June 12, 1956

Hi Bruce, et al/

Just got your tape; here's an impromptu reply.

Still struggling with the ms. I've decided to follow you on terminology and forget about catenate. So I'm back to line (emph. uni-linear vs. multi-linear) for chain, and so forth. No new data at all, and I hope not to do any more experiments on this for a while. Only interesting thing gleaned out of my notes that I haven't already told you is a pedigree in which a sib to a swarm gave 22 lines. The boundary between E and non-E will have to be rather arbitrary. Still hope to have this in a 1956 Genetics. Haven't actually done any experiments myself on lines since last summer. Did I tell you then that motile lines isolated at an intermediate generation from Fla+ H_1^a —x SW666 were not all inhibited by anti-a a though all the initials were at least partially inhibited. Does this agree with you know (I hope I have that straight; don't have my notes here this PM). I burned my fingers with the i—x SW-666 experiments, as the anti-i serum also inhibited b—x SW666, much to my surprise. (This effect only on trails and isolated initials, not in agglutination or swarms; I hope you have such a control and have some idea what the reaction is. The serum of course is Colindale's).

Clive—

With n binary tests of course 2^n classes should be distinguishable, but Bruce is right that you ought to leave out 000000 and 111111 for confirmation. We used to rundown nutritional requirements of auxotrophs that way, but the trouble was that each unit reaction has to be perfect for the compound to be reliable, and I would be very leary of relying on Salmonella agglutinations so trustingly. I suppose the method should be ok for preliminary screening, but I feel you should have a redundancy of information in practical typing. When you serotype with individual, or with pooled reagents in discrete groups, you know that you must get a code like 00001000 and that 01000100 means something is wrong. One can, of course compromise between maximum information and maximum security, but I suspect that considerations like this have discouraged earlier applications. Do you think you can get away with a completely efficient code for Salmonella typing? By the way, a binary numbering scheme should make it easier to set up and to translate the efficient codes; as I suppose you've worked out in one form or another. Pool A should define the first digit (e.g., for 16 reagents should contain # 1-8), pool B the second (#1-4, 9-12) pool C the 3d (1-2, 5-6, 9-10, etc) and pool D the 4th (odd #'s). Then a code like 0101 is readily translated as #5, except I have it backwards and 0 = +, 1 = - reaction. I found this type of numbering very handy in summarizing genotypes; after translating to decimal, it's much easier to scan for how many type 7's than for how many -+++'s in a table full of + and -'s.

I'll be interested to hear how this works out in practice.