High-resolution SNP mapping by denaturing HPLC

Knud Nairz*, Hugo Stocker*, Benno Schindelholz†, and Ernst Hafen*‡

*Zoologisches Institut der Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland; and †The Genetics Company, Incorporated, Wagistrasse 27, CH-8952 Schlieren, Switzerland

Edited by Michael S. Levine, University of California, Berkeley, CA, and approved June 13, 2002 (received for review March 7, 2002)

With the availability of complete genome sequences, new rapid and reliable strategies for positional cloning become possible. Single-nucleotide polymorphisms (SNPs) permit the mapping of mutations at a resolution not amenable to classical genetics. Here we describe a SNP mapping procedure that relies on resolving polymorphisms by denaturing HPLC without the necessity of determining the nature of the SNPs. With the example of mapping mutations to the *Drosophila nicastrin* locus, we discuss the benefits of this method, evaluate the frequency of closely linked and potentially misleading second site mutations, and demonstrate the use of denaturing high-performance liquid chromatography to identify mutations in the candidate genes and to fine-map chromosomal breakpoints. Furthermore, we show that recombination events are not uniformly dispersed over the investigated region but rather occur at hot spots.

SNP detection | recombination mapping | DHPLC | nicastrin | recombination hot spots

Genome-wide genetic screens in model organisms like *Drosophila* facilitate the identification of genetic loci involved in biological processes (1). However, the isolation of the affected gene is a tedious process, particularly because the most commonly used mutagen, EMS, primarily induces point mutations. Standard procedures to localize point mutations involve (*i*) identification of the affected chromosome (arm); (*ii*) mapping by noncomplementation with chromosomal deficiencies; and (*iii*) recombination mapping relative to visible markers. In this way, it is generally possible to map mutations to a region of a few hundred kilobase pairs (kb). Limiting factors are the resolution of the meiotic map, the accuracy of determined deficiency breakpoints, and misleading effects caused by second site lethal mutations.

To meiotically map mutations to the level of a single gene, sequence polymorphisms such as single-nucleotide polymorphisms (SNPs), nucleotide insertions, or deletions are exploited as genetic markers. These are widespread, even in regions devoid of phenotypic markers, usually genetically inert, and dominant. An all-molecular approach to recombination mapping is aided by genome-wide SNP maps that have been established for some *Drosophila* strains (2, 3). However, the resolution of these maps (114–1,000 kb) is in many cases not sufficient to localize the candidate gene and also is not much higher than the resolution attained by classical techniques. Moreover, in contrast to more recently introduced model organisms like *Arabidopsis thaliana* and *Caenorhabditis elegans*, for which clearly defined lines exist, the genetic backgrounds of *Drosophila* strains are often heterogeneous and difficult to trace, such that SNP maps cannot readily be applied to other strains. Furthermore, the complete reliance on sequence polymorphisms requires PCR amplification of the genomic DNA, which becomes a rate-limiting factor (4). A mapping strategy combining visible and molecular markers for recombination mapping can reduce efforts and costs (5).

When a genetic locus has been mapped at low resolution by any of the described classical methods, it is necessary to establish a high-resolution SNP map between mutant and tester strains. For this purpose, sequence polymorphisms have to be identified in the target region, which normally involves the amplification of genomic DNA, DNA sequence determination, identification of polymorphisms, and the subsequent development of a detection assay (4). As the exact nature of the polymorphism is irrelevant for mapping purposes, a method permitting the reliable detection of the mere difference in DNA sequence without determining the actual nature of the difference could save the sequencing and assay development steps.

Possibly the most advanced method for mutation detection without sequencing is denaturing HPLC, which permits the resolution of heteroduplex DNA in PCR fragments of up to ≈1,000 bp in length (6–8). The underlying principle of the technique is a slightly altered melting behavior of heteroduplexes versus homoduplexes, leading to a difference in retention time on ion-pair reversed-phase HPLC columns. On a chromatogram, DNA homoduplexes generally elute in one peak, whereas DNA heteroduplexes produce one or more additional peaks. The sequence difference is thus translated into an altered chromatogram that eliminates the need to obtain information about the change at the DNA sequence level.

Here we demonstrate a recombination mapping procedure combining visual and SNP markers achieving low and high resolution, respectively. SNP mapping was successfully carried out on a denaturing high-performance liquid chromatography (DHPLC)-based map with no information on the nature of the polymorphism. The strategy thus both minimizes the rate-limiting PCR step and drastically reduces the effort associated with SNP detection. This latter aspect is of general interest, because it also applies to any utilization of SNPs, e.g., for mapping of human disease loci. Furthermore, we fine mapped the breakpoints of recombinant chromosomes and found evidence for meiotic hot spots of recombination. The SNP map was also used to determine the breakpoints of a deficiency deriving from an unrelated strain background. Finally, we applied DHPLC for screening of candidate genes and found a surprisingly high rate of second-site mutations.

GENETICS

Materials and Methods

***Drosophila* Stocks and Generation of Recombinants.** The *FRT82B* strain containing *eyFLP* was obtained from B. Dickson (Institute of Molecular Pathology, Vienna) (9) and originates from the Rubin lab (10). Recombinants were generated by the crossing schemes depicted in Figs. 2 and 3 between the mutant *FRT82B* strain and either a strain carrying the w^+ marked *EP3455* inserted at 96B1 (11) or a strain harboring a phenotypically neutral y^+ marked enhancer promoter (EP) transposable element inserted at 96B20 (S. Breuer, personal communication). The *EP*y^+ element was constructed by D. Nellen (Institut für Molekularbiologie, Universität Zürich, Zürich) and is inserted in a *y w* strain background. *y w* flies have been kept in our lab for years and were provided by W. Gelbart (Department of Molecular and Cellular Biology, Harvard University) (12). The *rab1* mutation was found in a collection of P-element insertions (13) (F. Rintelen, personal communication) and derives from a *y w*

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: EMS, ethyl methanesulphonate; DHPLC, denaturing high-performance liquid chromatography; SNP, single-nucleotide polymorphism; EP, enhancer promoter element.

‡To whom reprint requests should be addressed. E-mail: hafen@zool.unizh.ch.

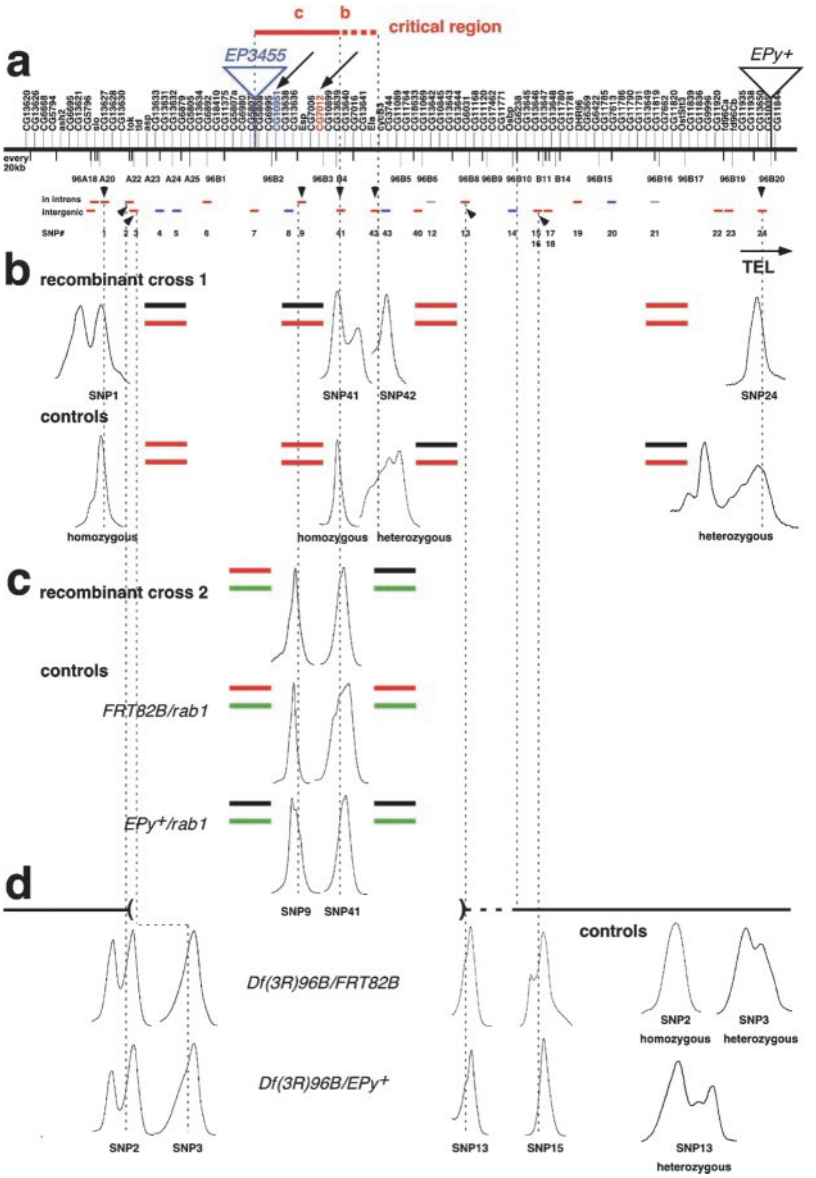

Fig. 1. The chromosomal organization of cytological region 96A18 to 96B20 and the outcome of the SNP-mapping experiments. (*a*) The order of the annotated genes from proximal to distal and the insertion sites of the markers used to generate recombinants are indicated. The map is not drawn to scale: thick vertical bars mark every 20 kb, and thin bars refer to the estimated cytological position. Upper horizontal bars symbolize PCR fragments that were amplified from intronic regions; lower bars are products from intergenic regions. Red symbolizes a DHPLC polymorphism, blue a nonmeaningful fragment, and gray a nonamplifiable fragment. SNP marker identifications are given below. Arrowheads point to SNPs shown in *b–d* that are also aligned with the map by dashed lines. The genes affected by mutagenesis are marked by an arrow and highlighted in red and blue, respectively, and the critical region remaining after SNP mapping is indicated. (*b*) The chromatographic profile of the closest recombinant from cross 1 is depicted. Note that between SNP41 and SNP42, there is a switch from a heterozygous to a homozygous pattern. The respective profiles derived from the *FRT82B* parental strain and the heterozygous *FRT82B/EPy*$^+$ controls are shown below. Bars next to the peaks symbolize the genotype at the site according to the color code from Fig. 2. (*c*) The closest recombinant from cross 2 switches from a *FRT82B/rab1* to a *EPy*$^+$*/rab1* pattern at the SNPs adjacent to the breakpoint. The profiles of control flies and the color-coded genotypes are indicated. (*d*) The deletion chromosome is polymorphic to *FRT82B* at SNP15 and exhibits a polymorphism at SNP2 with both *FRT82B* and *EPy*$^+$. However, at SNP3 and SNP13 the profile is homozygous against both parental strains indicating loss of heterozygosity. The breakpoints thus place between SNP2 and SNP3 and between SNP13 and SNP15. *CG6238/slingshot* is situated between the latter SNPs and the mutation complemented by the deficiency thus indicating that the breakpoint is more toward SNP13. Homozygous or heterozygous control profiles are shown and the uncovered region aligned to the map in *a* falls into the region between the parentheses.

PlacW strain. It is unclear whether the high incidence of polymorphisms at the investigated region in *y w* versus *y w PlacW* flies is because of the accumulation of spontaneous mutations or a different origin of the third chromosome. The *rab1*$^{737/14}$ allele, which proved to be cell lethal, was recombined onto a *FRT82B* w^+ chromosome. Crosses were set up at 25°C.

Screening and Classical Mapping. *y w; FRT82B* males were fed with 33 mM EMS according to standard protocols and crossed to females of the genotype *y w eyFLP*; *FRT82B* w^+ *cl3R3.7/TM6B* y^+. The F_1 progeny were scored for eyes and heads of abnormal size. The underlying principle of the clonal expansion of mutant head tissue is the exclusive action of the *eyFLP* recombinase in head progenitor cells and the unmutagenized *FRT82B* chromosome carrying the recessive cell lethal mutation *cl3R3.7*. Two cell populations are generated after *eyFLP*-induced recombination at the *FRT* sites: one homozygous for the newly induced mutation and the other homozygous for the cell lethal mutation,

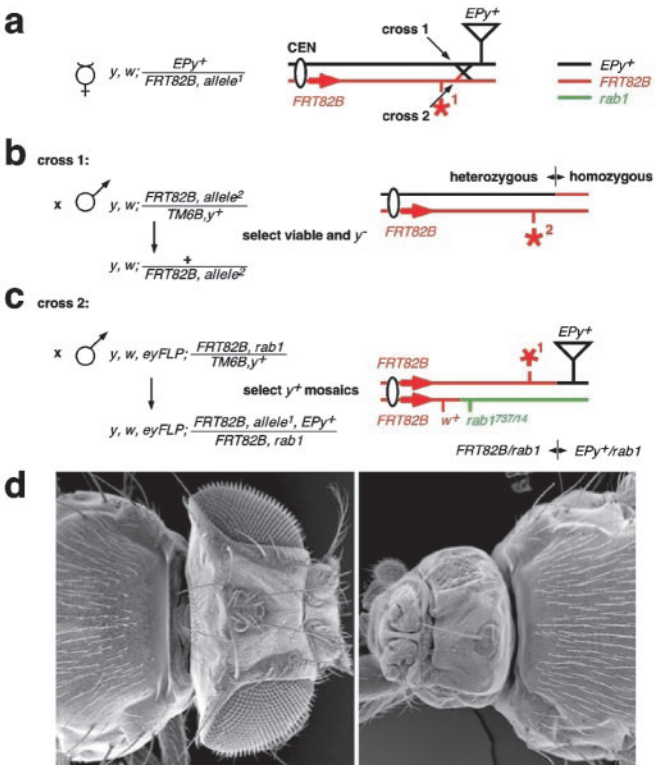

Fig. 2. The crossing scheme to recover recombinants with a distal, y^+ marked, and phenotypically neutral EP element. (*a*) A recombination event between the mutation (asterisk) on the red *FRT82B* chromosome and the black *EPy*$^+$ parental chromosome yields two complementary outcomes (see arrows), which can be selected for by suitable crosses depicted in *b* and *c*. By crossing back to a y^+-balanced other allele (to avoid effects of second hits), recombinants are distinguished as y^- flies. Distal to the breakpoint a recombinant is homozygous for the *FRT82B* background (*b*). (*c*) The reciprocal recombinant is discerned by the clonal plus the marker phenotype. To allow for expansion of the clonal tissue, we used an *FRT82B* chromosome onto which a cell-lethal transposon insertion at the *rab1* locus was recombined. At the investigated region, the chromosome thus provides a third strain background (*PlacW*) that is symbolized by green. Distal to the breakpoint, a recombinant has a *EPy*$^+$/*rab1* and proximal to it a *FRT82B*/*rab1* genotype. (*d*) A *nct*^{5P3} "pinhead" (*Right*) compared with a wild-type head.

which will be eliminated. By using the *FRT82B* w^+ y^+ marker chromosome (10), the loci were meiotically mapped relative to the visible markers w^+ (at cytological position 90E) and y^+ (96E).

DNA Isolation. Genomic DNA was prepared by homogenizing a single recombinant fly in 40 μl of TE (10 mM Tris/1 mM EDTA, pH 8.0), containing 25 mM NaCl, 0.2% Triton X-100, and 200 μg/μl of proteinase K, incubating the suspension for 30 min at 37°C, and finally inactivating proteinase K for 5 min at 95°C.

PCR. We searched the *Drosophila* genome (release 2) (14) for intergenic regions or introns from cytological position 96A13 to 96C2. Primer pairs were designed to amplify fragments at melting temperatures of ≈55°C. PCR reactions contained 1 μl of DNA, 1 μl primer each (10 pmol/μl), 0.2 mM each dNTP, 2 mM $MgCl_2$, 5 μl of 10 × buffer, and 1.25 units of *Taq* (Sigma). After being melted for 3 min at 95°C, they were cycled 32 times at 94°C for 20 sec, 55°C for 30 sec, 72°C for 1 min, and finally extended for 2 min at 72°C. To allow for efficient heteroduplex formation, the reactions were subsequently heated to 95°C for 30 sec and then slowly cooled to room temperature. However, we obtained comparable results when omitting this final step.

Table 1. *EPy*$^+$ and *FRT82B* strains are sufficiently polymorphic to aid SNP mapping

	Location of amplicons		
	Intronic	Intergenic	Total
No. of fragments analyzed by DHPLC	12	28	40
No. of polymorphisms found by DHPLC	10	17	27
DHPLC polymorphisms, %	83%	61%	67.5%
Frequency of polymorphisms at the sequence level	1/250	1/593	1/377

Total and relative numbers are given for number of analyzed and polymorphic fragments, and the rate of sequence polymorphism between the two strains determined for 29 of those 40 amplicons is listed. Values are also categorized according to the location of the fragments.

DHPLC. Analysis was carried out on a Transgenomic (Omaha, NE) HPLC apparatus. For the calibration step, we first analyzed the predicted melting profile of the fragments by the supplied WAVEMAKER (Transgenomic) software to determine the optimal temperatures and gradient conditions to be tested. We then directly loaded PCR reactions derived from homozygous and heterozygous flies and analyzed them at the determined temperatures, injecting 5 μl per run. Calibration was successful when the chromatogram of a heterozygous fragment deviated from the control homozygous fragment. For SNP mapping, amplicons derived from recombinants, homo-, and heterozygous control flies were run at the established conditions.

Polymorphism Density and DHPLC Accuracy. To test the accuracy of DHPLC, we sequenced 29 heterozygous fragments, 18 of which showed a polymorphism by DHPLC. The exact nature of the polymorphism was determined by "double-peak" analysis. In only one case, sequencing revealed a SNP that was not detected by DHPLC. Therefore, in our hands, DHPLC accuracy is 95% (18/19), which is comparable to numbers achieved in more elaborate studies (15). By amplifying a smaller segment surrounding the unrecognized SNP, the polymorphism could be detected by DHPLC as well (Fig. 1*d*, SNP3). In the ≈23.4-kb region tested, the average polymorphism rate between the two parental strains is 1 per 377 bp. More than 80% of the polymorphisms are SNPs, and the others are small insertions or deletions. Interestingly, introns were more than twice as polymorphic as intergenic regions (Table 1).

Statistical Analysis. If the recombination breakpoints in the investigated interval between *nct* and the *EPy*$^+$ at 96B20 were randomly dispersed, one would expect a Poisson distribution of recombination events. The average recombination ratio in the investigated region of 267.9 kb and 1.01 cM (33 y^+ flies among 3,263 "pinheads") physical and genetic distance, respectively, is 3.775 cM/Mb. The recombination rate per investigated interval in cM per Mb is given in Fig. 5. Those values were categorized into arbitrary classes, and the expected number of events was determined according to the Poisson formula $P(k,n) = n^k e^{-n}/k!$ (which yields the probability for k events when the mean is n) (Table 2). $\chi^2 = \Sigma(y_i - x_i)^2/x_i$ for those values amounts to 66,744, indicating an extreme deviation from the null hypothesis that the observed distribution is a Poisson distribution (3° of freedom).

Results

A Genetic Locus with a Growth Phenotype Resembling *Ras* Mutations. Our gene identification strategy is exemplified by mapping the recessive genetic locus *S5A*, identified in a F_1 mosaic screen for genes affecting cellular growth. Exploiting the *eyFLP* technique

GENETICS

Table 2. Actual versus expected frequencies of recombination rates classified into seven arbitrary intervals

Interval, cM/Mb	Events y_i		Expected x_i	
0–2	8		5.23	
2–4	4		6.21	
4–6	0	6	2.95	3.82
6–8	0		0.75	
8–10	6		0.12	
10–12	6		0.013	
12–14	8		0.001	

Three classes have to be joined, because two are not represented by any values.

(9), animals were generated that were largely homozygous for a mutagenized *FRT82B* chromosome in the head but heterozygous for the mutation in the body (*Materials and Methods*). Mutations affecting cell growth but not differentiation produce flies with smaller (pinheads) or larger heads. The *Drosophila Ras* gene influences both growth and differentiation, and *Ras* mutations produce a very small head with scarred eye tissue in this assay (16, 17). The *S5A* locus comprises four lethal alleles, two of which are Ras^{-}-like (Fig. 2*d*). A more detailed analysis of the screen and the clonal phenotype of the locus will be described elsewhere.

Low-Resolution Mapping. Mutations in *S5A* associated with the pinhead phenotype were placed 2.3 cM proximal to cytological position 96E by recombination mapping with a marker chromosome (*Materials and Methods*). The lethality associated with the four alleles was not complemented by deficiency *Df(3R)96B*, which reportedly uncovers 96A21 to 96C2. This cytological interval comprises approximately 440 kb of genomic DNA, thus illustrating that resolution achieved by classical techniques can be competitive to whole-genome SNP maps. We chose two markers situated within the candidate region EPy^{+} and *EP3455* (see below) to generate recombinants for fine-mapping.

Establishing an SNP Map in the 96 A–C Region. Assuming that noncoding regions exhibit a higher rate of polymorphism than coding DNA, we designed primer pairs to amplify PCR fragments from intergenic regions or from introns greater than 1 kb (*Materials and Methods*). The fragments chosen for amplification had an approximate size of 800 bp and were spaced at intervals of approximately 17 kb. In a control experiment, we determined by sequencing that 95% of the polymorphisms present in these fragments had been resolved by DHPLC (*Materials and Methods* and Table 1). We then tested whether PCR products derived from control homozygotes and from flies heterozygous for the parental mutagenized chromosome, *FRT82B*, and the marker chromosomes EPy^{+} or *EP3455* (Figs. 1–3) could be discriminated by DHPLC. Of 40 products amplified from $FRT82B/EPy^{+}$ heterozygotes, 27 (67.5%) exhibited an altered chromatogram when compared with homozygotic amplicons. The SNP map established for the $FRT82B/EPy^{+}$ strain combination spanned 594 kb (from 96A14 to 96C2) and had an average resolution of approximately 22 kb (Fig. 1*a*). The other heterozygote (*FRT82B/EP3455*) was not polymorphic enough to permit the construction of a high-resolution map but was nevertheless useful, because a single SNP can provide information about the relative position of marker and mutation (see below).

Mapping of Mutations. Recombinants between *S5A* and the markers EPy^{+} and *EP3455* chosen for fine mapping are rare but are efficiently recovered by the crossing schemes depicted in Figs. 2 and 3. Two recombinant products can be isolated from the cross

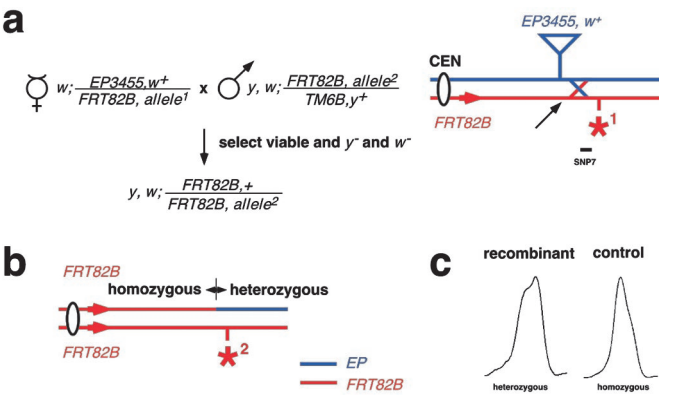

Fig. 3. (*a*) Crossing scheme (related to Fig. 2*b*) and SNP-mapping experiment to establish a left border for the critical region by recombination with a w^{+}-marked EP element. Recombinants will be w^{-} and y^{-}. (*b*) A polymorphism between the parental strains *EP3455* and *FRT82B* could be established in the SNP7 fragment directly distal to the insertion site. (*b*) In a recombinant, SNP7 showed the heterozygotic DHPLC profile (*c*), indicating that the gene is distal to the transposon.

with EPy^{+}, which is located distally to the locus (Figs. 1*a* and 2*a*). One product loses both the mutation and the marker and can be selected on the basis of the fact that it is viable over another allele of the same locus (Fig. 2*b*). The other product carries both the mutation and the marker and can be detected by the small head assay (Fig. 2*c*; see below). Such recombinants can be recovered only when the marker is distal to the gene, thus reducing the critical region to maximally 373 kb (96A21 to 96B20; Fig. 1*a*).

Recombinants of the first crossing scheme were directly assessed because the SNP map had been calibrated for this strain combination (*FRT82B* and EPy^{+}). Two predications can be made from the SNP data obtained from individual recombinants. First, the position of the crossing over can be mapped (Fig. 1*b*). Second, the relative position of the mutation and the marker can be deduced. A recombinant fly will be homozygous (no SNP detected) on one side of the crossing over and heterozygous (SNP detected) on the other, depending on whether the marker is situated distally or proximally to the gene (compare Figs. 2*b* and 3*b* and Fig. 6, which is published as supporting information on the PNAS web site, www.pnas.org). Thirty-two recombinant chromosomes that lost the mutation and the EPy^{+} marker (Fig. 2*a*) were recovered and analyzed by DHPLC (*Materials and Methods*). The most proximal recombination event (i.e., the recombination closest to the gene) was localized between two SNPs at 96B4–B5 (Fig. 1 *a* and *b*), and the critical region was thus reduced by 325 kb. Furthermore, they exhibited a homozygous profile distally and a heterozygous profile proximally, indicating that the EPy^{+} transposon is located distally to the locus (Figs. 1*b* and 6).

The applied crossing scheme depends on the following prerequisites: First, the mutation must be homozygous lethal or exhibit an easily scorable phenotype. Second, at least two alleles are required, because lethal second site mutations could be mapped instead. Therefore, recombinant chromosomes of single or nonlethal alleles should be isolated by the alternative scheme selecting for the pinhead and marker phenotypes. However, this approach poses the problem of introducing the third strain background of a *FRT82B* chromosome carrying a cell lethal *rab1* mutation that permits the expansion of the mutant tissue (Fig. 2*c*). For this third chromosome, the SNP map has not been calibrated. However, because SNPs are codominant, the calibrated sequence difference between *FRT82B* and EPy^{+} must unambiguously discriminate between *rab1/FRT82B* or $rab1/EPy^{+}$ heterozygotes, respectively. Thus, we recalibrated

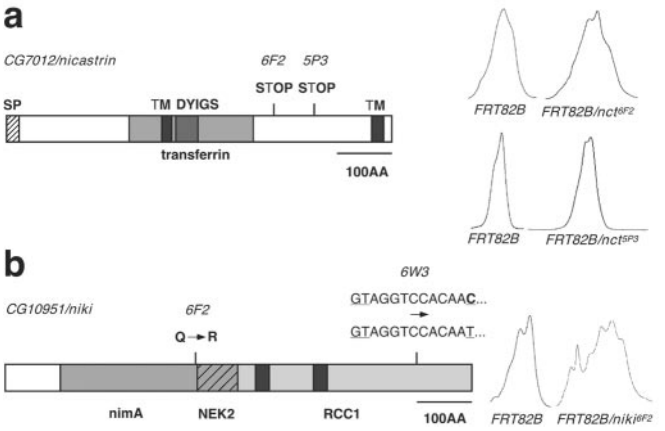

Fig. 4. Mutations of two *nicastrin* alleles and of *CG10951/niki*. (*a*) The domain organization of *nicastrin* and the DHPLC profiles of *FRT82B* control homozygotes and $nct^{6F2}/FRT82$ and $nct^{5P3}/FRT82$ heterozygotes, respectively. SP, signal peptide (hatched); DYIGS, conserved motif (dark gray); transferrin domain in gray. Note the existence of two transmembrane domains (TM, black). STOP, the site of the mutations in the protein sequence. (*b*) The predicted niki protein and the DHPLC profiles of a *FRT82B* control homozygote and a $niki^{6F2}/FRT82B$ heterozygote. One EMS-induced polymorphism causes an amino acid exchange in the kinase domain ($niki^{6F2}$); the other is located in a small intron included in the survey of PCR fragments ($niki^{6W3}$). nimA and NEK2, domains homologous to the mitotic kinase nimA and its mammalian homologue NEK2 (dark gray and hatched). RCC1, C terminus homologous to the Ran guanine exchange factor RCC1 (gray). Black boxes indicate RCC1 domains.

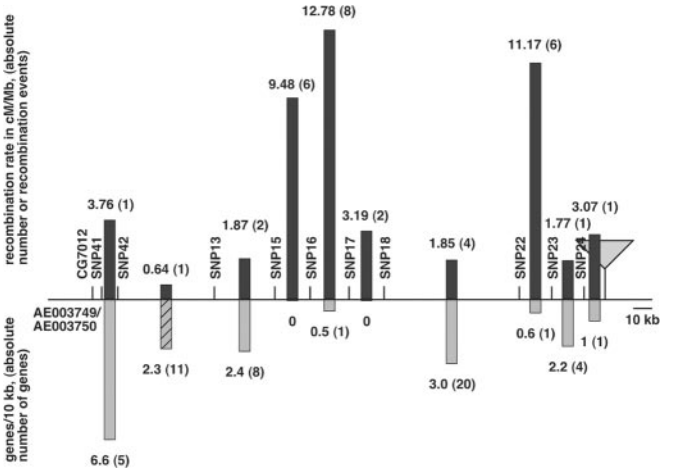

Fig. 5. The recombination rate and the distribution of breakpoints between *CG7012/nct* and the EP element at 96B20 determined for 32 recombinants. The genetic distance between those two markers is 1.0 cM and the physical distance is 0.268 Mb. The map is drawn to scale, the scale bar corresponding to 10 kb. Black bars between two SNPs indicate the ratio of cM per Mb for a given interval. The number of genes per 10 kb is shown by gray bars. Absolute numbers are given in parentheses. One gray bar is hatched because of a gap of unknown length in the reference genome. The regions between SNP15 and SNP18 and SNP22 and SNP23 are meiotic hot spots, which fall into regions of very low gene density.

the local SNP map for the third strain background by assessing the profiles of the two new heterozygous combinations. Now the breakpoints in the recombinants could be directly mapped, because they are characterized by a switch from one heterozygous profile to the other (Fig. 1*c*). The most proximal recombinant recovered from this analysis was located only 4 kb apart from the *S5A* locus (Fig. 1 *a* and *c*).

Although the density of polymorphisms in *FRT82B/EP3455* heterozygotes is low, the relative position of the *EP3455* marker and the mutation could be determined. We found a polymorphism between the two parental strains directly distal to the transposon insertion site (SNP7, Figs. 1*a* and 3). A recombinant between the mutant and *EP3455* exhibited a heterozygous profile, indicating that *EP3455* is situated proximally to the gene (Fig. 3*b* and Fig. 6). Therefore, the insertion point sets the left border of the candidate region, which is thus reduced by another 64 kb and contains only 10 genes in a 45-kb interval (Fig. 1*a*).

Candidate Gene Approach. Three of the 10 genes in the critical region, *CG7012*, *CG13638*, and *CG10951*, were considered primary candidates on the basis of the function of homologs in other species. As EMS, the mutagen used, primarily induces point mutations (18), we crossed the mutant chromosomes back to the parental *FRT82B* strain, amplified the coding regions from these heterozygotes, and searched for sequence changes by DHPLC. *CG7012/nct*, the *Drosophila* homolog of *nicastrin* (19), exhibited a DHPLC-resolvable polymorphism in all four alleles. By subsequent sequence and genetic analysis, it was shown that these polymorphisms are responsible for the phenotype (Fig. 4*a* and data not shown). On two of the four chromosomes, we also found mutations in *CG10951*, coding for a *ni*mA-like *ki*nase (*niki)* with a C terminus homologous to the *RCC1* Ran guanine exchange factor (Fig. 4*b*). The lethality of the *niki* heteroallelic combination could not be rescued by either a *tub-niki* or a *niki* genomic rescue construct, suggesting that the lethality is not associated with the *niki* mutations but is because of the closely linked *nct* mutations on the same chromosome. The *nct* phenotype is unaffected by the second site mutations in *niki* (data not shown).

These data show that chemically induced mutations can be efficiently scanned by DHPLC. Furthermore, they suggest that closely linked second site mutations are more common than previously appreciated.

Mapping the Breakpoints of a Deficiency. During this study, we realized that one of the breakpoints of deficiency *Df(3R)96B* has been mapped inaccurately. To determine the breakpoints at high resolution, we made use of our local high-density SNP map. However, in the unrelated *Df(3R)96B/TM6B* strain background, this map is not applicable. We therefore crossed the deficiency chromosome into the parental backgrounds of EPy^{+} and *FRT82B*, respectively. Because SNPs are codominant, the calibrated sequence difference must appear in either of the two heterozygotes. The hemizygous region uncovered by the deficiency, on the other hand, retains a homozygous profile in both combinations (Fig. 1*d*). In this way, we could place one chromosomal breakpoint between *tok* and *tld* at 96A22 and the other between *CG6031* and *CG13646* or between 96B8 and B11. The right border is shifted toward 96B10, because the deficiency complements a lethal P-element insertion in *CG6238/slingshot* (Fig. 1*d*).

Gene-Poor Regions at 96B Are Recombination Hot Spots. We were interested in whether recombination events are randomly dispersed, because the presence of hot spots would limit the resolution of SNP mapping. The sites of recombination were determined in all 32 recombinants, revealing that 22 events are located in two major intervals. These cover a region of 76 kb, whereas the other events occurred in the remaining 192 kb (Fig. 5). At the given resolution, the difference between meiotically inert and active intervals is up to 20-fold. The skewed distribution of crossing-over events deviates significantly from a Poisson distribution expected for a random distribution of recombination (see *Materials and Methods*, statistical analysis). Plotting the

GENETICS

recombination rate against the number of genes in the intervals (Fig. 5) reveals a negative correlation of recombination events and gene density. Presently, the scarcity of data and the relatively low resolution of the SNP map do not permit the distinction of whether it is the low gene-density *per se* or another feature within that gene-poor region that makes those sites more recombinogenic.

Discussion

The mapping strategy described here combines recombination mapping using visible marker for low-resolution and SNP markers for high resolution to locate a mutation to a genomic region of a few kilobase pairs. This "mixed method" is akin to the principle described by Martin *et al.* (5), who argue that the rate-limiting SNP detection reactions are reduced at least by a factor of four. Aside from the SNP assays, the difference between the two strategies is that Martin *et al.* describe a one-step process resulting in medium resolution, whereas our method is divided into a low- and high-resolution step. The two-step strategy has also been applied by Berger *et al.* (2), who followed an all-molecular sequence-based approach. Our protocol guarantees maximal resolution with a minimal number of recombinants, because choosing closely linked markers provides a recombination event in the region of interest. The SNP map is then calibrated for the same strain combination, so that recombinants can be directly analyzed. This procedure allows saving one generation compared with the strategy of crossing recombinant chromosomes to standard SNP chromosomes (3).

Here we show that "SNP mapping without knowing the SNP" by merely resolving sequence differences between two strains is an exact and very efficient mapping strategy. Three advantages of this method can be put forward. First, calibrating SNPs by DHPLC is very accurate. Ninety-five percent of the SNPs determined by sequencing in a 23.4-kb region were detected by DHPLC. Second, the method is semiautomatic. Third, and most important, once a SNP has been established, there is no need to further develop SNP detection assays. SNP detection by DHPLC has been demonstrated previously (8, 20), and DHPLC has been used for SNP mapping recently (3, 21), but in these cases it was introduced as an assay system to detect an already established sequence-verified SNP.

Screening candidate genes in the critical region by DHPLC [also shown by Cargill *et al.* (22)] for EMS-induced point mutations led to the surprising discovery that two of the four *nct* alleles carry a second site mutation in a closely linked gene, which we termed *niki*. These mutations are not responsible for the growth phenotype. The mutation rate for high doses of EMS (50 mM) has been calculated by Bentley *et al.* (23) to be 1/209 kb. The chromosomes investigated in this study were mutagenized with 33 mM EMS. Therefore *niki*, a gene with a 2-kb coding region, has a less than 1% likelihood of being affected and a less than 1 in 10,000 chance of being hit twice. Pastink *et al.* (18) found two double substitutions among 52 EMS-induced mutations in the vermilion locus. Similarly, with the related mutagen MMS, we induced two point mutations within a single yeast gene (24). It is thus conceivable that a chemically induced mutation favors the occurrence of second site mutations in the vicinity. Such closely linked mutations may pose serious problems when one is trying to map single alleles, because it cannot be taken for granted that even in small candidate regions an identified mutation is causative for the phenotype. Generating recombinants by selecting for the phenotype (Fig. 2*c*) rather than against the associated lethality (Fig. 2*b*) can reduce, but not eliminate, this risk.

A potential limitation of gene mapping by recombination is the nonrandomness of recombination events. In the course of mapping mutations to the *nct* locus, we found that the distribution of crossover sites between *nct* and EPy^{+} in the 32 recombinants was nonrandom. Two recombination hot spots (responsible for 0.44 and 0.19 cM, respectively) were identified. These hot spots fall within regions of low gene density. Comparatively few recombination events were identified in gene-rich regions. Collecting similar high-resolution data throughout the genome could help in addressing several interesting issues: Does the observed negative correlation of gene density and meiotic activity apply as a general rule, or is it valid only for the 96B region? Is the recombination rate within such an interval uniform, or can it be localized to a single site? In other words, is it the low gene density *per se* that renders a region meiotically active, or can the apparent hot spot be attributed to another feature of the chromosome, for example a promoter? In yeast, meiotic hot spots generally coincide with promoters of transcriptionally active genes (25), whereas human hot spots may reside in promoters, inter-, and intragenic regions (26). In the 96B region, the *CG13647* promoter could be the primary site of recombination, because it contains SNP16, which is located in the center of 14 recombination events (Figs. 1*a* and 5). The high incidence of polymorphisms (an average of 1 in 377) and the high recombination rate (0.44 cM between SNP15 and SNP17) should allow the generation of an ultrahigh-density SNP map to test this prediction.

We are grateful to Felix Rintelen for the *rab1* mutation, to Sebastian Breuer and Denise Nellen for the EP line, to Rita Bopp for selecting recombinants, to Barbara Flückiger for assisting the DHPLC, to Olaf Nairz and Roman Kaelin for statistical advice, and to Michael Spörri and Giancarlo Tomio for sequencing. We also acknowledge Peter Gallant and Sean Oldham for critically reading the manuscript. K.N. was supported by fellowships from the Human Frontiers Science Program, European Molecular Biology Organization, and a European Union-sponsored Training and Mobility of Researchers grant (to E.H.).

1. St Johnston, D. (2002) *Nat. Rev. Genet.* **3,** 176–188.
2. Berger, J., Suzuki, T., Senti, K. A., Stubbs, J., Schaffner, G. & Dickson, B. J. (2001) *Nat. Genet.* **29,** 475–481.
3. Hoskins, R. A., Phan, A. C., Naeemuddin, M., Mapa, F. A., Ruddy, D. A., Ryan, J. J., Young, L. M., Wells, T., Kopczynski, C. & Ellis, M. C. (2001) *Genome Res.* **11,** 1100–1113.
4. Syvanen, A. C. (2001) *Nat. Rev. Genet.* **2,** 930–942.
5. Martin, S. G., Dobi, K. C. & Johnston, D. S. (2001) *Genome Biol.* **2,** 1–12.
6. Underhill, P. A., Jin, L., Zemans, R., Oefner, P. J. & Cavalli-Sforza, L. L. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 196–200.
7. Oefner, P. J. (2000) *J. Chromatogr. B Biomed. Sci. Appl.* **739,** 345–355.
8. Wolford, J. K., Blunt, D., Ballecer, C. & Prochazka, M. (2000) *Hum. Genet.* **107,** 483–487.
9. Newsome, T. P., Asling, B. & Dickson, B. J. (2000) *Development (Cambridge, U.K.)* **127,** 851–860.
10. Xu, T. & Rubin, G. M. (1993) *Development (Cambridge, U.K.)* **117,** 1223–1237.
11. Rorth, P., Szabo, K., Bailey, A., Laverty, T., Rehm, J., Rubin, G. M., Weigmann, K., Milan, M., Benes, V., Ansorge, W. & Cohen, S. M. (1998) *Development (Cambridge, U.K.)* **125,** 1049–1057.
12. Smith, D., Wohlgemuth, J., Calvi, B. R., Franklin, I. & Gelbart, W. M. (1993) *Genetics* **135,** 1063–1076.
13. Deak, P., Omar, M. M., Saunders, R. D., Pal, M., Komonyi, O., Szidonya, J., Maroy, P., Zhang, Y., Ashburner, M., Benos, P., *et al.* (1997) *Genetics* **147,** 1697–1722.
14. Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., *et al.* (2000) *Science* **287,** 2185–2195.
15. Spiegelman, J. I., Mindrinos, M. N. & Oefner, P. J. (2000) *Biotechniques* **29,** 1084–1092.
16. Prober, D. A. & Edgar, B. A. (2000) *Cell* **100,** 435–446.
17. Halfar, K., Rommel, C., Stocker, H. & Hafen, E. (2001) *Development (Cambridge, U.K.)* **128,** 1687–1696.
18. Pastink, A., Heemskerk, E., Nivard, M. J., van Vliet, C. J. & Vogel, E. W. (1991) *Mol. Gen. Genet.* **229,** 213–218.
19. Yu, G., Nishimura, M., Arawaka, S., Levitan, D., Zhang, L., Tandon, A., Song, Y. Q., Rogaeva, E., Chen, F., Kawarai, T., *et al.* (2000) *Nature (London)* **407,** 48–54.
20. Giordano, M., Oefner, P. J., Underhill, P. A., Cavalli Sforza, L. L., Tosi, R. & Richiardi, P. M. (1999) *Genomics* **56,** 247–253.
21. Spiegelman, J. I., Mindrinos, M. N., Fankhauser, C., Richards, D., Lutes, J., Chory, J. & Oefner, P. J. (2000) *Plant Cell* **12,** 2485–2498.
22. Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Shaw, N., Lane, C. R., Lim, E. P., Kalyanaraman, N., *et al.* (1999) *Nat. Genet.* **22,** 231–238.
23. Bentley, A., MacLennan, B., Calvo, J. & Dearolf, C. R. (2000) *Genetics* **156,** 1169–1173.
24. Nairz, K. & Klein, F. (1997) *Genes Dev.* **11,** 2272–2290.
25. Nicolas, A. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 87–89.
26. Jeffreys, A. J., Kauppi, L. & Neumann, R. (2001) *Nat. Genet.* **29,** 217–222.