NATIONAL ACADEMY OF SCIENCES

COLUMBUS CLARK COCKERHAM

1921-1996

A Biographical Memoir by

BRUCE S. WEIR

C. Clark Cockerham, item number 0006835, Special Collections Research Center, North Carolina State University Research Libraries, Raleigh, North Carolina

COLUMBUS CLARK COCKERHAM

December 21, 1921–November 4, 1996

BY BRUCE S. WEIR

C. CLARK COCKERHAM IS REMEMBERED for his pioneering formulation of new statistical theory in quantitative genetics and for establishing principles for the understanding of heredity and the genetic improvement of animals and plants. His related work in population-genetic formulations of the variance of allele frequencies provides the means to characterize population structure, and is used by both human geneticists and molecular ecologists. His work is marked by elegance and evidence of a deep intuition.

Clark Cockerham was born on December 21, 1921, in Mountain Park, North Carolina, and grew up on the family farm next to the mountains of western North Carolina. He received a B.S. in animal production from North Carolina State College in 1943 and then joined the U. S. Marines. He saw service in the Pacific and would have gone to Japan had it not been for the end of the war following the dropping of the atomic bombs. He returned to North Carolina and managed a lumber mill before returning to graduate school at NC State with support from the GI Bill. With an M.S. in animal industry awarded in 1949 he went on to Iowa State College to study with renowned quantitative geneticist Jay Lush. Iowa State was a center for quantitative genetics, and the thesis for Clark's 1952 Ph.D. in animal breeding and genetics laid the foundation for much subsequent development in the field.

Before leaving Raleigh, Clark built a set of modular furniture that could be disassembled to fit into a small trailer. He and his young family then drove to Ames. His love and skill for woodworking remained throughout his life, and were put to good use later when he built his Raleigh home and when his wife Joyce ventured into the antiques business and her finds needed repairs or restoration. My own family benefited from my being able to use his workshop to construct bookcases and deck furniture when we were setting up house in Raleigh.

In 1952 Clark returned to North Carolina as an assistant professor in the Department of Biostatistics at the University of North Carolina at Chapel Hill. He did not enjoy his experience teaching medical students, especially "after they had seen blood," and he was grateful to return to an agricultural and biological setting as associate professor of statistics at NC State the following year. Clark was not a natural lecturer, and it is little wonder that his UNC experience was short lived. The Department of Experimental Statistics at NC State was still young and had established a supportive environment for interdisciplinary studies that would last for several years. Ralph Comstock and H. F. ("Cotton") Robinson, known for their formulation of the "North Carolina Designs" for quantitative genetic experiments, had received funding for their work from the Rockefeller Foundation. Clark thrived in this atmosphere, and he elected to stay in the statistics department when Comstock and Robinson moved to the new Department of Genetics. By 1963 he was able to compete successfully for a National Institute of General Medical Sciences program project grant in quantitative genetics. He directed this grant until his retirement in 1990 and was involved in the successor program project grant that I was able to direct until 2005.

This continuous 42-year funding period supported a broad approach to quantitative genetics that included applications to *Drosophila,* mice, and maize, as well as basic statistical methodology. Under Clark's leadership NC State became recognized as the leading center for this field, and this was made clear when the university hosted the Second International Conference on Quantitative Genetics to mark his 65th birthday in 1986. The proceedings volume from that conference provides a valuable snapshot of the field at a moment that just preceded the impact of the explosion in molecular genetic technology.

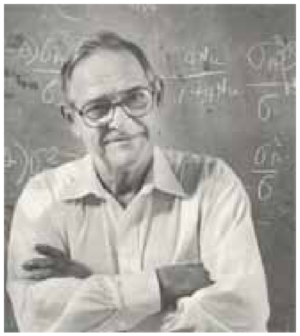

Photo courtesy North Carolina State University.

Clark remained active after retirement and published his last paper in 1996. His final years were made difficult by ill health that probably owed something to heavy smoking throughout his life. When I asked him how he was doing on my last visit, he replied with characteristic bluntness, "Deteriorating, but not quickly enough." We lost a wonderful colleague on November 4, 1996, and quantitative genetics lost its leader. Clark was survived by his wife, the former Joyce Evelyn Allen, and three children, C. Clark Cockerham Jr., Jean Davis, and Bruce A. Cockerham.

Clark was not a natural lecturer, but he was a natural mentor and leader. He hosted a monthly evening meeting at his home for quantitative geneticists from NC State and Research Triangle Park. An invitation to speak at this gathering was both coveted and feared, as Clark relished debate and delighted in exposing weaknesses in the speakers' arguments. Visual aids or handouts were frowned on; use of a blackboard was encouraged instead. My Ph.D. thesis work was developed on the board in Clark's office and the photograph on page 4 shows Clark standing in front of that board with his own writing showing clearly. Clark worked either at the board or at his desk with pencil and yellow legal pad. I have yet to meet anyone else who could remain seated and concentrating for as many hours at a stretch as he did.

Clark Cockerham's early work was in theoretical quantitative genetics. When there is natural variation for a continuous character in a population, plant or animal breeders can select the best individuals and so improve the quality of that character in subsequent generations. The selection protocol requires the trait variation to be partitioned into hereditary and environmental components; prediction of the likely gain in economic traits requires the further partitioning of the genetic variation into components reflecting the actions of alleles acting singly or jointly. In Cockerham (1954) he derived this partitioning into additive components and dominance or epistatic components for pairs of alleles of one gene or sets of alleles from more than one gene. His interest in the theory of multiple alleles is reflected in much of his later work.

Some of the dependencies between effects or frequencies of alleles at different genetic loci results from their physical proximity and the linkage between them. Cockerham (1956) showed how the covariance of trait values for pairs of relatives was affected by linkage of the

The work for which Clark Cockerham has been most cited began with his 1969 paper on the variance of allele frequencies. The paper is difficult to read, but it brings the quantitative genetic language of variance components to the population-genetic building blocks of allele frequencies.

underlying genes, and then Cockerham (1963) laid out the methodology for estimating the genetic components of variance. The essence of the method was to translate statistical covariances of trait values for specified classes of relatives into the additive, dominance, and epistatic components. This allowed the genetic components to be estimated from the observed statistical components. The duality of genetic and statistical formulations is the hallmark of quantitative genetics, and Cockerham had the biological and mathematical insights to master both.

Cockerham was always careful not to claim overly much credit for the impact of his work, but his theory on genetic variance components, the correlation of relatives, and the prediction of selection response certainly coincided with the move away from multi-line crosses in maize breeding toward two-line hybrids. The dominance of this strategy must surely owe something to his work and that of others at NC State.

Quantitative genetics is concerned with complex characters affected by many genes. Initially attention was focused on traits of economic importance although now there is attention increasingly on those affecting human health. Population genetics, on the other hand, has generally been concerned with the frequencies of alleles and so can be regarded as applying to discrete characters. The distinction is artificial in the sense that both fields rest on the same underlying principles. Clark Cockerham's major contributions to population genetics began with his 1967 paper on group inbreeding. The inbreeding coefficient of an individual is the probability that it receives identical copies of the same gene from each parent, and Cockerham extended this concept to apply to groups of individuals. In 1971 he carried these ideas forward to sets of three or four alleles and so provided a mechanism for describing the probabilities that two individuals had specific genotypes, or carried specific sets of alleles. These probabilities find great application in current methods for mapping disease genes when affected relatives are studied. An application that was certainly not anticipated in 1971 was in the forensic use of DNA profiles. Twenty-five years later, with his help, I was able to use his approach to give the probabilities that two people, an innocent suspect and an unknown perpetrator, would have matching profiles.

The work for which Clark Cockerham has been most cited began with his 1969 paper on the variance of allele frequencies. The paper is difficult to read, but it brings the quantitative genetic language of variance components to the population-genetic building blocks of allele frequencies. In a very elegant way Clark assigned each allele an indicator variable that identified the individual carrying the allele and the population to which the individual belongs. The variable was assigned the value of one if the allele was of a specific type, and zero otherwise. The statistical variances and covariances


C. Clark Cockerham, undated photo.

of these indicator variables can be expressed in terms of parameters, confusingly known as "F-statistics," that serve to describe both the covariances of alleles within populations and the variances of alleles over populations—a vivid marrying of statistical and genetic concepts. In essence Clark broke the variance of allele frequencies into three components: those for alleles within individuals, for individuals within populations, and between populations. Sample variances of the first two components may be reduced by increasing the sample numbers for individuals or populations, but the third component is driven by the evolutionary process and its variance cannot be reduced. This is a subtlety that is often not appreciated.

In 1974—with the partitioning of genetic variance for quantitative traits and the partitioning of variance for allele frequencies published—Clark Cockerham was elected to the National Academy of Sciences. This national recognition preceded his recognition at the state level (North Carolina Medal for Science, 1976), at the university-system level (O. Max Gardner Award, 1980), and at the university-level (Holladay Medal, 1994). Clark was also recognized by his department (Mason Award, 1983), by his fellow alumni (NC State Alumni Award for Outstanding Research, 1986), and by his professional associations (Gamma Sigma Delta Award of Merit, 1964; American Society of Agronomy, fellow, 1972; and Crop Science Society of America, fellow, 1985). It was without his knowledge—and certainly against what he would have wished—that NC State nominated him and other members of the program project group for the USDA Superior Service Award for Scientific Research in 1990. That award was a fitting recognition for his career of scientific leadership.

Cockerham's contributions did not cease with National Academy of Sciences membership. His 1973 paper on the analysis of allele frequencies was intended to clarify the 1969 variance paper. The utility of allele frequency-variance components to measure the evolutionary distance between populations was described in 1983, and then his "bestseller" was published in 1984. This paper clarified some of the estimation procedures for the components of allele-frequency variance, and in 2005 this paper was hailed as one of the 25 most-cited papers in statistics. It has been interesting to watch the adoption of the methodology of that paper by different genetic communities. Initially there was interest from those studying multiple populations of the same species (very often fish species), then by ecologists wishing to estimate migration rates, then by forensic scientists wishing to strengthen their DNA profile match probabilities, and recently by human geneticists wishing to select single nucleotide polymorphisms (SNPs) for disease association studies.

.While being interviewed by the *Raleigh News and Observer* as Tarheel of the Week in 1986 he answered the question “How do you explain to lay people what you do?” with, “I don't.” But then he relented a little to say, “You work on things which you think might be useful…”

My own role in the F-statistics work was complemented by studies Clark guided for my thesis and that we continued after I became his faculty colleague at NC State. We described a set of descent measures that summarized the history of sets of alleles at two genetic loci. Clark used these measures in a series of papers in 1983, 1984, and 1985 that described the trajectory of quantitative trait values in plant populations that propagate by self-fertilization. The importance of this work follows from the fact that many of our major crop species are full or partial selfers Cockerham's papers allow the prediction of the response to artificial selection, and so allow the comparison of alternative strategies for crop improvement. In the 1985 paper he found conditions under which long-term gain is improved by evaluations of selfed rather than outcrossed progeny of selected individuals. In a more general development, by breaking away from the classical two-allele models of quantitative genetics, Cockerham showed how the descent measures served as coefficients of components of variance in addition to the usual additive and dominance terms for single-gene models.

Clark Cockerham's contributions were to the theory of quantitative and population genetics, but he had several collaborations with experimentalists, many of whom were supported on the program project grant. In addition to applications of his work for selfing species of plants, and his early student work on swine, he worked on *Drosophila*. He suggested that statistical models for quantitative genetic characters be expanded to allow for nongenetic paternal as well as maternal influences on offspring. This led, in a 1978 paper, to surprisingly high estimates of paternal influences on *Drosophila* viability. In a 1977 paper he provided a method for estimating mutation rates from massive experiments with fruit flies.

Clark Cockerham had a deep understanding of his field, and a deep passion for honesty and excellence. This did not always make for an easy personality. While being interviewed by the *Raleigh News and Observer* as Tarheel of the Week in 1986 he answered the question “How do you explain to lay people what you do?” with, “I don't.” But then he relented a little to say, “You work on things which you think might be useful . . . I work with inheritance in an area which has some utility in plant and animal breeding. But the details I don't try to explain at all.” He saw little need for textbooks in his field or for other introductory writing, and he did not allow such activities to distract him from his research. He could be harsh in print when criticizing others, but his own writing has never required a correction, not only because of the quality of his work but also because of the care he took: “We don't write ‘instant' papers.” His ethical standards were of the very highest quality. On the other hand, he was the most generous of colleagues.

The nicest summary of the work of C. Clark Cockerham is in the citation to his Gardner award:

> For many people of the world your work is the difference between food and hunger. For those who probe the mysteries of life for the sake of knowledge, your work is the triumph of the human mind.

Undoubtedly, he would have preferred it to use his own words, from the *Raleigh News and Observer* interview:

> Some of the best satisfaction I have had is to be able to find fairly simple solutions to reasonably complex problems.

This memoir is based on my recollections of an association with Clark Cockerham that began when he invited me to come from New Zealand to study with him in 1965 and lasted until his death in 1996. I have augmented these recollections with material in the various nomination packages for Clark's awards, many of which were written with the help of Major M. Goodman, whose association with Clark predates my own. I shall forever be in Clark's debt.

NOTES

1. T. Ryan and W. H. Woodall. The most cited statistical papers. *J. Appl. Stat.* 32(2005):461-474.

SELECTED BIBLIOGRAPHY

1954

An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics* 39:859-882.

1956

Effects of linkage on the covariances between relatives. *Genetics* 41:138-141.

1959

Partitions of hereditary variance for various genetic models. *Genetics* 44:1141-1148.

1963

Estimation of genetic variances. In *Statistical Genetics and Plant Breeding*, eds. W. D. Hanson and H. F. Robinson, pub. 982, pp. 53-94. Washington, D.C.: National Academy of Sciences/National Research Council.

1967

Group inbreeding and coancestry. *Genetics* 56:89-104.

1969

Variance of gene frequencies. *Evolution* 23:72-84.

1971

Higher order probabilities of identity of alleles by descent. *Genetics* 69:235-246.

1973

Analyses of gene frequencies. *Genetics* 74:679-700.

With B. S. Weir. Descent measures for two loci with some applications. *Theor. Popul. Biol.* 4:300-330.

1977

With T. Mukai. Spontaneous mutation-rates at enzyme loci in *Drosophila melanogaster. Proc. Natl. Acad. Sci. U.S.A.* 74:2514-2517.

SELECTED BIBLIOGRAPHY

1978

With T. Mukai. Effects of marker chromosomes on relative viability. *Genetics* 90:827-849.

1980

Random and fixed effects in plant genetics. *Theor. Appl. Genet.* 56:119-131.

1983

Covariances of relatives from self-fertilization. *Crop Sci.* 23:1177-1180.

With J. Reynolds and B. S. Weir. Estimation of the coancestry coefficient: Basis for a short-term genetic distance. *Genetics* 105:767-779.

1984

Additive by additive variance within breeding and linkage. *Genetics* 18:487-500.

With B. S. Weir. Estimating F-statistics for the analysis of population-structure. *Evolution* 28:1358-1370.

1985

With D. F. Matzinger. Selection response based on selfed progenies. *Crop Sci.* 25:483-488.

1996

With Z. B. Zeng. Design III with marker loci. *Genetics* 143:1437-1456.