

Perspectives

Anecdotal, Historical and Critical Commentaries on Genetics

Edited by James F. Crow and William F. Dove

Haldane, Bailey, Taylor and Recombinant-Inbred Lines

James F. Crow¹

Laboratory of Genetics, University of Wisconsin, Madison, Wisconsin 53706

THAT mouse Mecca, the Jackson Laboratory, has repeatedly pioneered in bringing mouse genetics to its present state. There was George Snell's Nobel-Prize-winning work on histocompatibility, Roy Stevens' work on embryonal carcinoma, Tibby Russell's on hematopoiesis, and many others (reviewed by PAIGEN 2003a,b). It has also been the site of several important methodological innovations. First and most important, C. C. Little had the foresight to establish inbred lines (CROW 2002). His first line was started in 1909; by 1980, there were >300 (PAIGEN 2003a). Another innovation was the development of congenic strains—inbred lines with a small foreign chromosomal region introgressed by repeated backcrossing into the line (SNELL 1948). A third was chromosome substitution (consomic) strains. These have a single chromosome introgressed into an inbred line (SINGER *et al.* 2004). The fourth innovation, in many ways the cleverest, was recombinant inbred (RI) lines. These innovations each required many years of advance work before they could be utilized effectively. Such projects certainly would not fare well as grant applications today. Only in an organization with a long-time commitment, such as the Jackson Laboratory, could such projects be carried through.

The idea of RI lines arose sometime in the 1950s or 1960s in the fertile mind of Donald Bailey. Don is a quiet, low-key scientist who has not made a big splash in the genetic world at large. But within the Jackson Laboratory and with others who know his work, he has long been revered. He is knowledgeable and creative—the person to go to for help with a technical problem or to search for a new idea.

The principle of RI lines is simple (BAILEY 1971). In retrospect it has a “why didn't I think of it” quality: Two inbred lines are crossed and the hybrids are intercrossed to produce F₂ progeny. Pairs of the F₂ mice are then

mated to establish inbred lines through repeated sib-mating. The genomes of each of these lines are a homozygous mosaic of chromosomal regions from the two founding inbreds. These RI lines are then typed for the genotypes and phenotypes that differed between the two founders.

Sets of RI lines have a number of advantages. Since each RI line is nearly homozygous, its genotype is reproducible and individual genetic variation is minimized. Replications average out the effects of environmental influences and measurement errors. Furthermore, once a line has been genotyped, this information can be used over and over. Unlinked loci largely randomize during the process, even though inbred lines can show “linkage disequilibrium” for loci on different chromosomes (GRABER *et al.* 2006), but linked genes retain some of the linkage disequilibrium that characterized the two founding inbred strains. Furthermore, there are several meioses in the F₂ and during the inbreeding stage, with the result that the amount of recombination is increased fourfold; this is now called map expansion and is very advantageous for mapping closely linked loci. For the history of linkage studies in the mouse, see LYON (1990).

Bailey started with 12 RI lines from a cross of BALB/cBy and C57BL/6By (designated C × B6). Of these, 7 survived for 30 generations of sib-mating. Bailey identified 11 loci and classified them as to the strain of origin. Three were coat-color genes and 8 were histocompatibility factors. The power of the method was shown by the immediate discovery that some phenotypically similar histocompatibility factors mapped to different locations. Despite the small number of RI lines, Bailey and his associates were able to discover some 20 linkages in the next 5 years (TAYLOR 1978).

The next person to enter the RI story was Ben Taylor. Ben joined the Jackson Laboratory in 1969 and immediately started generating RI lines and developing the theory. He, like Don, is soft spoken and reticent, with a manner that belies his sharp mind.

¹Author e-mail: jfcrow@wisc.edu



Donald W. Bailey (courtesy of the Jackson Laboratory).

Some of the background mathematics had been done by HALDANE and WADDINGTON (1931), who worked out the detailed consequences of repeated brother–sister mating. This involved the kind of extensive algebraic manipulations that most people hate, but which Haldane loved. When I read the article I was overwhelmed. For two linked autosomal loci, this involved no less than 22 linear recurrence equations, which Haldane and Waddington were able to solve for the equilibrium values. Remember that this work, published in 1931, involved hand calculations. This was long before the development of high-speed computers.

HALDANE and WADDINGTON (1931) includes only four references, all from 1921 and earlier, and the summary reads: “Formulae are given for the amount of crossing over which is found in the final population when organisms heterozygous for linked genes are inbred according to various systems” (p. 374). Readers who were seduced by the innocent-sounding title, “Inbreeding and Linkage,” and the three-line summary did not realize what an algebraic morass they were getting into. It is interesting and perhaps significant that Waddington later left transmission genetics and studied development. He frequently expressed the opinion that mathematical population genetics was a fruitless endeavor. I wonder if experience with the exhausting and tedious algebra in this article sensitized him forever against any further work in this field.

Ben Taylor was familiar with the Haldane and Waddington article, having studied it as a graduate student at the University of Wisconsin in connection with a research project on radiation effects in rats. Following Haldane and Waddington, he pointed out that the map expansion was a factor of 4 for sib-mating, 3 for an X-linked locus, and 2 for selfing. He also found a remarkable result. Despite very strong positive interference in meiosis, the number of exchanges in RI lines agreed closely with the Poisson distribution; in other words, there was no measurable interference. This can be understood by noting that in RI lines there are several

meioses in succession. Following an exchange, if a different exchange occurs in the same chromosome in a later meiosis, in effect this is a double crossover. But the two exchanges are largely independent of each other and produce an interferenceless double exchange. This means that the simple Haldane mapping function is appropriate for RI lines. Later we shall see that the situation is more complicated.

Taylor made use of another Haldane idea. HALDANE (1956) undertook to develop a method for measuring the recessive lethal mutation rate in mice following irradiation. The idea was to discover a lethal linked to a known recessive marker, detected by the absence of the marker phenotype when the mating system rendered the linked lethal homozygous. As the probability of remaining linked to the marker decays exponentially with distance, Haldane asked for an equivalent region with the probability of detecting the lethal and called this the “swept distance.” He hoped that this would be roughly comparable to the powerful *Drosophila* methods, such as Muller’s CIB. (See the APPENDIX for an account of a curious Haldane mistake.) Taylor calculated the swept distance on either side of a marker in RI crosses. He found that with seven RI strains the swept distance within which no exchanges occur is 9.3 cM compared to 23.9 cM in a corresponding backcross. Actually Haldane’s idea of a swept distance has found only limited usage for mutation studies in the mouse (CARTER and FALCONER 1951). But similar schemes for finding recessive lethals in a specified chromosomal region, such as using appropriately spaced markers and taking advantage of the near-complete interference for short distances, have been fruitfully applied (*e.g.*, SHEDLOVSKY *et al.* 1988).

Taylor developed a number of RI sets, which were used for a variety of molecular traits. One of the earliest uses was identification of genes affecting the group-specific antigen of the murine leukemia virus (TAYLOR *et al.* 1971).

In the ensuing years, 20 or more sets of RI lines were developed in the mouse and hundreds of markers were mapped. Of course, any known DNA sequence can now be mapped by reference to the mouse genome sequence. But this technique is not applicable to many phenotypes. RI lines are particularly useful for genes causing phenotypes whose molecular basis is not known, including components of quantitative traits. Some sets of RI lines have been cryopreserved.

The techniques have spread to other species. RI lines have been useful for studying insecticide resistance (COCHRANE *et al.* 1998). Of course *Drosophila melanogaster* is not to be left out. RI lines have been especially helpful in identifying quantitative trait loci (GIBSON and MACKAY 2002). As the fields of genetics and genomics have moved from gene identification to gene expression, so have applications of RI lines. For example, they have been used to study gene expression in brain tissue

in mice, hematopoietic tissues also in mice, and in fat and kidney tissues in the rat (summarized by BROMAN 2005b).

Plants have a number of advantages for RI analysis. Two of the most important are the ready availability of large numbers and the possibility of self-fertilization, which greatly shortens the necessary time of inbreeding. For example, to lower heterozygosity to 0.016 of its original amount requires only 6 generations of selfing but 20 generations of sib-mating. RI lines have been developed for rice, sunflowers, soybeans, tomatoes, wheat, maize, Brassica, and undoubtedly others. Not surprisingly, RI lines have been extensively used in *Arabidopsis*. By 1993 a set of 100 RI lines involving 64 RFLP probes at ~ 20 -cM intervals were being used to map 500 loci (LISTER and DEAN 1993). Finally, one maize group has created a set of 5000 RI lines by crossing 25 diverse maize inbreds to a single common inbred and deriving 200 RI lines from each cross. The set is being genotyped at 1500 marker loci and the 26 parents are all being sequenced (http://www.panzea.org/info/RIL_phenotyping_press_rel.html).

Early in the game, two modifications of RI lines were suggested. One was to increase the number of foundation inbred lines from two to four or eight; for the mating diagram, see TEUSCHER and BROMAN (2007). A drawback of having only two lines is that the analysis is restricted to genes in the two parent inbreds; to some extent this defect is repaired by having a number of sets of RI lines. This does not compare alleles in different sets, but by expanding the RI set to more lines this difficulty is partially overcome. This also increases greatly the opportunity for study of epistatic interactions. A complication is that epistasis in RI lines may differ from that in crosses of the parental strains. The second suggestion is, in the interest of further map expansion, to interpose one or more generations of random mating before inbreeding starts. Yet the effect is slight. With eight foundation lines and sib-mating, the map expansion increases only from 7.0- to 7.5-fold (TEUSCHER and BROMAN 2007).

The COMPLEX TRAIT CONSORTIUM (2004) recommended developing a set of eight-way RI lines for mice and this has been started. It would require 23 generations of sib-mating to achieve a 99% reduction of heterozygosity. This is a heroic undertaking and would involve some 1000 strains, each needing to be typed. If accomplished, this could provide a valuable tool for mouse research, especially for difficult phenotypes or quantitative traits. It would also permit study of many two- and three-way interactions. A limitation of all RI lines is that they do not give any information on heterozygotes without additional crosses. But intercrosses between RI lines are also proposed by the Consortium. Whether this program will be accomplished and, if so, whether it will live up to its great expectations will be decided in the future.

Meanwhile, the theoretical work has gone on apace. BROMAN (2005a) extended the Haldane and Waddington method to four and eight lines, for both sib-mating and selfing, giving expressions for map expansion, interference, and clustering of breakpoints. With a three-point cross of tightly linked loci, there is actually negative interference. (This was actually foreshadowed by Haldane and Waddington.) The coincidence for tightly linked loci in two-way RI lines, where the effect is most pronounced, is ~ 1.75 .

The negative interference, at first glance, is surprising. Yet it has a ready explanation. It is comparable to the negative interference found in bacteria and phage crosses (ROTHFELS 1952; VISCONTI and DELBRÜCK 1953). The number of single and double crossovers occurs randomly and, if divided by the number of individuals in which these occur, would show a coincidence of one. But there are a number of individuals in the RI line that have become homozygous for this region and exchanges are irrelevant. When these are included in the denominator, there is a seemingly high coincidence.

More recently, TEUSCHER and BROMAN (2007) have discovered a remarkable simplification, a real *tour de force*. They were able to reduce Haldane and Waddington's 22 equations to a much smaller set that is readily solved. They also solved three-point haplotype probabilities for four- and eightfold RI lines, which previously had been done numerically by large computer runs. The theory is in good shape; the task ahead—to put the theory to good use—is harder.

For several years Don Bailey has led a quiet life in retirement. More recently, Ben Taylor has also retired. I hope and trust that both of them are finding satisfaction in the recent progress and great popularity of the methods that they pioneered a quarter century ago.

I am thankful to Alan Attie, Karl Broman, John Doebley, Christina Kendzierski, and Millard Susman, who read the manuscript and offered suggestions that have greatly improved it.

LITERATURE CITED

- BAILEY, D. W., 1971 Recombinant-inbred strains: an aid to finding identity, linkage, and function of histocompatibility and other genes. *Transplantation* **11**: 325–327.
- BROMAN, K. W., 2005a The genomes of recombinant inbred lines. *Genetics* **169**: 1133–1146.
- BROMAN, K. W., 2005b Mapping expression in randomized rodent genomes. *Nat. Genet.* **37**: 209–210.
- CARTER, T. C., and D. S. FALCONER, 1951 Stocks for detecting linkage in the mouse and the theory of their design. *J. Genet.* **50**: 307–323.
- COCHRANE, B. J., M. WINDELSPECHT, S. BRANDON, M. MORROW and L. DRYDEN, 1998 Use of recombinant inbred lines for the investigation of insecticide resistance and cross resistance in *Drosophila simulans*. *Pesticide Biochem. Physiol.* **61**: 95–114.
- COMPLEX TRAIT CONSORTIUM, 2004 The collaborative cross, a community resource for the genetic analysis of complex traits. *Nat. Genet.* **36**: 1133–1137.
- CROW, J. F., 1989 Concern for environmental mutagens: some personal reminiscences. *Environ. Mol. Mutagen.* **14** S16: 7–10.

- CROW, J. F., 1995 Quarreling geneticists and a diplomat. *Genetics* **140**: 421–426.
- CROW, J. F., 2002 C. C. Little, cancer and inbred mice. *Genetics* **151**: 1357–1361.
- GIBSON, G., and T. MACKAY, 2002 Enabling population and quantitative genomics. *Genet. Res.* **80**: 1–6.
- GRABER, J. H., G. A. CHURCHILL, K. J. DIPETRILLO, B. L. KING, P. M. PETKOV *et al.*, 2006 Patterns and mechanisms of genome organization in the mouse. *J. Exp. Zool. A Comp. Exp. Biol.* **305**: 683–688.
- HALDANE, J. B. S., 1956 The detection of autosomal lethals in mice induced by mutagenic agents. *J. Genet.* **54**: 327–342.
- HALDANE, J. B. S., and C. H. WADDINGTON, 1931 Inbreeding and linkage. *Genetics* **16**: 357–374.
- LISTER, C., and C. DEAN, 1993 Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana*. *Plant J.* **4**: 745–750.
- LYON, M. F., 1990 L. C. Dunn and mouse genetic mapping. *Genetics* **125**: 231–236.
- PAIGEN, K., 2003a One hundred years of mouse genetics: an intellectual history. I. The classical period (1902–1980). *Genetics* **163**: 1–7.
- PAIGEN, K., 2003b One hundred years of mouse genetics: an intellectual history. II. The molecular revolution (1981–2002). *Genetics* **163**: 1227–1235.
- ROTHFELS, K. H., 1952 Gene linearity and negative interference in crosses of *Escherichia coli*. *Genetics* **37**: 297–311.
- RUSSELL, W. F., 1989 Reminiscences of a mouse specific-locus test addict. *Environ. Mol. Mutagen.* **14** S16: 16–22.
- SHEDLOVSKY, A., T. R. KING and W. F. DOVE, 1988 Saturation germ line mutagenesis of the murine *t* region including a lethal allele at the quaking locus. *Proc. Natl. Acad. Sci. USA* **85**: 180–184.
- SINGER, J. B., A. E. HILL, L. C. BURRAGE, K. R. OLSZENS, J. SONG *et al.*, 2004 Genetic dissection of complex traits with chromosome substitution strains of mice. *Science* **304**: 445–448.
- SNELL, G. D., 1948 Methods for the study of histocompatibility genes. *J. Genet.* **49**: 87–108.
- TAYLOR, B. A., 1978 Recombinant inbred strains: use in gene mapping, pp. 423–438 in *Origins of Inbred Mice: Proceedings of a Workshop, Bethesda, Maryland, February 14–16, 1978*, edited by H. C. MORSE. Academic Press, New York.
- TAYLOR, B. A., H. MEIER and D. D. MYERS, 1971 Host-gene control of C-type RNA tumor virus: inheritance of the group-specific antigen of murine leukemia virus. *Proc. Natl. Acad. Sci. USA* **68**: 3190–3194.
- TEUSCHER, F., and K. W. BROMAN, 2007 Haplotype probabilities for multiple-strain recombinant inbred lines. *Genetics* **175**: 1267–1274.
- VISCONTI, N., and M. DELBRÜCK, 1953 The mechanism of genetic recombination in phage. *Genetics* **38**: 5–33.

APPENDIX: HALDANE'S ERROR

For many years, William Russell, at the Oak Ridge National Laboratory, studied radiation-induced mutations occurring at specific mouse loci. This mega-mouse

project involved enormous numbers and great expense, and it yielded few mutations, although it eventually led to important findings, both quantitative and qualitative. In the mid-1950s, Haldane wrote an article for the *Bulletin of the Atomic Scientists* stating that this approach was all wrong; what Russell should be doing, Haldane wrote, is measuring *all* the mutations in a chromosome or chromosome segment, as was done with *Drosophila*. Haldane, following his standard practice of insulting people, went on to write that undoubtedly geneticists and statisticians at Oak Ridge were kept in separate cages or they would have followed the *Drosophila* pattern. He suggested a mating scheme by which recessive lethals linked to a marker gene could be identified. He wrote that, from this mating system, he could define a chromosomal region around the marker that was “swept” in that all the lethals in that region could be identified. Haldane also wrote that space did not permit his deriving the method, but “Sewall Wright could do it in 20 minutes.” The article made the rounds at a meeting of the Biological Effects of Atomic Radiations (BEAR) committee (CROW 1995). Needless to say, Bill Russell was incensed (RUSSELL 1989). The article was not published.

Wright eagerly accepted the challenge. He and I attended the BEAR meeting and occupied adjacent berths on a sleeper train back to Wisconsin from the New York meeting. We both worked on the problem. Wright solved it first, but it took much longer than 20 min. I got an answer later by a different method. Wright and I agreed, but differed from Haldane. It turned out that Haldane had had a rare mental lapse. He had made a rather elementary conceptual error (CROW 1989).

Wright wrote Haldane, giving the correct solution. Haldane sent back a casual acknowledgment; it was clear that he had not read Wright's letter. The reason was obvious. As several Wisconsin secretaries can attest, Wright's handwriting was atrocious. Aware of this, he sent a second letter, this time typewritten, whereupon Haldane answered appreciatively and gave the correct formula in his published paper (HALDANE 1956). This is the article that Ben Taylor studied to determine swept distance for markers in RI lines.