Supplement to "Fifty years of *Theoretical Population Biology*" by Noah A. Rosenberg

In preparing the 50th anniversary issue of *Theoretical Population Biology*, I wrote to Eric Charnov and Warren Ewens with a series of questions about their iconic papers. Both kindly responded with details on the origin of their papers. Answers from Prof. Charnov have been incorporated into the introductory essay for the special issue; Prof. Ewens shared the following recollections.

-Noah Rosenberg

Recollections from Warren Ewens on the genesis of Ewens (1972)

I did a double major in Pure Mathematics and Mathematical Statistics at the University of Melbourne, followed by a Master's degree in Statistics at the same institution. In 1961 I moved to the Australian National University as a lecturer (= assistant professor) - yes, you could get such a position in those post-Sputnik days with only a Master's degree, since academic positions in the sciences were opening up at an astonishing rate. Upon arrival at ANU, I decided to do a PhD under Pat Moran concurrently with my lecturing work - yes, in those days you could do that. I decided to work with Pat since I had been advised that genetics was the coming area, and Pat, who changed his research field every five years or so, was then working in population genetics.

Pat followed the Cambridge tradition of "sink or swim" by leaving me pretty much alone. I almost sank but eventually swam, albeit weakly. After completing my degree I did a post-doc with Sam Karlin at Stanford. Two days after arriving at Stanford in 1964 I went to a seminar given by Jim Crow, also attended by Jim McGregor, Walter Bodmer and Luca Cavalli-Sforza. Crow's talk was on the material in his well-known 1964 paper (with Kimura) on the number of alleles maintained in a finite population, as it happens received for publication on the same day as his seminar (January 3). This paper was inspired by the effectively infinite number of possible alleles at any site if an allele is taken as a DNA sequence of possibly 5,000 bases, and the neutral theory part of the paper was based on the thinking of the infinitely many alleles model of Malécot and Wright. The "neutral" part of that paper is well known. Crow also discussed the associated selective theory part of that paper, and I took my life in my hands and said that what Crow was doing was wrong, since he used Wright's stationary distribution formula (Equation (6) in the 1964 paper) where there was no concept of the stationary distribution for the frequency of any allele in the infinitely many alleles model. This part of the paper was never subsequently developed or referred to in the literature. Crow took no exception to this behavior and invited me to visit him in Madison later that year. As we walked back after the seminar Sam said that I appeared to have much to say for myself, and invited me to give the next few lectures in the regular Monday evening seminar series on population genetics which he ran jointly with Walter Bodmer. It was well known that, for the Wright-Fisher Markov chain model, the expressions for fixation probabilities in the selective case and mean fixation times in both the neutral and selective cases were extremely difficult if not impossible to calculate, and part of my PhD thesis consisted of finding the first-order corrections to the diffusion theory approximations for these quantities. This required a knowledge of the Feller boundary criteria for diffusion processes. When I discussed this in my first lecture in this seminar series Sam claimed that the boundary criterion I used for the model I discussed was wrong, and we had a heated argument about this. (I was very aggressive in those days.) However the next day Sam came to say that I was right. He actually appreciated this aggressive behavior and this, I think, led to a transforming (to me) lifelong close friendship, the memory of which I treasure to this day.

Moving forward, in Spring 1971 I was at the Department of Zoology, University of Texas at Austin, on sabbatical leave from La Trobe University, Australia. My host there was Ken Kojima, who was, tragically, to die later that year in a car accident. The neutral theory was very much under discussion at the time, and the empirical population geneticist Bob Selander told me that an important step in assessing the validity of the neutral theory was to find the gene frequency sampling formula under this theory. It was not obvious how to proceed, and after several false starts eventually I used the concept of a frequency spectrum (an expression which I introduced in my 1972 paper) and found that I could find this formula, for the infinitely many alleles model, provided that one assumption held. This was that in taking a sample of genes one by one, the probability that the next gene sampled would be of a new allelic type not so far seen in the sample was independent of the allelic composition of the sample so far, whatever that composition might be. (Chuck Langley, at Austin at the time, emphasized to me the importance of the "whatever" aspect of this assumption.) This assumption seemed unlikely to be true to all but one of the population geneticists to whom I send a draft of the paper describing my work. However I was quite convinced, on the basis of algebraic and other considerations, that it was true. Only Sam, then in Israel, trusted my judgement. He invited me to Stanford later in 1971 and after much discussion with me, he and Jim McGregor were able to prove that the assumption was true. Thus the sampling formula was proven.

Although I had submitted a paper on my work to *GENETICS*, Sam convinced me to withdraw that submission and submit the paper to *TPB*, where he would referee it and write an addendum with Jim McGregor proving the assumption referred to above. In effect the refereeing process of the paper consisted of one or two comments by Sam on points of presentation - none of that nonsense about a formal refereeing procedure! I was extremely happy that the eventual home of the paper was the then-new journal *TPB*, with which I had a long subsequent involvement as Associate Editor.

I did not follow up on testing for neutrality using the sampling formula with actual data since I felt that the infinitely many alleles assumption on which the formula was based was not appropriate for the then-prevalent electrophoretic data. The calculations also assumed a population of fixed size and that, in Markov chain terms, a stationary gene frequency situation had been reached, neither of which seemed realistic to me. Instead, for most of the next twenty years I mainly worked in human disease genetics. Two memorable experiences in that field stand out. The first was my work with my PhD student Nereda Shute on ascertainment sampling, and the second my association with Rich Spielman and Ralph McGinnis in the Medical School at Penn in formulating the transmission-disequilibrium test (TDT).

I did however keep in touch with colleagues in evolutionary genetics. The most important of these was Geoff Watterson. In 1975 Geoff made the crucial observation that the Poisson-Dirichlet distribution, which John Kingman had recently developed in the context of storage systems, was, quite serendipitously, central to population genetics theory. I shudder to think what would have happened if Geoff had not noticed that. I had already met Kingman and introduced him to current problems in population genetics where I felt that his mathematical powers would be needed. (Population genetics was not new to Kingman: as an undergraduate he had published papers on the question of the conditions under which the mean population fitness would increase.) Geoff's observation led to a second transforming relationship, namely a continuing three-way interaction between the three of us. It was simply astonishing to watch Kingman's powerful mind as, for example, he showed the relation between the sampling formula and the Poisson-Dirichlet distribution and as he formulated the concept of partition structures, now so important in probability theory. Most of all, of course, was the experience of reading a letter from him jointly to Geoff and me starting with something like: "I have developed and am sending you a new idea which I think will be useful in population genetics. I call it the coalescent."

It will be clear that I owe much to friends and colleagues. Jim Crow's generous character and depth of thought made a strong impression on me, and I treasure the visits I made to Madison to work with him. I have already mention the debt that I owe to Sam Karlin. More recently my association with Geoff Watterson, John Kingman, Peter Donnelly, Simon Tavaré and Bob Griffiths has been essential to my work. I have been lucky in my colleagues, my work, and not least in my association with *TPB*.