Probabilistic walks with Graeme

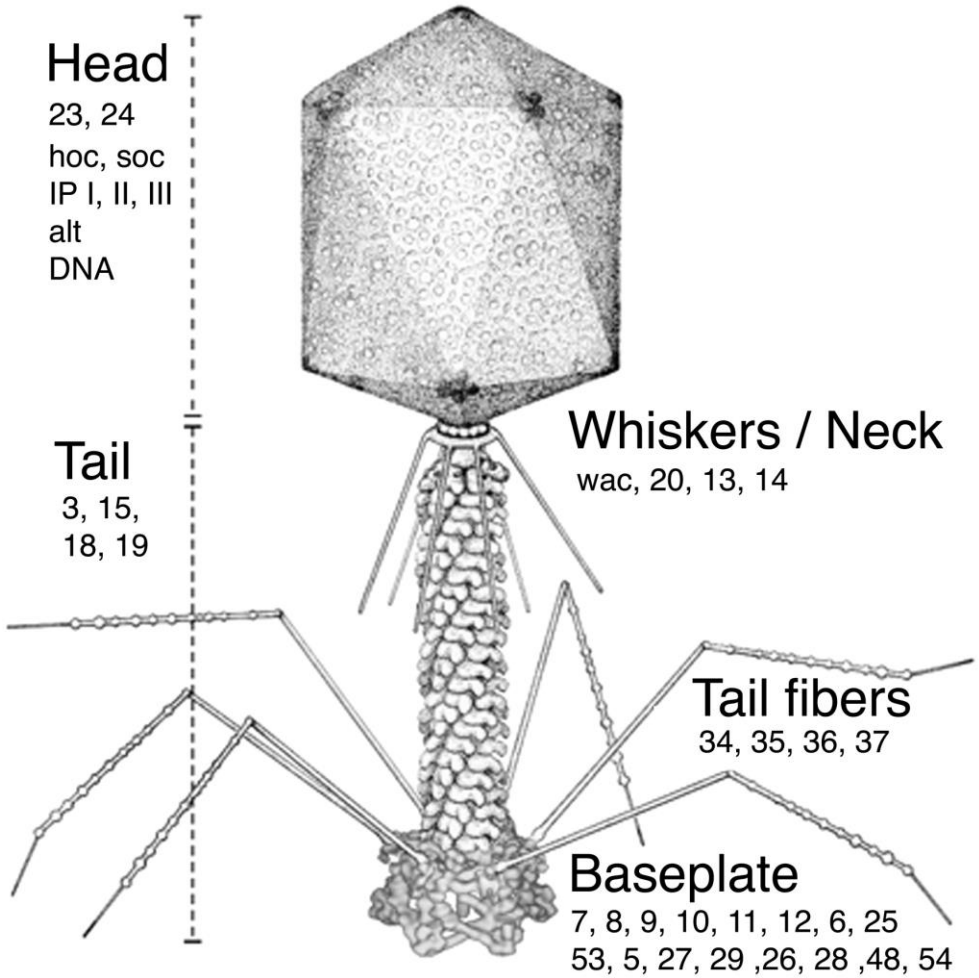
Sean Eddy
Molecular & Cellular Biology, and Applied Mathematics
HHMI and Harvard University
eddylab.org

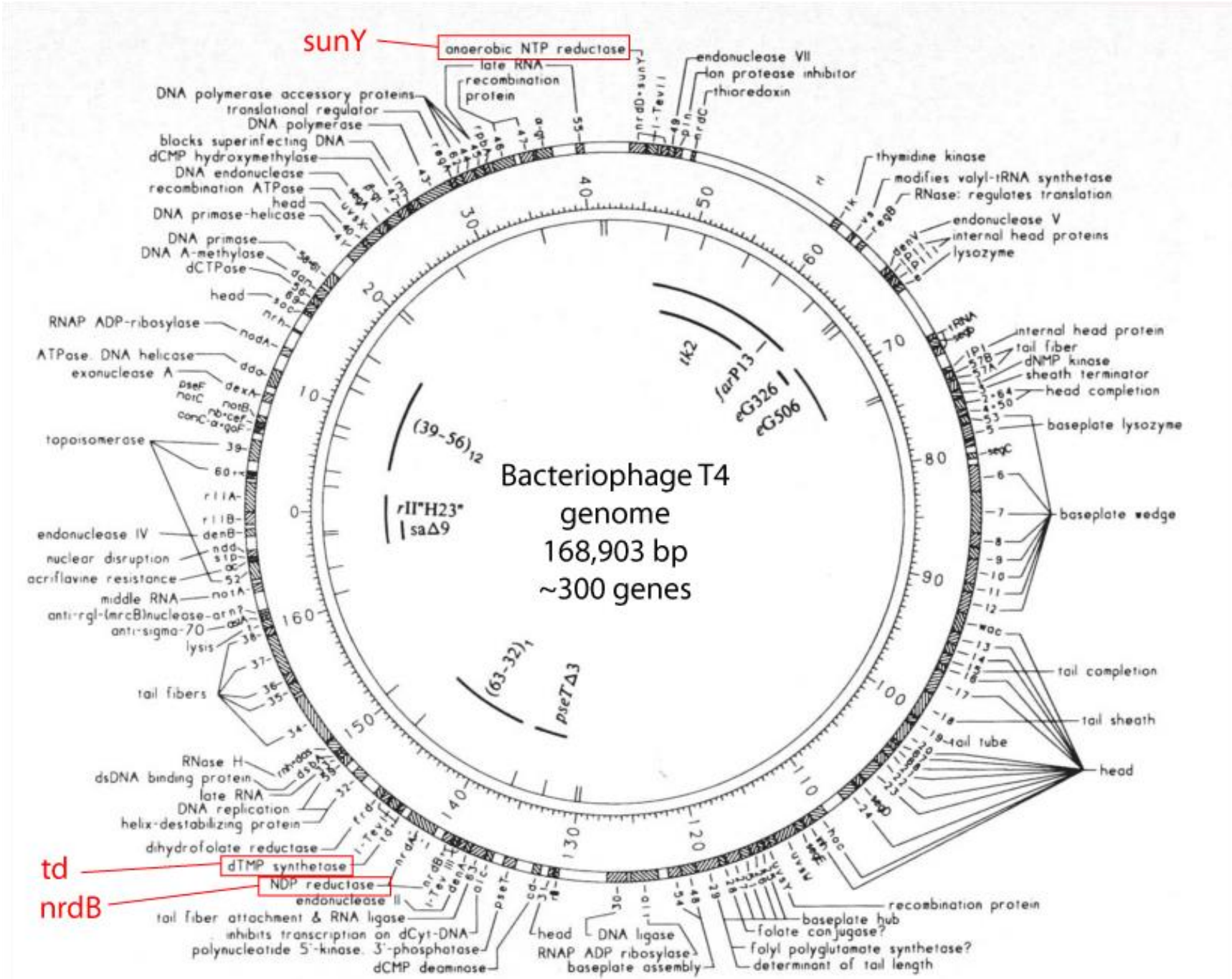

Cambridge Celts FC
The Perse Upper School
A1134
High St
Hills Rd
Queen Edith's Way
RED CROSS
A1307
Fendon Rd
Nightingale Recreation Ground
Long Road Sixth Form College
Cambridge Academy for Science and Technology
Fawcett Primary School
Cancer Research UK Cambridge Institute
Cambridge Blood Donor Centre
The Green Man
CPDC Cambridge Professional...
MRC Laboratory of Molecular Biology
The University of Cambridge School of...
Addenbrooke's Hospital
Cambridge Biomedical Campus
Bell Cambridge
Hudson's Ale House
Church Ln
Hobson's Brook
The Rosie Hospital
Babraham Rd
The Bun Shop (Trumpington Post...
Maris Ln
Cambridge Institute of Public Health
Hauxton Rd
Abcam

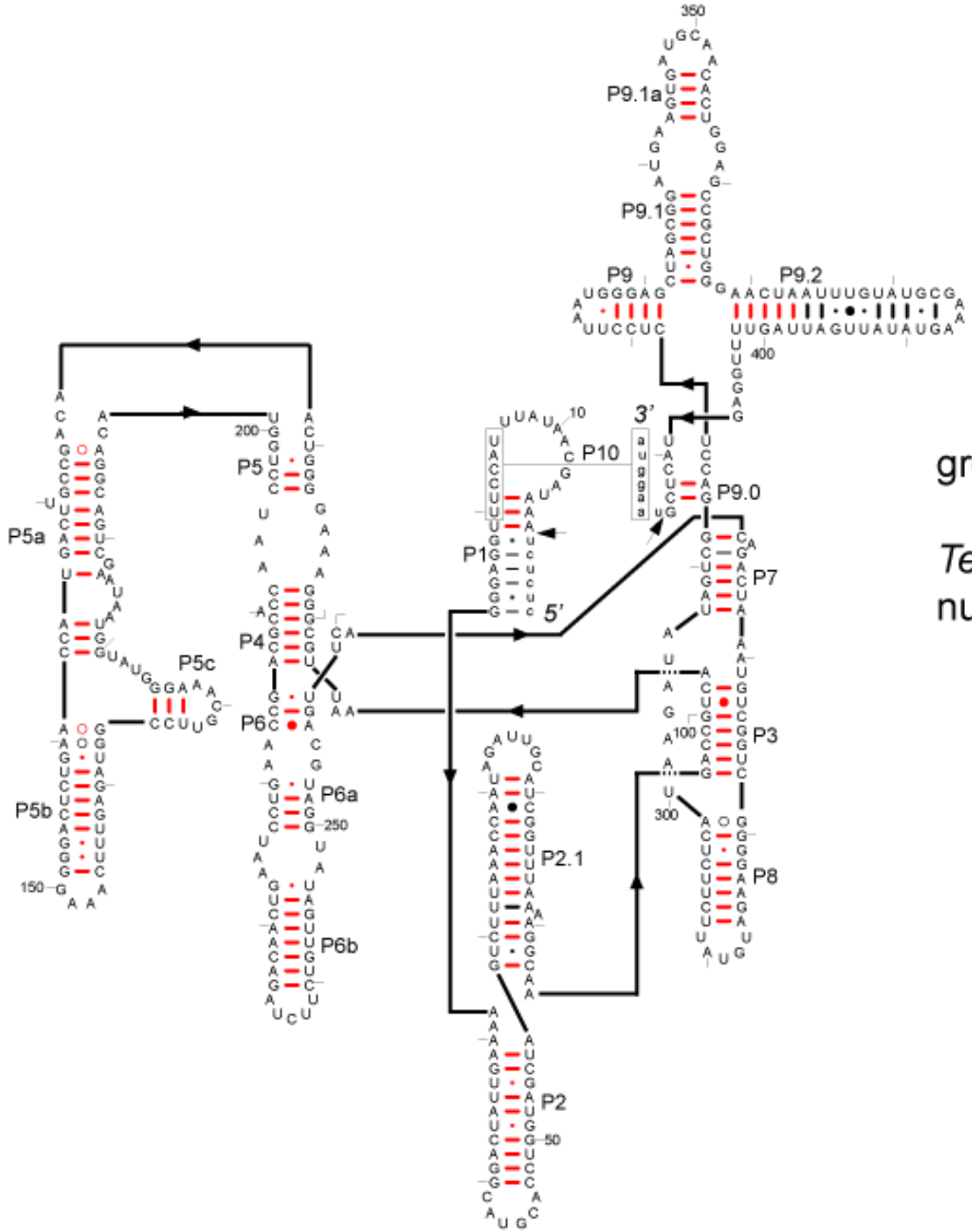
Cell, Vol. 47, 81-87, October 10, 1986, Copyright © 1986 by Cell Press

Multiple Self-Splicing Introns in Bacteriophage T4: Evidence from Autocatalytic GTP Labeling of RNA In Vitro

Jonatha M. Gott,* David A. Shub,*
and Marlene Belfort †
* Department of Biological Sciences
State University of New York, Albany
Albany, New York 12222
†Wadsworth Center for Laboratories and Research
New York State Department of Health
Albany, New York 12201




sunY
anaerobic NTP reductase
Bacteriophage T4
genome
168,903 bp
~300 genes
td
nrdB
dTMP synthetase
NDP reductase
endonuclease II
late RNA
recombination protein
DNA polymerase accessory proteins
translational regulator
DNA polymerase
blocks superinfecting DNA
dCMP hydroxymethylase
DNA endonuclease
recombination ATPase
head
DNA primase-helicase
DNA primase
DNA A-methylase
dCTPase
head
RNAP ADP-ribosylase
ATPase. DNA helicase
exonuclease A
topoisomerase
endonuclease IV
nuclear disruption
acriflavine resistance
middle RNA
anti-rgI-(mrcB)nuclease-anti-sigma-70
lysis
tail fibers
RNase H
dsDNA binding protein
late RNA
DNA replication
helix-destabilizing protein
dihydrofolate reductase
tail fiber attachment & RNA ligase
inhibits transcription on dCyt-DNA
polynucleotide 5'-kinase. 3'-phosphatase
dCMP deaminase
head
DNA ligase
RNAP ADP ribosylase
baseplate assembly
endonuclease VII
ion protease inhibitor
thioredoxin
thymidine kinase
modifies valyl-tRNA synthetase
RNase: regulates translation
endonuclease V
internal head proteins
lysozyme
internal head protein
tail fiber
dNMP kinase
sheath terminator
head completion
baseplate lysozyme
baseplate wedge
tail completion
tail sheath
tail tube
head
recombination protein
baseplate hub
folate conjugase?
folyl polyglutamate synthetase?
determinant of tail length
(39-56)12
rII*H23*
saΔ9
(63-32)1
pseTΔ3
1k2
farP13
eG326
eG506



group I self-splicing intron

Tetrahymena thermophila
nuclear LSU rRNA

source: Robin Gutell
Comparative RNA Website
http://www.rna.icmb.utexas.edu/

Hidden Markov Models in Computational Biology: Applications to Protein Modeling UCSC-CRL-93-32

Anders Krogh*†, Michael Brown†, I. Saira Mian§,
Kimmen Sjölander†, David Haussler†

† Computer and Information Sciences
§ Sinsheimer Laboratories
University of California, Santa Cruz, CA 95064, USA.

email: krogh@nordig.ei.dth.dk, haussler@cse.ucsc.edu

August 17, 1993

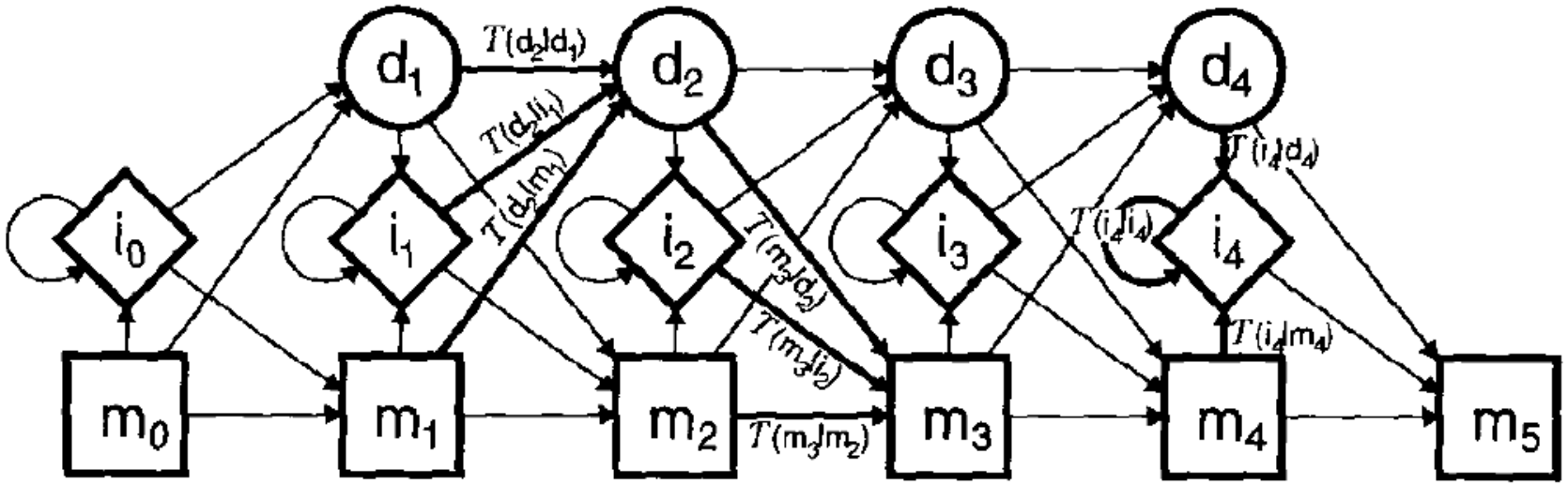
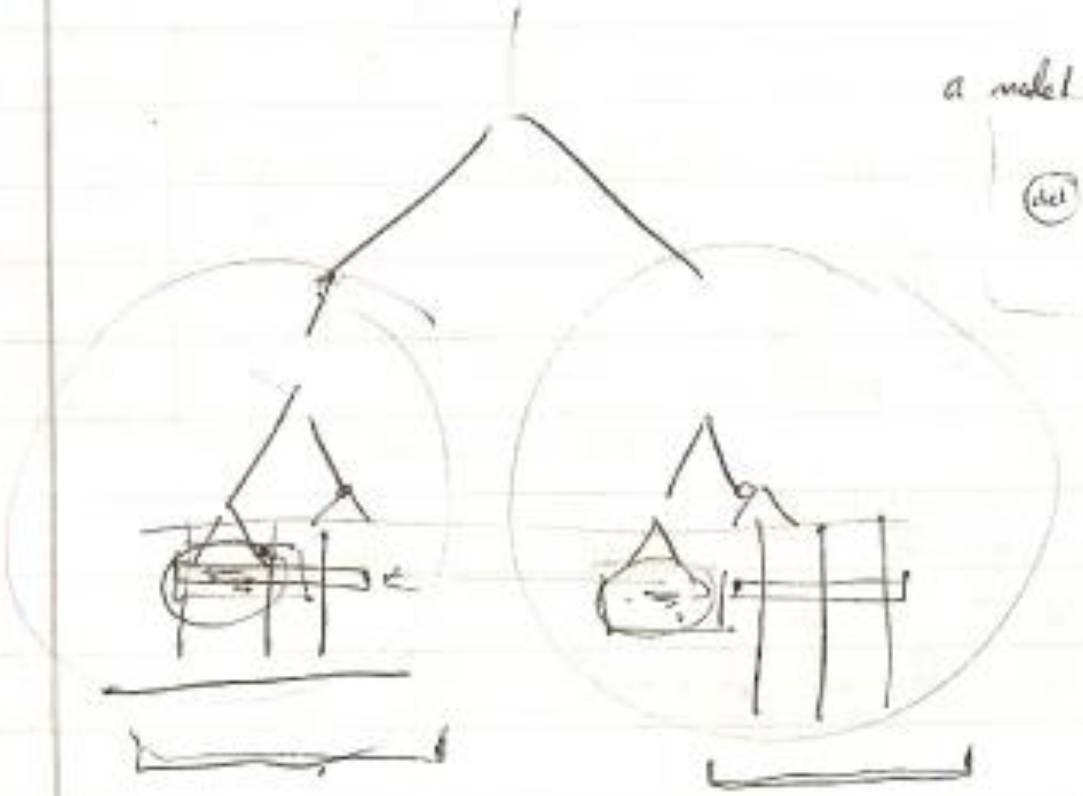


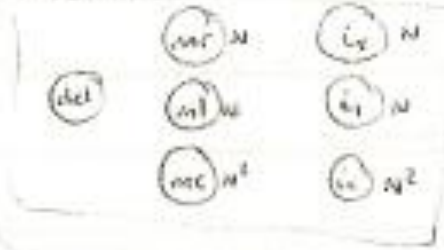
Figure 1. The model.

Attempt to repeat some analytic method that is considered unreliable and difficult until patience and hard work yield results similar to those published by the author. Pleasure derived from success, especially if it has come without the supervision of an instructor (that is, working alone), is a clear indication of aptitude for experimental work.

Santiago Ramon y Cajal (1916)



a model is:



100 Ps/state

7 substates.

49 transition Ps.

& 8 × 49 emission Ps

tRNA val has 34 ss

21 pairs

∴ needs 55 states

= 5500 Ps.

16s rRNA 1600 nt 600

1600 pairs = 800

1100 ~~states~~ states

experiment 1. construct a tRNA model by hand
use it to search sequences.

needs 7 × 55 = 385 states

× 70

100

× 100

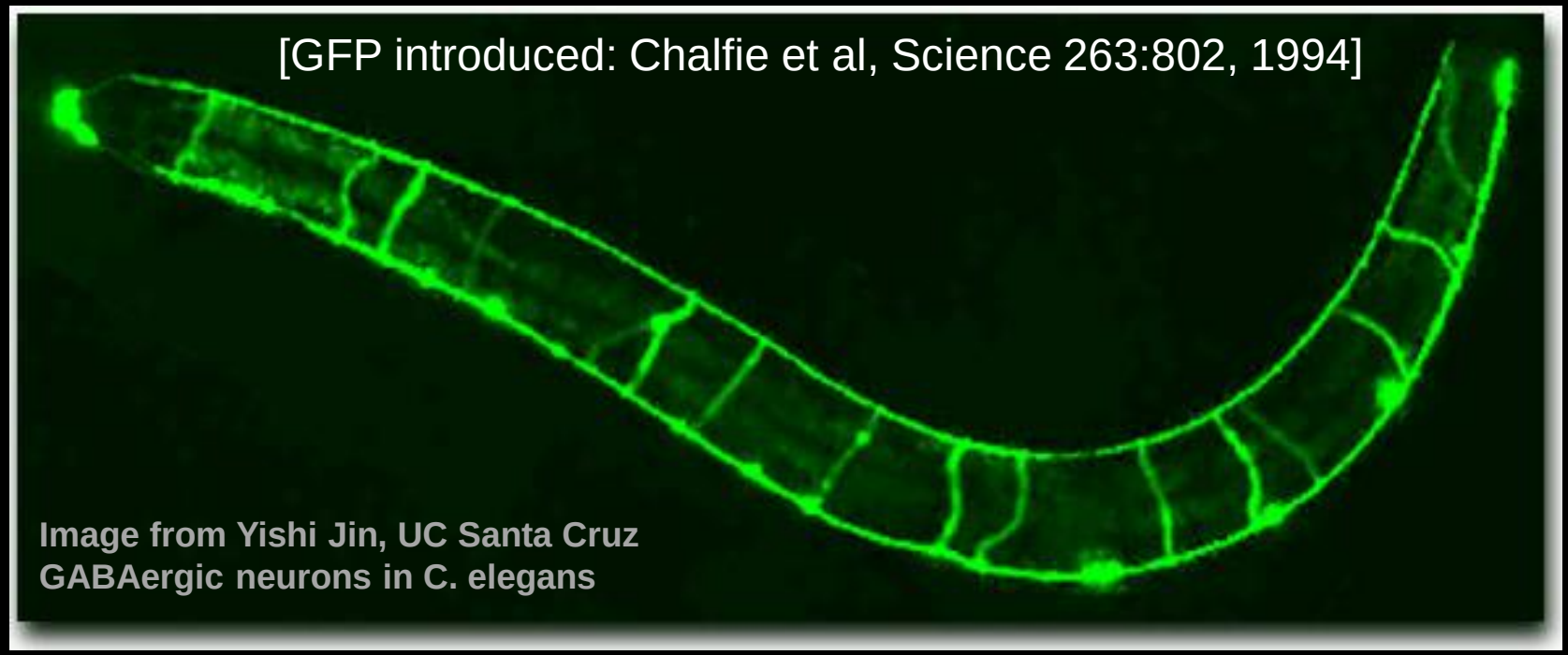
3.8 Mb N^3 in g

April, 1993

I had proposed to develop reporter fusions to neural-specific promoters as a tool to visualize axonal processes in live animals, facilitating genetic screens...

I have started to play with... green fluorescent protein (GFP) from the jellyfish...

I found out at the worm meeting in June that Marty Chalfie's lab is onto the same idea. Chalfie has already obtained bright fluorescence in the axons of the six touch neurons...



Somewhat embarrassingly, my most productive work has been unrelated to the original proposal, resulting from some moonlighting as a computational biologist... and a lingering interest in RNA structure from my thesis work.... I invented a new kind of statistical model, related to HMMs, which can model the two-dimensional structure consensus of RNAs. - **progress report for my postdoc grant, 1993**


Lyttelton
Playing Fields
HAMPSTEAD
GARDEN SUBURB
A598
Golders Green
Crematorium
Hampstead Golf Club
Hampstead
Heath
Extension
Finchley Rd
EEN
Highgat

From sequence to RNA structure analysis

Goal	HMM algorithm (sequence)	SCFG algorithm (RNA structure)
optimal alignment	Viterbi	CYK
P(sequence | model)	Forward	Inside
EM parameter estimation	Forward-Backward	Inside-Outside
memory complexity:	$O(MN)$	$O(MN^2)$
time complexity (general):	$O(M^2N)$	$O(M^3N^3)$
time complexity (as used):	$O(MN)$	$O(MN^3)$

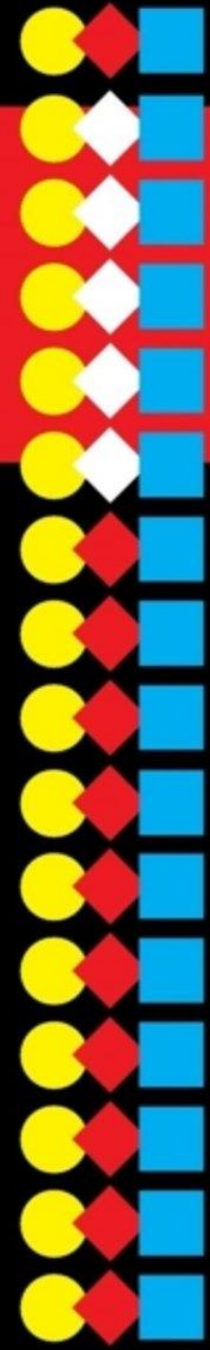
- we can analyze target sequences with secondary structure models;
- but the algorithms are computationally expensive**.**

Biological sequence analysis

Probabilistic models of proteins and nucleic acids

R. Durbin
S. Eddy
A. Krogh
G. Mitchison

CAMBRIDGE



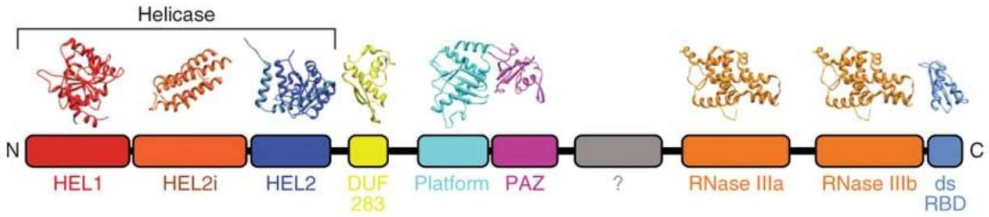

Carlton Way
Milton Rd
Cambridge North
The Plough
Arbury Rd
Green End Rd
Gilbert Rd
Milton Rd
River Cam
Fen Ditton
Ditton Meadows
B1047
Scotland Rd
High St
A1134
Gilbert Rd
A1134
High St
Stourbridge Common
CHESTERTON
Cambridge Museum Of Technology
River Cam
A1134
Cambridge City Cemetery
Newmarket Road Park & Ride
McDonald's
A1134
A1303
Cambridge United
Tesco Superstore
B&Q
Marshall FordStore Cambridge
Jesus Green
Victoria Ave
Currys PC World Featuring Carphone...
Marshall Aerospace and Defence Group
A1303
A1303
Midsummer Common
A1134
Coldham's Brook
Homebase - Cambridge
The Round Church
A1134
Newmarket Rd
Cambridge International Airport
Christ's College Cambridge
TK Maxx
The Grafton Centre
Coldhams Common
Primark
Asda Cambridge Superstore
Christ's Pieces
A603
A1134
Cambridge
Anglia Ruskin University
Coldhams Ln
Mill Road Cemetery
Sedgwick Museum of Earth Sciences

Software tools for homology search and alignment

HMMER

protein homology search: profile HMMs
http://hmmer.org

55K lines source code
>40,000 downloads/yr

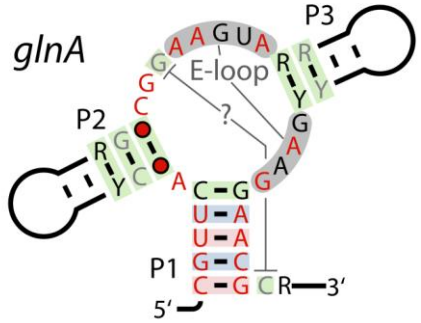


human Dicer | Lau et al, Nature Struct Mol Biol 19:436 (2012)

Infernal

RNA homology search: profile SCFGs
http://eddylab.org/infernal

90K lines source code
>10,000 downloads/yr



glnA glutamine riboswitch | RF01739

Pfam

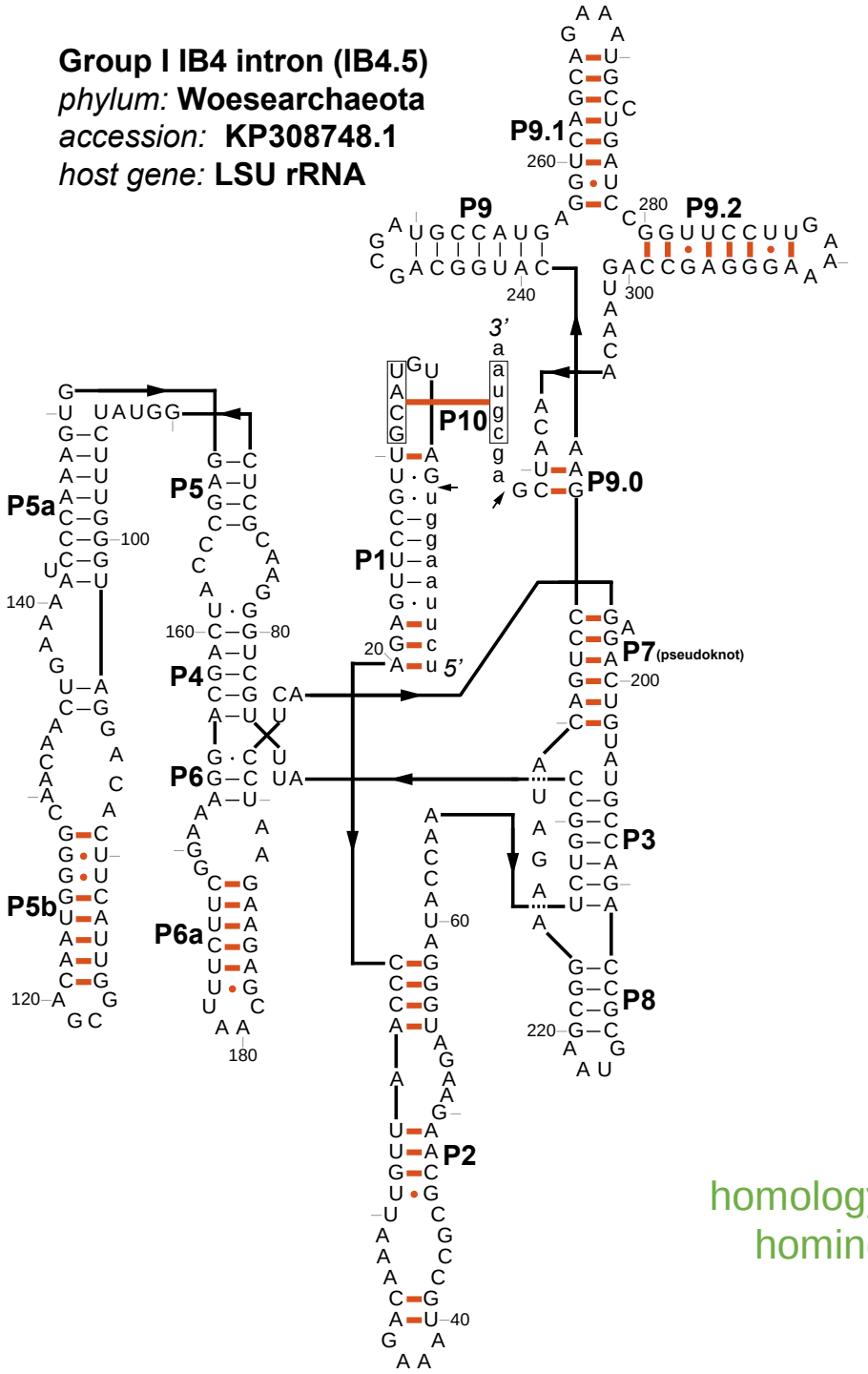
17929 protein domain families
http://pfam.xfam.org

Rfam

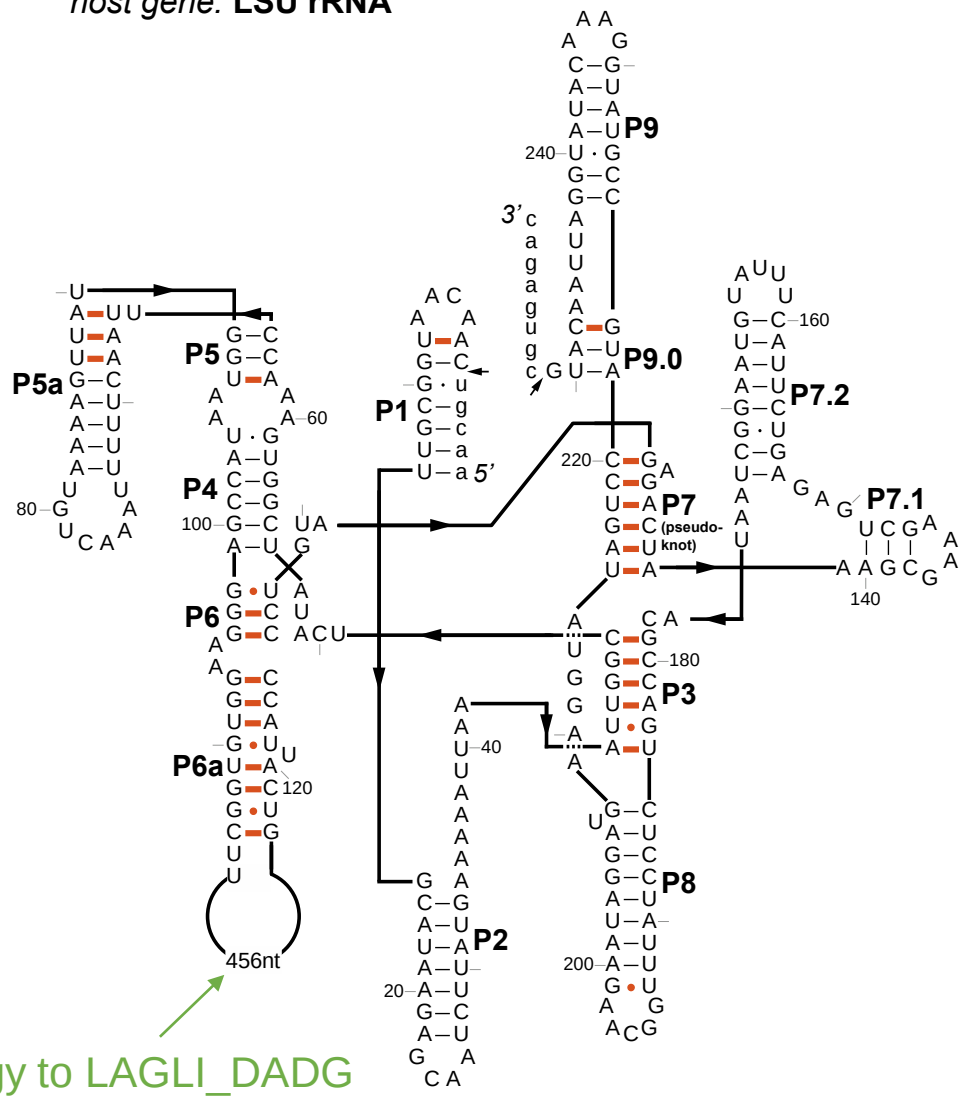
3016 RNA structure families
http://rfam.xfam.org

Xfam Consortium and HMMER server team: Rob Finn, Alex Bateman
European Bioinformatics Institute, Cambridge UK

Group I IB4 intron (IB4.5)
phylum: **Woesearchaeota**
accession: **KP308748.1**
host gene: **LSU rRNA**

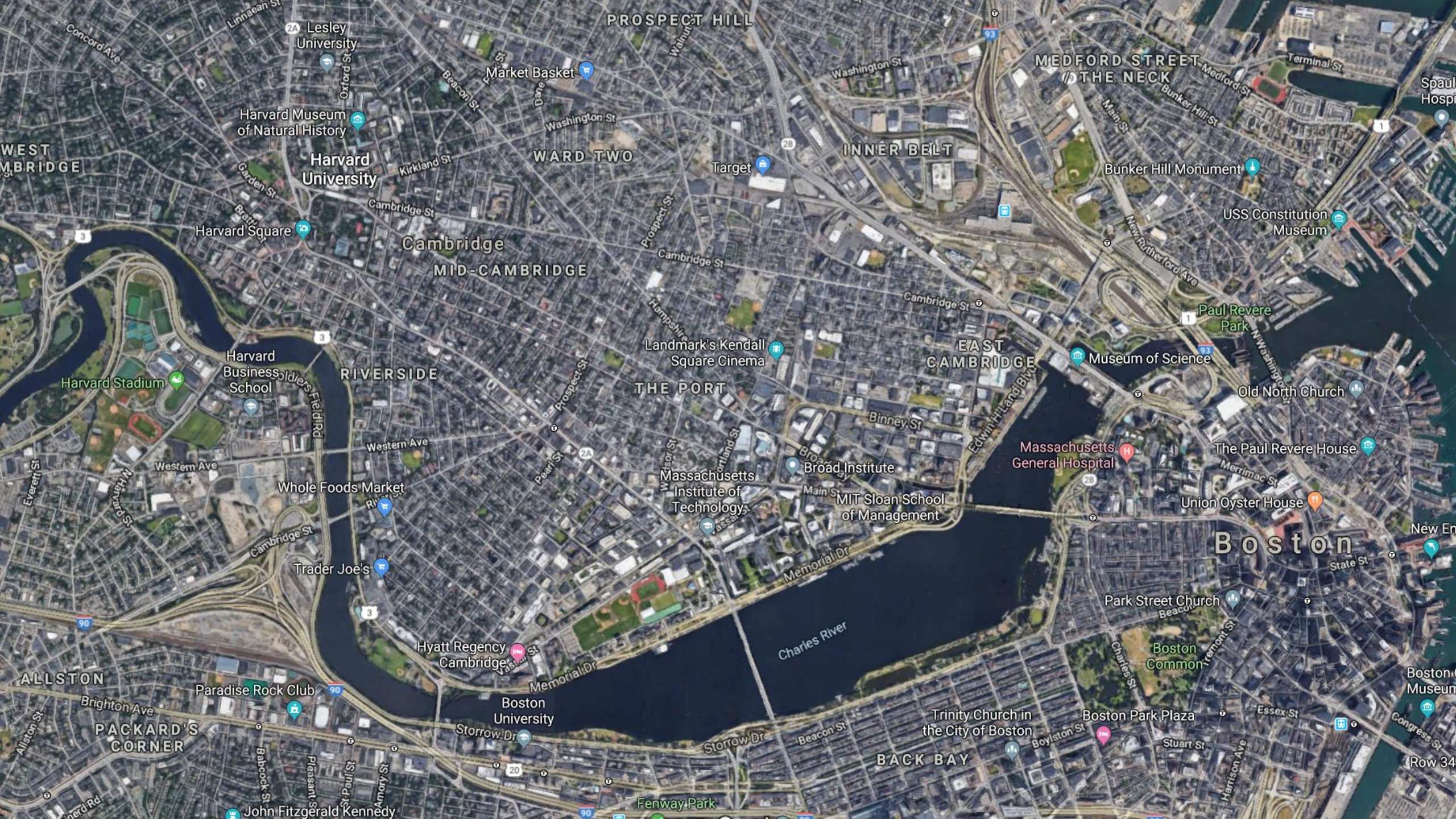


Group I IA3 intron (IA3.1)
phylum: **Woesearchaeota**
accession: **CP010426.1**
host gene: **LSU rRNA**



homology to LAGLI_DADG homing endonuclease
(PF00961)

Nawrocki et al, NAR 46:7970 (2018)


PROSPECT HILL
Lesley University
Concord Ave
Linnaean St
Oxford St
Market Basket
Beacon St
Dana St
Walnut St
Washington St
MEDFORD STREET / THE NECK
Medford St
Terminal St
Bunker Hill St
Main St
Harvard Museum of Natural History
Harvard University
WARD TWO
Kirkland St
Garden St
Target
INNER BELT
Bunker Hill Monument
USS Constitution Museum
New Rutherford Ave
Harvard Square
Cambridge St
Cambridge
MID-CAMBRIDGE
Prospect St
Hampshire St
Landmark's Kendall Square Cinema
EAST CAMBRIDGE
Museum of Science
Paul Revere Park
Harvard Business School
Harvard Stadium
Soldiers Field Rd
RIVERSIDE
THE PORT
Old North Church
Binney St
Edwin H Land Blvd
Massachusetts General Hospital
The Paul Revere House
Merrimac St
Broad Institute
Western Ave
Pearl St
Whole Foods Market
Massachusetts Institute of Technology
Main St
MIT Sloan School of Management
Union Oyster House
Everett St
Harvard St
Trader Joe's
Memorial Dr
Boston
State St
Park Street Church
Charles River
Hyatt Regency Cambridge
Boston Common
Charles St
Tremont St
ALLSTON
Paradise Rock Club
Brighton Ave
Boston University
Storrow Dr
Trinity Church in the City of Boston
Boston Park Plaza
Essex St
Boylston St
Stuart St
Beacon St
BACK BAY
PACKARD'S CORNER
Allston St
Babcock St
Pleasant St
Paul St
Amory St
Harrison Ave
Congress St
Boston Museum
Row 34
John Fitzgerald Kennedy
Fenway Park

“In my view, biology needs numbers; not after the fashion of physics, but in a good engineering, computational spirit.”