

Sample mix-ups and mixtures in microbiome data in DO mice

Karl Broman

Biostatistics & Medical Informatics
Univ. Wisconsin–Madison

kbroman.org

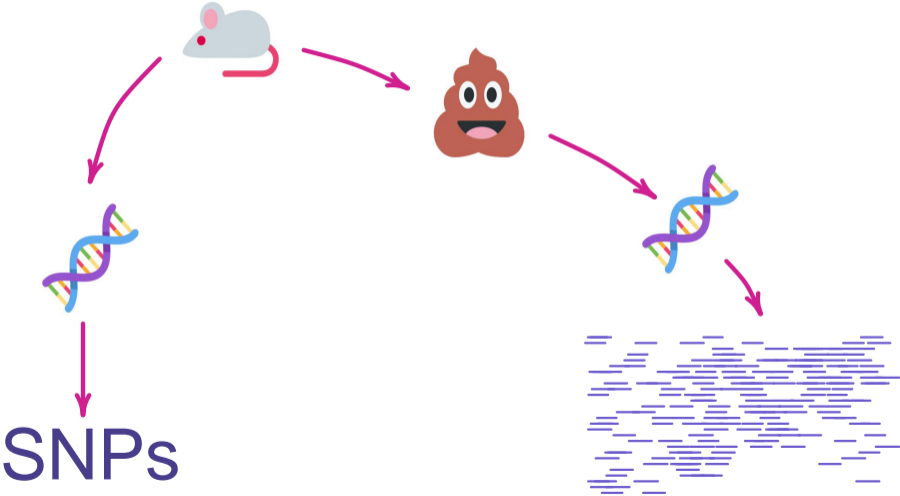
github.com/kbroman

@kwbroman

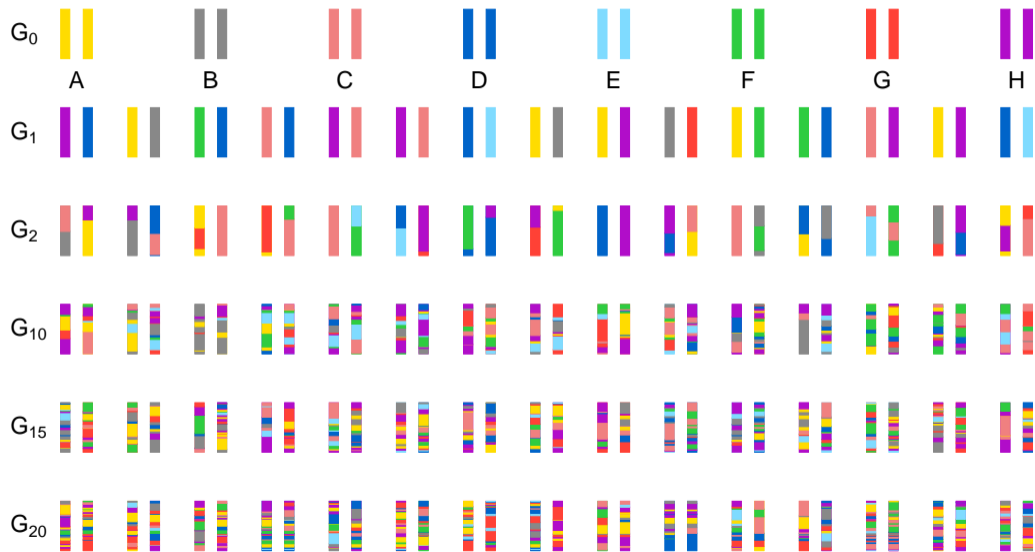
Slides: bit.ly/2019CTC



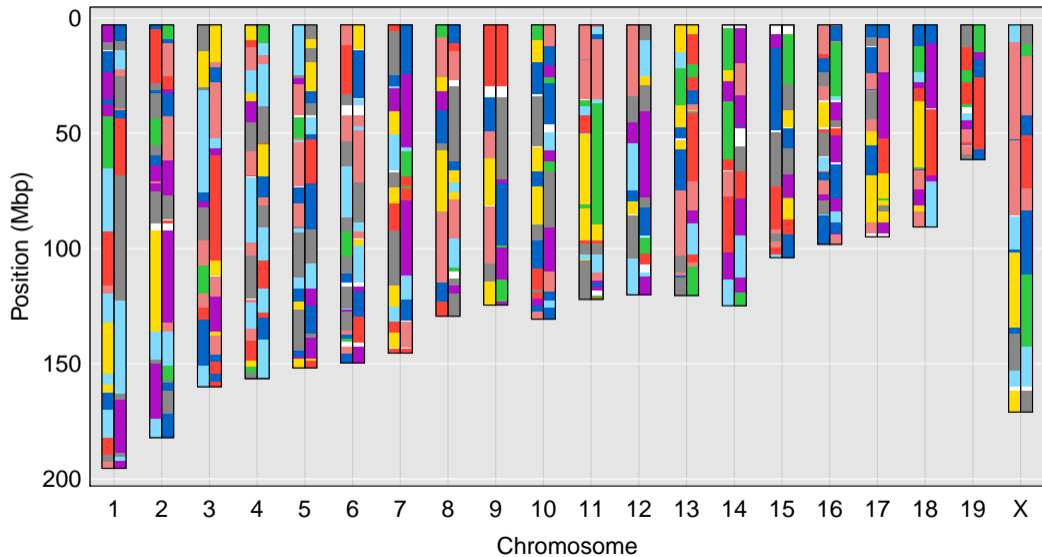
Microbiome genetics data



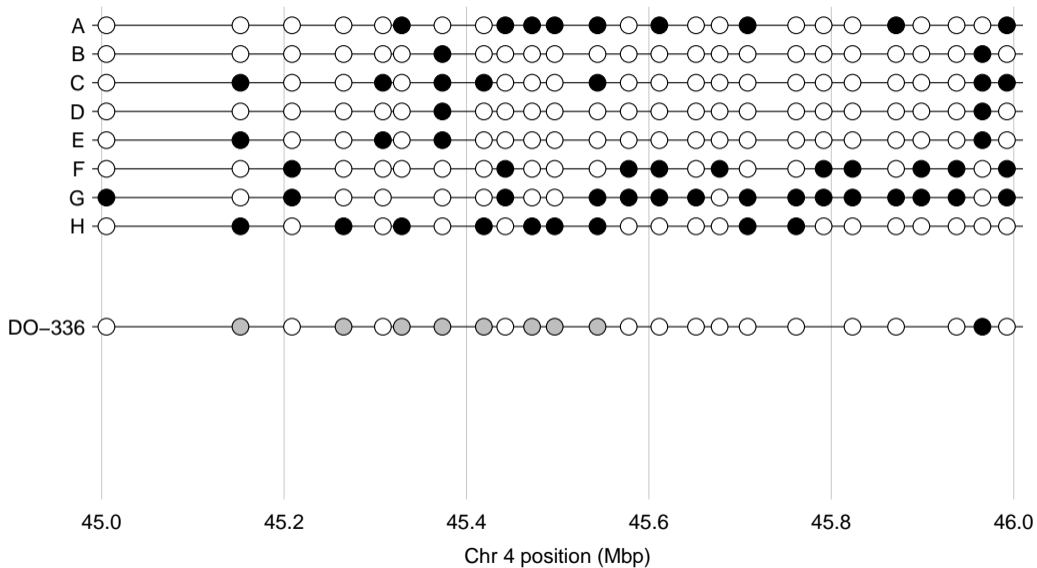
Multi-parent advanced intercross population



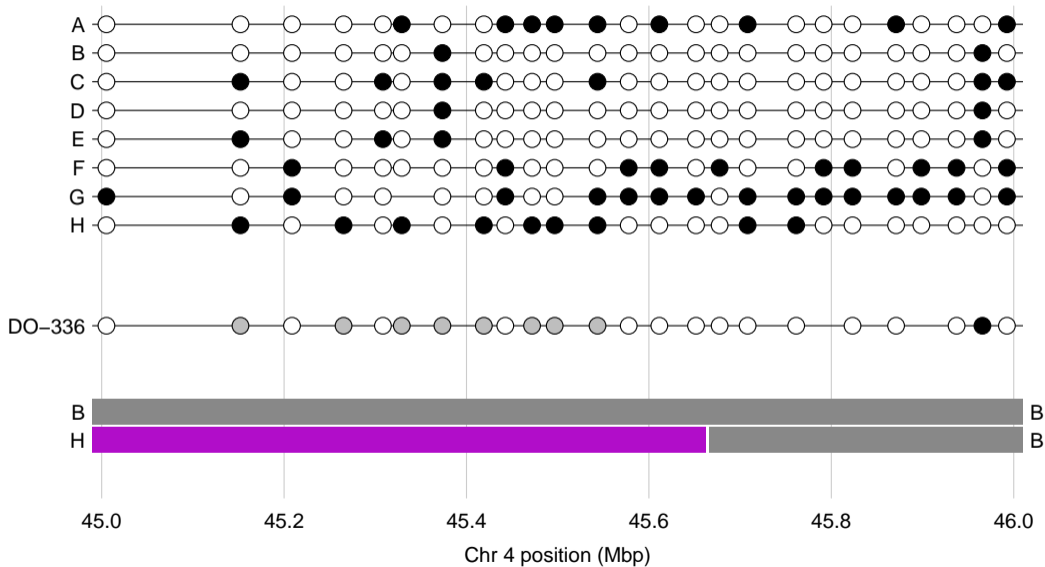
Genome of a diversity outbred mouse



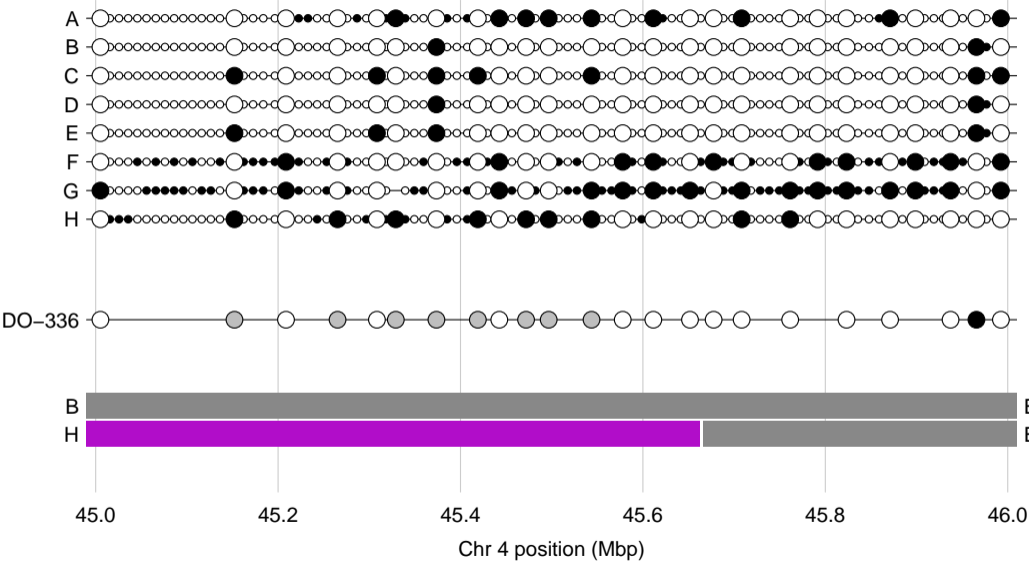
Genotype reconstruction



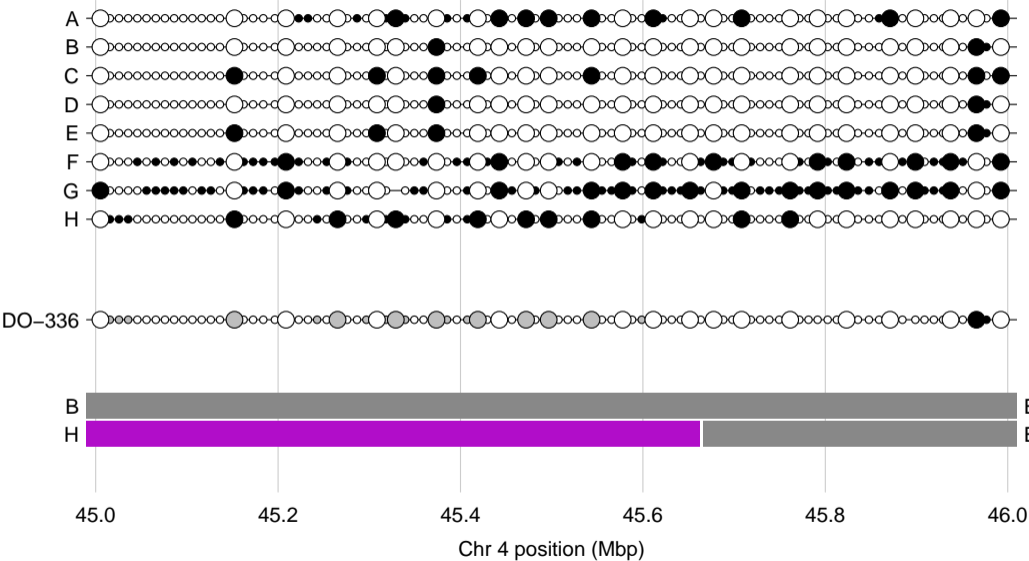
Genotype reconstruction



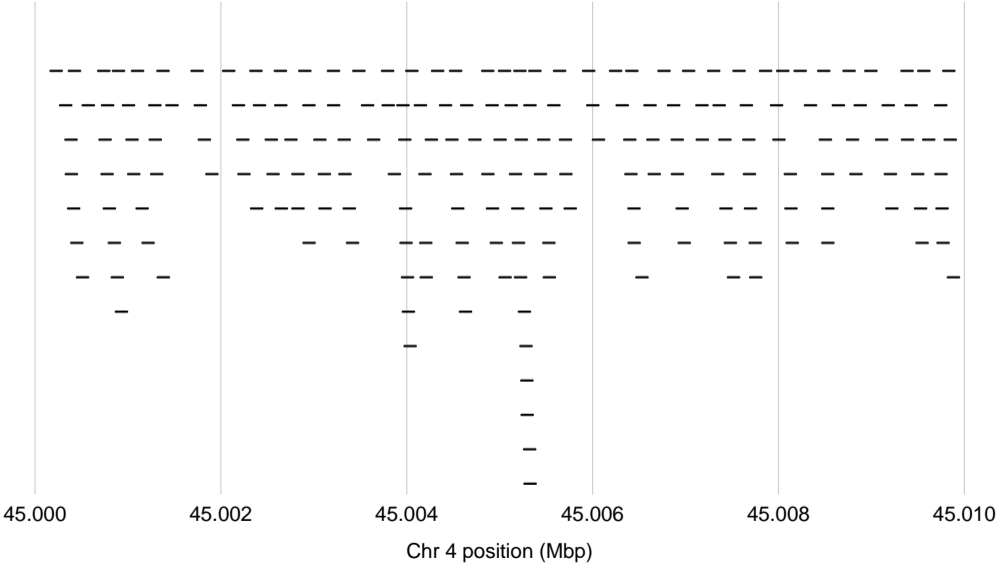
Genotype reconstruction



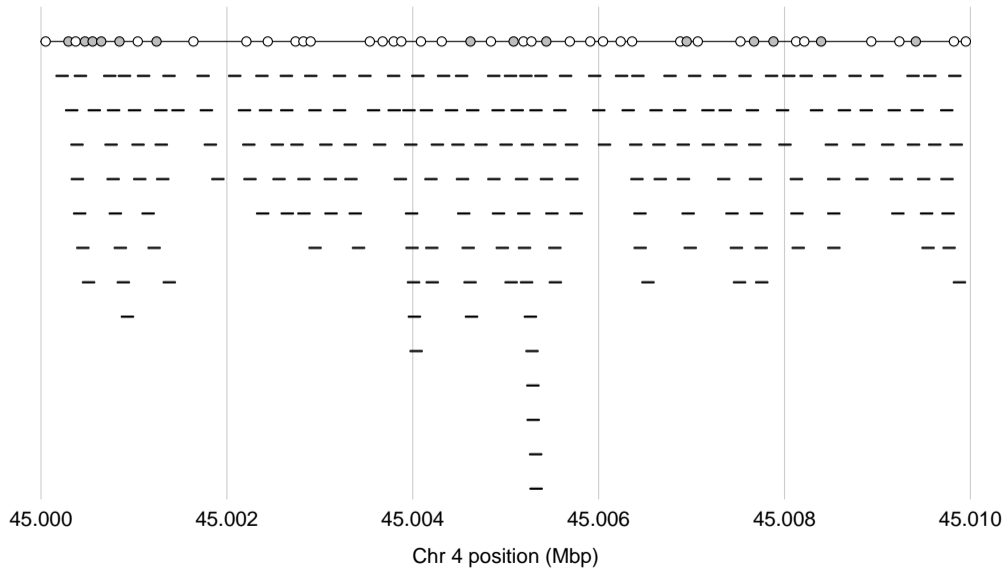
Genotype reconstruction



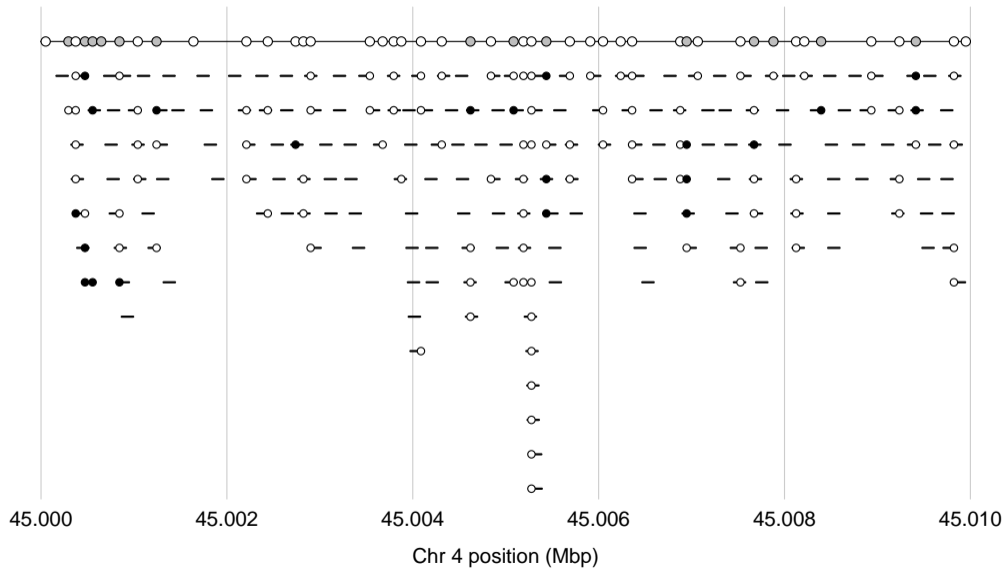
Mapped reads



Mapped reads



Mapped reads



Genomic DNA vs microbiome reads

	microbiome DNA	
genomic DNA	A	B
AA		
AB		
BB		

Genomic DNA vs microbiome reads

genomic DO-381 vs microbiome DO-381

genomic DNA	microbiome DNA	
	A	B
AA	2,762,341	7,303
AB	606,312	578,017
BB	2,128	375,559

Genomic DNA vs microbiome reads

genomic DO-381 vs microbiome DO-381

genomic DNA	microbiome DNA	
	A	B
AA	2,762,341	7,303
AB	606,312	578,017
BB	2,128	375,559

percent mismatch = 0.3%

Genomic DNA vs microbiome reads

genomic DO-360 vs microbiome DO-360

genomic DNA	microbiome DNA	
	A	B
AA	8,863,572	1,520,169
AB	2,870,063	1,075,126
BB	671,722	536,010

percent mismatch = 19%

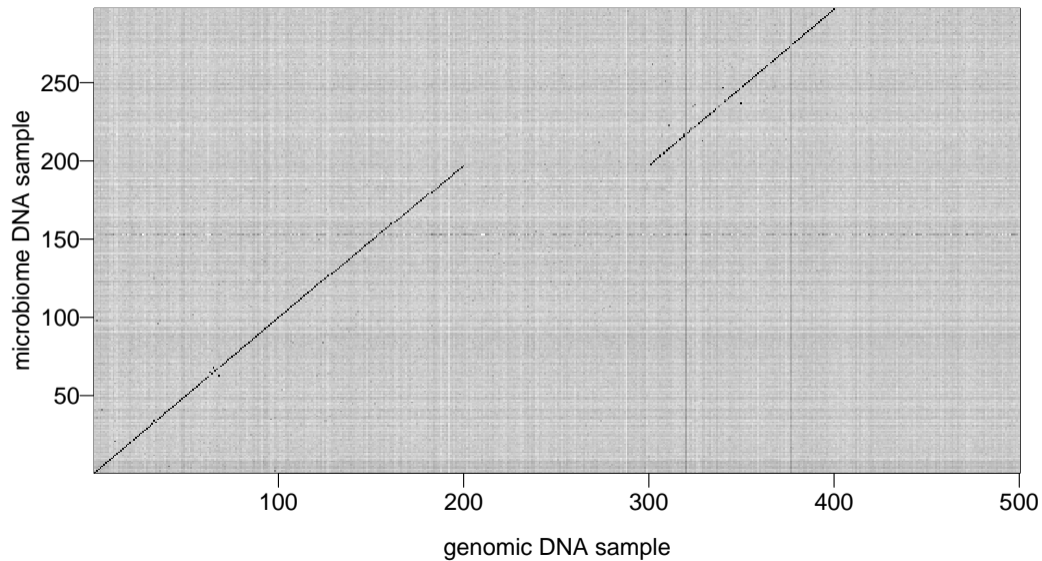
Genomic DNA vs microbiome reads

genomic DO-370 vs microbiome DO-360

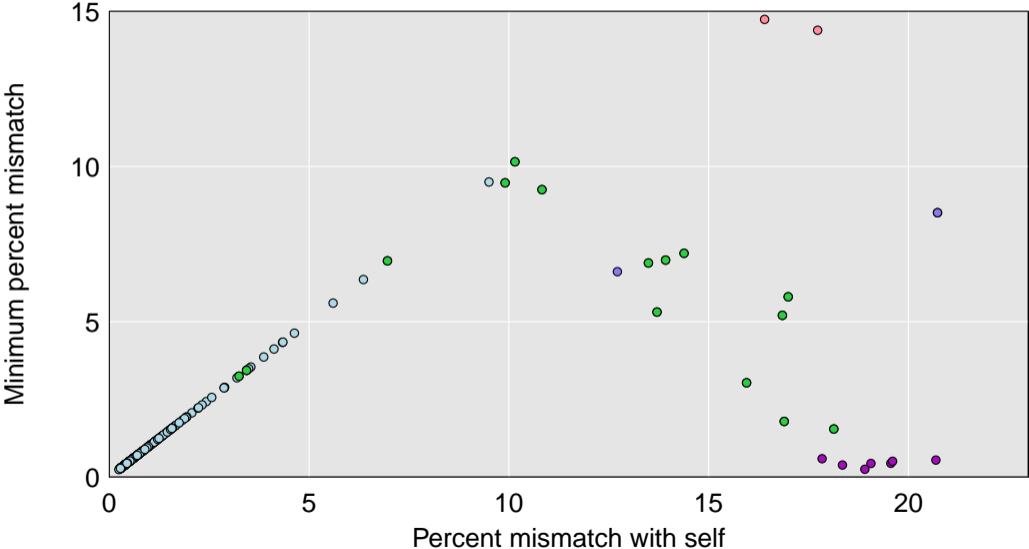
genomic DNA	microbiome DNA	
	A	B
AA	10,324,265	23,256
AB	2,083,947	1,986,380
BB	5,347	1,117,994

percent mismatch = 0.2%

Distance matrix

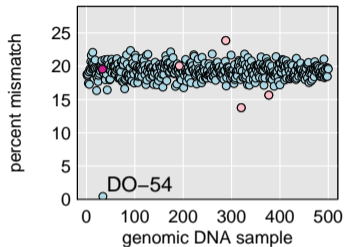


Minimum vs. self distance

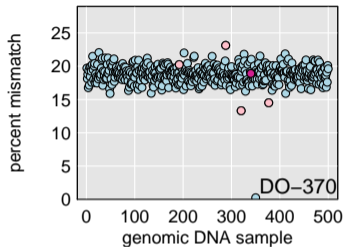


Selected samples

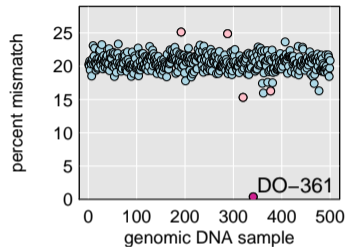
microbiome sample DO-53



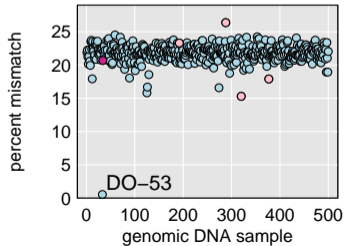
microbiome sample DO-360



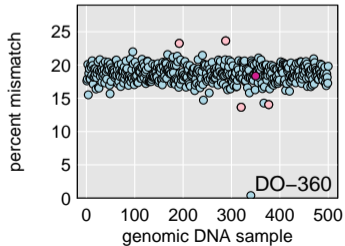
microbiome sample DO-361



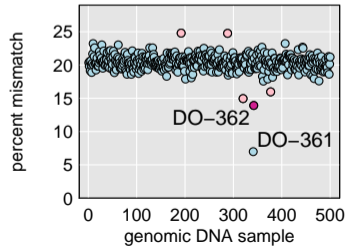
microbiome sample DO-54



microbiome sample DO-370



microbiome sample DO-362



Genotype pair vs microbiome reads

genomic DO-362 and DO-361 vs microbiome DO-362

DO-362:

AA

AB

BB

DO-361

A

B

DO-361

A

B

DO-361

A

B

AA

1,306,406

3,688

AA

395,562

75,326

AA

41,593

19,813

AB

339,397

162,432

AB

135,825

128,895

AB

25,798

49,487

BB

33,661

63,318

BB

14,836

71,817

BB

529

93,162

Genotype pair vs microbiome reads

genomic DO-362 and DO-361 vs microbiome DO-362

DO-362:

AA

AB

BB

DO-361

A

B

DO-361

A

B

DO-361

A

B

AA

99.7%

0.3%

AA

84.0%

16.0%

AA

67.7%

32.3%

AB

67.6%

32.4%

AB

51.3%

48.7%

AB

34.3%

65.7%

BB

34.7%

65.3%

BB

17.1%

82.9%

BB

0.6%

99.4%

Genotype pair vs microbiome reads

genomic DO-361 and DO-362 vs microbiome DO-361

DO-361:

AA

AB

BB

DO-362

A

B

DO-362

A

B

DO-362

A

B

AA

2,162,059

6,503

AA

428,229

405,190

AA

1,623

155,276

AB

788,011

2,806

AB

223,680

215,833

AB

1,126

140,739

BB

104,511

511

BB

64,082

62,658

BB

1,103

151,334

Genotype pair vs microbiome reads

genomic DO-361 and DO-362 vs microbiome DO-361

DO-361:

AA

AB

BB

DO-362

A

B

DO-362

A

B

DO-362

A

B

AA

99.7%

0.3%

AA

51.4%

48.6%

AA

1.0%

99.0%

AB

99.6%

0.4%

AB

50.9%

49.1%

AB

0.8%

99.2%

BB

99.5%

0.5%

BB

50.6%

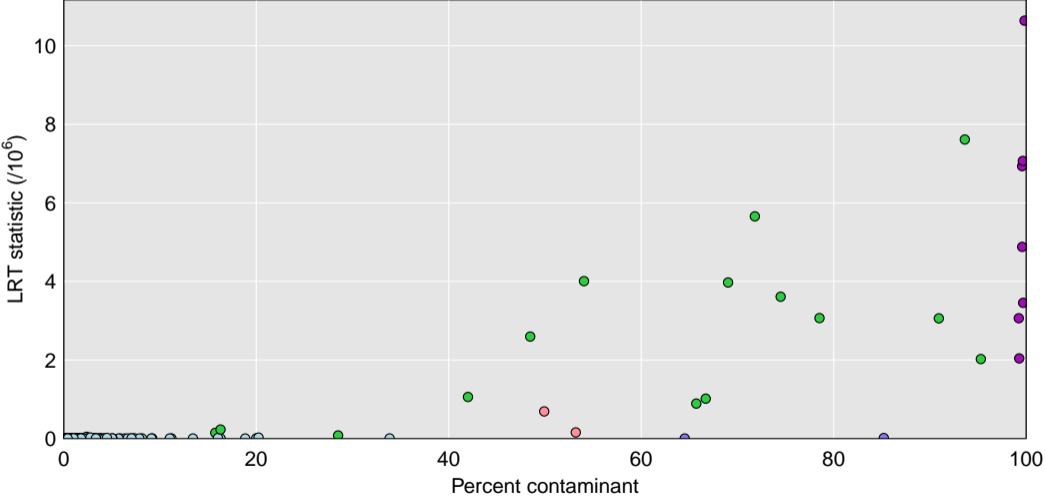
49.4%

BB

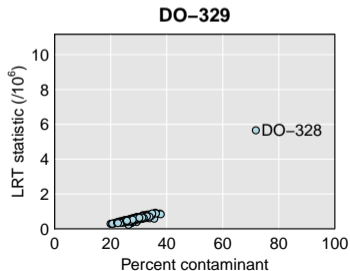
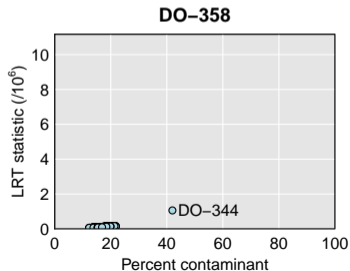
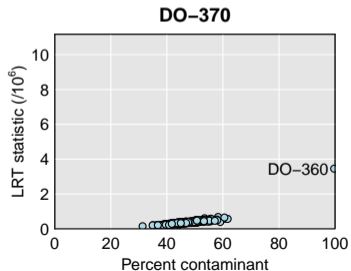
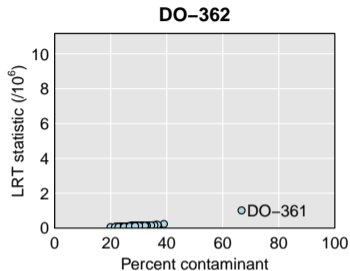
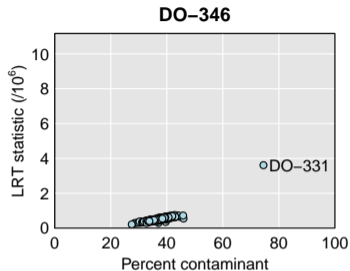
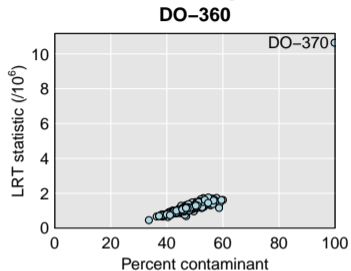
0.7%

99.3%

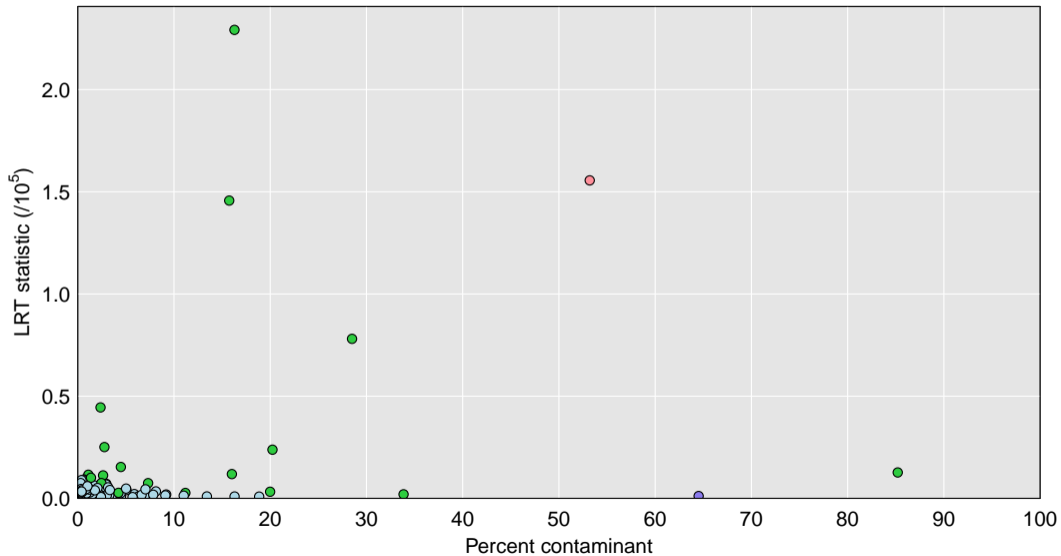
log likelihood vs percent contaminant



Selected samples



Is everything contaminated?



Summary

- ▶ Microbiome shotgun reads include reads from the host
- ▶ With such data, sample mix-ups can be identified
- ▶ Simple method:
 - Impute genotype at all SNPs
 - Count alleles in reads overlapping SNPs
 - Focus on homozygous SNPs and calculate percent discordant reads
- ▶ We also saw strong evidence for many samples being mixtures

Acknowledgments

- ▶ Lindsay Traeger
- ▶ Alexandra Lobo
- ▶ Federico Rey
- ▶ Alan Attie, Mark Keller, Gary Churchill, Brian Yandell
- ▶ NIH: NIDDK, NIGMS

Slides: bit.ly/2019CTC
bit.ly/2019CTC_notes



bioRxiv manuscript: doi.org/10.1101/529040

kbroman.org

github.com/kbroman

@kwbroman