Barcode data preparation

This document summarizes the data stored in Data_store_barcode and how they were processed. Consensus sequences were previously called for isolates considered monoclonal (isolates with 1-5 heteroallelic genotyping calls - Fig 1), replacing any het calls with 'N' (Fig 2), akin to a sequencing error (Nkhoma et al. 2013). Hereafter minor/major SNPs are encoded 0 and 1, respectively, where minority/majority was based on proportions discarding mixed and ignoring missing.

Nkhoma et al. (2013) report

- 1731 infections overall and here we have: 1731
- 629 from Wang Pha and here: 629 from WPA
- 396 from Maela and here: 396 from MLA
- 146 from Mae Kon Ken and here: 146 from PLU
- 560 from Mawker-Thai and here: 560 from MKT

To conclude: MLA (Maela); WPA (Wang Pha); PLU (Mae Kon Ken); MKT (Mawker Thai)

Of the data imported from TableS3_Nkhoma_et_al.txt

- Percentage declared multiclonal: 32.24
- Percentage with missing: 16.18
- Number with missing: 280
- Number with missing excluding the samples with ≤ 5 het calls: 71

In the analyses reported in the main manuscript, which are all based on either Data_store_barcode.RData, Barcode93.txt or Barcode24.txt (generated in S4Appendix.Rmd), we use single-genotype isolates only.

Sample sizes

Table 1: Isolate counts over sites and years, excluding those with missing data

	01	02	03	04	05	06	07	08	09	10	Total
MKT	10	104	88	28	21	11	4	95	79	40	480
MLA	49	67	54	38	2	0	2	47	32	26	317
MKK	0	0	0	0	0	0	2	53	45	17	117
WPA	0	1	0	13	0	0	26	275	151	71	537
Total	59	172	142	79	23	11	34	470	307	154	1451

Table 2: Single-genotype isolate counts over sites and years

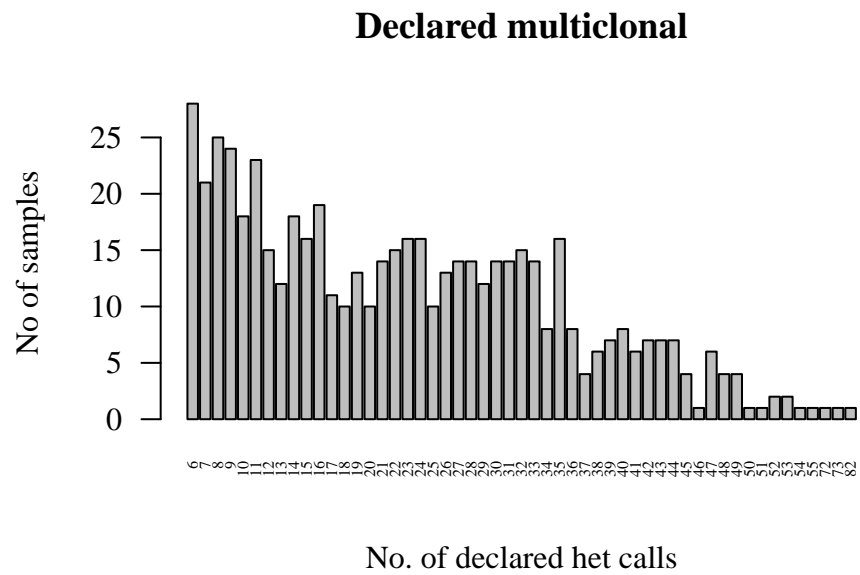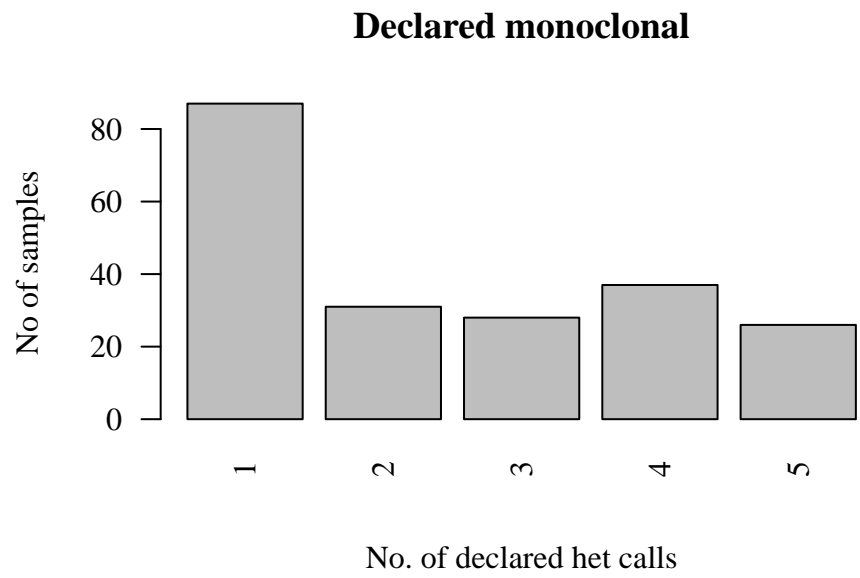|	01	02	03	04	05	06	07	08	09	10	Total
MKT	12	58	67	22	19	12	3	88	70	37	388
MLA	17	39	35	28	2	0	5	37	26	23	212
MKK	0	0	0	0	0	0	4	57	40	15	116
WPA	0	0	0	12	0	0	20	228	120	77	457
Total	29	97	102	62	21	12	32	410	256	152	1173

Figure 1: Number of heteroallelic calls for isolates considered mono and multi-genotype.
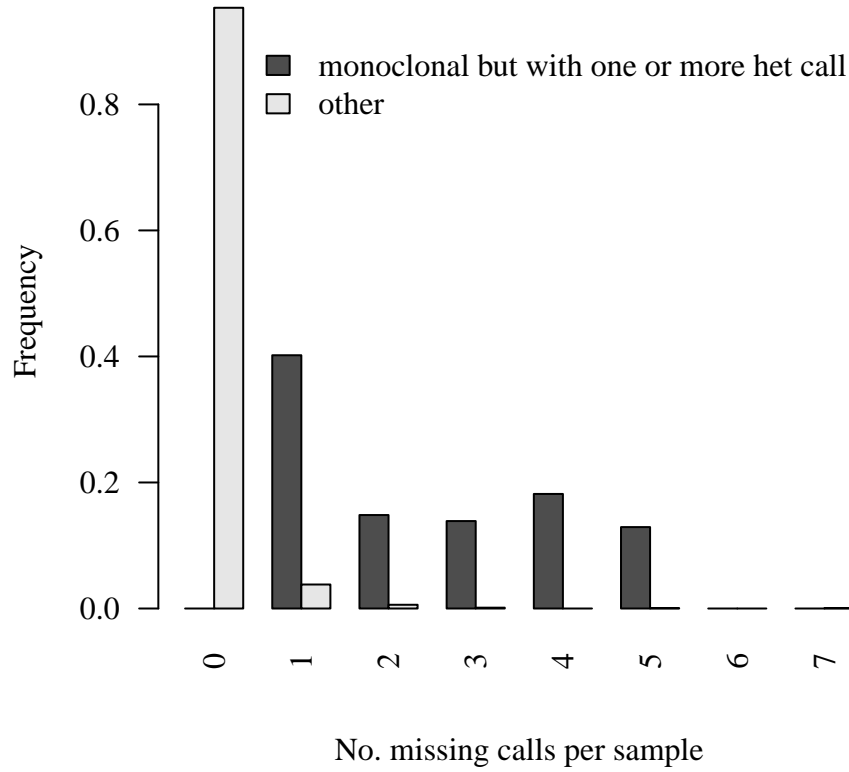
Figure 2: Number of 'N' calls for isolates considered monoclonal but with 1 or more het calls versus all other isolates.

Table 3: Single-genotype isolate counts over sites and years, excluding those with missing data

	01	02	03	04	05	06	07	08	09	10	Total
MKT	8	49	51	19	14	9	2	71	57	34	314
MLA	11	29	21	18	0	0	1	31	23	22	156
MKK	0	0	0	0	0	0	2	42	30	14	88
WPA	0	0	0	6	0	0	17	189	103	59	374
Total	19	78	72	43	14	9	22	333	213	129	932

Table 4: Multi and single-genotype isolate counts over sites and years

	01	02	03	04	05	06	07	08	09	10	Total
MKT	15	115	107	31	26	14	5	112	92	43	560
MLA	63	81	76	51	4	0	6	53	35	27	396
MKK	0	0	1	0	0	0	4	68	55	18	146
WPA	0	1	0	19	0	0	30	316	174	89	629
Total	78	197	184	101	30	14	45	549	356	177	1731

Nkhoma, Standwell C., Shalini Nair, Salma Al-Saai, Elizabeth Ashley, Rose McGready, Aung P. Phyo, Francois Nosten, and Tim J C Anderson. 2013. "Population genetic correlates of declining transmission in a human pathogen." *Molecular Ecology* 22 (2): 273–85. doi:10.1111/mec.12099.