetic mosaics, for which few visible markers are available and transheterozygous complementation tests are impossible. Third, the EP elements carry a dominant *white*$^{+}$ marker and, unlike many other transposon insertions, they have not been preselected for any mutant phenotype[10]. Fourth, flies carrying both a proximal FRT and a distal EP insertion on the same chromosome arm can readily be identified, greatly facilitating the recovery of a set of recombination events spanning almost an entire chromosome arm. We selected paired FRT and EP strains for each of the four major autosomal arms. The X chromosome was omitted from this analysis, as it is less frequently included in mutagenesis screens involving genetic mosaics. Recombination is exceedingly rare on the small fourth chromosome, and so it too was excluded.

We amplified and sequenced PCR products approximately 1 kb in size from each pair of FRT and EP homozygotes. We designed primers to amplify nonrepetitive, noncoding regions spaced at regular intervals across each autosomal arm. We then screened for polymorphisms in sequences from the two strains using POLYBAYES[16] and limited visual inspection. We screened a total of 821 pairs of PCR products. Of these, 787 (95.9%) contained at least one sequence polymorphism and thus define a set of marker loci useful for genetic mapping experiments (Fig. 1*a*). These marker loci are evenly distributed across each of the four major autosomal arms, with an average resolution of 1 marker every 114 kb. The largest gap is 578 kb. We identified a total of 8,615 polymorphisms (Web Tables A–D): 7,223 SNPs (83.8%; Fig. 1*b*) and 1,392 InDels (16.2%; Fig. 1*c*).

We detected sequence polymorphisms between FRT and EP strains, on average, once every 200 bp; and between each of these strains and the Release 2 genome sequence once every 157 and 153 bp, respectively (Fig. 1*a*). The density of polymorphisms is not uniform across each arm, being generally lower nearer the centromere and, to a lesser extent, towards the telomere (Fig. 1*d*). Polymorphism frequency thus correlates with recombination rate (Figs. 1*d*,*e*)[17], consistent with the prediction that selective sweeps should eliminate polymorphisms more effectively in regions with lower recombination rates[17]. Nevertheless, we detected sequence polymorphisms even in regions of low recombination. The largest gap between polymorphisms that we identified in any two strains was 1.86 Mb; the second largest, just 804 kb. Within any 2-Mb window, the average polymorphism frequency between any pair of strains is at least one polymorphism per kb. In existing laboratory strains of *Drosophila*, sequence polymorphisms are thus sufficiently abundant and widespread to support high-resolution genome-wide genetic mapping.

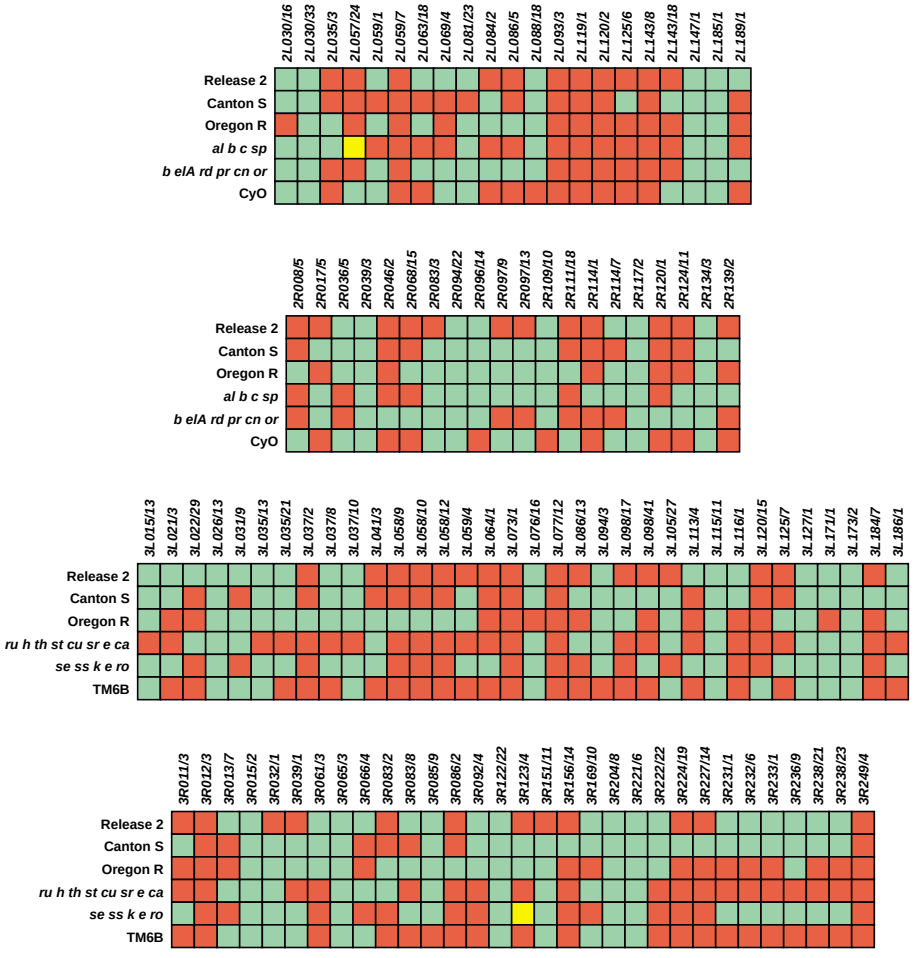


**Fig. 3** Genotypes of additional useful laboratory strains. We genotyped the indicated strains using PLP assays for 107 InDels. Numbers indicate the marker name, followed by the polymorphism number within that marker. Green indicates the same allele as the corresponding FRT strain, red indicates the EP allele and yellow indicates two cases in which a new product size was observed. Most FRT-versus-EP polymorphisms also discriminate other pairs of strains. These genotyped strains provide additional opportunities for generating and analyzing recombinants. For example, recombination events within specific chromosomal regions might be generated using the chromosomes carrying multiple recessive markers. In some cases it may also be desirable to score genotypes by analyzing DNA prepared from balanced flies.

To facilitate mapping using the sequence polymorphisms we identified, we next sought to establish a set of assays for rapidly genotyping a subset of those polymorphisms that discriminate between the FRT and EP strains (Fig. 2). We established PLP assays for 110 of the 134 InDels of 7 or more base pairs (Fig. 2 and Web Table E). These PLP assays can be used to genotype 93 markers. We then examined all SNPs for restriction-site polymorphisms and identified 1,592 predicted RFLP-SNPs (46.6% of all FRT-versus-EP SNPs). As most marker loci contain several SNPs, nearly all markers (83.0%) can potentially be scored using a restriction digest of PCR products. We have verified 239 of these RFLP-SNPs by a restriction digest of PCR products from FRT and EP homozygotes and FRT/EP heterozygotes (Fig. 2 and Web Table F). These verified RFLP assays can be used to genotype 162 markers. Using either of these established PCR assays, genotypes can be determined for 215 markers on the FRT-versus-EP map at an average resolution of 419 kb.

The SNP map and PCR assays we have established for the FRT and EP strains will be useful immediately for mapping the many mutations that have been induced on FRT chromosomes in screens involving genetic mosaics. If mutations induced on other genetic backgrounds are to be mapped using these markers and assays, a sufficient number of these markers must also be polymorphic in other pairs of strains. To assess this, we used the PLP assays to genotype 107 InDels in several additional strains: two wildtype strains, two strains carrying multiple recessive markers and one 'balancer' strain for each autosome (Fig. 3). Eighty-two (76.6%) of these InDels were indeed polymorphic between at least one other pair of strains in addition to the FRT and EP strains. Most of these markers and assays should therefore also be useful for mapping mutations using different pairs of strains.

To assess the feasibility of mapping using these markers and assays, we next sought to localize two mutations, *3L-H1135* and *3R-I256*, that had been isolated in a mosaic screen for abnormal patterns of neuronal connectivity in the adult visual system (ref. 13 and T. Newsome, B. Åsling, G. Gahmon, K.-A.S. and B.J.D., unpublished data). These two mutations had been induced by chemical mutagenesis on the FRT80B and FRT82B chromosomes, respectively. We first isolated sets of 45 and 94 recombinants, respectively, between the proximal FRT and distal EP insertions using an eye-specific source of FLP recombinase[13] so

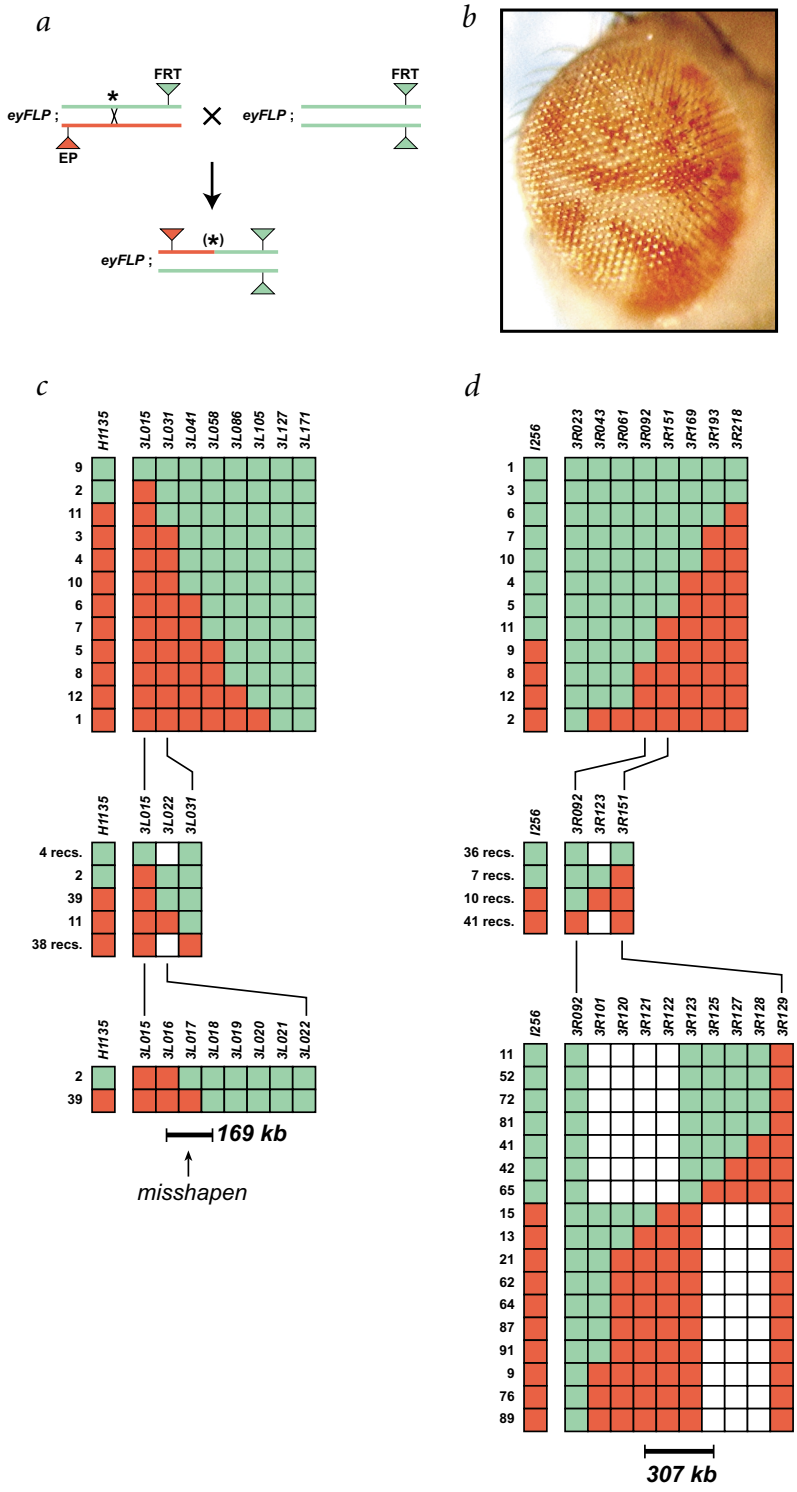


**Fig. 4** Genetic mapping with SNP markers. ***a***, Schematic of the crossing strategy used to isolate a set of recombinants spanning an entire chromosome arm. The asterisk indicates a mutation initially present on the parental FRT chromosome, and on all recombinant chromosomes for which the crossover has occurred distal (left) to the mutation. ***b***, In the presence of the *eyFLP* transgene, eye-color mosaicism can be used to identify recombinant chromosomes carrying both the FRT and EP insertions. ***c***,***d***, Mapping of the *3L-H1135* (*c*) and *3R-I256* (*d*) mutations. Mapping proceeded in three stages. First, we obtained a low-resolution map position by analyzing a random selection of 12 recombinants (top panels). The left columns indicate the recombinant number and the presence (green) or absence (red) of the visual system connectivity defect. The columns on the right show the results of PLP assays used to genotype each of the indicated SNP markers (green, FRT; red, EP). This initial stage placed each mutation between a pair of markers separated by 1–3 Mb. We next screened all remaining recombinants for informative crossover events that had occurred between these two markers (middle panels). We also scored one intervening marker for each mutation (white, not determined). We then used these informative recombinants to obtain a high-resolution map position (lower panels). In this final stage, we determined SNP genotypes by DNA sequencing.

that recombinants could be identified by their mosaicism for the *white*$^+$ eye-color marker located on the EP element (Fig. 4*a*,*b*). We then scored each recombinant for the connectivity phenotype and determined marker genotypes by either PLP or RFLP assays or by direct sequencing of PCR products. This analysis placed the *3L-H1135* and *3R-I256* mutations within 169-kb and 307-kb intervals, respectively (Fig. 4*c*,*d*). The region containing *3L-H1135* also includes the gene *misshapen*, which has previously been shown to function in photoreceptor axon pathfinding[18]. Complementation and phenotypic analysis identified the *3L-H1135* mutation as a new allele of *misshapen*, confirming the accuracy of this mapping strategy.

This mapping procedure proved to be extremely rapid. Once the recombinant progeny had been generated and scored for the connectivity defect, it took only two to three days to map each mutation. The rate-limiting step in this procedure was the histological analysis of neuronal connectivity in each of the recombinant progeny. Clearly, more visible phenotypes could be mapped with even greater efficiency. In such cases it may also be desirable to carry out PCR genotyping assays on DNA samples prepared from several rather than individual recombinants. Although this approach sacrifices the ability to localize a mutation unambiguously between a specific pair of markers, it would allow an approximate map location to be determined within just a few hours of obtaining the recombinants.

With an average density of one sequence polymorphism every 150–200 bp, it should in principle be possible to map mutations down to the sub-genic level. To assess the feasibility of such high-resolution genetic mapping, we examined a third mutation from the same connectivity screen, *2R-R971*. Both the mosaic visual system connectivity defect and homozygous lethality associated with this mutation had previously been localized by traditional methods to the cytological interval 59EF. We first constructed a high-resolution SNP map within this interval, comparing again the parental FRT and the distal EP chromosomes. To localize the *2R-R971* mutation on this map, we generated a set of 71 recombination events between the *2R-R971* mutation and a P{*white*$^+$}insertion within the 59D–F region (Fig. 5). Of 11 markers tested, 8 were found to discriminate between the parental *2R-R971* and P{*white*$^+$} chromosomes. We scored these markers on the recombinant progeny, placing the *2R-R971* mutation within a 2-kb interval that includes the 3′ region of the *dead ringer* gene (*dri*; Fig. 5). Complementation, phenotypic and sequence analyses confirmed that *2R-R971* is

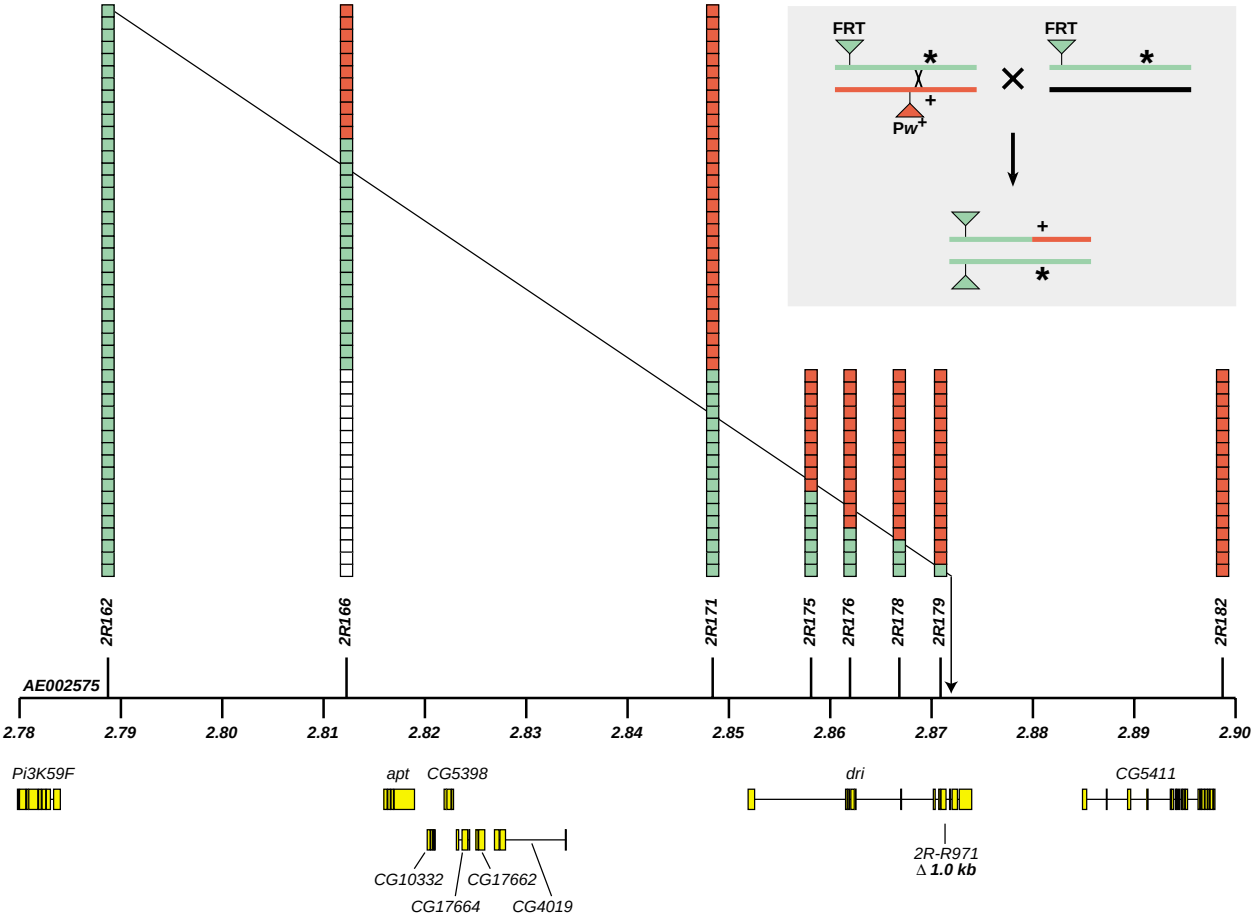


**Fig. 5** High-resolution mapping. The inset in the upper right illustrates the strategy used to recover a set of recombinants between the *2R-R971* mutation (asterisk) and a closely linked P{*white*$^+$} insertion. Green indicates the FRT chromosome carrying the *2R-R971* mutation. Red indicates the P{*white*$^+$} chromosome that carries a wildtype allele at this locus (+). Black indicates the CyO balancer chromosome. The map shows nucleotide positions in the AE002575 contig, in Mb. Proximal is to the left. Boxes indicate SNP genotypes determined for each recombinant (rows) at the indicated SNP marker (columns). All SNP genotypes shown here were determined by sequence analysis (green, FRT; red, P{*white*$^+$}; white, not determined). Of 71 recombination events, 24 occurred proximal to the *2R162* marker (not shown), 46 occurred in the 83-kb interval between *2R162* and *2R179* and one occurred distal to *2R179*. The *2R-R971* mutation thus lies distal to *2R179* but probably less than 2 kb from it (given the average of one recombination event every 1.8 kb). Sequencing the entire gene *dri* from the *2R-R971* chromosome revealed a 1,005-bp deletion, 661 bp distal to the *2R179* marker. No evidence of recombination hotspots could be detected at this resolution.

indeed an allele of *dri*. The practical limit of SNP mapping is thus likely to be imposed only by the density of identified SNPs, rather than by a lack of available polymorphisms or the presence of recombination hot spots.

Our SNP map and PCR assays provide a means for fast, inexpensive and high-resolution genetic mapping using existing laboratory strains of *Drosophila*. We have demonstrated both the feasibility and accuracy of this method. In conjunction with the many methods currently being developed for high-throughput SNP genotyping[19], our SNP map should now allow high-throughput, high-resolution genetic mapping in *Drosophila*. Positional cloning should no longer be the rate-limiting step in the forward genetic analysis of complex biological phenomena.

## Methods

***Drosophila* stocks.** The FRT and EP strains we selected are shown in Fig. 2. We isogenized both sets of strains before constructing the SNP map. We obtained Canton S, Oregon R and the strains with multiple recessive markers from the Bloomington *Drosophila* Stock Center. The CyO and TM6B balancer chromosomes each carried a P{*yellow*$^+$} insertion. The generation of these strains and the isolation of the visual system connectivity mutants *2R-R971*, *3L-H1135* and *3R-I256* were previously described[13].

**Strategy for SNP identification.** Consider the general problem of constructing a SNP map for two strains, A and B, given a reference genome, G. One approach is to carry out light shotgun sequencing of A, align these sequences to G and determine the genotype of B for any polymorphic site thus identified. Alternatively, G might be used to design primers for the amplification and sequencing of selected loci from both A and B, which can then be directly compared. The cost per validated SNP for the shotgun ($S$) and directed ($D$) approaches can be estimated (see Web Note A) by:

$$S=\frac{1}{f}\left(\frac{C_r}{(1-e^{-p_{AG}L})(1-r)}+2(C_p+C_a)\right)$$

$$D=2\left(\frac{C_p+2C_r}{s(1-e^{-2p_{AB}L})}+C_a\right)$$

where:

$C_r$=the cost of a sequencing reaction,
$C_p$=the cost of a primer for PCR and sequencing,
$C_a$=the cost of a genotyping assay (excluding primers),
$p_{XY}$=SNP frequency between strains X and Y,
$L$=read length,
$r$=repetitive fraction of the genome,
$s$=the success rate for genotyping assays using the original pair of primers, and
$f$=the probabilty of any A versus G polymorphism also being an A versus B polymorphism, given by the formula:
$f=(1+(p_{AB}-p_{BG})/p_{AG})/2$.

We have calculated $D$ and $S$ for a wide range of values for these parameters. Extreme values can be found under which, irrespective of polymorphism frequency ($p_{AB}$), either $D$ is always less than $S$ (such as a highly repetitive genome ($r\approx1$)) or $S$ is always less than $D$ (if genotyping assays are difficult to design using the original primers ($s\approx0$), B is largely isogenic with G ($p_{BG}\approx0$) or primers and assays are much cheaper than sequencing reactions ($C_p$,$C_a\approx0$)). Over a broad range of realistic values, however, there is a cut-off polymorphism frequency, above which $D<S$ and below which $S<D$. Specifically, for the FRT and EP *Drosophila* strains, we determined that the directed approach is favorable if one is interested in any SNP or only RFLP-SNPs, whereas the shotgun approach would be more useful if one were interested only in larger (≥7 bp) InDels.

**SNP identification.** We searched for sequence polymorphisms in each of the ten autosomal contigs in Release 2 of the *Drosophila* genome with a length of at least 1 Mb (ref. 20 and Berkeley *Drosophila* Genome Project/Celera, 4 October 2000). Together, these contigs comprise 90,043 kb. We selected loci for amplification from noncoding and nonrepetitive regions evenly distributed across each contig. We defined noncoding regions by the lack of any known or predicted open reading frame and nonrepetitive regions by the presence of a single high-scoring BLAST hit on the Release 2 genome sequence.

We carried out PCR amplification with genomic DNA prepared from FRT or EP homozygotes, primers (0.4 μM), $MgCl_2$ (1.5 mM), dNTPs (0.2 mM each) and Taq DNA polymerase (TaKaRa, 1.25 U). Cycle conditions were 94 °C for 5 min, followed by 35 cycles at 94 °C for 30 s, 62 °C for 30 s and 72 °C for 2 min and a final extension at 72 °C for 7 min. We carried out PCR analysis using either a GeneAmp 9700 (Perkin-Elmer) or Primus HT (MWG Biotech). PCR products were treated with ExoSAP-IT (US Biochemicals) to remove primers and excess nucleotides and sequenced on an ABI377 or ABI3100 in two separate reactions, each using one of the PCR primers as a sequencing primer.

Individual sequence traces were called with PHRED 0.000925.c (refs. 21,22). Sequences were then preclustered into 1-Mb genomic intervals according to the initial PCR primer selection and pair-wise alignments to the corresponding Release 2 genome sequence were prepared for POLYBAYES using CrossMatch (P. Green, unpublished data). We then multiply re-aligned the sequences using the POLYBAYES 3.0 (release 2001-03-27) anchored-alignment algorithm[16]. We used the default parameters, except that the paralog filtering was unnecessary because of the preselection of unique regions and was therefore supressed. The Release 2 genome sequence was assigned a PHRED score of 40.

We retained polymorphisms reported by POLYBAYES only if the PHRED quality score of each variant nucleotide exceeded 20 and the $P_{SNP}$ value assigned by POLYBAYES exceeded 0.5 (corresponding to a PHRED score of 33 in a pair-wise alignment). We also aligned and visually inspected FRT and EP sequence traces using the SeqMan program (DNA-Star). We used these alignments both to make minor adjustments to the POLYBAYES alignment and for the detection of most InDels larger than 8 nt. We screened SNPs for the presence of recognition sites for a panel of 186 restriction endonucleases. A list of these enzymes is available at http://www.imp.univie.ac.at/groups/dickson/dickson.html.

**PCR primer design.** We designed all PCR primers using Primer3 (Whitehead Institute/MIT Center for Genome Research; http://www-genome.wi.mit.edu). For SNP identification and PLP assays, we selected primers with melting temperatures of 67–73 °C, maximum $T_m$ difference of 1 °C, length of 20–30 nt, 20–80% GC content, maximum self-complementarity of 4.0 and maximum 3′ self-complementarity of 1.0. Primers for SNP identification were chosen to give a product of 1.0–1.3 kb. For PLP assays, the product size was typically 130–230 bp (for InDels of up to 20 bp) or 200–500 bp (for larger InDels). Primers used for SNP identification were also used for RFLP assays, although they are not necessarily the optimal choice for distinguishing the alternative fragment sizes. All oligonucleotides were synthesized by MWG Biotech.

**PLP and RFLP assays.** We prepared genomic DNA from either 20–50 flies in 1-ml reaction tubes or 1–2 flies in 96-well plates. For the former, DNA was resuspended in 50 μl, of which 1–2 μl was used for PCR. For preparation in the 96-well format, we manually crushed flies in 50 μl of extraction buffer (10 mM *Tris*-Cl, 1 mM EDTA, 25 mM NaCl, 200 μg/ml Proteinase K) and incubated them at 37 °C for 30 min. We then inactivated proteinase K by incubation at 95 °C for 5 min and used 1 μl for PCR. PCR conditions were identical to those described above, except that the extension times were reduced to 1 min for PLP assays. For these assays, we separated amplification products on 2.5% agarose gels. For RFLP assays, we digested 5 μl PCR product with the appropriate restriction enzyme in a 20-μl reaction and separated the resulting fragments on a 1% agarose gel.

We selected a set of 274 predicted RFLP-SNPs for verification based on the presence of a recognition site for a readily obtainable restriction endonuclease. We confirmed experimentally 87% of these predicted RFLP-SNPs. Negative results may be due to either the presence of multiple recognition sites for the selected enzyme, similarities in the fragment sizes from the two strains or a false prediction. We expect that, with optimal primer design, RFLP assays could be designed for more than 90% of the predicted RFLP-SNPs, allowing genotyping of 98% of those markers that contain at least one RFLP-SNP (89% of all markers).

**High-resolution mapping of *2R-R971*.** We recovered recombinants between the *2R-R971* mutation and one of four selected P{*white*$^+$} insertions in the 59DE region: *l(2)k07136*, *l(2)s4830*, *l(2)k06908*, or *l(2)k05606*. We identified recombinants by screening for chromosomes that lacked the *white*$^+$ marker but were viable in *trans* to *2R-J1609*, an independent mutation that has the same visual system connectivity defect, maps to 59DF and fails to complement the lethality of *2R-R971*. We recovered 71 recombinants from approximately 30,000 progeny.

*Note: Supplementary information is available on the Nature Genetics web site (http://genetics.nature.com/supplementary_info/).*

**Acknowledgments**

*We thank F. Eisenhaber and S. Maurer-Stroh for providing assistance in bioinformatics, A. Graf for technical assistance and M. Hohl and I. Botto for DNA sequencing. We are also grateful to the Berkeley Drosophila Genome Project, Celera Genomics and FlyBase for providing the wealth of resources that made this project possible. This work was funded by Boehringer Ingelheim GmbH.*

Received 30 August; accepted 17 October 2001.

1. Collins, F.S., Guyer, M.S. & Charkravarti, A. Variations on a theme: cataloging human DNA sequence variation. *Science* **278**, 1580–1581 (1997).
2. Winzeler, E.A. *et al.* Direct allelic variation scanning of the yeast genome. *Science* **281**, 1194–1197 (1998).
3. Cho, R.J. *et al.* Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. *Nature Genet.* **23**, 203–207 (1999).
4. Wicks, S.R., Yeh, R.T., Gish, W.R., Waterston, R.H. & Plasterk, R.H. Rapid gene mapping in *Caenorhabditis elegans* using a high density polymorphism map. *Nature Genet.* **28**, 160–164 (2001).
5. Hoskins, R.A. *et al.* Single-nucleotide polymorphism markers for genetic mapping in *Drosophila melanogaster*. *Genome Res.* **11**, 1100–1113 (2001).
6. Lindblad-Toh, K. *et al.* Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nature Genet.* **24**, 381–386 (2000).
7. Kimmerly, W. *et al.* A P1-based physical map of the *Drosophila* euchromatic genome. *Genome Res.* **6**, 414–430 (1996).
8. Teeter, K. *et al.* Haplotype dimorphism in a SNP collection from *Drosophila melanogaster*. *J. Exp. Zool.* **288**, 63–75 (2000).
9. Xu, T. & Rubin, G.M. Analysis of genetic mosaics in developing and adult *Drosophila* tissues. *Development* **117**, 1223–1237 (1993).
10. Rørth, P. *et al.* Systematic gain-of-function genetics in *Drosophila*. *Development* **125**, 1049–1057 (1998).
11. Xu, T., Wang, W., Zhang, S., Stewart, R.A. & Yu, W. Identifying tumor suppressors in genetic mosaics: the *Drosophila lats* gene encodes a putative protein kinase. *Development* **121**, 1053–1063 (1995).
12. Liu, Y. & Montell, D.J. Identification of mutations that cause cell migration defects in mosaic clones. *Development* **126**, 1869–1878 (1999).
13. Newsome, T.P., Asling, B. & Dickson, B.J. Analysis of *Drosophila* photoreceptor axon guidance in eye-specific mosaics. *Development* **127**, 851–860 (2000).
14. Benlali, A., Draskovic, I., Hazelett, D.J. & Treisman, J.E. *act up* controls actin polymerization to alter cell shape and restrict Hedgehog signaling in the *Drosophila* eye disc. *Cell* **101**, 271–281 (2000).
15. Pichaud, F. & Desplan, C. A new visualization approach for identifying mutations that affect differentiation and organization of the *Drosophila* ommatidia. *Development* **128**, 815–826 (2001).
16. Marth, G.T. *et al.* A general approach to single-nucleotide polymorphism discovery. *Nature Genet.* **23**, 452–456 (1999).
17. Begun, D.J. & Aquadro, C.F. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* **356**, 519–520 (1992).
18. Ruan, W., Pang, P. & Rao, Y. The SH2/SH3 adaptor protein dock interacts with the Ste20-like kinase misshapen in controlling growth cone motility. *Neuron* **24**, 595–605 (1999).
19. Kwok, P.-Y. Methods for genotyping single nucleotide polymorphisms. *Annu. Rev. Genomics Hum. Genet.* **2**, 235–258 (2001).
20. Adams, M.D. *et al.* The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195 (2000).
21. Ewing, B., Hillier, L., Wendl, M.C. & Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**, 175–185 (1998).
22. Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**, 186–194 (1998).
23. Petrov, D.A. & Hartl, D.L. Patterns of nucleotide substitution in *Drosophila* and mammalian genomes. *Proc. Natl Acad. Sci. USA*. **96**, 1475–1479 (1999).