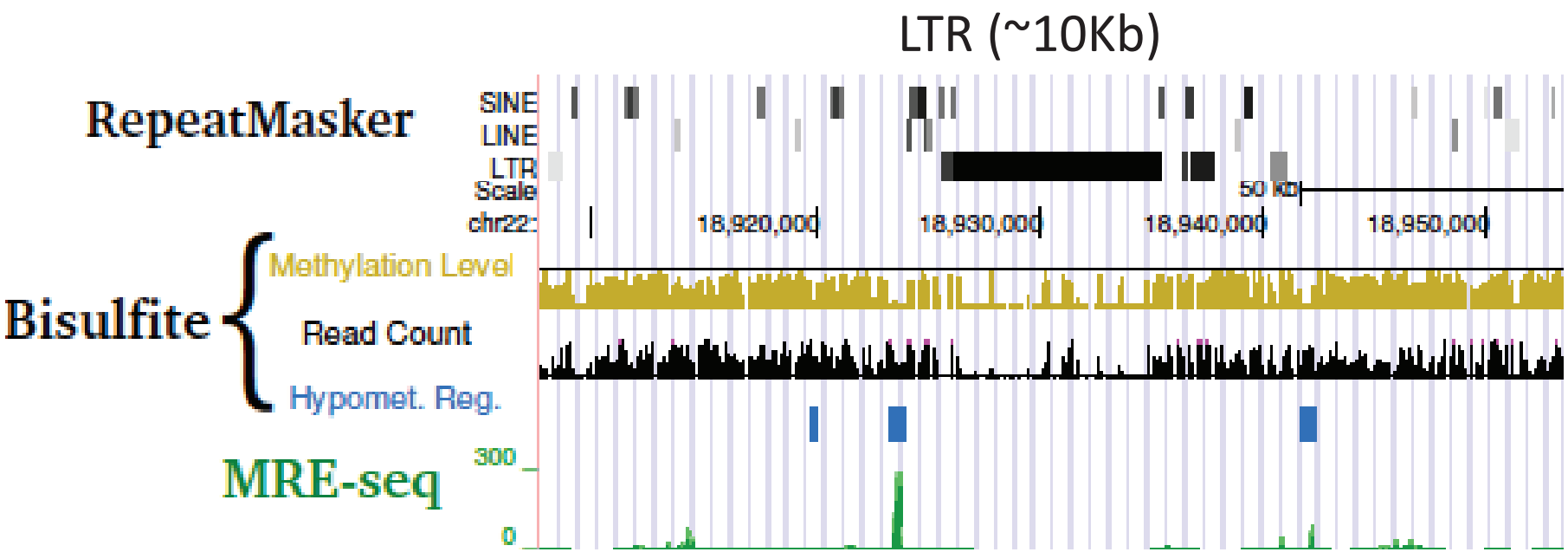
Understanding methylome, metagenome, structural variants using SMRT sequencing

Shinichi Morishita
University of Tokyo

Can we characterize methylome in repeats?

Bisulfite-treatment often divides DNA into ~600bp fragments.

We extended SMRT sequencing.

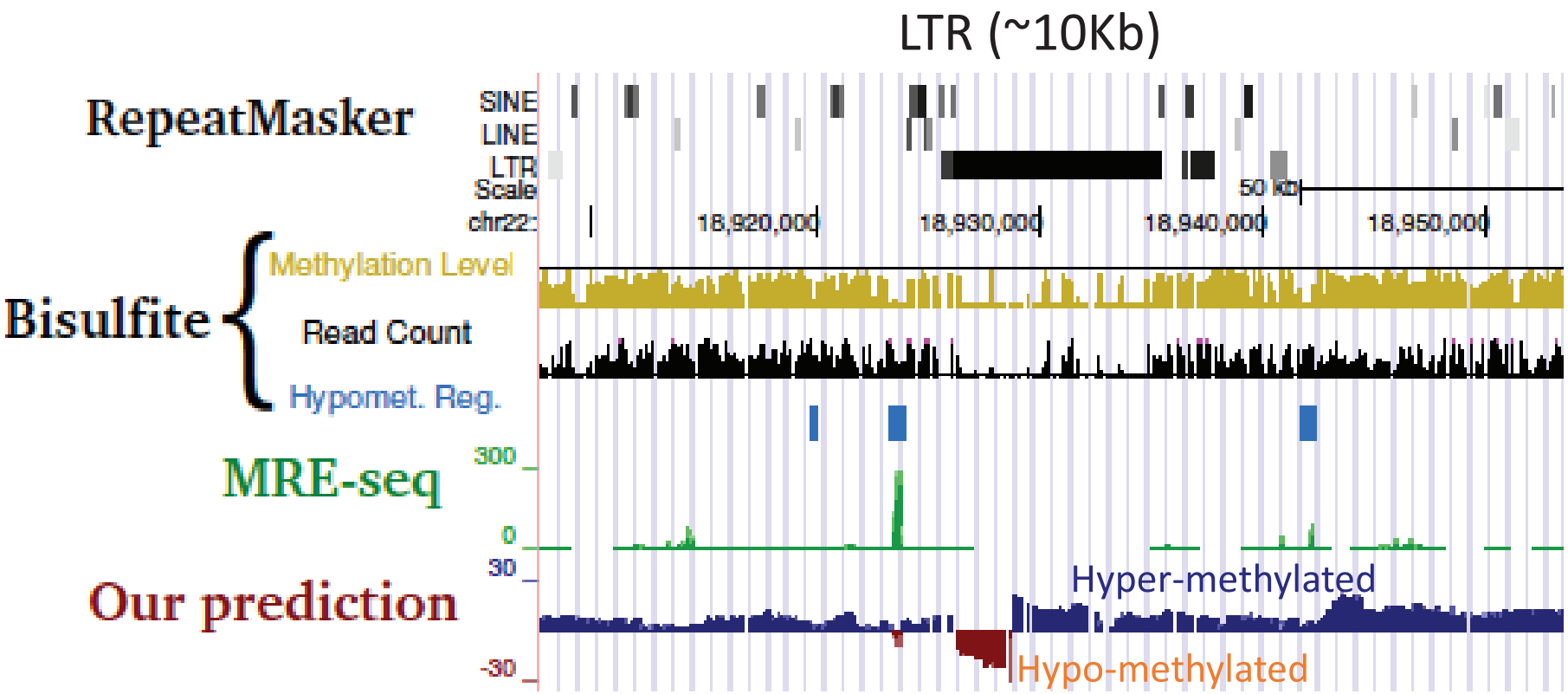


Yuta Suzuki

Can we characterize methylome in repeats?

Bisulfite-treatment often divides DNA into ~600bp fragments.

We extended SMRT sequencing.

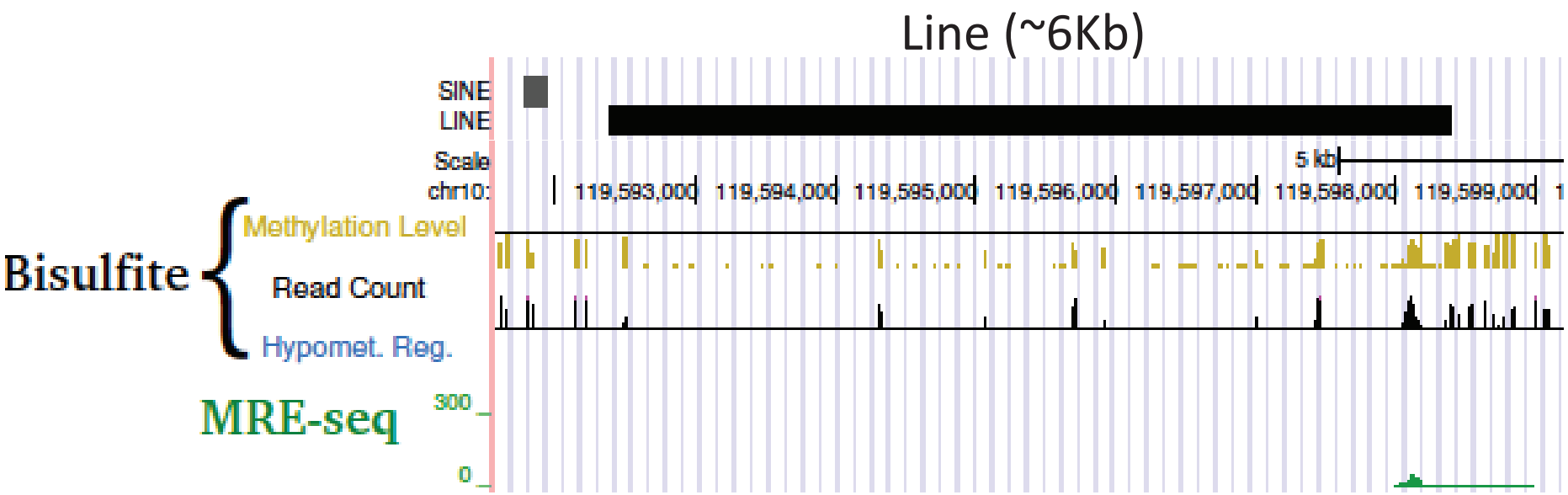


Yuta Suzuki

Can we characterize methylome in repeats?

Bisulfite-treatment often divides DNA into ~600bp fragments.

We extended SMRT sequencing.

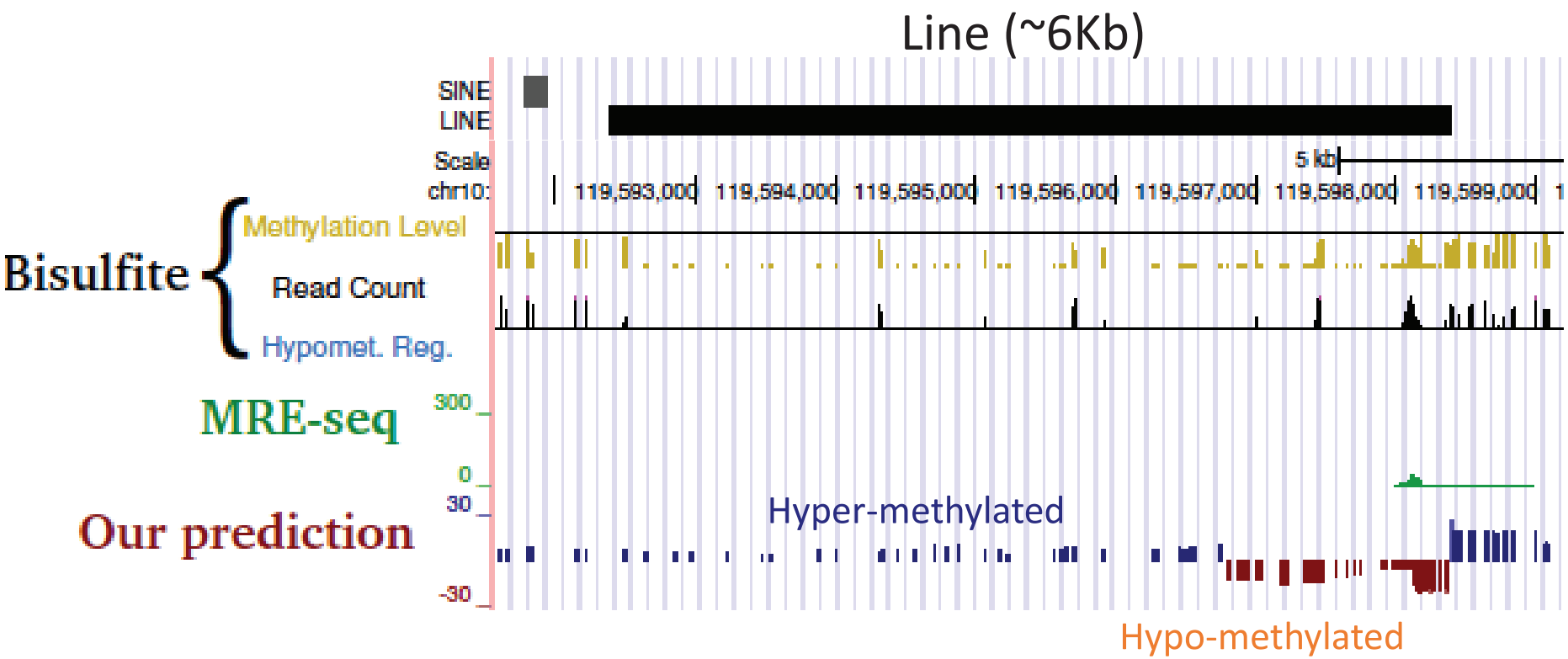


Yuta Suzuki

Can we characterize methylome in repeats?

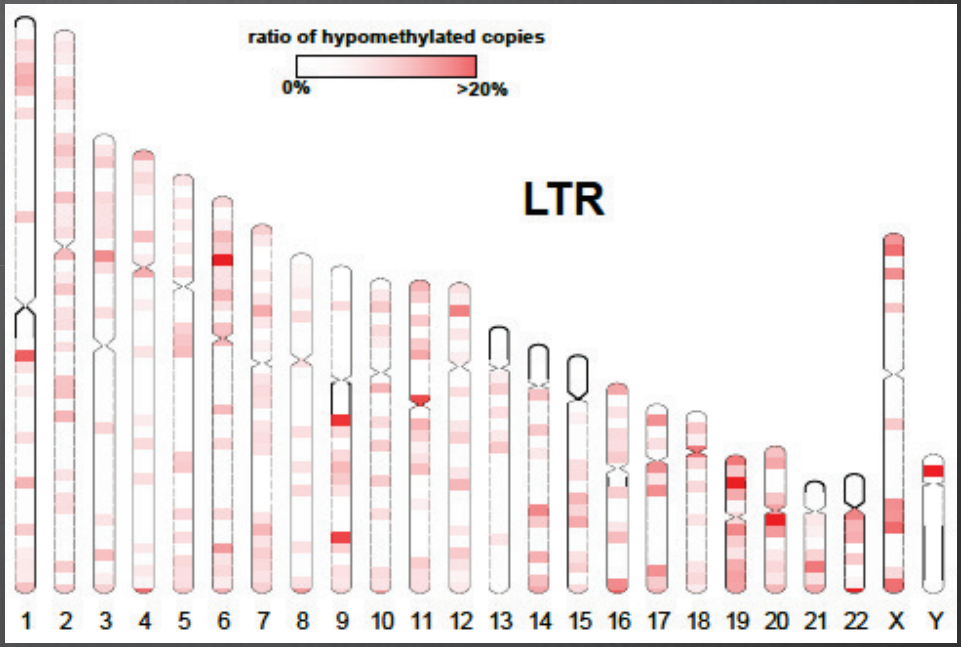
Bisulfite-treatment often divides DNA into ~600bp fragments.

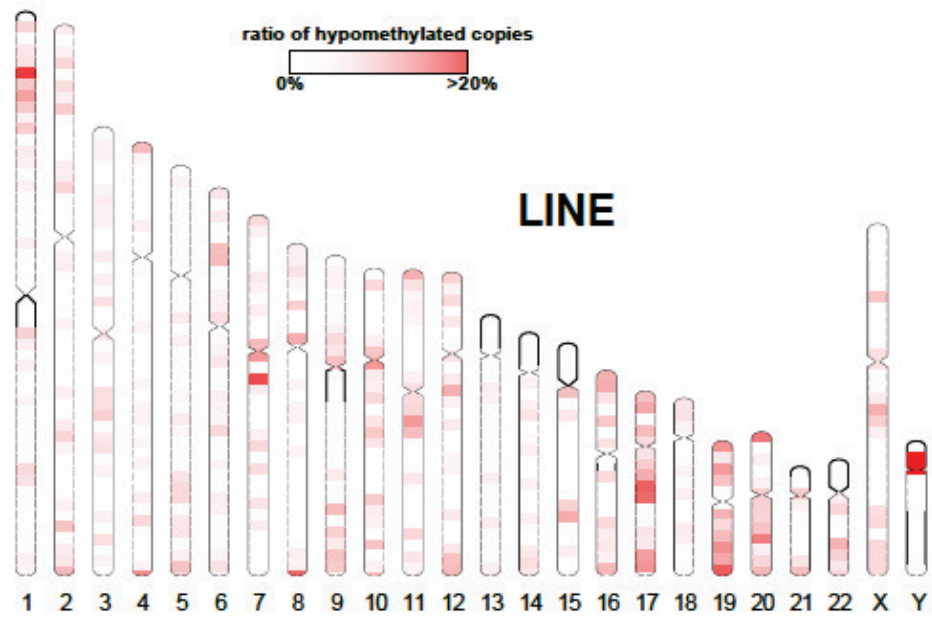
We extended SMRT sequencing.

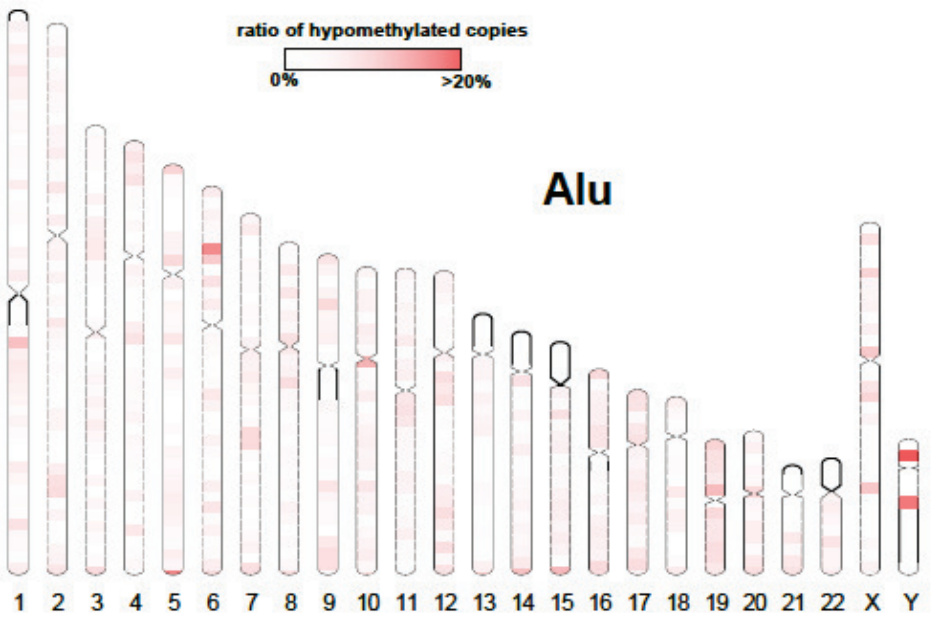


Yuta Suzuki

Landscape of hypomethylated repeats in the human genome

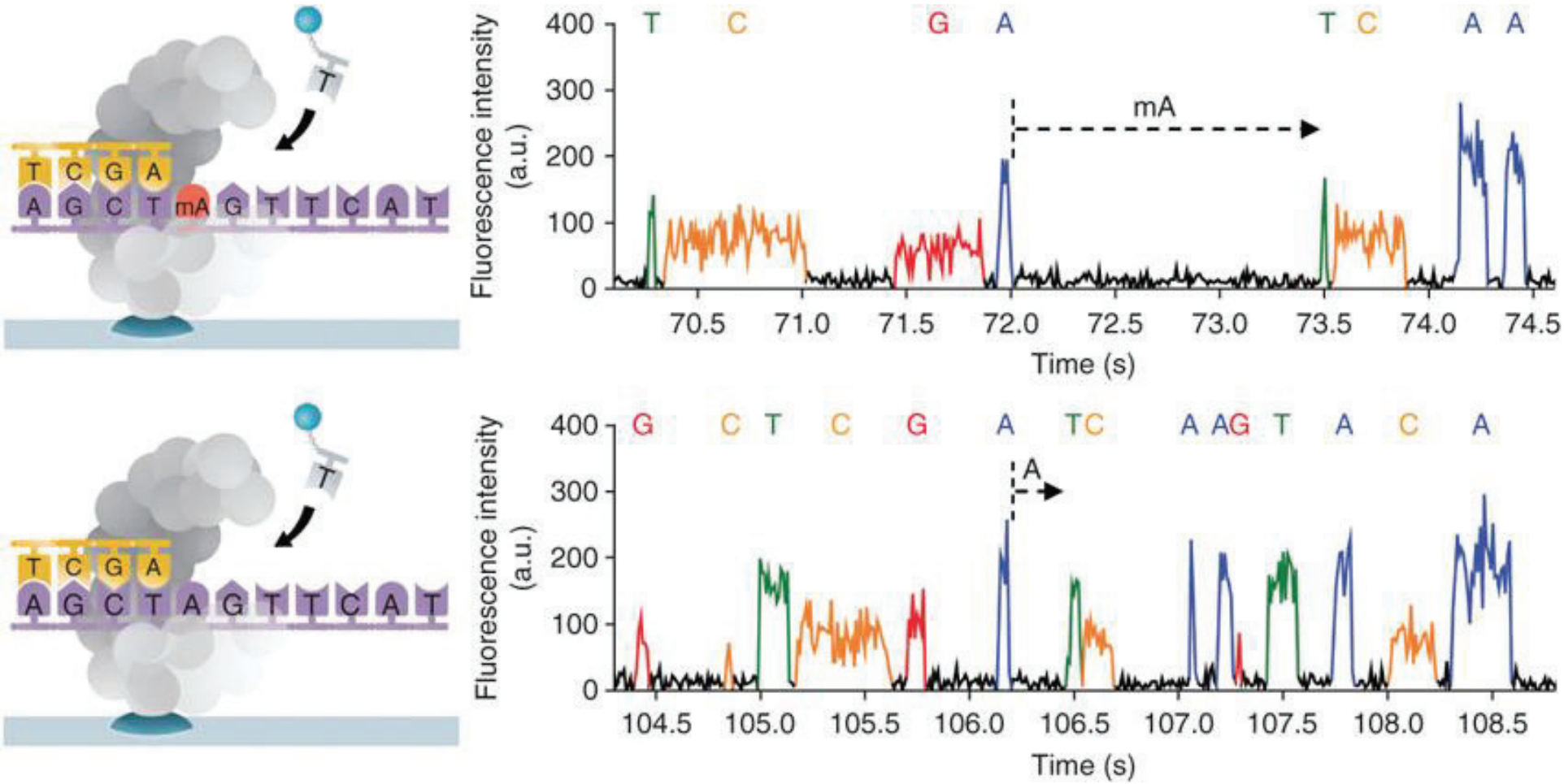






Yuta Suzuki

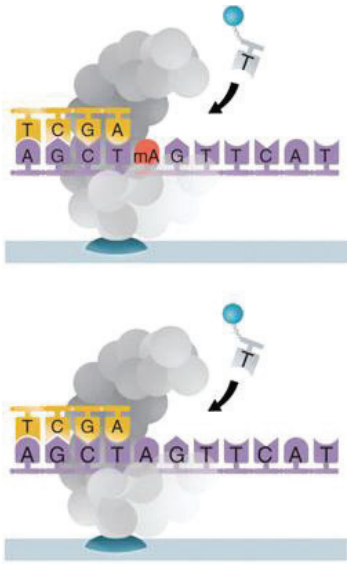
Estimating CpG methylation using SMRT sequencing



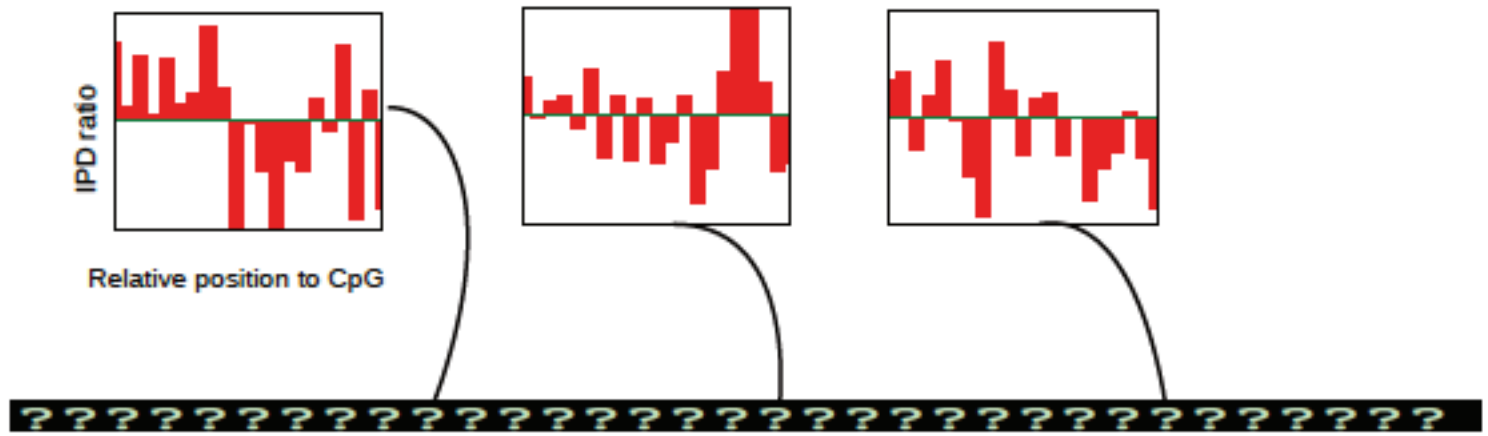
Benjamin A Flusberg, Dale R Webster, Jessica H Lee, Kevin J Travers, Eric C Olivares, Tyson A Clark, Jonas Korlach & Stephen W Turner. Direct detection of DNA methylation during single-molecule, real-time sequencing. Nature Methods 7, 461 - 465 (2010)

Yuta Suzuki

Estimating CpG methylation using SMRT sequencing



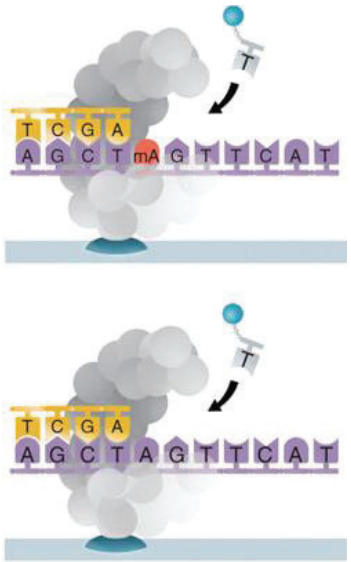
Subtle signals around CpGs…



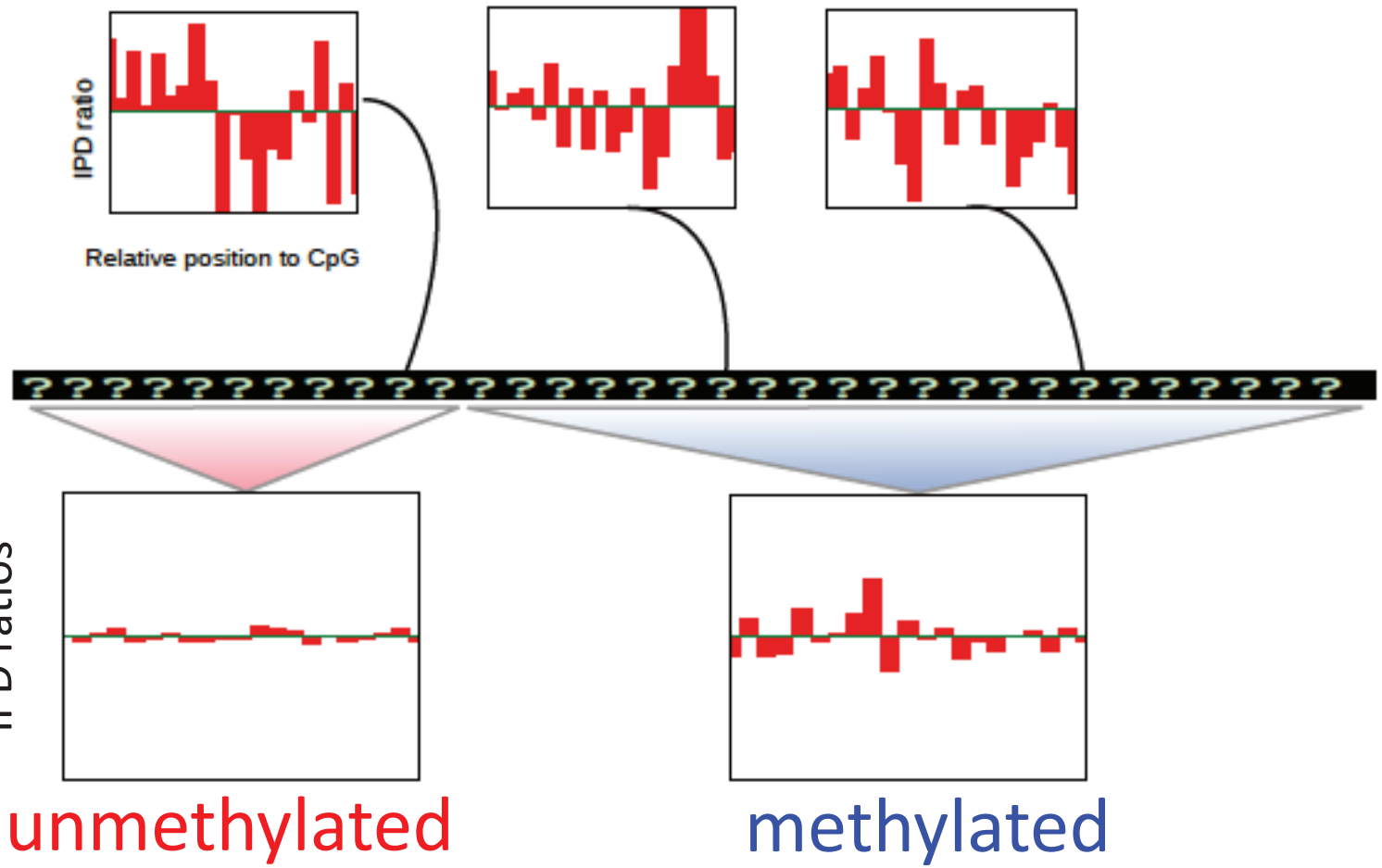
CpG sites

Yuta Suzuki

Estimating CpG methylation using SMRT sequencing



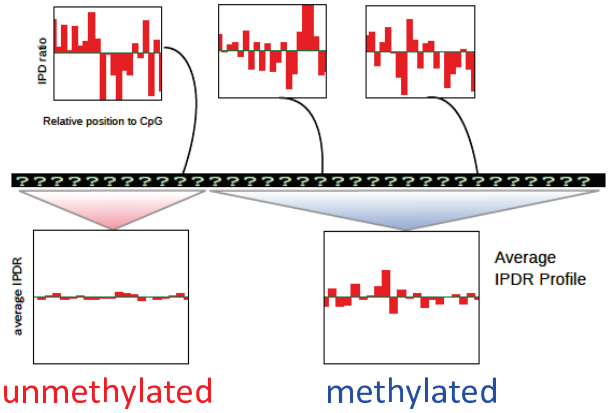
Subtle signals around CpGs…

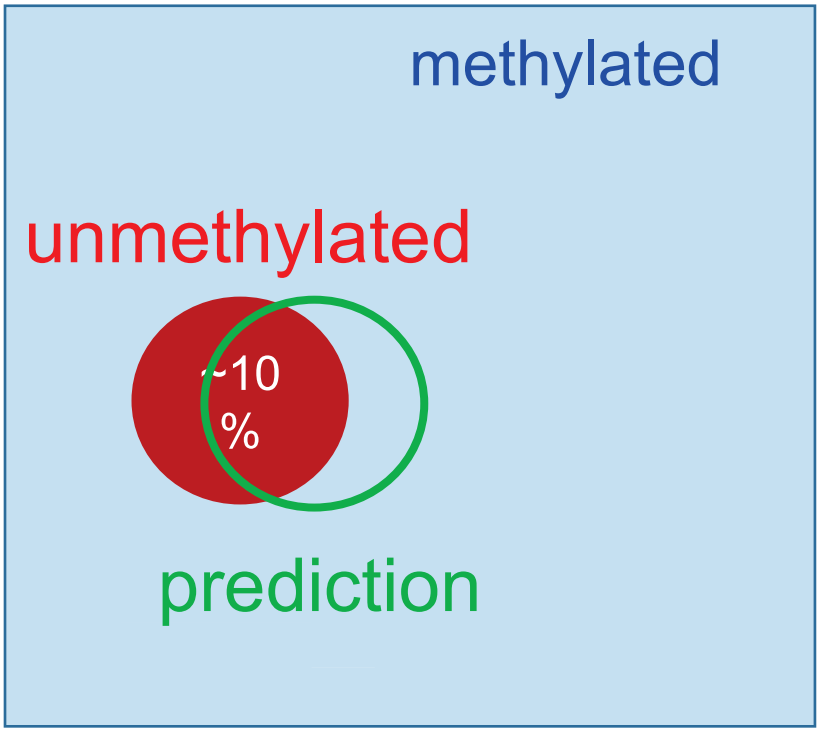


A linear-time optimization algorithm (2.265 s to process 1 Mbp on an Intel i7-3612QM processor)

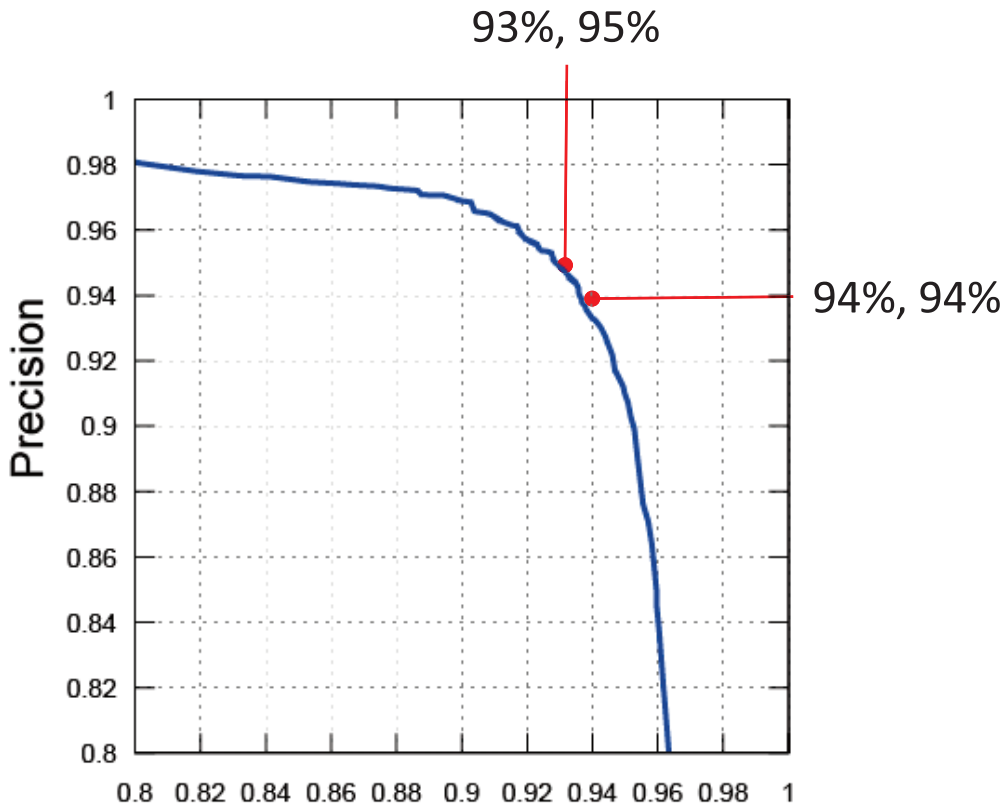
Yuta Suzuki

Estimating CpG methylation using SMRT sequencing





Bisulfite sequencing



minimum window size = 35bp
~20-fold SMRT subread coverage

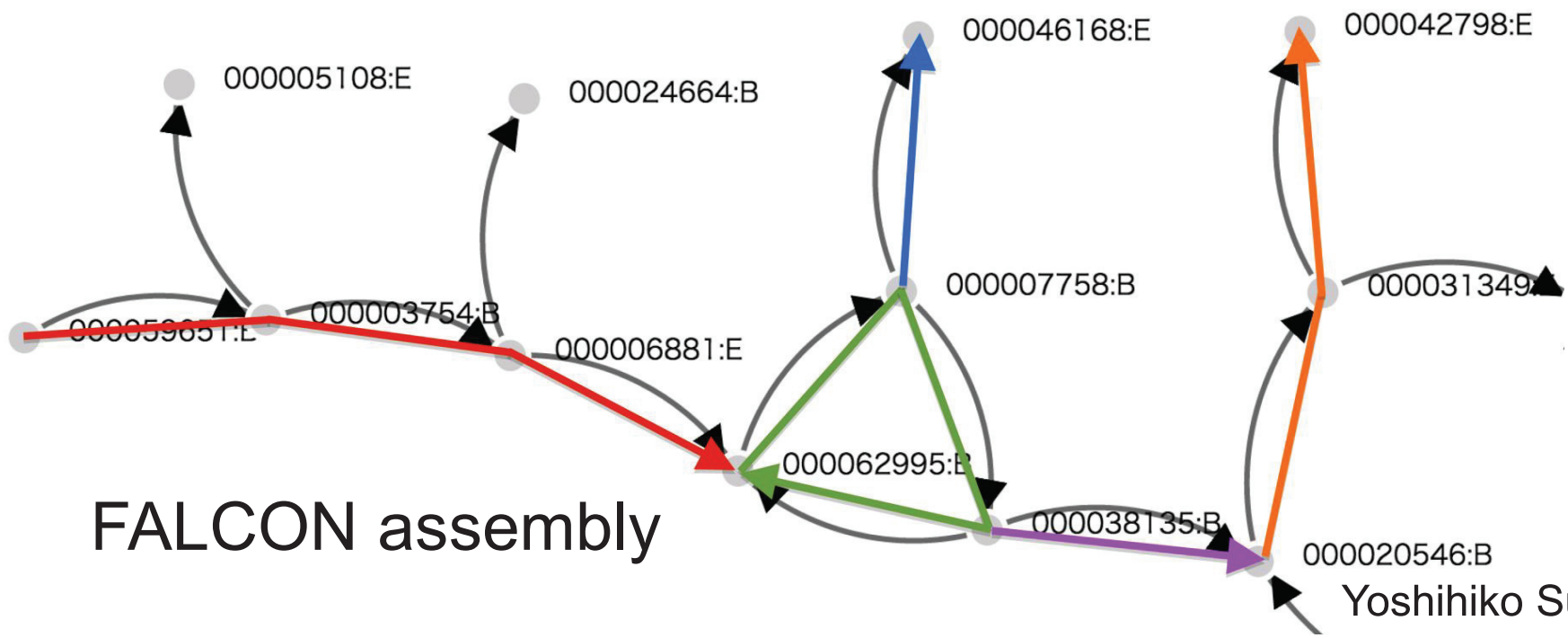
Yuta Suzuki

SMRT sequencing of un-isolated bacteria in flora

How much of bacterial genomes are unknown?

- Used lysozyme instead of QIAamp DNA Stool Mini Kit
  (Morita et al. Microbes and Environments, 2007)
- Avg. subread length = 7-9Kb, N50 = 13-15Kb (4 samples)
- 61 contigs of >1Mb in size, 67% are unknown

Assembly considering abundance of organisms



FALCON assembly

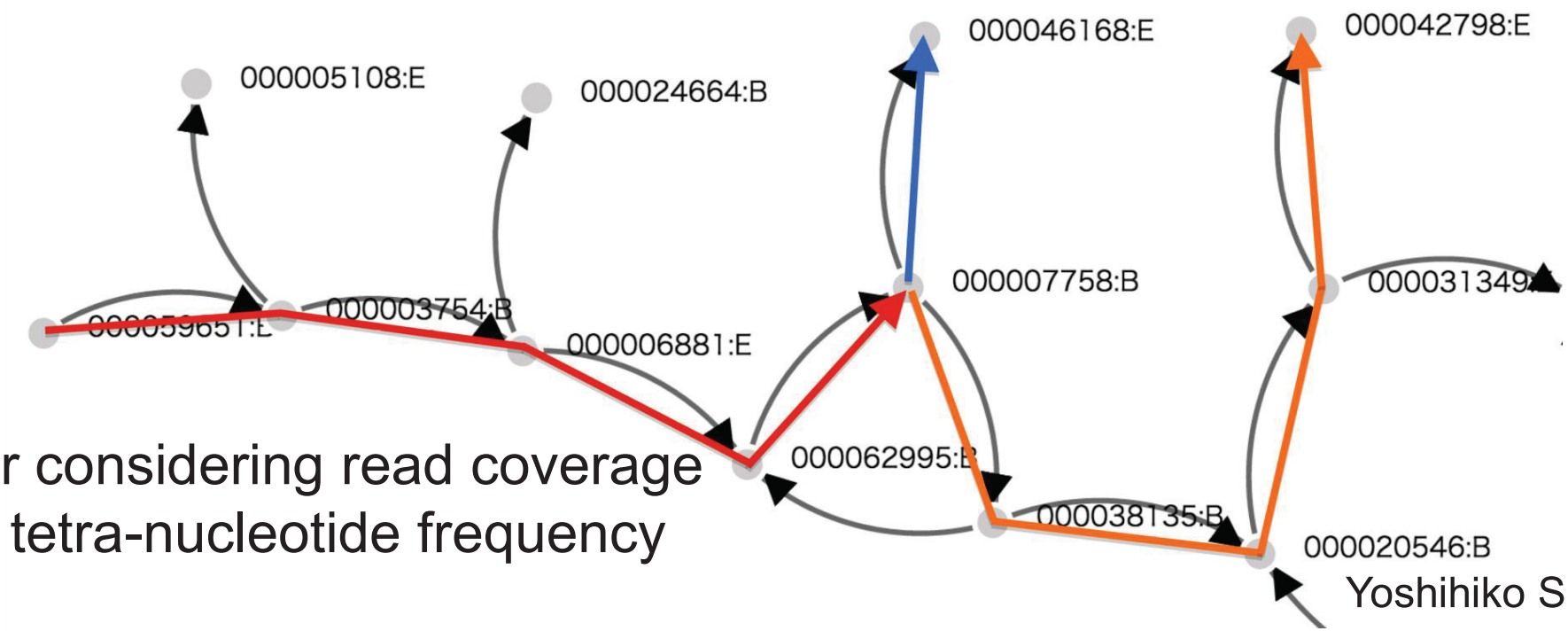
Yoshihiko Suzuki

SMRT sequencing of un-isolated bacteria in flora

How much of bacterial genomes are unknown?

- Used lysozyme instead of QIAamp DNA Stool Mini Kit
  (Morita et al. Microbes and Environments, 2007)
- Avg. subread length = 7-9Kb, N50 = 13-15Kb (4 samples)
- 61 contigs of >1Mb in size, 67% are unknown

Assembly considering abundance of organisms

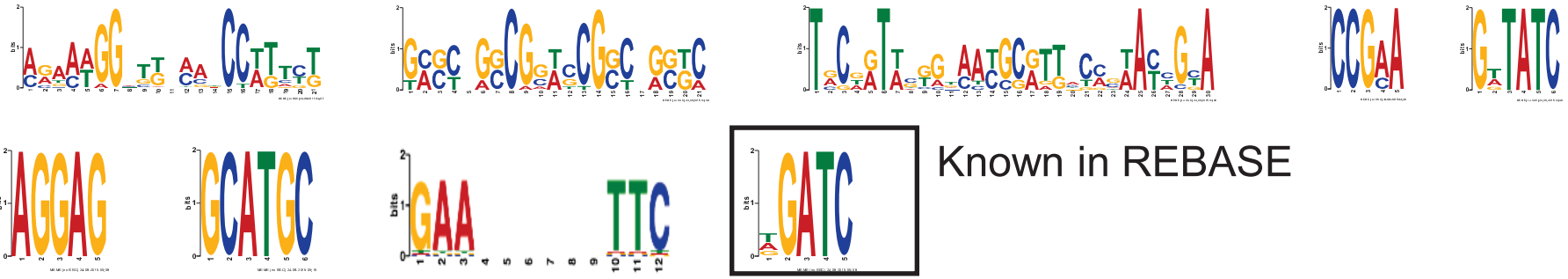


After considering read coverage and tetra-nucleotide frequency

Yoshihiko Suzuki

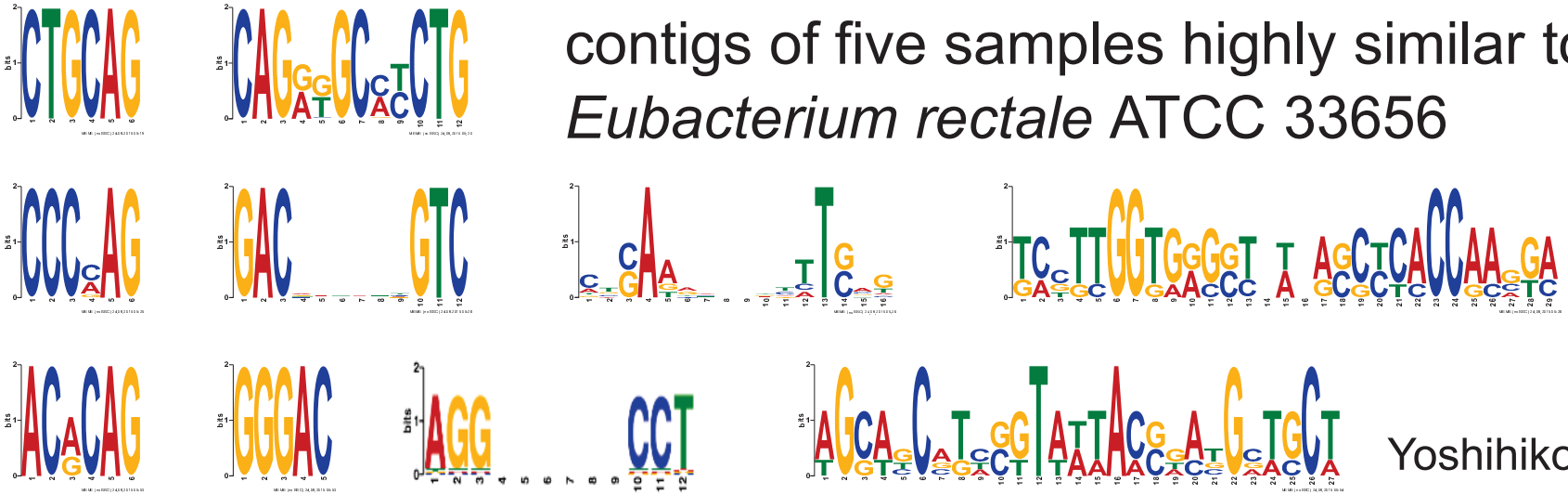
SMRT sequencing of un-isolated bacteria in flora

Might DNA modification patterns be different between un-isolated and isolated bacteria?



Known in REBASE

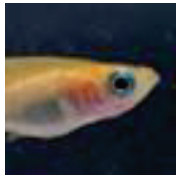
Motifs with candidate DNA modifications in contigs of five samples highly similar to *Eubacterium rectale* ATCC 33656



Yoshihiko Suzuki

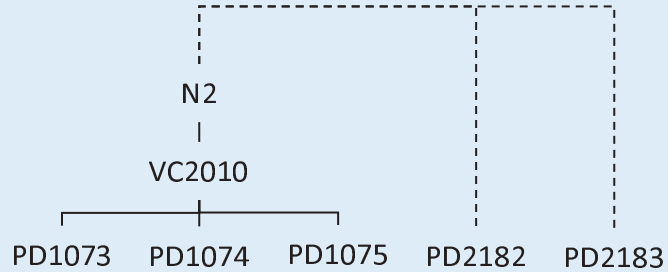
Can we correct deficits in reference genomes using SMRT sequencing?

x 3

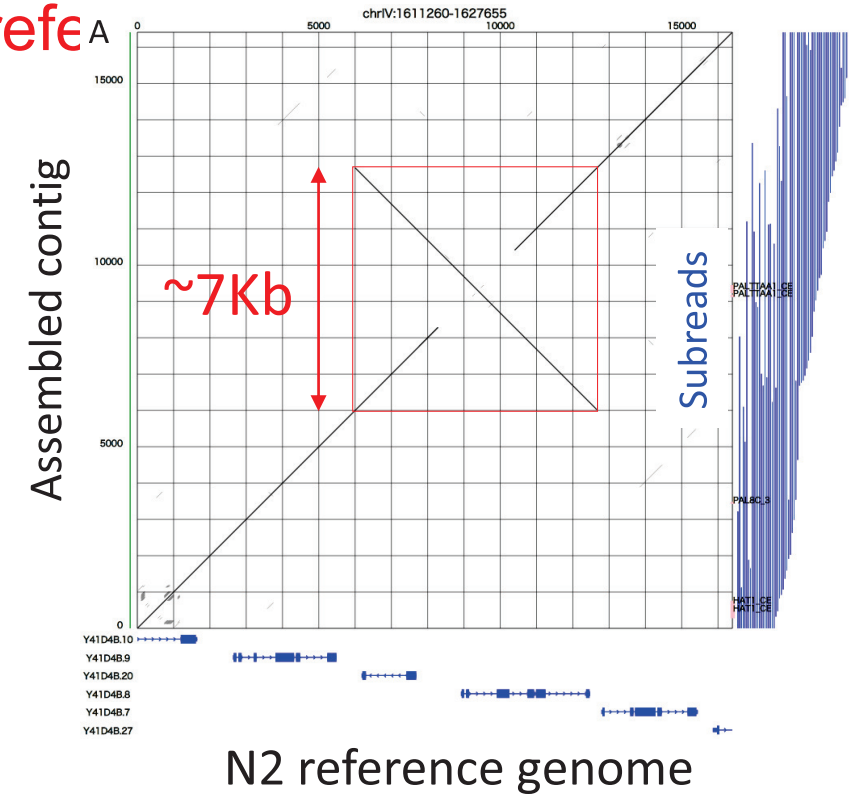


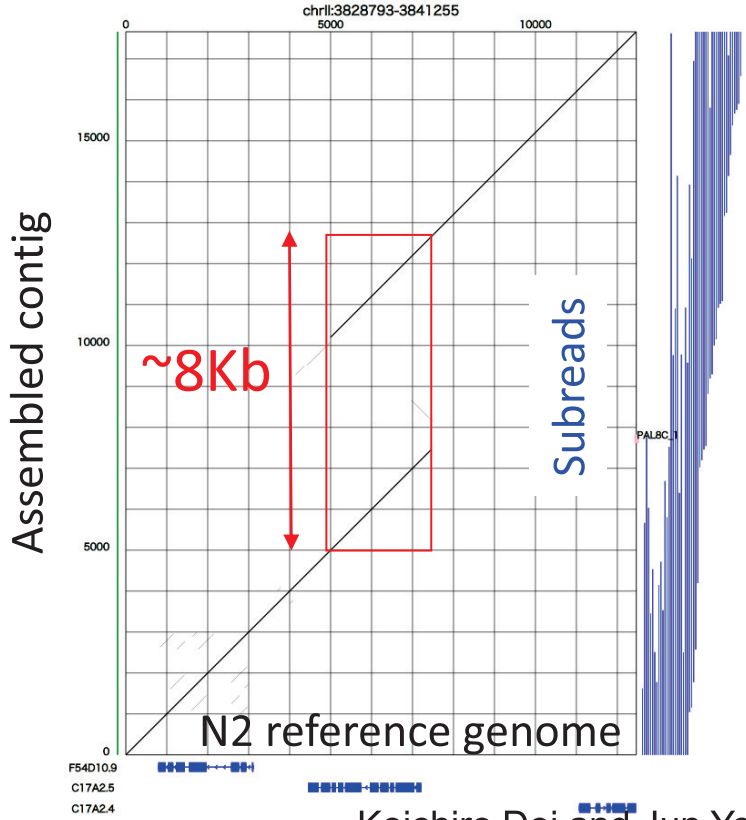
x 3

x 5



Detected 30 duplications and 24 inversions absent in the worm refe





Koichiro Doi and Jun Yoshimura

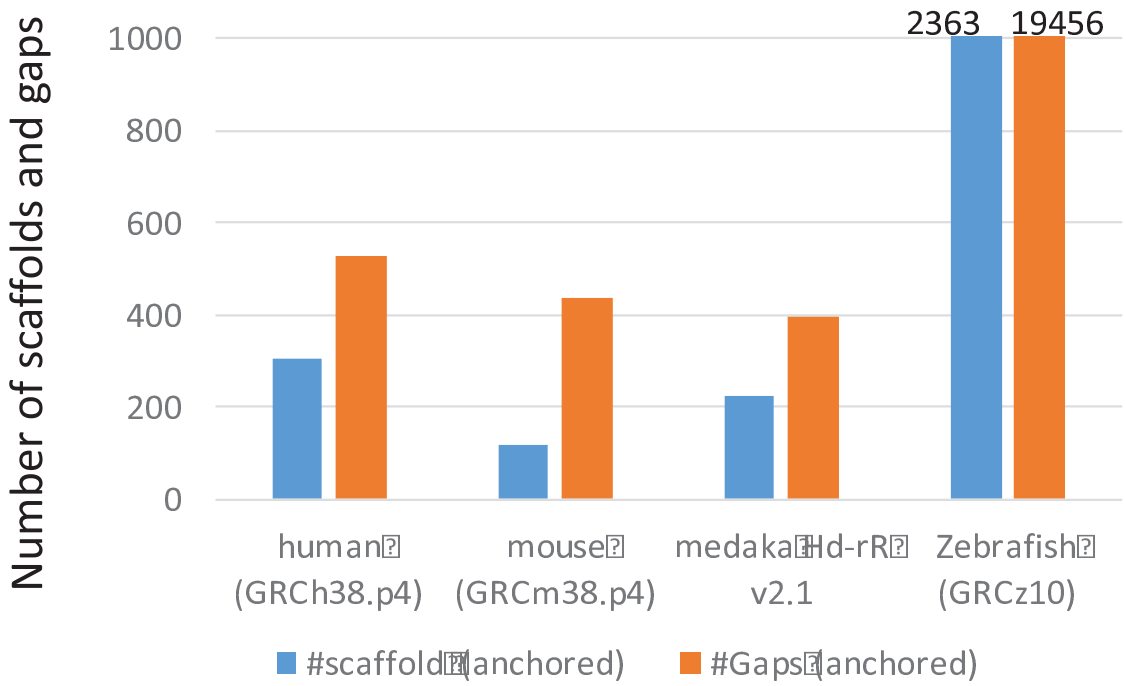
Can we correct deficits in reference genomes using SMRT sequencing?

High concordance of PacBio contigs with other information (medaka)

2,510 genetic markers (only 4% contigs had deficits)

25,148 BAC ends (only 0.04% are inconsistent)

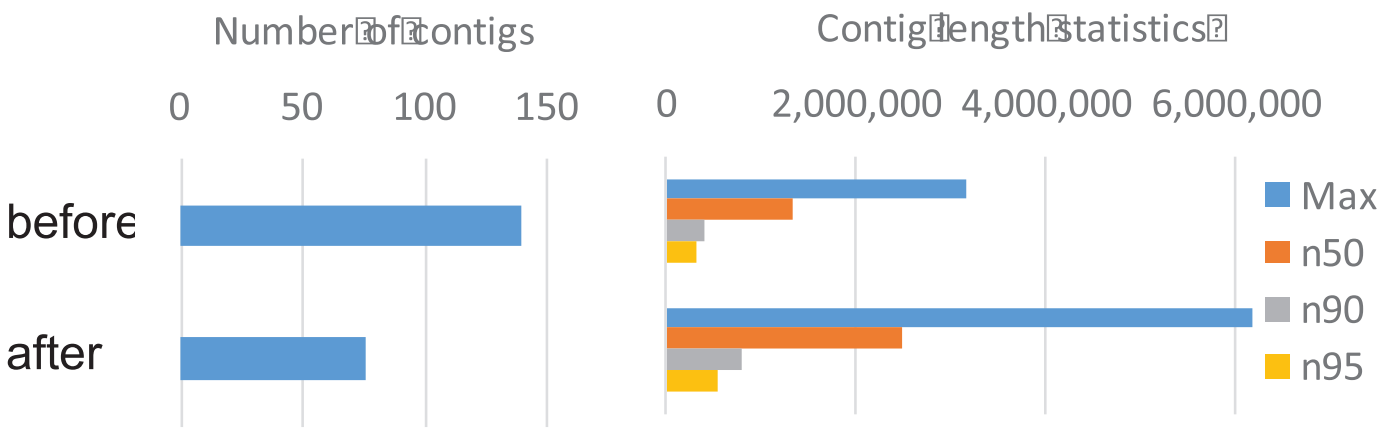
~90% of contigs on chromosomes (N50 contig = ~2.5Mb)

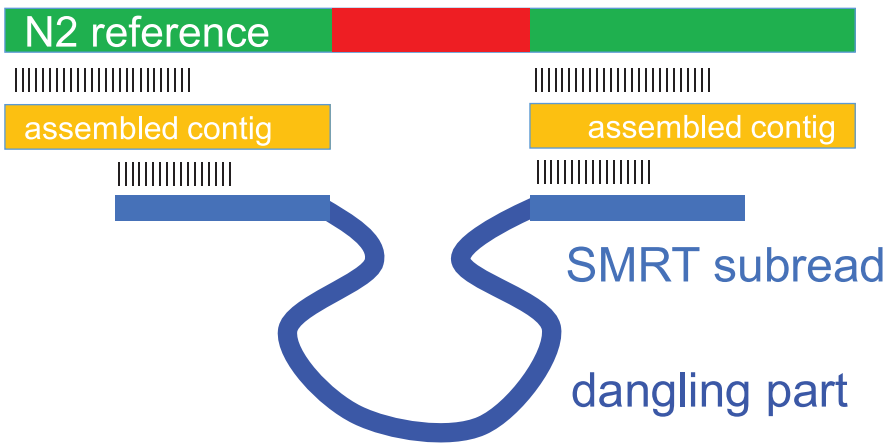


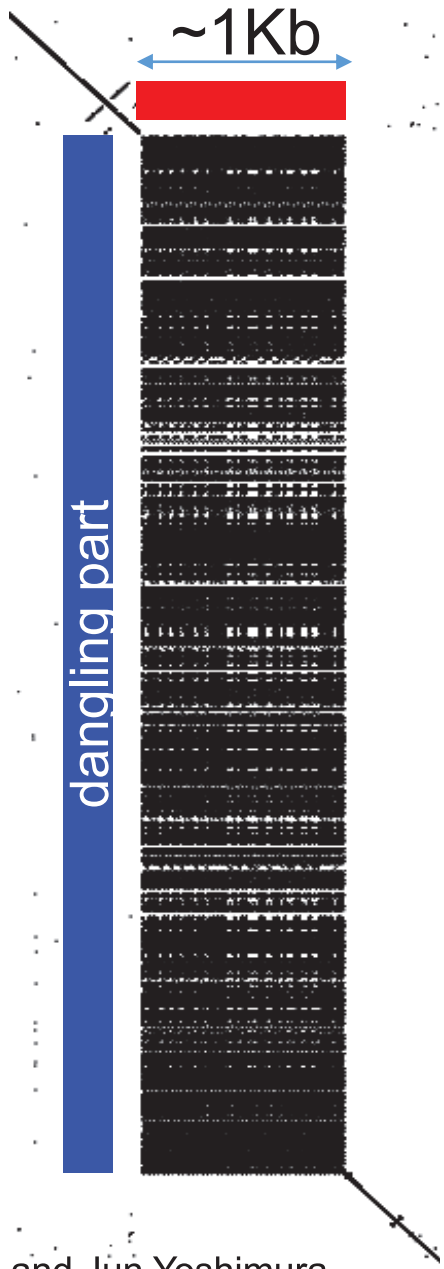
Kazuki Ichikawa and Jun Yoshimura

Can we correct deficits in reference genomes using SMRT sequencing?

Filling gaps in the worm assembly using PBJelly







Koichiro Doi and Jun Yoshimura

Acknowledgments

Pacific Biosciences	Jonas Korlach, Steve Turner
Stanford University	Andrew Fire
Washington University	Evan Eichler
MIT	Yoshikazu Furuta
University of Tokyo	
School of Medicine	Shoji Tsuji
School of Science	Hiroyuki Takeda, Ryohei Nakamura
Hattori Group	Masahira Hattori, Suguru Nishijima
Morishita Group	Koichiro Doi, Jun Yoshimura, Kazuki Ichikawa, Junko Taniguchi, Yuta Suzuki, Yoshihiko Suzuki, Shingo Tomioka