

# David Sankoff's projects: a biased sample of the first 25 years



Vieques, Puerto Rico

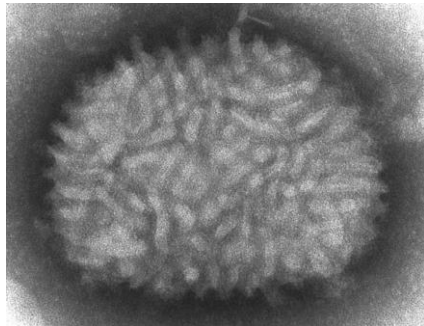
# 1963

## Radiobiological Studies of Vaccinia Virus<sup>1</sup>

J. D. FRIESEN,<sup>2</sup> D. SANKOFF, AND L. SIMINOVITCH

*Department of Medical Biophysics, University of Toronto, and The Ontario Cancer Institute, Toronto*

*Accepted July 11, 1963*



Vaccinia virus is well known for its role as a vaccine (its namesake) that eradicated the smallpox disease, making it the first human disease to be **successfully eradicated** by science.

# 1963



**SANKOFF, DAVID  
LAWRENCE**

"You don't like the  
country? Go hang  
yourself!"

—*Michel Picard*

Activities: Editor, Arts  
and Science News;  
Physics Society; Ma-  
thematics Society; So-  
ciology and Anthropo-  
logy Society; Film So-  
ciety; **Fencing** Club.

# 1967



"Working for the Bureau of Statistics of Papua and New Guinea in 1967, I carried out a territory-wide population projection, based on a complete collection of village patrol reports and some birth records from a small sample of hospitals."

David Sankoff, *How to predict the evolution of a bilingual community*, in *Social lives in languages* (2008).

# 1969

HISTORICAL LINGUISTICS AS  
STOCHASTIC PROCESS

by

David Sankoff

On page 110:

quite feasible. Accordingly, I started a project of extracting all relevant information in Buck's book and transferring it to punch cards. First, I set up the following guidelines:

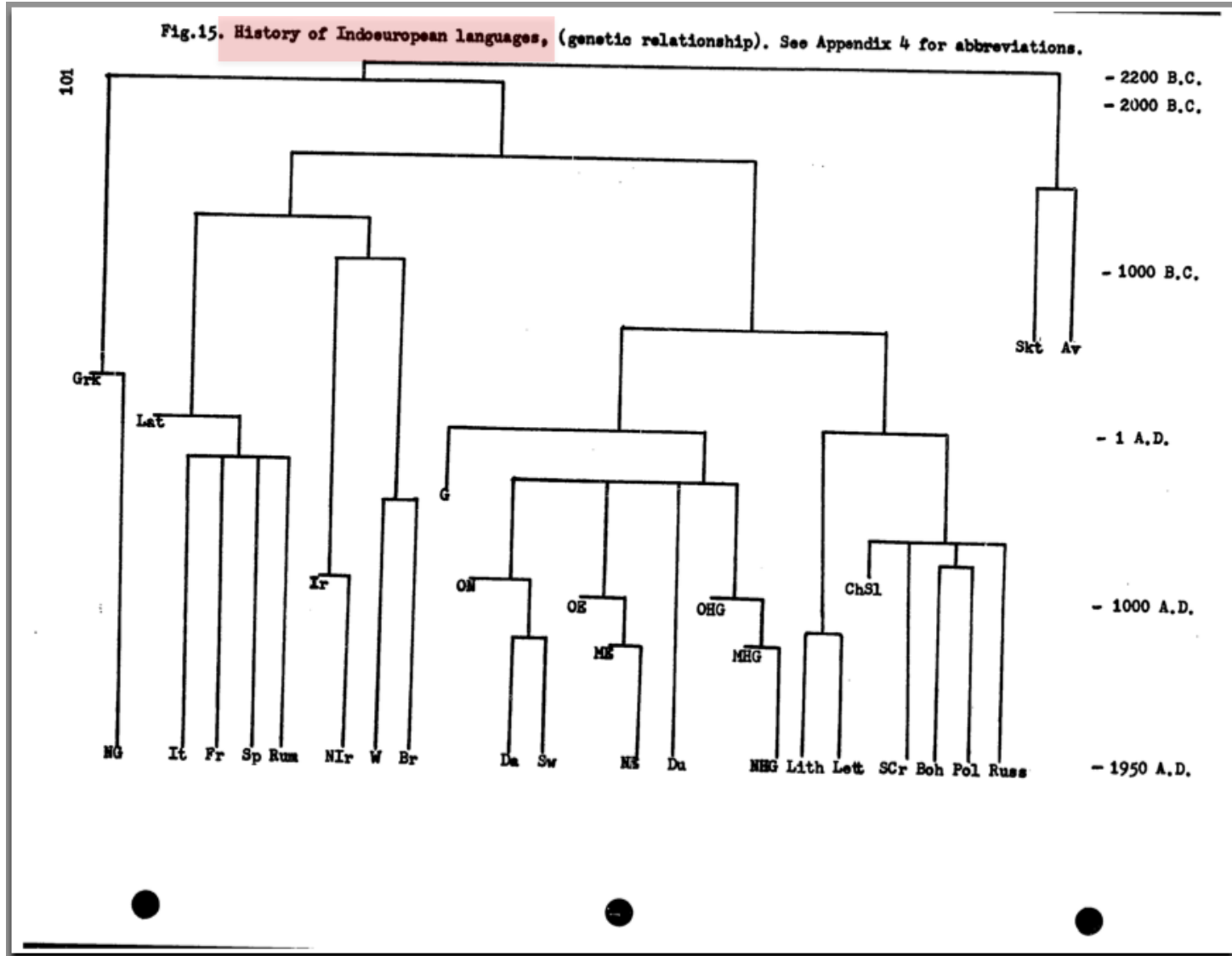
A thesis submitted to the Faculty of  
Graduate Studies and Research in  
partial fulfillment of the requirements  
for the degree of Doctor of Philosophy

McGill University

Montreal

July, 1969

# 1969





# 1969



# 1972

*Proc. Nat. Acad. Sci. USA*  
Vol. 69, No. 1, pp. 4-6, January 1972

## Matching Sequences under Deletion/Insertion Constraints

(algorithm/genetic homology/nucleotide sequence/amino-acid sequence)

DAVID SANKOFF

Centre de recherches mathématiques, Université de Montréal, c.p. 6128, Montréal 101

*Communicated by Mark Kac, October 18, 1971*

**ABSTRACT** Given two finite sequences, we wish to find the longest common subsequences satisfying certain deletion/insertion constraints. Consider two successive terms in the desired subsequence. The distance between their positions must be the same in the two original sequences for all but a limited number of such pairs of successive terms. Needleman and Wunsch gave an algorithm for finding longest common subsequences without constraints. This is improved from the viewpoint of computational economy. An economical algorithm is then elaborated for finding subsequences satisfying deletion/insertion constraints. This result is useful in the study of genetic homology based on nucleotide or amino-acid sequences.

and  $\{b_i\}_1^n$  we define a matrix  $V$  where  $V(i,j)$  is the highest value possible for a path to  $(i,j)$ . Making the convention

$$V(i,0) = V(0,j) = 0 \quad (2)$$

for all  $i \in \{0, \dots, m\}$ ,  $j \in \{0, \dots, n\}$ ,  
we have

$$V(i,j) = \max_{\substack{h \leq i-1 \\ k \leq j-1}} V(h,k) + \delta(a_i, b_j) \quad (3)$$

since a highest value path to  $(i,j)$  must contain  $(i,j)$ , as well as



# 1973

NATURE NEW BIOLOGY VOL. 245 OCTOBER 24 1973

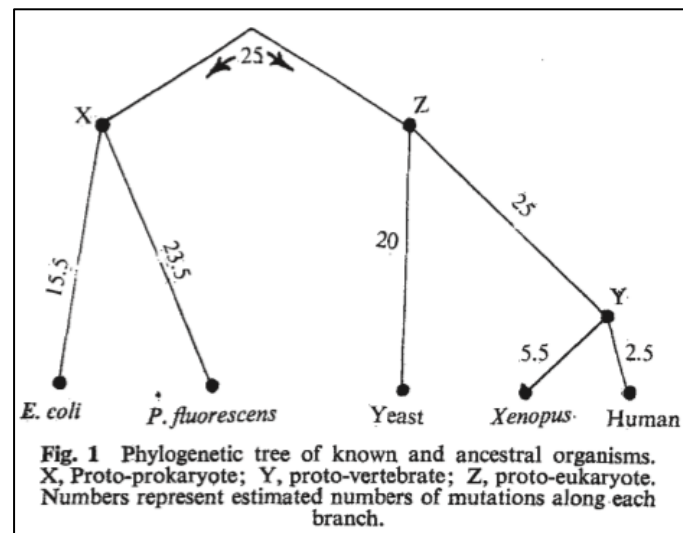
## Evolution of 5S RNA and the Non-randomness of Base Replacement

This work was supported in part by le Ministère de l'Éducation du Québec.

DAVID SANKOFF  
CRISTIANE MOREL  
ROBERT J. CEDERGREN

EC	UGC	CUGGCGGCCG	UAGC	GCGGUGGUCCACCUGACC	CCAUGCC
PP	UG <sup>?</sup>	CU GCGGCCA	UAGCAGC <sup>?</sup>	UUGGAA <sup>?</sup>	CACCUGAUC CCAU <sup>?</sup> CC
PF	UGUUCU	UUGACGAGUAGUAGCAUUGGAA	CACCUGAUC	CCAUC	CC
PE	G <sup>?</sup>	CU GCGGCCA	UACCA <sup>?</sup> C	UUGAAAGCACC <sup>?</sup>	GAUCUCCGU CC
YT	GG	UU GCGGCCA	UACCAUC	UAGAAAGCACC	GUUCUCCGU CC
KB	GU	CU ACGGCCA	UACCACC	CUGAACGCGCCCGAUCU	CGU CU
PV	G <sup>?</sup>	CU ACGGCCA	UACCACC	CUGAAAGCGCCCGAUCU	CGU CU
XL	GC	CU ACGGCCA	CACCACC	CUGAAAGUGCCCGAUCU	CGU CU
		10	20	30	40

$$d(i, j, k) = \min \left\{ \begin{array}{l} d(i-1, j-1, k-1) \begin{cases} +0 \text{ if } a_i = b_j = c_k \\ +1 \text{ if } a_i = b_j \neq c_k, a_i = c_k \neq b_j \\ \text{or } b_j = c_k \neq a_i \\ +2 \text{ if } a_i \neq b_j \neq c_k \neq a_i \end{cases} \\ d(i-1, j-1, k) \begin{cases} +1 \text{ if } a_i = b_j \\ +2 \text{ if } a_i \neq b_j \end{cases} \\ d(i-1, j, k-1) \begin{cases} +1 \text{ if } a_i = c_k \\ +2 \text{ if } a_i \neq c_k \end{cases} \\ d(i, j-1, k-1) \begin{cases} +1 \text{ if } b_j = c_k \\ +2 \text{ if } b_j \neq c_k \end{cases} \\ d(i, j, k-1) \quad +1 \\ d(i, j-1, k) \quad +1 \\ d(i-1, j, k) \quad +1 \end{array} \right.$$



# 1973

NATURE NEW BIOLOGY VOL. 245 OCTOBER 24 1973

## Evolution of 5S RNA and the Non-randomness of Base Replacement

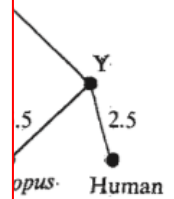
This work was supported in part by le Ministère de l'Éducation du Québec.

DAVID SANKOFF  
CRISTIANE MOREL  
ROBERT J. CEDERGREN

EC	UGC	CUGGCGGCCG	UAGC	GCGGUGGUCCCACCUGACC	CCAUGCC		
PP	UG?	CU	CCCCCA	UAGCAGC?	UUGGAA?	CACCUGAUC	CCAUGCC
PF	UGUUCU						
PE	G?	CU					
YT	GG	UU					
KB	GU	CU					
PV	G?	CU					
XL	GC	CU					

All of our trade tools are there... Comparative genomics, as we know it, will soon turn 40!

$$d(i, j, k) = \begin{cases} d(i-1, j, k-1) & +2 \text{ if } a_i \neq c_k \\ d(i, j-1, k-1) & +1 \text{ if } b_j = c_k \\ & +2 \text{ if } b_j \neq c_k \\ d(i, j, k-1) & +1 \\ d(i, j-1, k) & +1 \\ d(i-1, j, k) & +1 \end{cases}$$

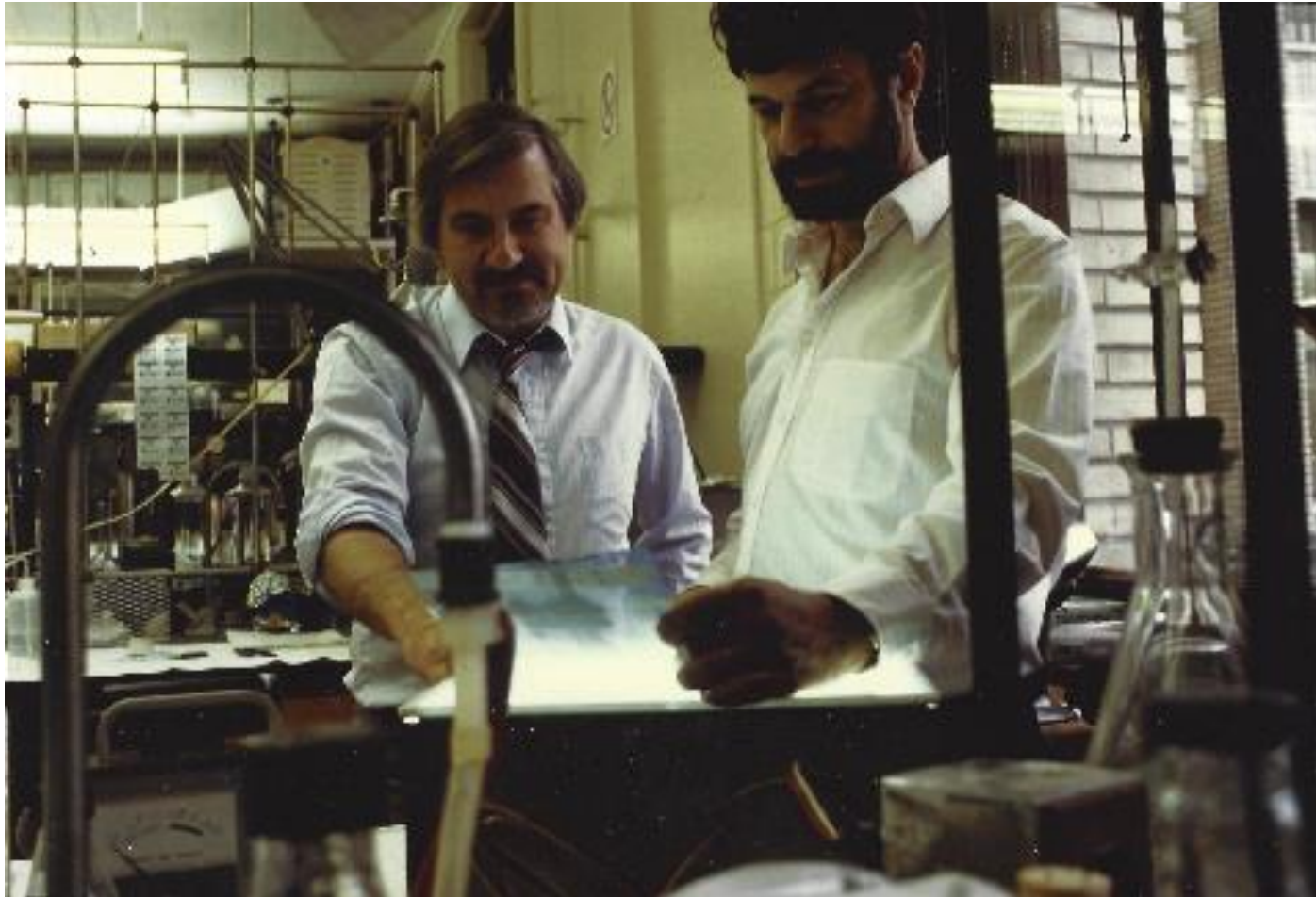


Numbers represent estimated numbers of mutations along each branch.

organisms. to-eukaryote.

# 1977?

Robert J. Cedergren and David Sankoff (reading gels?)



# 1974

## Variable rules: Performance as a statistical reflection of competence

HJ Cedergren, D Sankoff - Language, 1974 - JSTOR

Speech performances are here considered as **statistical** samples drawn from a probabilistic language **competence**. This **competence** is modeled in conventional generative terms, except that optional **rules** are assigned application probabilities as functions of the ...

[Cited by 652](#) [Related articles](#) [All 4 versions](#) [Cite](#)

VARBRUL: Software implementing a key methodological tool of variationist sociolinguistics: the variable rule program. Still in use after almost 40 years!

Example

[1] Est-ce vrai?	<i>(Is it true?)</i>
[2] C'est-tu vrai?	

# 1981

## A FORMAL GRAMMAR FOR CODE-SWITCHING <sup>1</sup>

David Sankoff

Centre de Recherches Mathématiques

Université de Montréal

and

Shana Poplack

Center for Puerto Rican Studies, CUNY

### ABSTRACT

Code-switching in situations of language contact has been studied largely from the point of view of its social determinants. This paper will propose formal means for describing the syntax of code-switching with examples from Puerto Rican Spanish and English.

Example

Pèse sur la *clutch*, mets-toi au neutre,  
*stoppe*, embraye...

# 1988?



During the 1980's, biochemist Robert Cedergren, seconded by David Sankoff, formed the first research group in bioinformatics at the Université de Montréal.



# 1984

Meeting at Fondation des Treilles, Tourtour, France,  
(photo borrowed from the website of Mike Waterman)



Gene Myers and  
David Sankoff

# 1990?

Meeting at Fondation des Treilles, Tourtour, France,  
(photo borrowed from the website of Mike Waterman)



Gene Myers and  
David Sankoff

# 1987

## "Query-by-Humming"

The Mongeau-Sankoff Algorithm for the comparison of musical sequences.



Dmitri Shostakovich – Waltz No. 2