Motoo Kimura and James Crow on the Infinitely Many Alleles Model

Warren J. Ewens¹

Department of Biology, The University of Pennsylvania, Philadelphia, Pennsylvania 19104

ORIGINAL CITATION

The Number of Alleles That Can Be Maintained in a Finite Population

Motoo Kimura and James F. Crow

GENETICS April 1964 **49**: 725–738

Kimura and Crow's 1964 article is justly regarded as one of the foundational articles of evolutionary molecular genetics. It is based on what they called the “infinite alleles” model (now called the “infinitely many alleles” model). This model was motivated by the recognition that a gene is a sequence of perhaps several thousand nucleotides, implying that an astronomically large number of alleles (equal to nucleotide sequences) is possible at any locus. In the later sections of their article Kimura and Crow discuss aspects of this model when selection exists. However, the theory for this case is still unresolved and here only the first section, which in any event has been by far the most influential section, is discussed. In that section only selectively neutral alleles (“isoalleles”) were considered.

Properties of the selectively neutral infinitely many alleles model were introduced by Malécot (1948). The essentially infinite number of possible alleles at a locus leads, in this model, to the assumption that a mutation creates an allele of an entirely novel type. Malécot's analysis assumed a simple Wright–Fisher evolutionary model (Fisher 1922; Wright 1931) in which the genes in any offspring generation are assumed to be chosen at random and independently from the genes in the parental generation (binomial sampling). He showed that, in a diploid population of fixed size N and with mutation rate u , the probability that two genes taken at random from the $2N$ genes in any (stationary) generation are of the same allelic type is very close to $(4Nu + 1)^{-1}$. This may be taken as a measure of the genetic diversity in the population at that locus.

Kimura and Crow were no doubt motivated by the fact that information about the extent of genetic variation in natural populations was becoming available in the 1960s, so that

Malécot's result was relevant for a theoretical analysis of the reasons for this variation. They made the straightforward generalization of Malécot's formula to $(4N_e u + 1)^{-1}$, where N_e is the effective population size. [There are at least four concepts of effective population size (Ewens 2000); Kimura and Crow used the “inbreeding” effective population size.] From this they defined the “effective number of alleles” as $n = 4N_e u + 1$, which would be the mean number of alleles present in the population at any time if all alleles had the same frequency. However, in the infinitely many alleles model the alleles present at any one time usually have widely varying frequencies, and as a result the mean number of alleles present is much larger than n . For example, the “effective number of alleles” n is 4 when $N_e = 250,000$ and $u = 4 \times 10^{-6}$, whereas the mean number of alleles present in the population is about 42 (Ewens 1964). Despite the title of their article, Kimura and Crow did not give a formula for the mean number of alleles present and claimed that it should be close to the effective number, which the above example shows is not the case. As a result the effective number of alleles concept never gained traction.

Malécot obtained his equation $(4Nu + 1)^{-1}$ by considering a retrospective analysis, looking backward in time (as opposed to the traditional evolutionary prospective analysis, looking forward in time). The retrospective approach has dominated population genetics theory for the last few decades. For example, Kimura's (1968) neutral theory and the investigation of the properties of a sample to test for neutrality (Ewens 1972; Watterson 1974, 1978) were directly influenced by Kimura and Crow's article. In another direction, Kimura (1971), also a foundational article of evolutionary molecular genetics, explicitly took into account the gene as a sequence of nucleotides. Watterson (1975) found many important properties of this model, again focusing on a retrospective analysis by considering the ancestry of a sample of genes. This form of analysis culminated in Kingman's (1982)

concept of the coalescent, in which the entire ancestry of a sample of genes, or all the genes in a population, is traced back to a common ancestor, leading to a revolution in population genetics theory. All of these developments, and many others, can be traced back to Kimura and Crow's foundational article.

Literature Cited

- Ewens, W. J., 1964 The maintenance of alleles by mutation. *Genetics* 50: 891–898.
- Ewens, W. J., 1972 The sampling theory of selectively neutral alleles. *Theor. Popul. Biol.* 3: 87–112.
- Ewens, W. J., 2000 *Mathematical Population Genetics*. Springer-Verlag, New York.
- Fisher, R. A., 1922 On the dominance ratio. *Proc. R. Soc. Edinb.* 42: 321–341.
- Kimura, M., 1968 Evolution at the molecular level. *Nature* 217: 624–626.
- Kimura, M., 1971 Theoretical foundations of population genetics at the molecular level. *Theor. Popul. Biol.* 2: 174–208.
- Kingman, J. F. C., 1982 The coalescent. *Stoch. Proc. Appl.* 13: 235–248.
- Malécot, G., 1948 *Les Mathématiques de l'Hérédité*. Masson, Paris.
- Watterson, G. A., 1974 The sampling theory of selectively neutral alleles. *Adv. Appl. Probab.* 6: 463–488.
- Watterson, G. A., 1975 On the number of segregating sites in genetic models without recombination. *Theor. Popul. Biol.* 7: 256–276.
- Watterson, G. A., 1978 The homozygosity test of neutrality. *Genetics* 88: 405–417.
- Wright, S., 1931 Evolution in Mendelian populations. *Genetics* 16: 97–159.
- Crow, J. F., 1948a A consequence of the dominance hypothesis of hybrid vigor. *Genetics* 33: 101.
- Crow, J. F., 1948b Differences in susceptibility to ether in the *virilis* group of *Drosophila*. *Genetics* 33: 102.
- Crow, J. F., 1948c Alternative hypotheses of hybrid vigor. *Genetics* 33: 477–487.
- Crow, J. F., 1979 Minor viability mutants in *Drosophila*. *Genetics* 92: s165–s172.
- Crow, J. F., 1987 Seventy years ago in *Genetics*: H.S. Jennings and inbreeding theory. *Genetics* 115: 389–391.
- Crow, J. F., 1988a A diamond anniversary: the first chromosomal map. *Genetics* 118: 1–3.
- Crow, J. F., 1988b The ultraselfish gene. *Genetics* 118: 389–391.
- Crow, J. F., 1988c Eighty years ago: the beginnings of population genetics. *Genetics* 119: 473–476.
- Crow, J. F., 1988d The genesis of dysgenesis. *Genetics* 120: 315–318.
- Crow, J. F., 1989 Fortunes of war. *Genetics* 122: 467–469.
- Crow, J. F., 1990a Anecdotal, historical and critical commentaries on genetics. R. A. Fisher, a centennial view. *Genetics* 124: 207–211.
- Crow, J. F., 1990b Mapping functions. *Genetics* 125: 669–671.
- Crow, J. F., 1991 Our diamond birthday anniversary. *Genetics* 127: 1–3.
- Crow, J. F., 1992a Centennial: J. B. S. Haldane, 1892–1964. *Genetics* 130: 1–6.
- Crow, J. F., 1992b Erwin Schrodinger and the hornless cattle problem. *Genetics* 130: 237–239.
- Crow, J. F., 1992c Twenty-five years ago in *Genetics*: identical triplets. *Genetics* 130: 395–398.
- Crow, J. F., 1992d Sixty years ago: the 1932 International Congress of Genetics. *Genetics* 131: 761–768.
- Crow, J. F., 1993a Felix Bernstein and the first human marker locus. *Genetics* 133: 4–7.
- Crow, J. F., 1993b N. I. Vavilov, martyr to genetic truth. *Genetics* 134: 1–4.
- Crow, J. F., 1993c Francis Galton: count and measure, measure and count. *Genetics* 135: 1–4.
- Crow, J. F., 1994 Hitoshi Kihara, Japan's pioneer geneticist. *Genetics* 137: 891–894.
- Crow, J. F., 1995 Quarreling geneticists and a diplomat. *Genetics* 140: 421–426.
- Crow, J. F., 1996 A golden anniversary: cattle twins and immune tolerance. *Genetics* 144: 855–859.
- Crow, J. F., 1997 Birth defects, Jimson weeds and bell curves. *Genetics* 147: 1–6.
- Crow, J. F., 1998 90 years ago: the beginning of hybrid maize. *Genetics* 148: 923–928.
- Crow, J. F., 1999 Hardy, Weinberg and language impediments. *Genetics* 152: 821–825.
- Crow, J. F., 2000 Thomas H. Jukes (1906–1999). *Genetics* 154: 955–956.
- Crow, J. F., 2001a Plant breeding giants. Burbank, the artist; Vavilov, the scientist. *Genetics* 158: 1391–1395.
- Crow, J. F., 2001b Shannon's brief foray into genetics. *Genetics* 159: 915–917.
- Crow, J. F., 2002 C. C. Little, cancer and inbred mice. *Genetics* 161: 1357–1361.
- Crow, J. F., 2003 C. C. Tan: a life of peaks and valleys. *Genetics* 164: 1–4.
- Crow, J. F., 2006 H. J. Muller and the “competition hoax”. *Genetics* 173: 511–514.
- Crow, J. F., 2007 Haldane, Bailey, Taylor and recombinant-inbred lines. *Genetics* 176: 729–732.
- Crow, J. F., 2008 Just and unjust: E. E. Just (1883–1941). *Genetics* 179: 1735–1740.
- Crow, J. F., and S. Abrahamson, 1997 Seventy years ago: mutation becomes experimental. *Genetics* 147: 1491–1496.

Further Reading in *GENETICS*

- Abrahamson, S., 2012 James F. Crow: his life in public service. *Genetics* 190: 1–4.
- Crow, J. F., 1987 Twenty-five years ago in genetics: Motoo Kimura and molecular evolution. *Genetics* 116: 183–184.
- Crow, J. F., 1989a Twenty-five years ago in genetics: the infinite allele model. *Genetics* 121: 631–634.
- Crow, J. F., 1995 Motoo Kimura (1924–1994). *Genetics* 150: 1–5.
- Ewens, W. J., 2012 James F. Crow and the stochastic theory of population genetics. *Genetics* 190: 287–290.
- Hartl, D. L., 2011 James F. Crow and the art of teaching and mentoring. *Genetics* 189: 1129–1133.
- Hartl, D. L., 2012 James F. Crow (1916–2012) a remarkable geneticist, a remarkable man. *Genetics* 190: 1149–1150.
- Nagylaki, T., 1989 Gustave Malécot and the transition from classical to modern population genetics. *Genetics* 122: 253–268.
- Susman, M., and R. Greeberg Temin, 2012 James F. Crow: storied teacher, leader, and colleague at the University of Wisconsin. *Genetics* 191: 1–5.
- Turelli, M., and C. Langley, 2011 Honoring our colleague James F. Crow, an outstanding gentleman, citizen, and scientist. *Genetics* 189: 1127.

Other *GENETICS* articles by M. Kimura and J. F. Crow

- Cavalli-Sforza, L. L., M. Kimura, and I. Barrai, 1966 The probability of consanguineous marriages. *Genetics* 54: 37–60.

- Crow, J. F., and W. Bender, 2004 Edward B. Lewis, 1918–2004. *Genetics* 168: 1773–1783.
- Crow, J. F., and Y. J. Chung, 1967 Measurement of effective generation length in *Drosophila* population cages. *Genetics* 57: 951–955.
- Crow, E. W., and J. F. Crow, 2002 100 years ago: Walter Sutton and the chromosome theory of heredity. *Genetics* 160: 1–4.
- Crow, J. F., and W. F. Dove, 1987 Sewall Wright and physiological genetics. *Genetics* 115: 1–2.
- Crow, J. F., and W. F. Dove, 1998 Birds' eye view: a decade of perspectives. *Genetics* 148: 1405–1407.
- Crow, J. F., and J. Kermicle, 2002 Oliver Nelson and quality protein maize. *Genetics* 160: 819–821.
- Crow, J. F., and R. D. Owen, 2000 Kay Wilson and the NIH genetics study section. *Genetics* 155: 1–5.
- Crow, J. F., and W. C. Roberts, 1950 Inbreeding and homozygosis in bees. *Genetics* 35: 612–621.
- Crow, J. F., D. Lindsley, and J. Lucchesi, 2006 Edward Novitski: *Drosophila* virtuoso. *Genetics* 174: 549–553.
- Denniston, C., and J. F. Crow, 1990 Alternative fitness models with the same allele frequency dynamics. *Genetics* 125: 201–205.
- Drake, J. W., and J. F. Crow, 1996 Recollections of Howard Temin (1934–1994). *Genetics* 144: 1–6.
- Drake, J. W., B. Charlesworth, D. Charlesworth, and J. F. Crow, 1998 Rates of spontaneous mutation. *Genetics* 148: 1667–1686.
- Greenberg, R., and J. F. Crow, 1960 A comparison of the effect of lethal and detrimental chromosomes from *Drosophila* populations. *Genetics* 45: 1153–1168.
- Hiraizumi, Y., and J. F. Crow, 1960 Heterozygous effects on viability, fertility, rate of development, and longevity of *Drosophila* chromosomes that are lethal when homozygous. *Genetics* 45: 1071–1083.
- Horowitz, N. H., P. Berg, M. Singer, J. Lederberg, M. Susman *et al.*, 2004 A centennial: George W. Beadle, 1903–1989. *Genetics* 166: 1–10.
- Kimura, M., 1962 On the probability of fixation of mutant genes in a population. *Genetics* 47: 713–719.
- Kimura, M., 1965 Attainment of quasi linkage equilibrium when gene frequencies are changing by natural selection. *Genetics* 52: 875–890.
- Kimura, M., 1969 The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics* 61: 893–903.
- Kimura, M., 1975 Mathematical contributions to population genetics. *Genetics* 79(Suppl.): 91–100.
- Kimura, M., and H. Kayano, 1961 The maintenance of supernumerary chromosomes in wild populations of *Lilium callosum* by preferential segregation. *Genetics* 46: 1699–1712.
- Kimura, M., and T. Maruyama, 1966 The mutational load with epistatic gene interactions in fitness. *Genetics* 54: 1337–1351.
- Kimura, M., and T. Ohta, 1969 The average number of generations until fixation of a mutant gene in a finite population. *Genetics* 61: 763–771.
- Kimura, M., and T. Ohta, 1973a Mutation and evolution at the molecular level. *Genetics* 73(Suppl. 73): 19–35.
- Kimura, M., and T. Ohta, 1973b The age of a neutral mutant persisting in a finite population. *Genetics* 75: 199–212.
- Kimura, M., and T. Ohta, 1969 The average number of generations until extinction of an individual mutant gene in a finite population. *Genetics* 63: 701–709.
- Kimura, M., and T. Ohta, 1970 Probability of fixation of a mutant gene in a finite population when selective advantage decreases with time. *Genetics* 65: 525–534.
- Kimura, M., and G. H. Weiss, 1964 The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* 49: 561–576.
- Kimura, M., T. Maruyama, and J. F. Crow, 1963 The mutation load in small populations. *Genetics* 48: 1303–1312.
- Kondrashov, A. S., and J. F. Crow, 1988 King's formula for the mutation load with epistasis. *Genetics* 120: 853–856.
- Langley, C. H., and J. F. Crow, 1974 The direction of linkage disequilibrium. *Genetics* 78: 937–941.
- Maas, W., and J. F. Crow, 2004 Leo Szilard: a personal remembrance. *Genetics* 167: 555–558.
- Mukai, T., S. I. Chigusa, L. E. Mettler, and J. F. Crow, 1972 Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster*. *Genetics* 72: 335–355.
- Mukai, T., R. A. Cardellino, T. K. Watanabe, and J. F. Crow, 1974 The genetic variance for viability and its components in a local population of *Drosophila melanogaster*. *Genetics* 78: 1195–1208.
- Ohta, T., and M. Kimura, 1969 Linkage disequilibrium at steady state determined by random genetic drift and recurrent mutation. *Genetics* 63: 229–238.
- Ohta, T., and M. Kimura, 1970 Statistical analysis of the base composition of genes using data on the amino acid composition of proteins. *Genetics* 64: 387–395.
- Ohta, T., and M. Kimura, 1971a Linkage disequilibrium between two segregating nucleotide sites under the steady flux of mutations in a finite population. *Genetics* 68: 571–580.
- Ohta, T., and M. Kimura, 1971b Behavior of neutral mutants influenced by associated overdominant loci in finite populations. *Genetics* 69: 247–260.
- Ohta, T., and M. Kimura, 1974 Simulation studies on electrophoretically detectable genetic variability in a finite population. *Genetics* 76: 615–624.
- Simmons, M. J., E. W. Sheldon, and J. F. Crow, 1978 Heterozygous effects on fitness of EMS-treated chromosomes in *Drosophila melanogaster*. *Genetics* 88: 575–590.
- Takahata, N., and M. Kimura, 1981 A model of evolutionary base substitutions and its application with special reference to rapid change of pseudogenes. *Genetics* 98: 641–657.
- Temin, R. G., H. U. Meyer, P. S. Dawson, and J. F. Crow, 1969 The influence of epistasis on homozygous viability depression in *Drosophila melanogaster*. *Genetics* 61: 497–519.
- Wagner, R. P., and J. F. Crow, 2001 The other fly room: J. T. Patterson and Texas genetics. *Genetics* 157: 1–5.

Communicating editor: C. Gelling