

Jack Schwartz (1930 –2009)

By: Michael Wigler

I met Jack Schwartz in 2001, just before he retired from NYU as founder and then chairman of the department of computer science at the Courant Institute of Applied Mathematics. He visited my lab on sabbatical for a period of two years and then subsequently as a visiting scientist, during which time he had a strong influence on me and my group as we turned increasingly to quantitatively and computationally complex methods. Jack was famous for his fundamental mathematical work on linear operators, computer architecture and language, robotics and proof verification. These notes chronicle a slice of his thinking post retirement about other subjects, and document Jack's far-ranging curiosity and creative analytical abilities, which traits endured far past the time he ceased to care about publishing.

Genomes as searchable strings: The parallels of an organism's genetic instructions to a programming language had great appeal to theoretical computer scientists (<http://wigler.cshl.edu/jack/Hopkins.pdf>). These genetic instructions, or the "genome", can be (incompletely) described as a string of characters drawn from a four nucleotide alphabet which ultimately instructs the cell how to make other macromolecules such as RNA and protein¹. The fundamental query of the genomic string is the "substring exact match", and Jack argued persuasively that fundamental to that was the "sort". This simple idea germinated in a young computer scientist named John Healy, who had joined the lab at about the same time as Jack, and both shared my office. John saw how to use a sort on a transformed string so that the original string could be searched rapidly for exact matches. Computational efficiency of genome searches is a significant problem when the genomes are 3×10^9 nucleotides long, the size of the human genome. Jack grasped John's difficult algorithm immediately, and was able to explain it to me. After researching his method, John discovered that he had not only re-invented the Burroughs-Wheeler algorithm (BWA) (Burrows, et al. 1994), created for string compression, but also reinvented how to index it for searches (Ferragina, et al, 2000). The algorithm became the basis for our microarray designs for several years the only paper on which Jack and I were co-authors (Healy et al. 2003). Today, the BWA is the core for many of the software programs used in DNA sequence analysis.

Phylogeny of Mitochondria: Beginning with the Human Genome Sequence Initiative there was a National Institutes of Health mandate to place DNA sequence data into the public domain. Jack loved data, especially free data. There was much that could be done with the deposited sequence data, but in particular Jack was interested in phylogeny, as it was natural to organize the genomes by descent. Most phylogenies are based on comparison of protein sequences and nucleic acid sequences, where the

distance between strings is measured roughly as the number of mutations, one nucleotide at a time, that convert one string into another. Jack devised different methods.

In the early 2000's he showed me his results based on study of the circular mitochondrial genome² (http://wigler.cshl.edu/jack/ibm_mitochondria.pdf). The mitochondrial genes are highly conserved, so conserved that using single nucleotide mutation is insufficient to establish a phylogeny. Jack observed that the order of the genes (coherent blocks of nucleotides) was not conserved, and suffered permutation during the long course of evolution. Furthermore, Jack saw that there was a minimizing "rule" to describe the permutations: excise a segment of genes from the circle, circularize that segment, and then reinsert that circle anywhere between the remaining genes such that the gene order in that circle is preserved or reversed. If reversed, there was also a reversal of the orientation of each gene³. There was a preference for reinsertion with reversal at the site of excision. Unknown to Jack at the time he made his observations, his rules obeyed precisely the expectations of recombination within a circular genome predicted from the resolution of Holliday structures. The minimum number of applications of Jack's rule to transform one genome into another was a measure of the evolutionary distance of the genomes.

Using his method, the genomes of vertebrates and those of crustaceans could be lumped into two groupings that were very clearly distinct but distantly related. The mitochondrial gene order of lobsters closely resembled mosquitoes. Hippos resembled alligators. There were minor variations off the vertebrate theme, or the crustacean theme, and these variations illustrated the minimal rule. But three types of crustaceans stood out: ticks, fleas and mites. These bore no obvious relation to each other or to the crustaceans. This surprised and delighted Jack, as he had thought that all the modern animals with exoskeletons were derived from the same explosion of metazoans. The work remained unpublished, although Jack spoke of it at various meetings on computational biology to which he was invited. With the assistance of Google, I learned of a much earlier paper that had examined mitochondrial gene order as a means to compute phylogeny, but that had used far fewer genomes (Sankoff et al., 1992). As far as I know, Jack's insight into the crustaceans was new, as was the precision he brought to the problem.

The Clocks of Evolution: Jack discovered other methods of phylogenetic mapping. One was based on the conservation of the intron-exon structure of transcripts¹. Overall, his various methods can be viewed as clocks that run at different time scales, as do the second, minute and hour hands of a clock. One hand might be useful for measuring fast events, and the other for slow events. What would it say for our confidence in phylogeny if these clocks depicted different trees? Or if the branches were of different duration, depending on the clock used to measure it? The synthesis of these ideas into a coherent theory never occurred under Jack's watch. He was more interested in the thrill of a logically solid novel insight than in exhausting its

implications. But these ideas will eventually be useful in understanding cancer evolution, where the rates of different types of genome instability (point mutation, rearrangements, and changes in copy number) become uncoupled, and each can be measured relative to each other, and to the absolute rate of mutation in normal cells. The rates of these clocks in individual tumors are likely to predict patient survival (see for example Hicks, et al., 2006).

Illusions: Jack had a keen interest in human cognition, especially vision⁴. This choice was dictated by several considerations: Jack was gifted in visualization, he had mastery of the design of computer images, and he had himself as a willing subject. The five areas he explored were illusions of motion, distortion in parallelism, and depth perception, as well as the ability to distinguish textures and perceive objects against various backgrounds. He was especially adept at crafting three dimensional illusions, provided the user wore red-green colored glasses. He found how to make a static object on the computer screen glisten with metallic sheen by separately controlling left and right visual inputs. In all these instances he was able to use his command of the computed image to explore how the illusions fared under manipulation of the underlying parameters (such as contrast, graininess or rotation). He had little difficulty recruiting other willing subjects, and used this opportunity to probe human visual capability. To his surprise, he found some robust differences in the human response. He never took this to the level of exploring patho-physiology (for example testing patients with strokes to the occipital lobe) or genetic variation. This would have required far greater resources. To my knowledge Jack never published his methods or observations. Overall, Jack believed that vision was a composition of separate hardwired visual input processing algorithms, with a discrete structure not apparent to the user.

Memory: The organization of memories was a frequent subject of our discussions, although neither of us knew neurobiology. Jack would of course make use of his knowledge of computer architecture and algorithms to reason about the connectivity of the brain. A few of his observations had strong resonance with me. Here is the most important one. Human memory has a feature, as he put it, similar to “filename completion”. If a Unix user specifies part of a filename, the rest of the filename is suggested by the computer system. Jack was in fact working on algorithms at the time that did much more than this: produce a list of candidate names given only a part of the last name of the person, perhaps the street name of their address, and perhaps a fragment of their home phone number (such as an area code plus a few digits). How would you organize your contact databases to execute such a task? Humans cannot do this particular task very well, but we do respond very well in similar situations, like to this tickler: “Who is the young actress, very pretty, the heroine in one of Woody Allen’s murder mysteries and the victim in another?” Jack thought this was a fundamental property of memory. In fact, it is easier to “imagine” how to set up such

memory out of “brain stuff” than out of programmable silicon circuits. First, the “clues” to the full memory are disseminated broadly throughout the brain, recruiting various neurons⁵ reactive to the clues. A pre-existent but dormant original circuit, created by prior experience, connects a subset of these recruited neurons along with many others. This circuit is prone to be activated by even a few of its members, including those activated by the dissemination of the original tickler. The “name” or missing parts of the full memory reside in the full circuit. In other words, a neural circuit that can keep a set of neurons stably stimulated and that be driven into that stable firing state by activation of only a few of its participants may be sufficient for associative memory. The creation and management of such highly connected circuits does not have parallels in common computer architecture. Jack chose not to think too deeply about this, because his years as Director at DARPA/ISTO funding research into neural nets left an indelibly negative impression. He was also keenly aware of the gulf between how Nature conducts herself and how humans imagine she does.

The End Game: While Jack loved Cold Spring Harbor, with its invigorating empirical tradition and its natural beauty, his sabbatical ended upon his first serious illness. Once he recognized that his time was limited, he returned home to the city, teeming with the ethnic restaurants he adored, to the project on the application of computational logic to the foundations of mathematics. He maintained his contacts with his colleagues, and continued to visit from time to time, as I did him, often bringing vexing problems or new results. We discussed a myriad of subjects: cellular architecture, recombinant DNA, cloning, principles of experimental design, olfaction, transcriptional networks, defenses to bio-terrorism, cancer biology, Galois theory, the nature of proofs, the relationship of science to mathematics, the limits of the latter, computer architecture, programming languages, principles of algorithm design, and the uses of computers in musical annotation⁶ among others.

Jack held that a mathematician’s emotional age, often measured in single digits or low teens, was set by the year in which he discovered his intellectual gifts. But Jack escaped this tyranny of arrested development. His intellect was matched by a graceful emotional maturity and an almost serene detachment. A grumpy shrug was the most he would concede to a temporary setback. He drew his ability to grow from his curiosity, his autonomous powers for discovery, and the joy he took in sharing his thoughts and findings with others. By such means, and with the support of a loving wife and family, and highly appreciative friends, he traversed a unique and wonderful life. He died in 2009 of the complications of cancer. At Jack’s memorial service, his granddaughter, Adrienne Fainman, related this. On being frustrated in her studies of math, Jack told her, “Be happy when you are confused, because you are about to learn something”.

ACKNOWLEDGMENTS

I thank Diana Robinson Schwartz for her help with Jack’s archives, for stimulating my memory, and for her patience. She and Bud Mishra helped with

editorial comments and Dennis Sullivan and Jeff Cheeger straightened my mathematics.

FOOTNOTES

¹Genes are substrings of DNA. DNA is double stranded, so each element of the genome “string” is really a complementary pair of nucleotides. Messenger or coding RNA is made by transcribing one strand of the DNA as template. Parts of the RNA are edited, and the sections of RNA removed are called “introns”, leaving intact but joining together the “exons”. Only then is the RNA is used to make protein.

²Mitochondria are the tiny intracellular “engines” within every eukaryotic cell that produce ATP, the energy currency. They have their own genomes.

³The gene unit itself, being a string, has an orientation.

⁴Some of Jack’s creations can be found at

http://www.settheory.com/Glass_paper/cafe_wall_study.html

<http://www.multimedialibrary.com/Articles/Jack/Illusions/index.asp>

<http://www.multimedialibrary.com/education/illusions/index.asp>

http://www.settheory.com/hole_illusion/dill.html

⁵Roughly speaking, a neuron can be resting and have a low rate of firing, or be recruited into an active state with a high rate of firing.

⁶See for example

http://www.settheory.com/drum_machine/drums.html

REFERENCES

Burrows, M. and Wheeler, D.J. (1994). A block sorting lossless data compression algorithm. Technical Report 124, Digital Equipment Corporation, Palo Alto, CA.

Ferragina, P. and Manzini, G. (2000). Opportunistic data structures with applications. In *41st IEEE Symposium on Foundations of Computer Science*, pp. 390–3981.

Healy, J., Thomas, E., Schwartz, J.T. and Wigler, M. (2003). Annotating large genomes with exact word matches. *Genome Research* 13: 2306-2315.

Hicks, J., Krasnitz, A., Lakshmi, B., Navin, N., Riggs, M., Leib, E., Esposito, D., Alexander, J., Troge, J., Grubor, V., Yoon, S., Wigler, M., Ye, K., Børresen-Dale, A-L., Naume, B., Schlicting, E., Norton, L., Hagerstrom, T., Skoog, L., Auer, G., Maner, S., Lundin, P., and Zetterberg, A. (2006). Novel Patterns of genomic rearrangement and their association with survival in breast cancer. *Genome Research* 16:1465–1479.

Sankoff, D., Leduc, G., Antoine, N., Paquin, B., Lang, B. F. and Cedergren, R. (1992). Gene order comparisons for phylogenetic inference: evolution of the mitochondrial genome. *Proc. Natl. Acad. Sci. USA*. *Proc Natl Acad Sci U S A*. 89(14): 6575–6579.