

Low Parasitemia Report

Jacob Westaway

Last updated on 2021-08-06

About

After sharing the previous comparisons with Matt G, it was noted that none of the samples in the previous subset had low parasitemia (< 5000). Thus, another subset of only low parasitemia samples (16) was created, to determine if ZB's methods capture enough depth and coverage at lower parasitemia. These samples range from ~1040 - 4157.

Key terms/abbreviations:

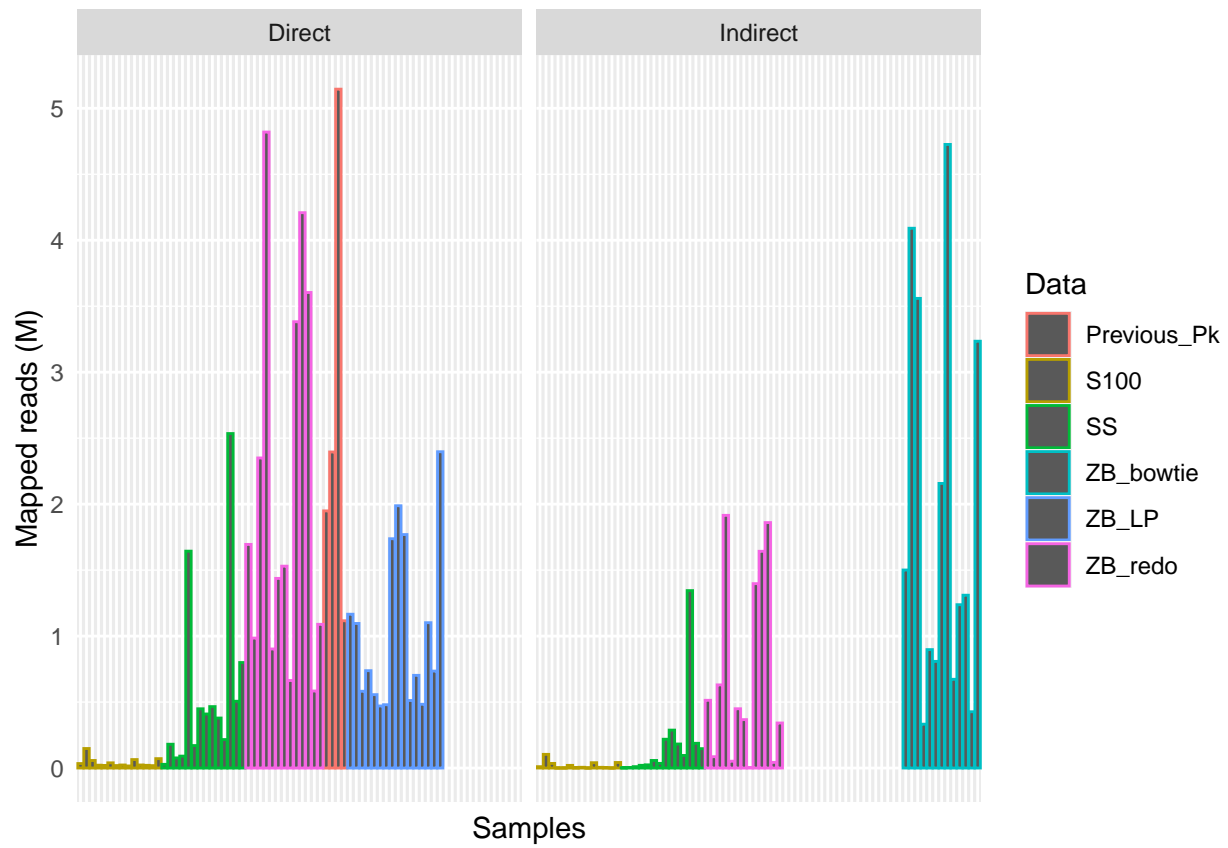
Data:

- ZB_LP: data from Singapore/Zbynek Bozdech with low parasitemia.
- ZB: data from Singapore/Zbynek Bozdech (truncated).
- ZB_redo: data from Singapore/Zbynek Bozdech (not truncated).
- SS: initial subset from Matt Grigg.
- S100: data from Sanger.
- Previous_Pk: data from a previous study provided by Ernest.
- Direct: aligning/mapping to Pk genome without removal of human contamination.
- Indirect: aligning/mapping to Pk after removal of human contamination.

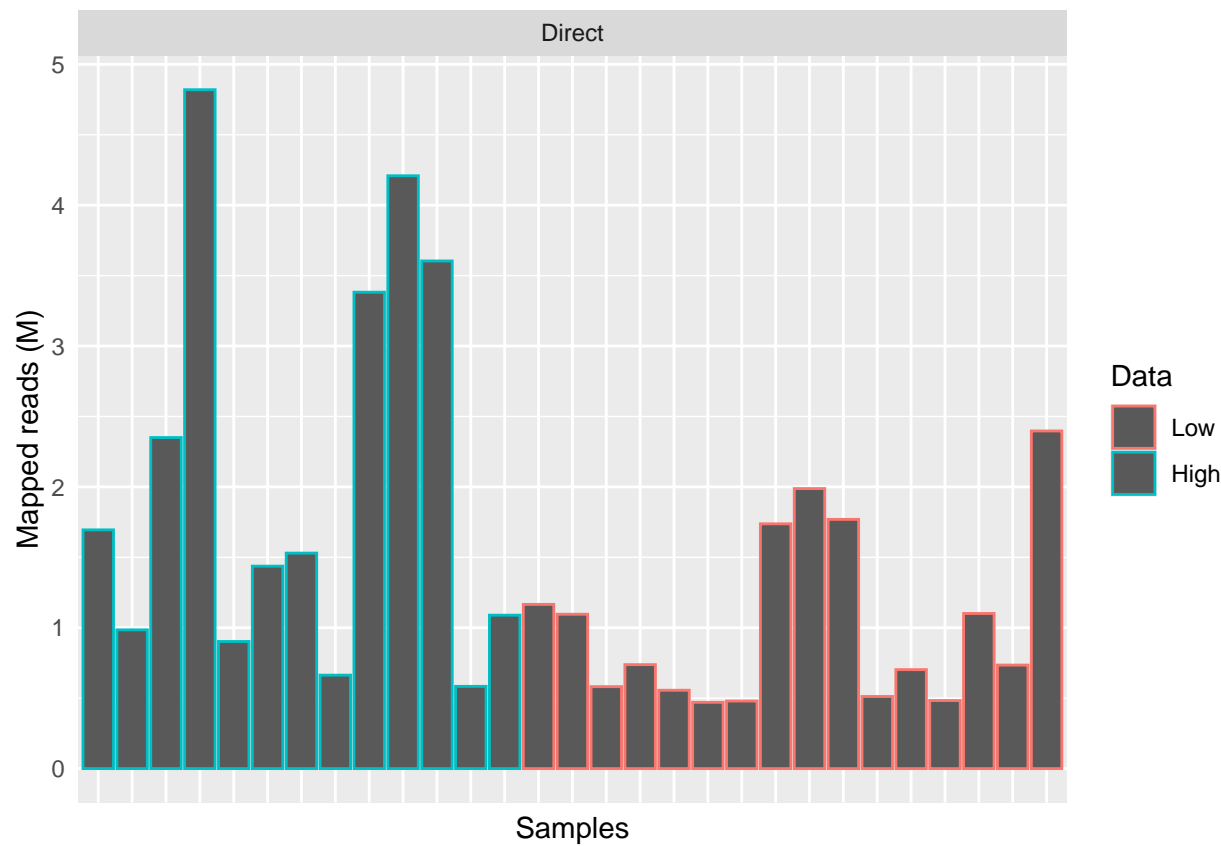
Metrics:

- Reads - number of reads from sequencing.
- Mapped reads - number of reads that are aligning to the reference genome.
- Depth - the average coverage for reads aligning to the genome.

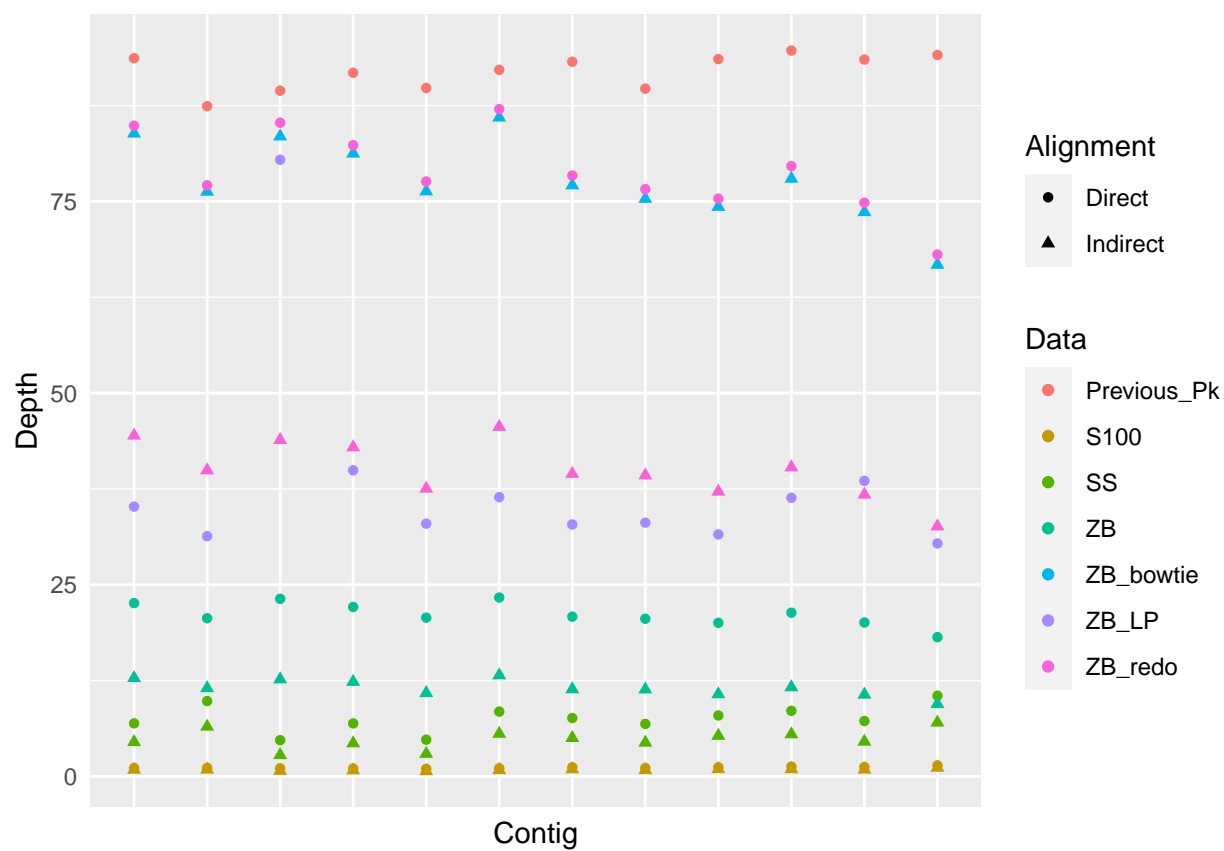
Comparison of the number reads mapping to Pk across all datasets and for different alignments



Plot comparing mapped reads for ZB data, and high and low parasitemia



Plot of the average read depth per contig



IGV

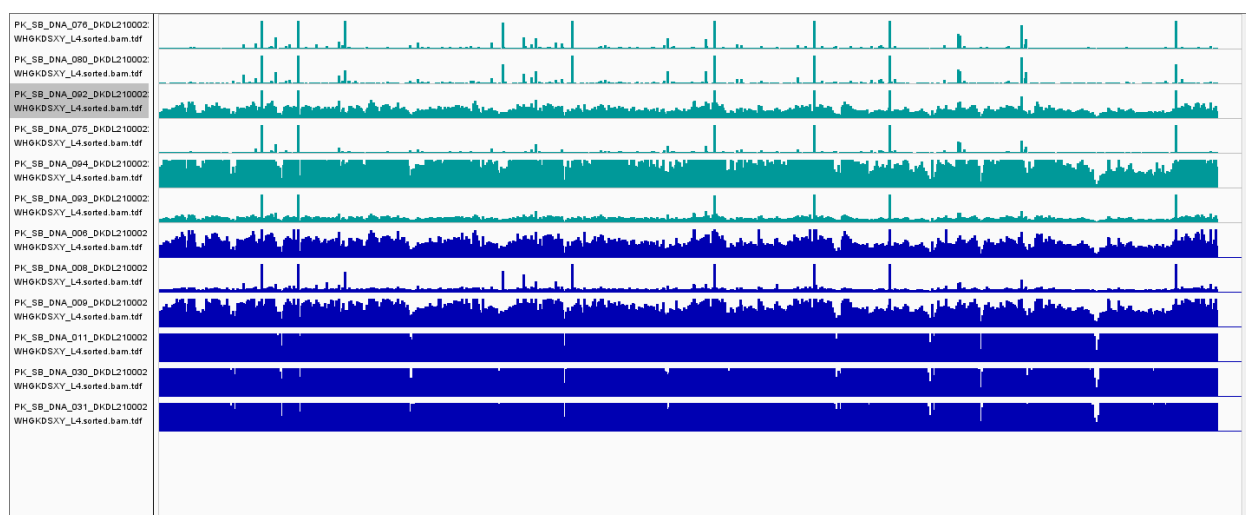
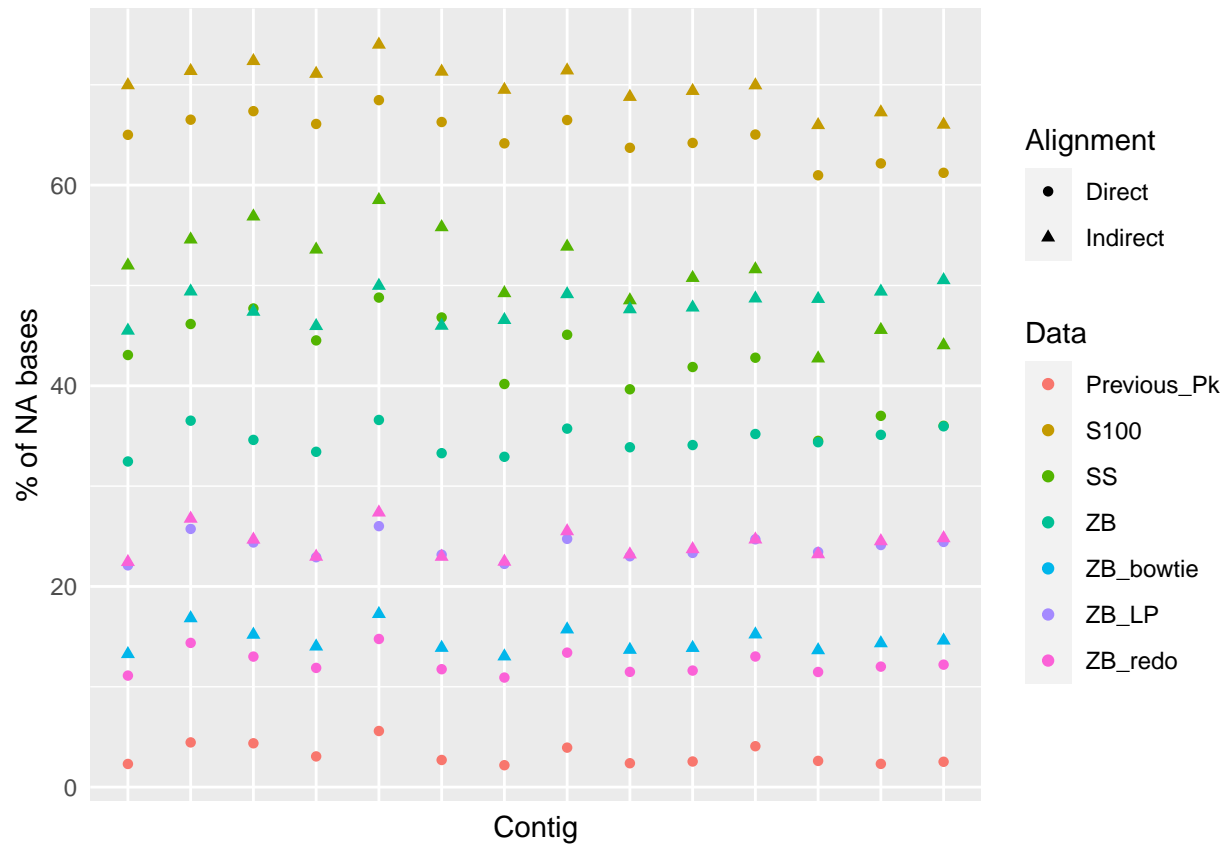


Figure 1: Comparison of the distribution of read depth across the genome for high and low parasitemia (direct alignments)

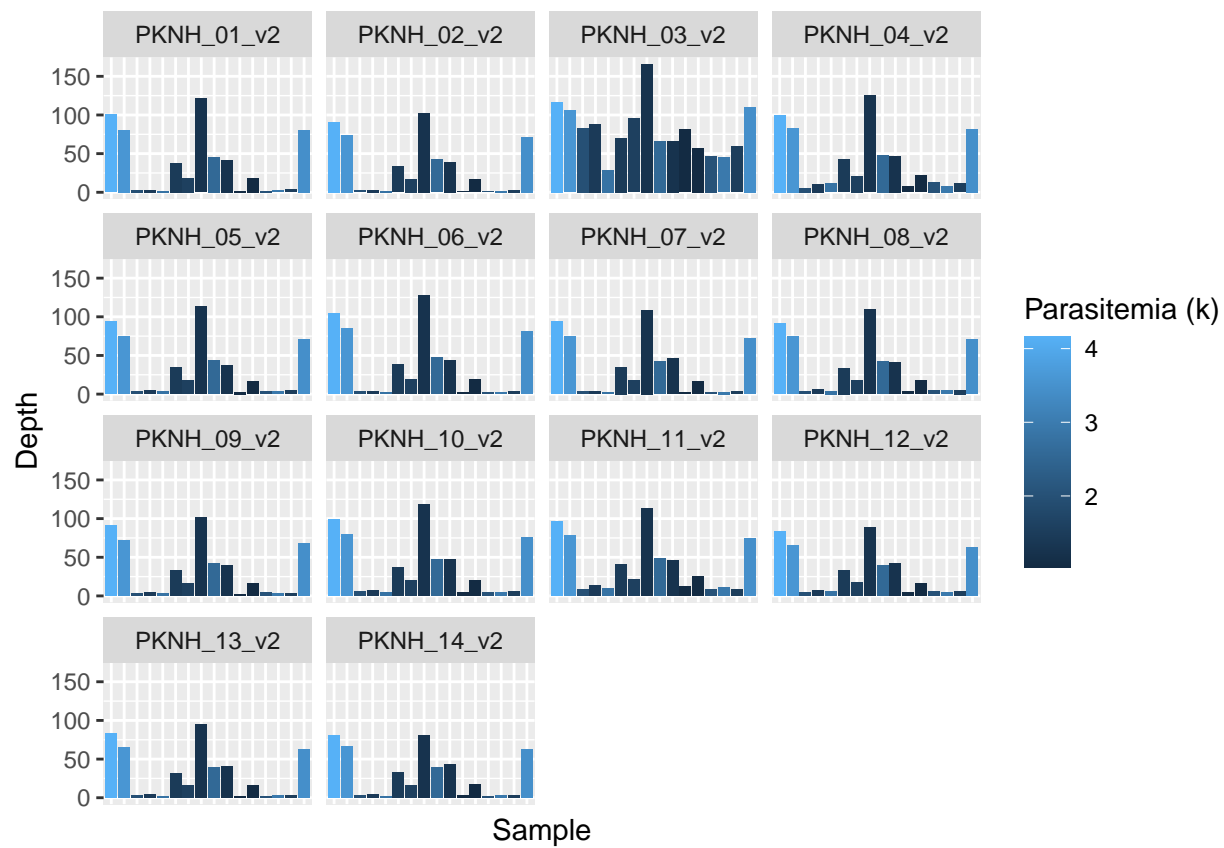
Coloured by data:

- Low parasitemia = light blue.
- High parasitemia = dark blue.

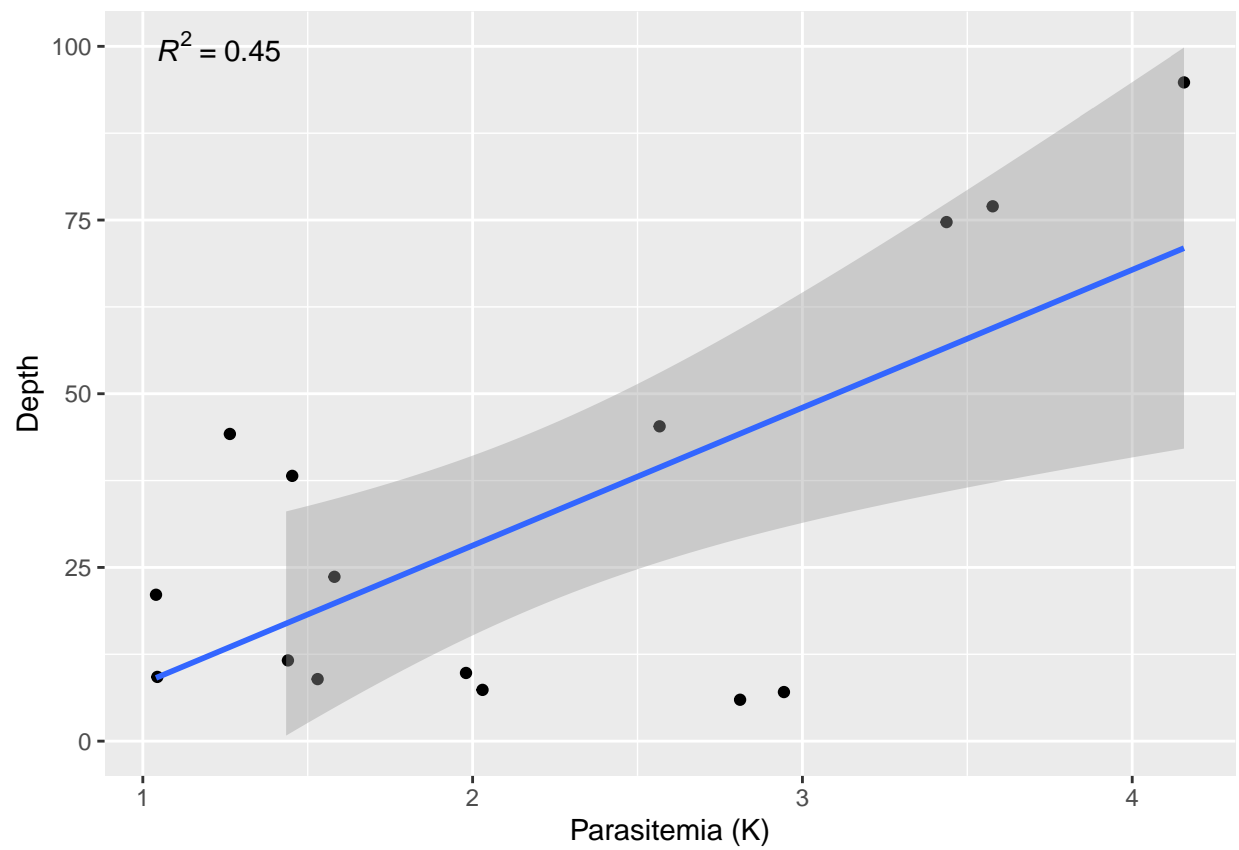
Plot of the percentage of bases **WITHOUT** coverage



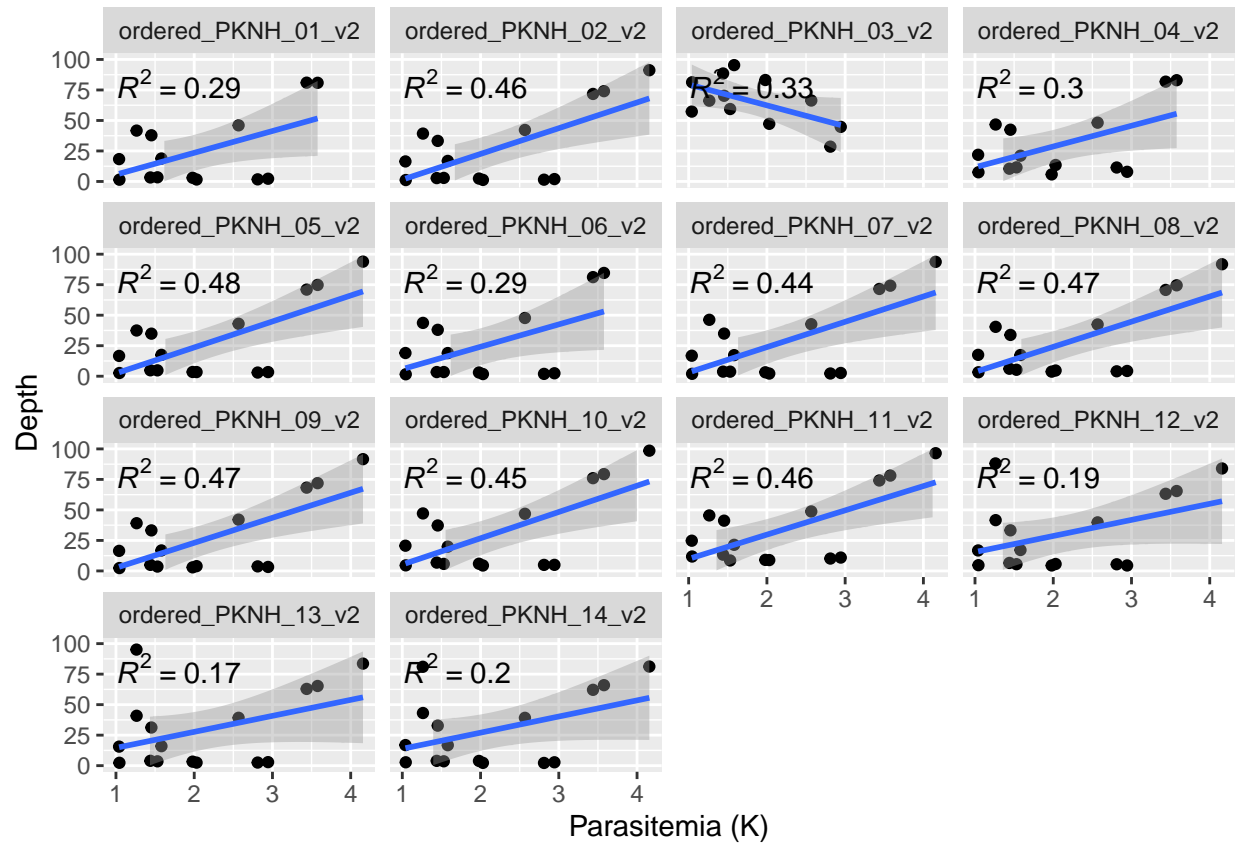
The depth across samples at each contig, coloured by parasitemia



Plot of the relationship between depth and parasitemia (low parasitemia) - outlier removed



Plot relationship between depth and parasitemia (low parasitemia),
 faceted by contig and outlier removed



Summary of the mapping statisitcs across samples, ordered by parasitemia

ID	Reads	Mapped.reads	Percent.mapped	Depth	Parasitemia
076	73858676	7024811	9.511	21.070746	1040.00
075	59958738	5118642	8.537	9.245902	1044.00
070	60914700	17379057	28.530	112.284528	1262.00
074	60758322	17694310	29.122	44.212491	1264.00
063	65971272	7379423	11.186	11.621332	1440.00
065	57116811	4705862	8.239	38.177380	1453.00
093	50249766	7348807	14.625	8.930064	1530.00
066	60582605	4797638	7.919	23.653912	1581.00
061	62202281	5818429	9.354	9.814160	1980.00
080	59974937	4831837	8.056	7.380827	2030.00
073	63718510	19879597	31.199	45.310426	2567.00
064	57174594	5562223	9.728	5.957668	2811.00
092	69368532	11013785	15.877	7.065444	2944.00
094	69643922	23971740	34.420	74.705439	3437.00
058	59641254	10958761	18.374	76.976119	3577.00
057	65699593	11658102	17.745	94.807859	4157.00
009	68641810	23501174	34.237	74.835392	5522.00
008	60161476	9850931	16.374	11.782071	6450.00
019	64389558	6634489	10.304	2.640097	6607.00
006	63478370	16948578	26.700	56.567076	17834.00
034	55339553	10892058	19.682	34.603467	19759.00
023	72279366	33823852	46.796	144.815271	21368.00
033	67368656	5838670	8.667	6.126351	34778.64
031	47086256	36040788	76.542	176.228271	48339.00
030	70575585	42089050	59.637	178.268707	60000.00
012	68025812	9023684	13.265	7.625737	123590.40
014	65440877	14374321	21.965	47.996346	249994.20
015	66491135	15299426	23.010	43.210637	340953.80
011	70371619	48198532	68.491	220.941714	584014.90

Avergaes for high vs low parasitemia

Data	High	Low
Reads	64588467	62302157
Mapped.reads	20962735	10321439
Percent.mapped	32.74385	16.40138
Average.coverage	83.08323	35.81437
Percent.bases.covered	88.19154	69.49062
Parasitemia	116862.380	2132.312
Depth	77.35701	36.95089

Key Points:

- The low parasitemia samples, on average, have roughly half the number of mapped reads and read depth, but still enough to call variants.
- The low parasitemia samples, on average, also have less coverage.
- Despite a general trend between read depth and parasitemia the metrics are not consistent across samples, with large amount of variability between samples.
- Thus, parasitemia does not appear to be a good metric for inclusion/exclusion of samples.
- It could be helpful to know how parasitemia is measured, how ZB processes his samples and how the metadata is collected, as this may help us to understand why we don't see a nice trend.