BACTERIAL DNA MODIFICATIONS

Yunfan Fan

Johns Hopkins University

@NanoporeConf | #NanoporeConf

BACTERIAL BASE MODIFICATIONS
Bacterial Immune System

Restriction-methylation system

Methyltransferases (MTases) methylate DNA at certain motifs.

Endonucleases digest only un-methylated DNA at these same motifs.

Methylation marks prevent digestion

Endonucleases digest phage DNA at recognition sites

MTases methylate bacterial DNA
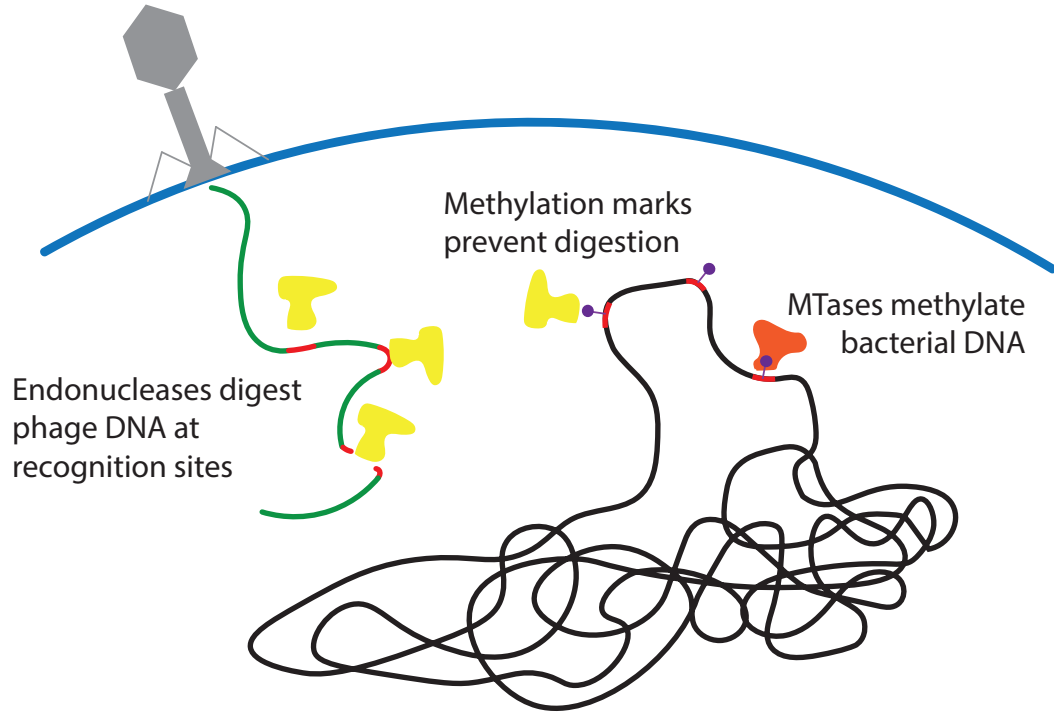
BACTERIAL BASE MODIFICATIONS
Methylation Motifs

Methylated *E. coli* *g*DNA

Kindly provided by NEB

MTase	Modification	Motif
PspJDRI	**4-methylcytosine**	(m4C)CGG
Sin395ORF667	**5-methylcytosine**	GAT(m5C)
Fnu4H	**5-methylcytosine**	G(5mC)NGC
M.SdeAII	**5-methylcytosine**	CCNGG(m5C)
M.HinfI	**6-methyladenine**	G(m6A)NTC
BstXII (dam style)	**6-methyladenine**	G(m6A)TC

E. coli ER2796 has all MTase genes knocked out. Selected MTases can then be transfected to control methylation motifs.

SIGNAL ALIGNMENT
nanopolish eventalign

Use the eventalign module of nanopolish to align signals to a reference

Associate reference genome positions with event means in the reads

HMM with the reference genome

Simpson *Nature Methods* (2015)
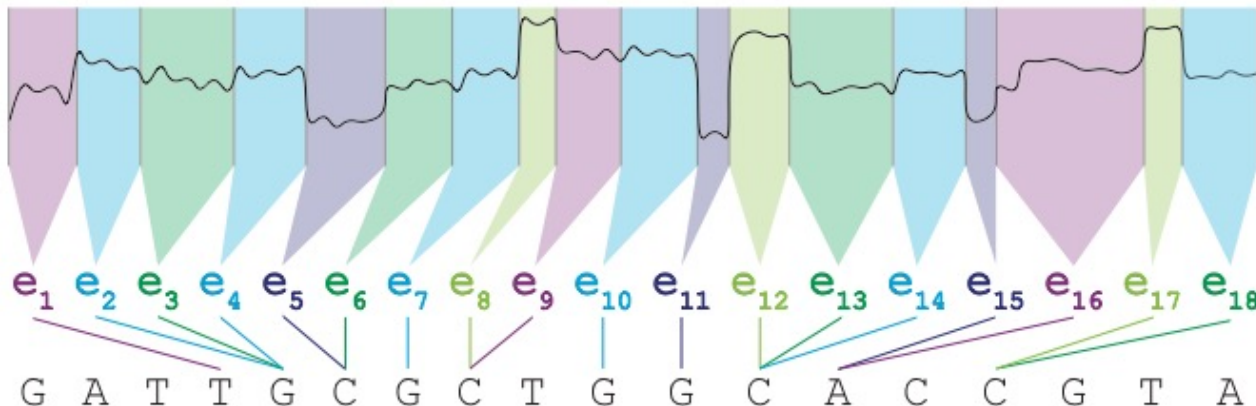
METHYLATION TRAINING
nanopolish methyltrain
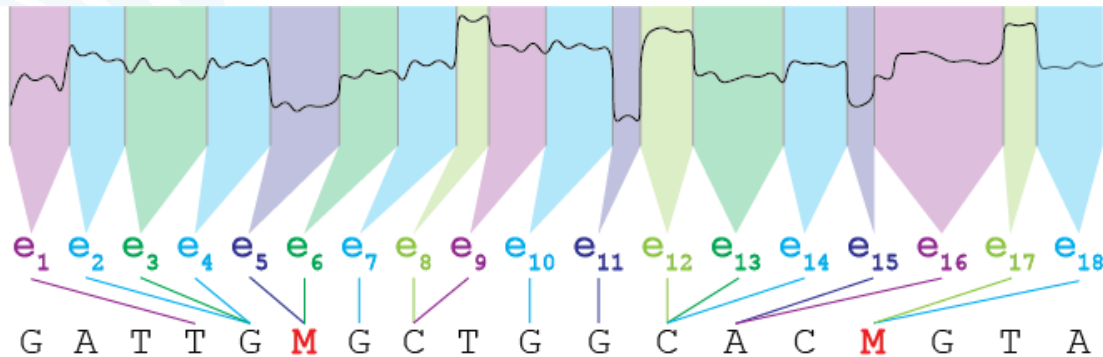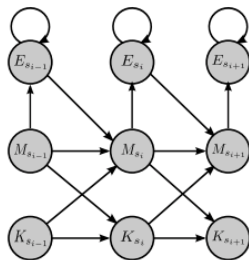
Train current signatures for methylated k-mers

Use an HMM as in eventalign

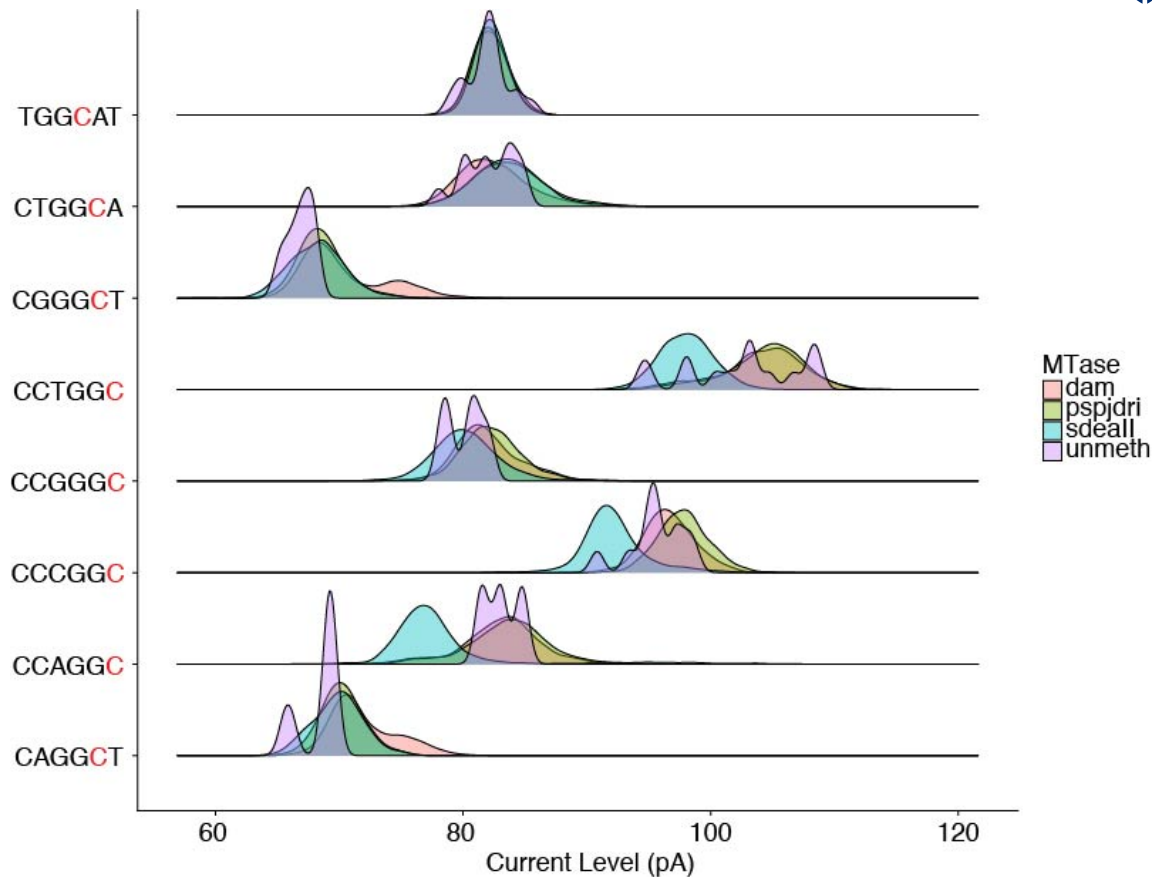Add methylation as a fifth base in the reference genome, and align events.

CURRENT LEVEL DISTRIBUTIONS
5mC – sdeaII (CCNGGCm)

Current distributions for 6-mers with modifications at selected motifs

Methylated base shown in red.

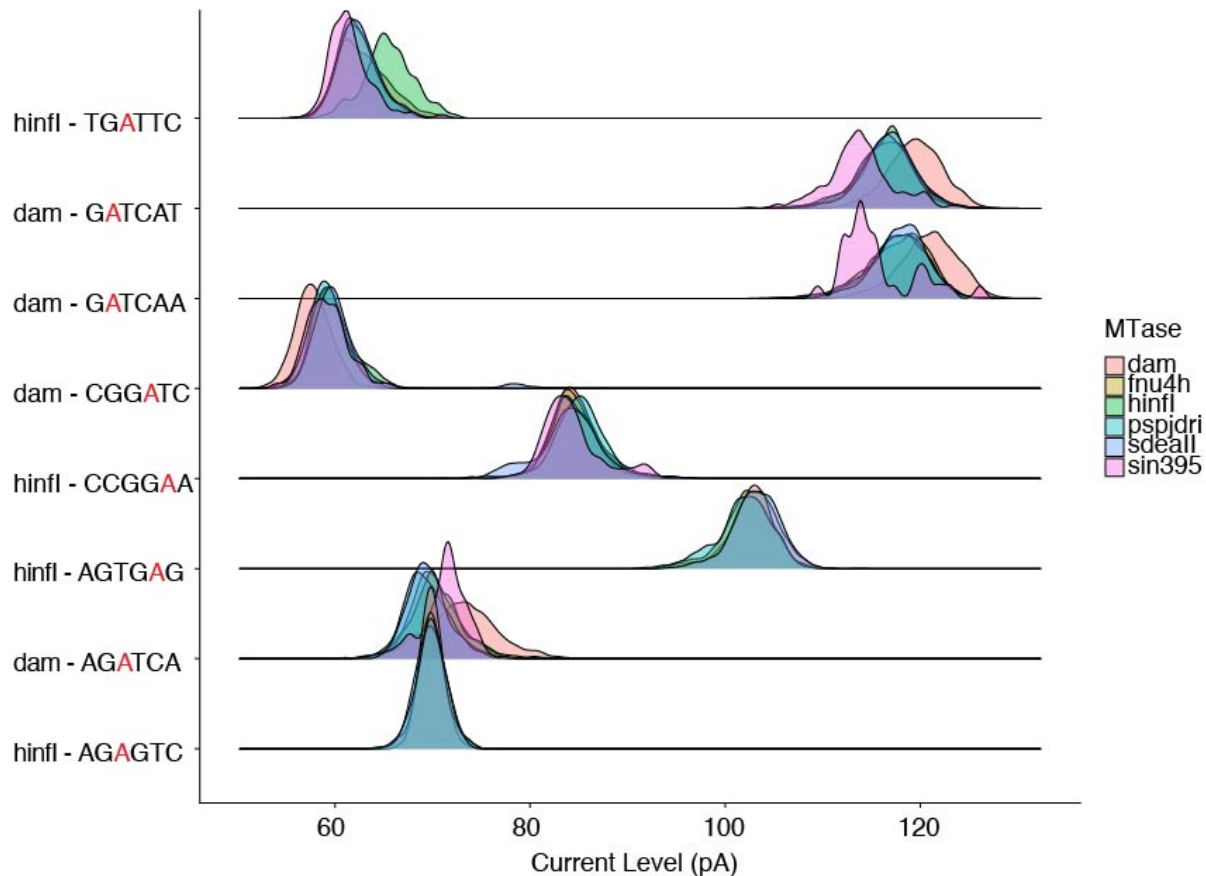Some signals are easily distinguishable, and some are not

CURRENT LEVEL DISTRIBUTIONS
6mA

Current distributions for 6-mers with modifications at selected motifs

The dam MTase shares a motif with the sin395 MTase (GATC). Distributions for both can be seen diverging.

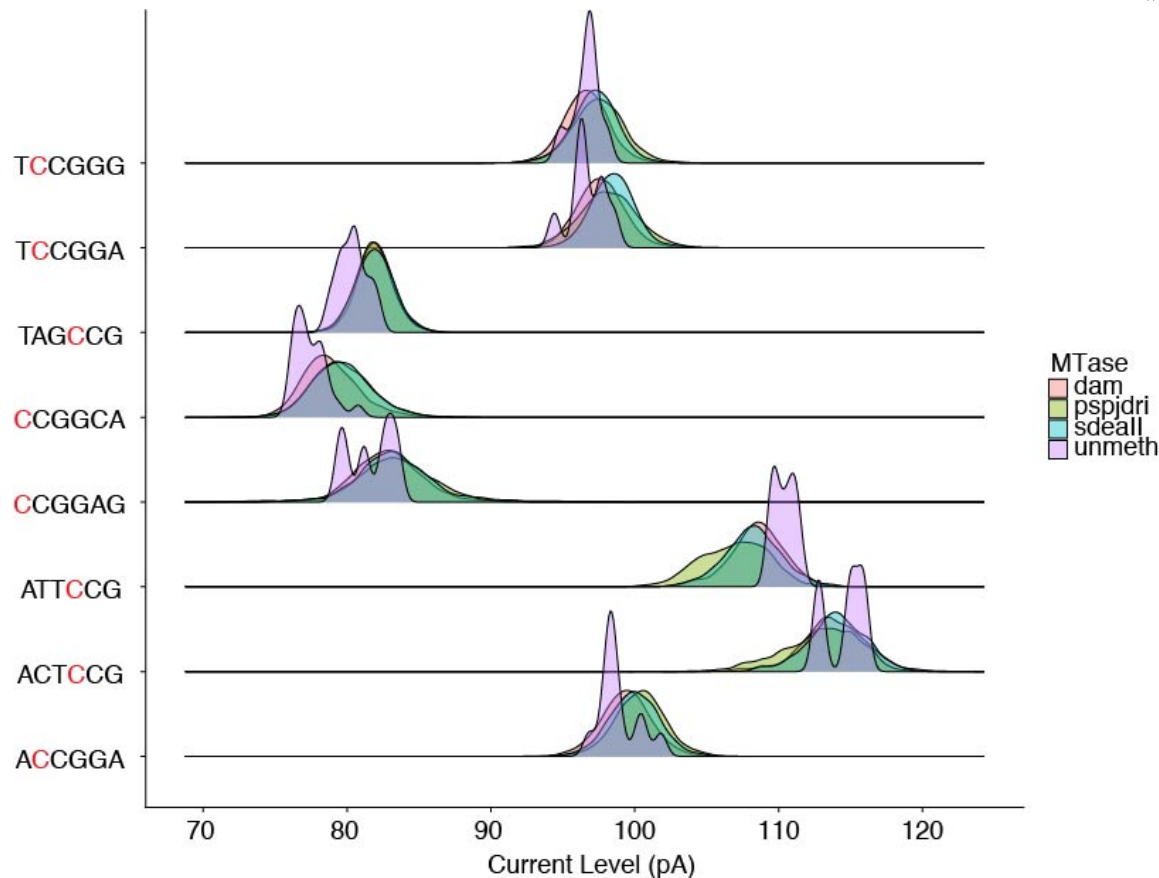hinfI methylation shows good separation for some k-mers, but not others.

CURRENT LEVEL DISTRIBUTIONS

4mC – pspjdri (mCCGG)

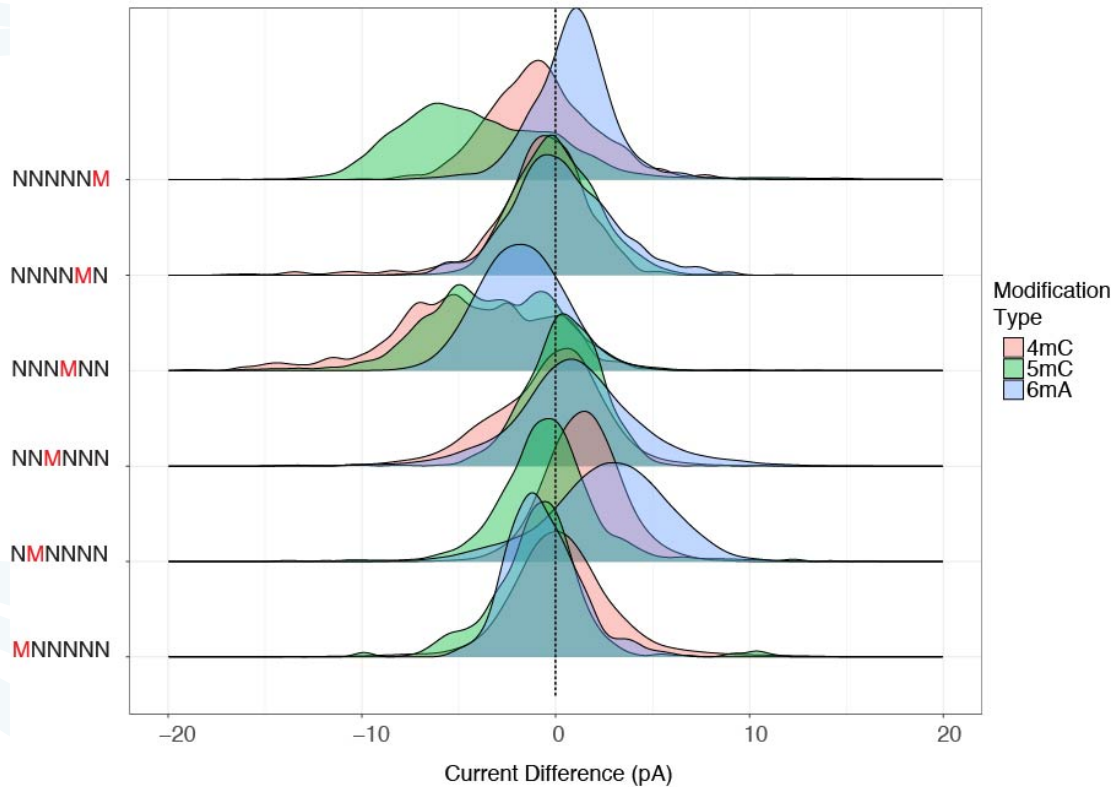**Current distributions for 6-mers with modifications at selected motifs**

Methylated base shown in red.

4mC appears to show a weaker signal.

CURRENT CHANGES

How does the signal change based on the location of the methyl mark?

Methyl mark locations variably shift the distribution of k-mer signals

Signal shift also appears to be modification dependent

Non-linear behaviours are perhaps due to effects of methylation on the biophysical properties of DNA, such as base stacking, persistence length, etc.

IN SUMMARY

Methylation Model

We've used nanopolish to train methylation models for a variety of enzyme motifs, and can observe shifts in event mean distributions.
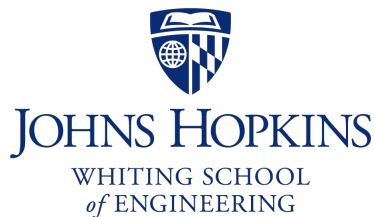
Next Step

Use these models to call methylation in the plasmids of these organisms.

ACKNOWLEDGEMENTS

Timp Lab, Johns Hopkins University
- Rachael Workman
- Stephanie Hao
- Isac Lee
- Winston Timp

- Jen Lu (Salzberg Lab)

Ontario Institute for Cancer Research
- Jared Simpson
- P.C. Zuzarte
- Matei David
- L. J. Dursi

New England BioLabs
- Alexey Fomenkov
- Rich Roberts

National Human Genome Research Institute
1R01HG009190-01A1

National Institute of Allergy and Infectious Diseases
1R21AI130608-01 (Trish Simner)

THANK YOU

The content contained in this presentation should not be reproduced without permission of the speaker.