

A model in two acts: a commentary on 'A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population' by Tomoko Ohta and Motoo Kimura

JODY HEY*

Department of Genetics, Rutgers University, Nelson Biological Labs (rm B326), 604 Allison Road, Piscataway, NJ 08854-8082, USA

In the early 1970s, protein electrophoresis was the primary tool geneticists used to discover and measure allelic variation in natural populations. It was a relatively simple and inexpensive technique and, most importantly, it permitted the detection of multiple alleles regardless of polymorphism levels. This was a critical point because before the age of protein electrophoresis, segregating alleles were usually discovered only in cases where clearly discrete patterns of phenotypic variation were first observed. With protein electrophoresis a geneticist's ability to identify multiple alleles did not depend on a prior indication of the presence of genetic variation (Hubby & Lewontin, 1966; Lewontin, 1974).

The purpose of the short paper that Tomoko Ohta and her mentor Motoo Kimura published in *Genetical Research* in 1973 was to devise a mutation model that was explicitly appropriate for protein electrophoretic data and that would permit such data to be analysed with regard to questions on the relative roles of natural selection and genetic drift. As data on electrophoretic alleles began to accumulate, it was discovered that individuals were heterozygous, and many species were polymorphic, at a substantial fraction of the proteins that could be surveyed. These numbers on heterozygosity and polymorphism immediately began to feed a long-standing hunger, that had built up from decades of sophisticated modelling, for data on such topics as mutation rates, genetic load, the rate of neutral mutations, and the relative roles of natural selection and genetic drift in shaping levels and patterns of variation.

Ohta and Kimura were the primary theoreticians of the neutral theory of molecular evolution and they had a very strong interest (as did most population geneticists of that age) in understanding how well the neutral theory explained the levels of polymorphism discovered by electrophoresis. The models they developed focused on amounts and patterns of genetic

variation, and they tended to include explicitly a neutral mutation rate as well as assumptions about the nature of the mutation process.

One prediction of the neutral theory was that the number of alleles in a population was expected to co-vary strongly with the effective population size. Earlier in 1964, Kimura and James Crow had developed the infinite alleles model, in which every mutation gives rise to a new allele (Kimura & Crow, 1964), and under this model the number of neutral alleles varies linearly with both effective population size and neutral mutation rate.

Ohta and Kimura's key idea in 1973 was a mutation model that explicitly gave rise to new allelic states in single steps that differed in net protein charge. Because four of the amino acids are normally charged at physiological pH, the surface of a soluble protein will carry a charge that affects its behaviour in gel electrophoresis, and mutations that raise or lower this charge will increase or decrease the rate of electrophoresis. In Ohta and Kimura's model, which later came to be known as the 'stepwise mutation model' (Kimura & Ohta, 1978) and also sometimes the 'ladder model', a protein may mutate to a different allelic state in +1 and -1 steps. Importantly, unlike the infinite sites model, it was possible under this model for two proteins to be identical in kind (i.e. have the same net charge), and not be identical by virtue of common descent from an ancestral gene of the same allelic state (i.e. identity by descent).

Ohta and Kimura were a virtual dynamic duo of differential equations, and in this paper as in many others they took a diffusion equation approach. Their primary target was an expression for the effective number of alleles in a population, n_e . Under the infinite sites model, Kimura and Crow had shown that $n_e = 1 + 4N_e u$ (Kimura & Crow, 1964). But under the stepwise model, Ohta and Kimura found that $n_e = \sqrt{1 + 8N_e u}$. The difference between the two models in predicted number of effective alleles is small for low values but becomes dramatic for high values of $N_e u$. Ohta and Kimura then drew upon these

* Tel: 732-445-5272. Fax: 732-445-5870. e-mail: hey@biology.rutgers.edu. <http://lifesci.rutgers.edu/~hey/lab/>

different predictions as a possible explanation for why species with very large effective population size (e.g. as expected for some *Drosophila* species) do not reveal hundreds or thousands of alleles, as might be expected under the infinite alleles model (Ayala *et al.*, 1972).

In reality, the electrophoretic mobility of a protein is affected by its shape as well as its charge, and both are affected by mutation. Yet empirical assessments of the fit of the model found that charge differences, and the resulting ladder-like pattern of electrophoretic mobilities, explained most of the mobility differences observed (Cobbs & Prakash, 1977; Fuerst & Ferrell, 1980).

1. Act II for the stepwise mutation model

The rate of use by investigators of the stepwise mutation model has had an unusual trajectory. Through the 1970s and 1980s, Ohta and Kimura's paper was highly cited, and their model was extensively used and tested. But thereafter, applications of the model declined as protein electrophoretic data began to be replaced by more direct assessments of genetic variation. As DNA sequences and the repetitive structure of genomes came under close examination, so came the discovery of common short tandem repeats (STRs) (Jeffreys *et al.*, 1985). In 1987, Levinson and Gutman discerned that slipped-strand mispairing was the likely mechanism for length changes in tandem repeats and that most altered repeats differed by 1 repeat unit from the parental strand (Levinson & Gutman, 1987*a, b*). Ohta and Kimura's little paper might be a footnote today were it not for the fact that DNA polymerase is highly susceptible to making unit-length mistakes when replicating tandem repeats. In early papers that considered the population genetic implications of STR polymorphism, the stepwise model was rejuvenated (Brookfield, 1989; Flint *et al.*, 1989; Chakraborty *et al.*, 1991). Today STR data are widely used in population genetics, and Ohta and Kimura's model is the starting point for modeling their mutation process.

References

- Ayala, F. J., Powell, J. R., Tracey, M. L., Mourao, C. A. & Perez-Salas, S. (1972). Enzyme variability in the *Drosophila willistoni* group. IV. Genic variation in natural populations of *Drosophila willistoni*. *Genetics* **70**, 113–139.
- Brookfield, J. F. Y. (1989). Analysis of DNA fingerprinting data in cases of disputed paternity. *Mathematical Medicine and Biology* **6**, 111–131.
- Chakraborty, R., Fornage, M., Gueguen, R. & Boerwinkle, E. (1991). Population genetics of hypervariable loci: analysis of PCR based VNTR polymorphism within a population. In *DNA Fingerprinting Approaches and Applications* (ed. T. Burke, G. Dolf, A. J. Jeffreys and R. Wolff), pp. 127–143. Basel, Switzerland: Birkhauser Verlag.
- Cobbs, G. & Prakash, S. (1977). An experimental investigation of the unit charge model of protein polymorphism and its relation to the esterase-5 locus of *Drosophila pseudoobscura*, *Drosophila persimilis*, and *Drosophila miranda*. *Genetics* **87**, 717–742.
- Flint, J., Boyce, A. J., Martinson, J. J. & Clegg, J. B. (1989). Population bottlenecks in Polynesia revealed by minisatellites. *Human Genetics* **83**, 257–263.
- Fuerst, P. A. & Ferrell, R. E. (1980). The stepwise mutation model: an experimental evaluation utilizing hemoglobin variants. *Genetics* **94**, 185–201.
- Hubby, J. L. & Lewontin, R. C. (1966). A molecular approach to the study of genic heterozygosity in natural populations. I. the number of alleles at different loci in *Drosophila pseudoobscura*. *Genetics* **54**, 577–594.
- Jeffreys, A. J., Wilson, V. & Thein, S. L. (1985). Hypervariable 'minisatellite' regions in human DNA. *Nature* **314**, 67–73.
- Kimura, M. & Crow, J. F. (1964). The number of alleles that can be maintained in a finite populations. *Genetics* **49**, 725–738.
- Kimura, M. & Ohta, T. (1978). Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proceedings of the National Academy of Sciences of the USA* **75**, 2868–2872.
- Levinson, G. & Gutman, G. A. (1987*a*). High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. *Nucleic Acids Research* **15**, 5323–5338.
- Levinson, G. & Gutman, G. A. (1987*b*). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution* **4**, 203–221.
- Lewontin, R. C. (1974). *The Genetic Basis of Evolutionary Change*. New York: Columbia University Press.