

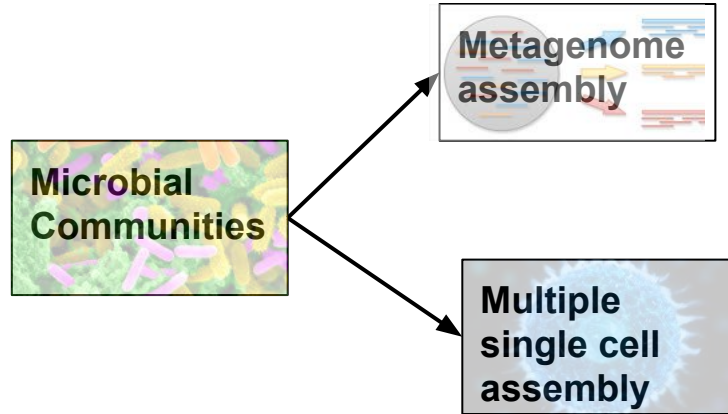


# Co-Assembly of Genomes from Multiple Single Cells

Vera Meleshko, Anton Bankevich  
Center for Algorithmic Biotechnology

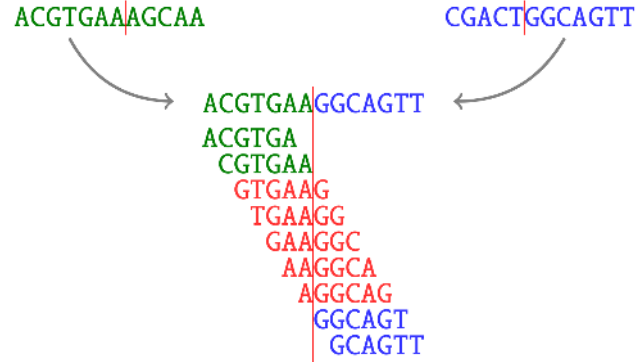
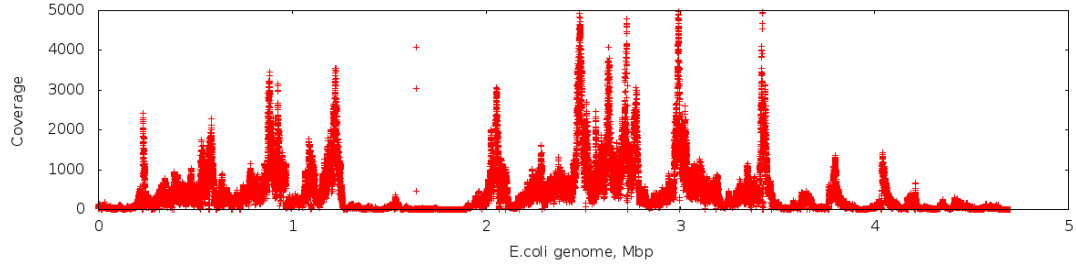


# Story



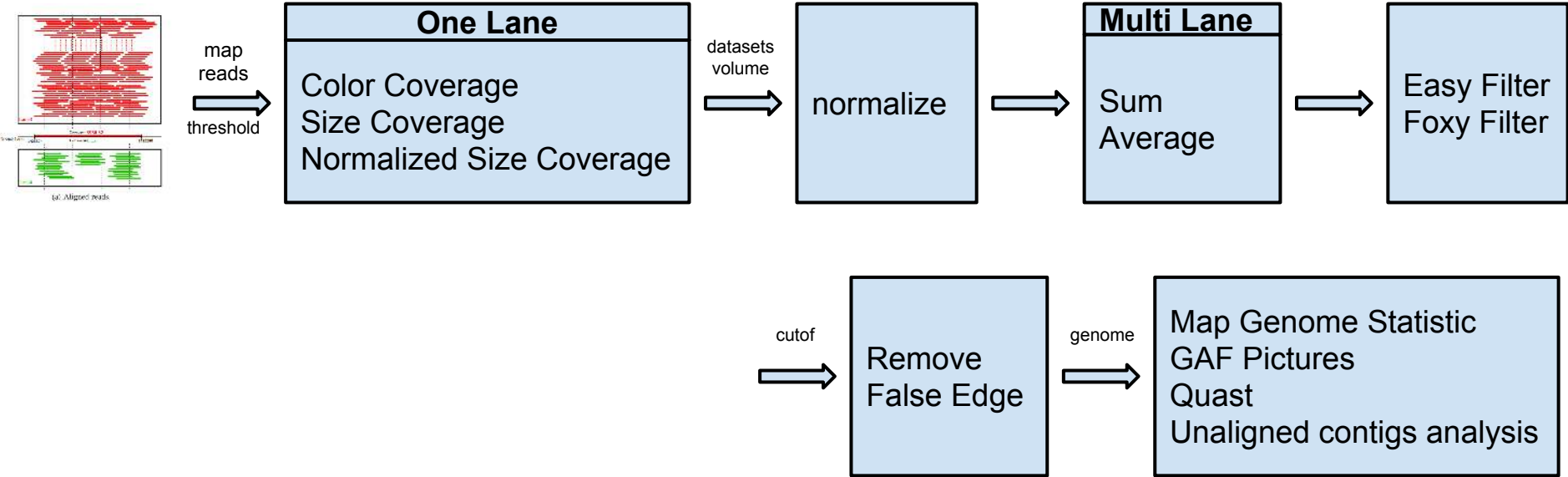
# Problem

- coverage bias
- errors caused by MDA

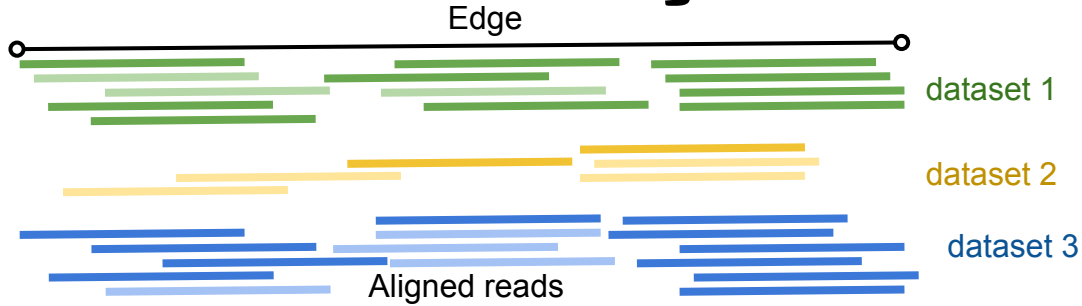


**Goal:** Algorithm for de novo co-assembly of genomes from multiple single cells.

# Algorithms & Methods



# Multi Lane Coverage



## One Lane Coverage

Color Coverage - number of reads aligned to edge for each dataset

Normalized Size Coverage - total aligned length normalized by edge length

## Multi Lane Coverage

Total - # datasets with Color Coverage > threshold

Average - average of One Lane Coverages excluding the most abundant

## Filter

Easy - Multi Lane Coverage

Foxy - most abundant / Average Multi Lane Coverage

# Results. Quast Report.

# Map Genome Statistic

Statistics without reference	master_single_top	master_mix_notop	master_mix_top	our
N50	105630	105582	117600	<b>133089</b>
<b>Misassemblies</b>				
# misassemblies	2	2	<b>0</b>	<b>0</b>
# fully unaligned contigs	97	225	199	<b>26</b>
<b>Genome statistics</b>				
Genome fraction (%)	94.96	98.286	98.294	<b>98.311</b>

#genomic	1158
#nongenomic	1568
true positive	1380/2726
false positive	16/2726
true negative	1142/2726
false negative	188/3729
sensitivity	0.88
specificity	0.986

# Results. Quast Report.

Statistics without reference	master_single_top	master_mix_top	our
N50	10339	112344	<b>132750</b>
<b>Misassemblies</b>			
# misassemblies	52	<b>0</b>	<b>0</b>
# fully unaligned contigs	34	222	<b>0</b>
<b>Genome statistics</b>			
Genome fraction (%)	33.344	<b>98.271</b>	97.653

# Map Genome Statistic

#genomic	2444
#nongenomic	3253
true positive	3103/5697
false positive	30/5697
true negative	2414/5697
false negative	150/5697
sensitivity	0.954
specificity	0.988

# Benchmark

Statistics without reference	SPAdes	HyDA
N50	133089	76428
<b>Misassemblies</b>		
# misassemblies	0	2
# fully unaligned contigs	26	117
<b>Genome statistics</b>		
Genome fraction (%)	98.311	97.766

Statistics without reference	SPAdes	HyDA
N50	132750	60662
<b>Misassemblies</b>		
# misassemblies	0	3
# fully unaligned contigs	0	7
<b>Genome statistics</b>		
Genome fraction (%)	97.653	96.97



# Future work

- Application of developed algorithms to variety of bacterial datasets (including metagenomic single-cell data)
- Application and tuning of developed algorithms for eukaryote (fungi) organisms
- Writing a paper



# Thank you!

[https://github.com/ablab/algorithmic-biology/tree/branch\\_v](https://github.com/ablab/algorithmic-biology/tree/branch_v)

