Perspectives

Anecdotal, Historical and Critical Commentaries on Genetics

Edited by James F. Crow and William F. Dove

Molecular Genetics of *Mus musculus*: Point Mutagenesis and Millimorgans

THE world of genetics resembles an intellectual food chain. Approaches and attitudes developed in phage genetics are used in turn to study yeasts, slime molds, nematodes, plants, Drosophila or mammals. Only rarely does this chain circularize, since investigators nowadays most commonly move from the simpler to the more complex organisms. I vividly recall one exception, a talk at a Cold Spring Harbor phage meeting in which the presenter spoke of "sacrificing" phage-infected bacterial cultures for an enzyme analysis. From rat liver to phage, I thought.

This month's column traces the movement of molecular genetics into mouse genetics along a chain in which phage, yeast and Drosophila genetics provide successive nourishment.

Natural variants with heterozygous phenotypes: Much of classical mouse genetics rests on natural variants, often mutations expressing heterozygous phenotypes that beget more severe homozygous phenotypes. The well known alleles *Agouti-yellow* (A^y), *Steel* (*Sl*) and dominant *White-spotting* (*W*) illustrate this proclivity. Haldane (1956) suggested the term "patent lethal" for alleles with dominant visible and recessive lethal phenotypes, usefully avoiding the ambiguous terms "dominant lethal" and "semidominant." The series of recessive *t* haplotypes is an intricate version of patent lethals: each displays a heterozygous visible phenotype, tailless, in repulsion to the *brachyury* mutation (*T*) and, commonly, a homozygous lethal phenotype. Starting with the discovery of *T* 60 years ago (Dobrovolskaïa-Zavadskaïa 1927), L. C. Dunn and his colleagues collected much of the natural variant *t* material defining the *t* complex, culminating in a paper in Genetics 25 years ago this spring in which the complementary interactions and rare recombinations with wild type were summarized (Dunn, Bennett and Beasley 1962). It is now becoming plain that the *t* haplotypes share a set of inversions with one another. The *t* inversions account for the suppression of recombination between *t* haplotypes and wild-type chromatin. The genetic isolation of *t* haplotypes from wild type (and even from one another) has permitted each to accumulate polymorphisms of both the deletion/addition and the base pair substitution variety. Sequencing particular regions of *t* chromatin indicates that substitutions have accumulated to a level of 0.007 per base pair (Figueroa *et al.* 1985; Uehara *et al.* 1987). Thus, a single transcribed 10-kb region may carry about 70 base-pair changes with respect to wild type. Two recent papers in Genetics (Mains 1986; Skow *et al.* 1987) deal with this highly polymorphic aspect of *t* haplotypes.

In summary, the natural variants of the mouse *t* complex accumulated under conditions of blocked recombination and segregation distortion (Lyon 1986) are a rich lode of developmental defects but an embarrasing plethora of changes at molecular resolution. For a geneticist working with a microorganism, or even with Drosophila, the reflex response might well be to induce new mutations on a standard background. In this issue of Genetics (Lee *et al.* 1987), the laboratories of Chovnick and Bender take just this approach to the controlling region of the *rosy* locus in Drosophila.

One powerful strategy employs mutagenesis. For mouse genetics, this approach is being successfully used over the whole genome by Jaenisch, Leder, Jenkins, and Copeland, and their collaborators. Mutations induced by cloned DNA elements give direct access to the affected region of the genome. However, detailed studies of retroviral insertions indicate that such elements exert position effects over at least kilobasepair distances (Breindl, Harbers and Jaenisch 1984; Rinchik *et al.* 1986). These position effects increase the target size for affecting a locus of interest but they confound the search for the DNA sequence that is responsible for the normal function of a gene. More crucial to the logistics of a mouse mutagenesis program, however, is the currently estimated efficiency of insertional mutagenesis, 10^{-4} to 10^{-5} per locus (Dove 1984; Jenkins and Copeland 1985). For a specific locus, it is vastly more difficult to screen

10,000 or more mouse pedigrees than the 1000–2000 required to find a new allele of the locus after chemical mutagenesis by ethylnitrosourea (ENU).

The second experimental paradigm comes directly from Drosophila genetics: to induce recessive mutations at a locus by using recessive visible markers and balancers to follow the pedigree of a mutated chromosome. The historical time scale is informative: MULLER and ALTENBURG published the first Drosophila experiments using recessive visible markers in 1919, SNELL carried out the first search for recessive lethals in the mouse in 1935, HALDANE (1956) worked out breeding systems for optimal scanning of the whole genome and CARTER (1957, 1959) used such systems in pilot experiments, detecting one X-ray-induced recessive lethal mutation. The efficiency of mutagenesis was clearly the limiting parameter, particularly for a mouse genetics facility of ordinary size that might screen only 300 mutagenized gametes per year.

ENU and 2000 mutagenized gametes: The nitrosamides have been championed by the Chester Beatty carcinogenesis workers LOVELESS, BROOKES and LAWLEY as penetrating alkylating agents needing no metabolic activation and more predisposed than is EMS to *O*-alkylation of guanine and thymine, generating base pair substitutions (LOVELESS 1969). In microorganisms, nitrosamides produce a favorable mutant frequency per unit of cell killing. Though these compounds are highly carcinogenic and mutagenic, they are so rapidly inactivated under mild conditions that safe handling requires only short-term chemical hazard containment.

Consistent with chemical expectations, mutations induced by ENU in the controlling region of the Drosophila *rosy* locus are predominantly G•C→A•T transitions (see LEE *et al.* in this issue). In contrast to the clustered mutations often found after nitrosoguanidine mutagenesis, only one of these *rosy* mutations was complex, an A•T→C•G transversion plus a nearby A•T deletion. RUSSELL and his colleagues (1979), working in the large-scale mutagen testing program at the Oak Ridge National Laboratory, established conditions in which ENU is a highly efficient germline mutagen for the mouse, particularly to spermatogonia. At visible loci, forward mutant frequencies average 0.0007 per locus. These frequencies make a crucial difference for a mouse laboratory that can handle only a few hundred parallel pedigrees emerging from mutagenized gametes.

BODE (1984) and JUSTICE and BODE (1986) have initiated a mutagenic attack on the *t* region by recovering induced alleles of its known visible loci, *tct*, *qk* and *tf*. As in the specific locus test, this involves only single-generation screens that detect recessive mutant phenotypes after mating a mutagenized wild-type male to a homozygous mutant female. Our Wisconsin laboratory and that of JEAN-LOUIS GUÉNET at the Institut Pasteur have collaborated on the efficient use of MULLER/HALDANE breeding protocols to recover numerous recessive lethal mutations in the *t* complex (SHEDLOVSKY *et al.* 1986). (One crucial component of this scheme is the fecundity of L. C. DUNN's mouse strain BTBR). This constitutes the first attempt to saturate a particular region of the mouse genome with induced mutant alleles and reflects a spirit spreading in mouse genetics: it is now feasible to induce mutations in a system of interest rather than merely to depend on natural variants.

The number of lethal complementation groups identified by ENU-induced mutations in the *t* region seems to be large, perhaps 50–100 (SHEDLOVSKY *et al.* 1986). These mutations must first be characterized by mapping rather than by complementation, because the number of mapping experiments increases linearly with the number of isolates rather than quadratically as do complementation tests. An increasing number of DNA markers (HERRMANN *et al.* 1986) and cDNA clones (WILLISON, DUDLEY and POTTER 1986) are becoming available for such mapping, with no limit in sight (POUSTKA and LEHRACH 1986). The question then becomes, what level of map resolution is necessary to assign point mutations to cloned segments?

Gene density in the mouse genome *vs.* the cosmid: Extrapolating the frequency of *t*-region recessive lethals induced by ENU suggests that there are 5,000–10,000 single-copy vital genes in the mouse genome (SHEDLOVSKY *et al.* 1986). [Previous measurements of lethals, with less efficient mutagenesis, had given estimates ranging from 1,000 (LYON, PHILLIPS and SEARLE 1964; RODERICK 1983; SHERIDAN 1983) to 10,000 (LYON 1956).] One imponderable in the ENU estimates is whether the average target size for generating a lethal allele is the same as that for producing a visible allele. A factor complicating interpretation of the X-ray induction of recessive lethals is the generally unknown size of deletions induced by X-rays. If the average deletion is large compared with the spacing between vital genes, then each induced recessive lethal mutation hits more than one vital locus and the number of vital loci will be underestimated (CHAKRABARTI *et al.* 1983).

The direct estimate of only 50–100 single-copy vital genes in the *t* region would correspond to a spacing no closer than 250 kb between such genes. The unique assignment of point mutations to loci would then require mapping to a resolution of 0.1–0.2 cM. There is reason to believe, however, that the frequency of transcribed regions in the mammalian genome is threefold to tenfold higher than this estimate of single-copy vital genes. Indeed, the number of known genes in the *t* region already approaches 100. OHNO (1986)

has estimated the average spacing of transcribed regions to be 35 kb. BIRD (1986) and BROWN and BIRD (1986) have found that unmethylated CpG clusters (*Hpa*II tiny fragments of "HTF islands") occur on average every 90 kb; however, HTF islands may only mark genes that are expressed constitutively, thus underestimating the total frequency of transcribed regions. Detailed analysis of the *H-2K* region of the *t* complex shows nine class I and class II genes in 600 kb of DNA (STEINMETZ, STEPHAN and LINDAHL 1986). If we take 50 kb as the spacing of active genes—that is, the size of a DNA cosmid clone—then mapping of point mutations to segments of cosmid size would require resolution to about 0.02 cM. The analysis of the *agouti* locus reported in the April issue (LOVETT *et al.* 1987) well illustrates this gap in mapping resolution: markers generated from a linked retroviral insertion may lie at least 1 cM (50 cosmid lengths) from the DNA segment of interest.

In summary, molecular genetic analysis in the mouse currently straddles a challenging chasm. On one side, the direct advantage of insertional mutagenesis is compromised by the unknown distances over which position effects are exerted and by low insertion frequencies per locus compared with the number of gametes that can realistically be screened in pedigrees. On the other side, the high efficiency and point character of nitrosamide mutagenesis must be coupled with heroic mapping efforts to achieve molecular assignments for any of the biologically interesting mutations that now regularly turn up in ENU mutagenesis programs. The intrinsic biological interest of the newly induced mutations must drive our efforts to bridge this chasm.

Number of vital genes *vs.* the number of active genes: I close by wondering why the estimated frequency of vital loci in the mammalian genome is lower, by up to an order of magnitude, than the frequency of transcribed regions or HTF islands. A prime possibility is the one recently presented for the streamlined genome of *Saccharomyces cerevisiae* by GOEBL and PETES (1986). Gene disruption of putative single-copy sequences leads to a detectable phenotype only 30% of the time, though at least 50% of the yeast genome is transcribed and introns are rare (KABACK, ANGERER and DAVIDSON 1979). The putative single-copy sequences may actually belong to functionally redundant sets of genes (multigene families in which sufficient neutral drift has occurred to eliminate hybridization under the stringencies used) or to genes unrelated in DNA sequence but convergent in function. Analogous possibilities for Drosophila and for the mouse are discussed by GOEBL and PETES. It seems possible that, for each of these genetically tamed organisms, only a minority of the active loci are identified by recessive mutations. The rest of the iceberg is invisible to us. Perhaps some useful variations can be played on the theme of the patent lethal from which mouse genetics began [see SUZUKI and PROCUNIER (1969) and KENNISON and RUSSELL in this issue].

I thank VERN BODE for cheerful company on the long walk from lambda to mouse genetics and MARY LYON for constructive suggestions.

WILLIAM F. DOVE
McArdle Laboratory
University of Wisconsin
Madison, Wisconsin 53706

LITERATURE CITED

BIRD, A. P., 1986 CpG-rich islands and the function of DNA methylation. Nature **321:** 209–213.

BODE, V. C., 1984 Ethylnitrosourea mutagenesis and the isolation of mutant alleles for specific genes located in the *t* region of mouse chromosome *17*. Genetics **108:** 457–470.

BREINDL, M., K. HARBERS and R. JAENISCH, 1984 Retrovirus-induced lethal mutation in collagen 1 gene of mice is associated with an altered chromatin structure. Cell **38:** 9–16.

BROWN, W. R. A. and A. P. BIRD, 1986 Long-range restriction site mapping of mammalian genomic DNA. Nature **322:** 477–481.

CARTER, T. C., 1957 Recessive lethal mutation induced in the mouse by chronic γ-irradiation. Proc. R. Soc. B **147:** 402–411.

CARTER, T. C., 1959 A pilot experiment with mice, using Haldane's method for detecting induced autosomal recessive lethal genes. J. Genet. **56:** 353–362.

CHAKRABARTI, S., G. STREISINGER, F. SINGER and C. WALKER, 1983 Frequency of γ-ray induced specific locus and recessive lethal mutations in mature germ cells of the zebrafish, *Brachydanio rerio*. Genetics **103:** 109–123.

DOBROVOLSKAÏA-ZAVADSKAÏA, N., 1927 Sur la mortification spontanée de la queue chez la souris nouveau-née et sur l'existence d'un caractère (facteur) héréditaire "non viable." C. R. Soc. Biol. **97:** 114–116.

DOVE, W. F., 1984 Developmental molecular genetics of the mouse and its embryonal carcinoma. Cell Differ. **15:** 205–213.

DUNN, L. C., D. BENNETT and A. B. BEASLEY, 1962 Mutation and recombination in the vicinity of a complex gene. Genetics **47:** 285–303.

FIGUEROA, F., M. GOLUBIĆ, D. NIŽETIĆ and J. KLEIN, 1985 Evolution of mouse major histocompatibility complex genes borne by *t* chromosomes. Proc. Natl. Acad. Sci. USA **82:** 2819–2823.

GOEBL, M. G. and T. D. PETES, 1986 Most of the yeast genomic sequences are not essential for cell growth and division. Cell **46:** 983–992.

HALDANE, J. B. S., 1956 The detection of autosomal lethals in mice induced by mutagenic agents. J. Genet. **54:** 327–342.

HERRMANN, B., M. BUĆAN, P. E. MAINS, A.-M. FRISCHAUF, L. M. SILVER and H. LEHRACH, 1986 Genetic analysis of the proximal portion of the mouse *t* complex: evidence for a second inversion within *t* haplotypes. Cell **44:** 469–476.

JENKINS, N. A. and N. G. COPELAND, 1985 High frequency germline acquisition of ecotropic MuLV proviruses in SWR/J-RF/J hybrid mice. Cell **43:** 811–819.

JUSTICE, M. J., and V. C. BODE, 1986 Induction of new mutations in a mouse *t*-haplotype using ethylnitrosourea mutagenesis. Genet. Res. **47:** 187–192.

KABACK, D. B., L. M. ANGERER and N. DAVIDSON, 1979 Improved

methods for the formation and stabilization of R-loops. Nucleic Acids Res. **6:** 2499–2517.

KENNISON, J. A. and M. A. RUSSELL, 1987 Dosage-dependent modifiers of homoeotic mutations in *Drosophila melanogaster*. Genetics **116:** 75–86.

LEE, C. S., D. CURTIS, M. MCCARRON, C. LOVE, M. GRAY, W. BENDER and A. CHOVNICK, 1987 Mutations affecting expression of the *rosy* locus in *Drosophila melanogaster*. Genetics **116:** 55–66.

LOVELESS, A., 1969 Possible relevance of *O*-6 alkylation of deoxyguanosine to the mutagenicity and carcinogenicity of nitrosamines and nitrosamides. Nature **223:** 206–207.

LOVETT, M., Z.-Y. CHENG, E. M. LAMELA, T. YOKOI and C. J. EPSTEIN, 1987 Molecular markers for the *agouti* coat color locus of the mouse. Genetics **115:** 747–754.

LYON, M. F., 1956 Some evidence concerning the "mutational load" in inbred strains of mice. Heredity **13:** 341–352.

LYON, M. F., 1986 Male sterility of the mouse *t*-complex is due to homozygosity of the distorter genes. Cell **44:** 357–363.

LYON, M. F., R. J. S. PHILLIPS and A. G. SEARLE, 1964 The overall rates of dominant and recessive lethal and visible mutation induced by spermatogonial x-irradiation of mice. Genet. Res. **5:** 448–467.

MAINS, P. E., 1986 The *cis-trans* test shows no evidence for a functional relationship between two mouse *t* complex lethal mutations: implications for the evolution of *t* haplotypes. Genetics **114:** 1225–1237.

MULLER, H. J. and E. ALTENBURG, 1919 The rate of change of hereditary factors in Drosophila. Proc. Soc. Exp. Biol. Med. **17:** 10–14.

OHNO, S., 1986 The total number of genes in the mammalian genome. Trends Genet. **2:** 8.

POUSTKA, A. and H. LEHRACH, 1986 Jumping libraries and linking libraries: the next generation of molecular tools in mammalian genetics. Trends Genet. **2:** 174–179.

RINCHIK, E. M., L. B. RUSSELL, N. G. COPELAND and N. A. Jenkins, 1986 Molecular genetic analysis of the *dilute-short ear* (*d-se*) region of the mouse. Genetics **112:** 321–342.

RODERICK, T. H., 1983 Using inversions to detect and study recessive lethals and detrimentals in mice. pp. 135–137. In: *Utilization of Mammalian Specific Locus Studies in Hazard Evaluation and Estimation of Genetic Risk*, Edited by F. J. DE SERRES and W. SHERIDAN. Plenum, New York.

RUSSELL, W. L., E. M. KELLY, P. R. HUNSICKER, J. W. BANGHAM, S. C. MADDUX and E. L. PHIPPS, 1979 Specific-locus test shows ethylnitrosourea to be the most potent mutagen in the mouse. Proc. Natl. Acad. Sci. USA **76:** 5818–5819.

SHEDLOVSKY, A., J.-L. GUÉNET, L. L. JOHNSON and W. F. DOVE, 1986 Induction of recessive lethal mutations in the *T/t-H-2* region of the mouse genome by a point mutagen. Genet. Res. **47:** 135–142.

SHERIDAN, W., 1983 The detection of induced recessive lethal mutations in mice. pp. 125–134. In: *Utilization of Mammalian Specific Locus Studies in Hazard Evaluation and Estimation of Genetic Risk*, Edited by F. J. DE SERRES and W. SHERIDAN. Plenum, New York.

SKOW, L. C., J. N. NADEAU, J. C. AHN, H.-S. SHIN, K. ARTZT and D. BENNETT, 1987 Polymorphism and linkage of the αA-crystallin gene in *t*-haplotypes of the mouse. Genetics **116:** 107–111.

SNELL, G. D., 1935 The induction by X-rays of hereditary changes in mice. Genetics **20:** 545–567.

STEINMETZ, M., D. STEPHAN and K. F. LINDAHL, 1986 Gene organization and recombinational hotspots in the murine major histocompatibility complex. Cell **44:** 895–904.

SUZUKI, D. T. and D. PROCUNIER, 1969 Temperature-sensitive mutations in *Drosophila melanogaster*. III. Dominant lethals and semilethals on chromosome 2. Proc. Natl. Acad. Sci. USA **62:** 369–376.

UEHARA, H., K. ABE, C.-H. T. PARK, H.-S. SHIN, D. BENNETT and K. ARTZT, 1987 The molecular organization of the *H-2K* region of two *t*-haplotypes: implications for the evolution of genetic diversity. EMBO J. **6:** 83–90.

WILLISON, K. R., K. DUDLEY and J. POTTER, 1986 Molecular cloning and sequence analysis of a haploid expressed gene encoding *t* complex polypeptide 1. Cell **44:** 727–738.