Results_summary

Aimee Taylor

2024-03-11

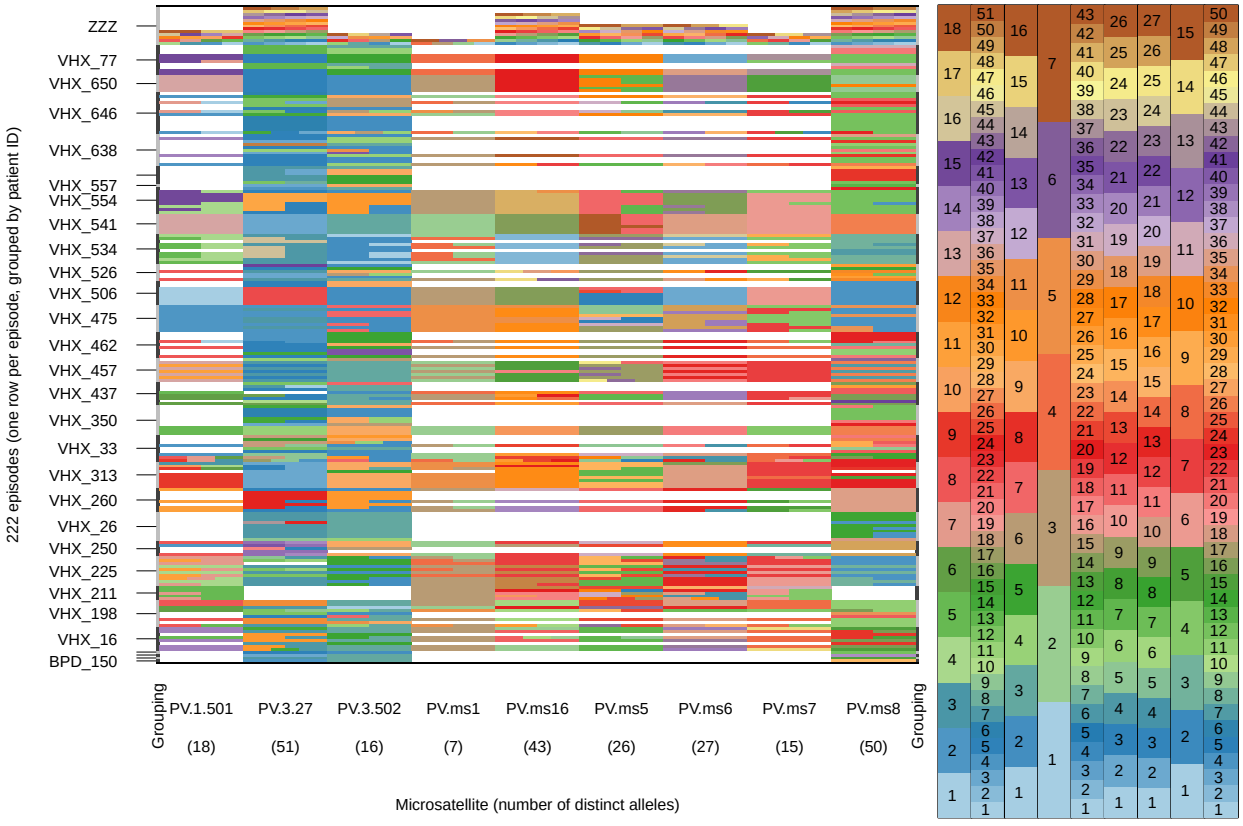
Could-do:

- This study needs updating to new code (e.g., plot_data) if it's going to be used.
- make ColorPlot_MSdata legend representative of allele frequencies
- Plot an example of high posterior relapse that would be classified as reinfection using established methods

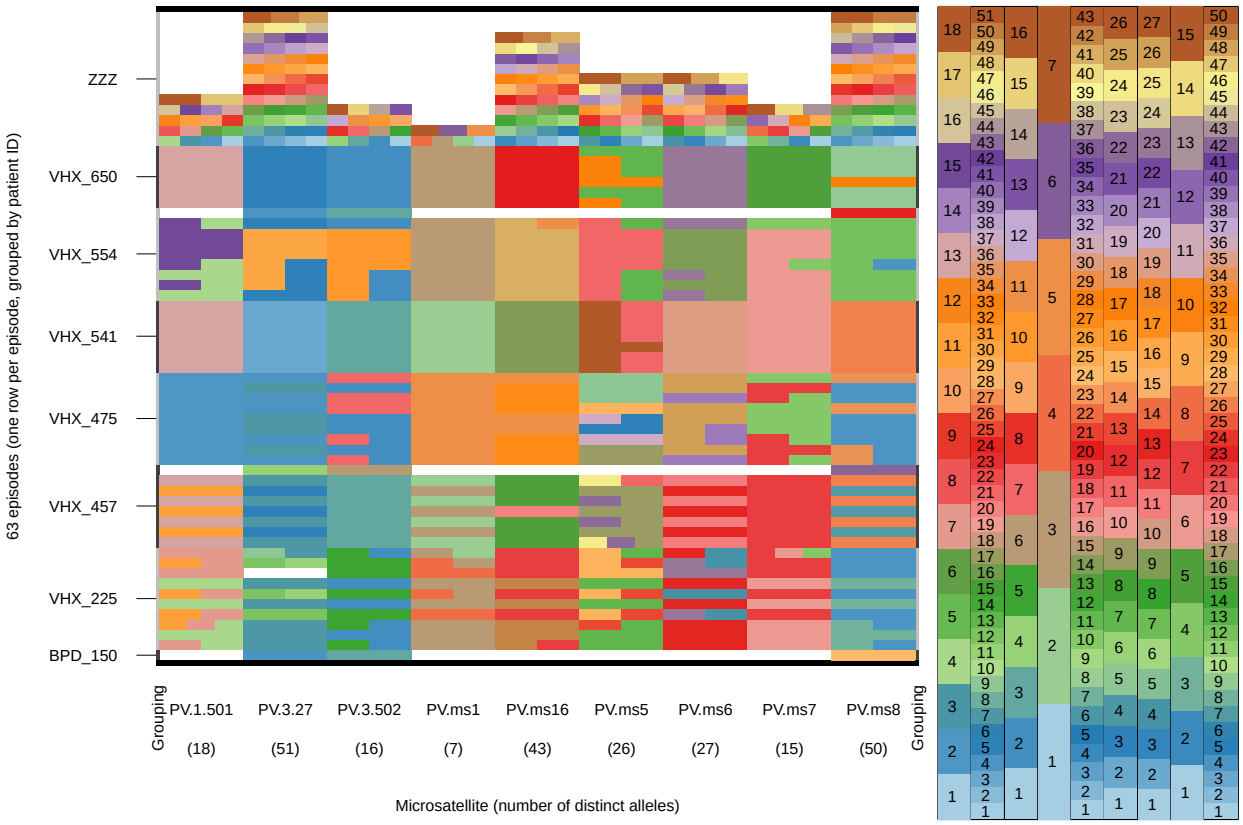
Unestimated

The new model was not used to estimate states for patients with evidence of more than eight genotypes within and across infections.

Data for which recurrent states are not directly estimatable:



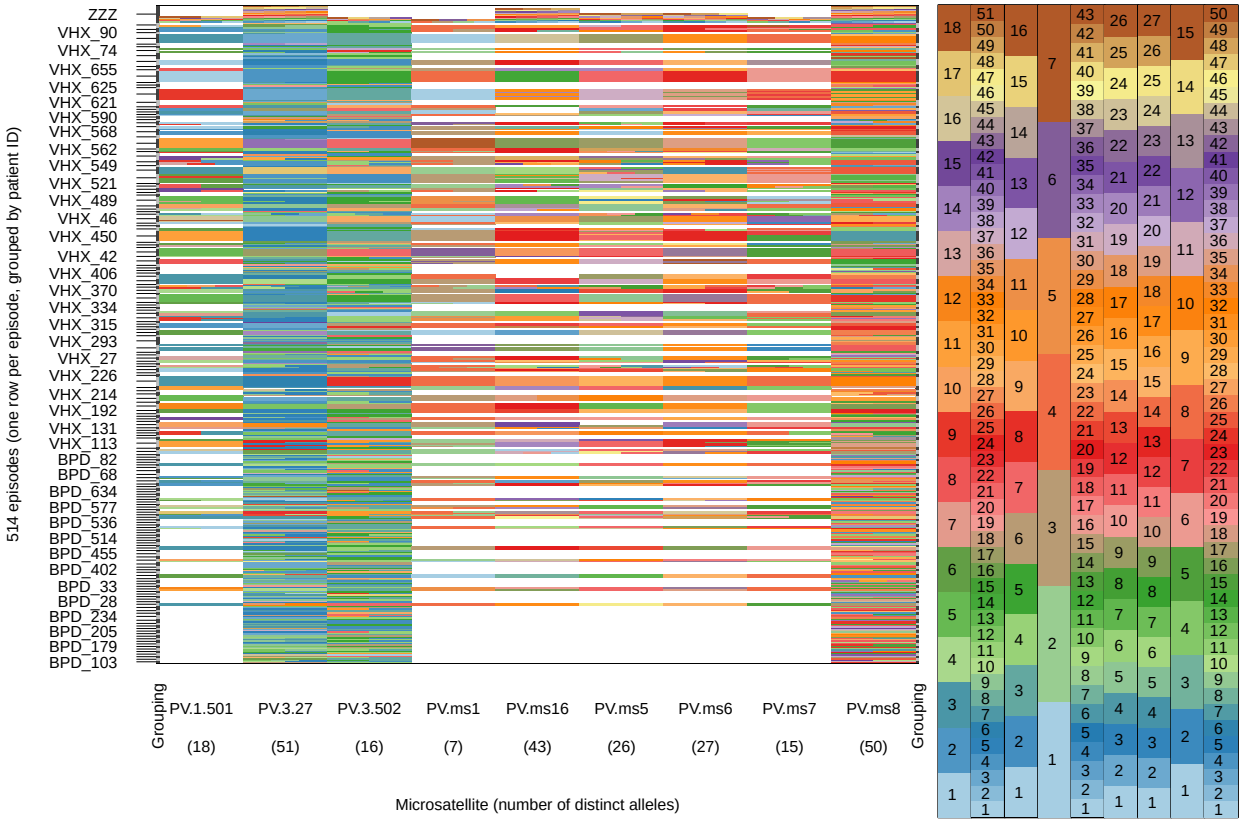
Total of 29 patients. They appear to be mostly relapsing (zooming in on some interesting cases identified by eye-balling the plot above):


63 episodes (one row per episode, grouped by patient ID)
ZZZ
VHX_650
VHX_554
VHX_541
VHX_475
VHX_457
VHX_225
BPD_150
Grouping
PV.1.501 (18)
PV.3.27 (51)
PV.3.502 (16)
PV.ms1 (7)
PV.ms16 (43)
PV.ms5 (26)
PV.ms6 (27)
PV.ms7 (15)
PV.ms8 (50)
Grouping
Microsatellite (number of distinct alleles)

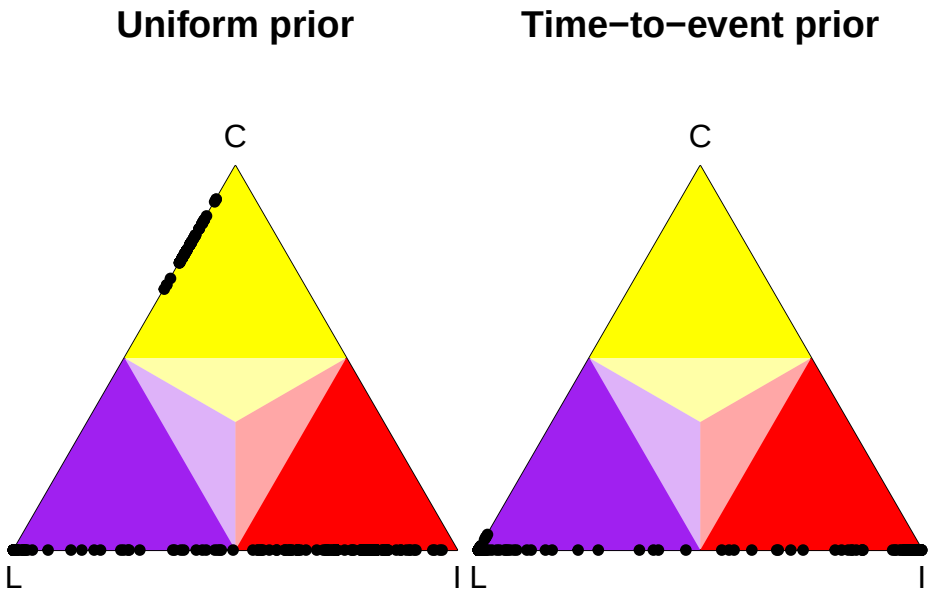
Newly estimated

These are all estimated directly. That is to say, we didn't use the pairwise hack used in Taylor & Watson et al. 2019.

Data for which recurrent states are directly estimated:

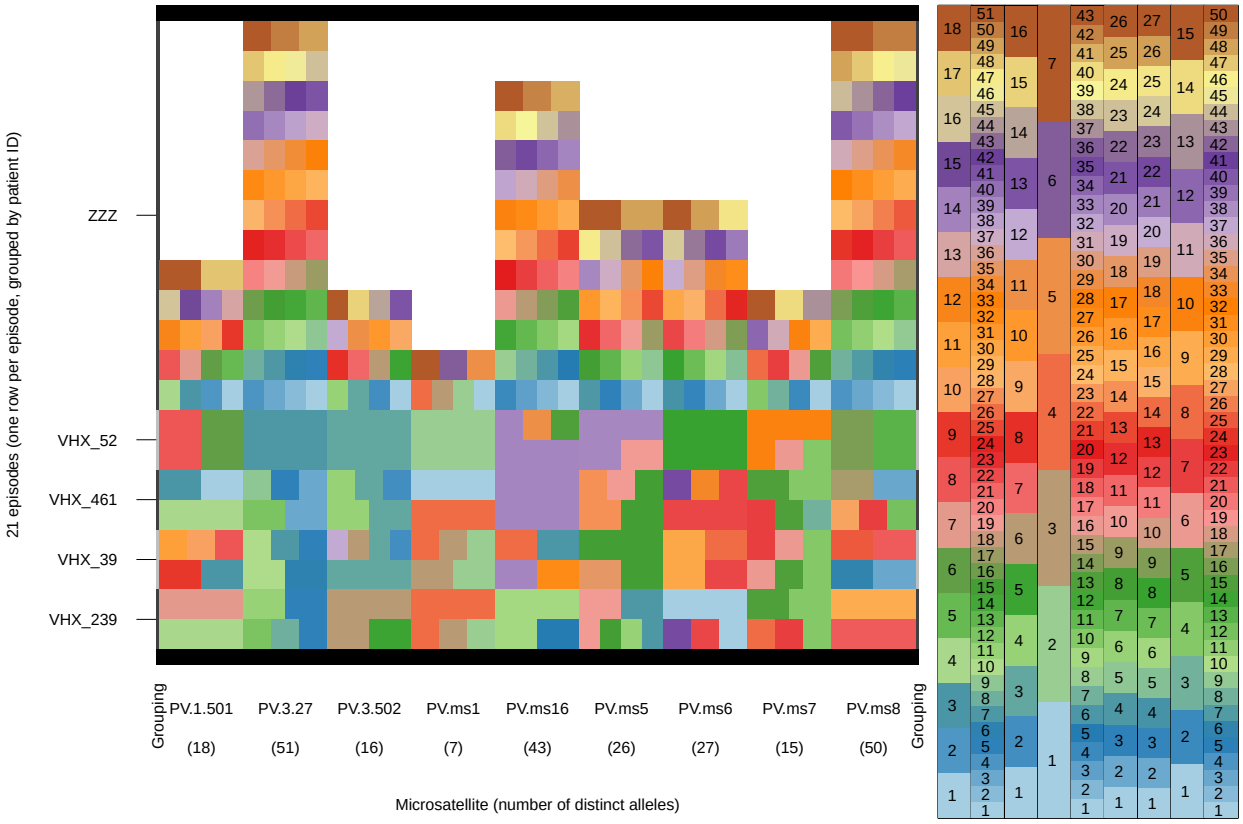


All estimates have weight on either one or two states, excluding reinfection-recrudescence (see simplex plots below). As such, the posterior relapse probability is a sufficient summary statistic.



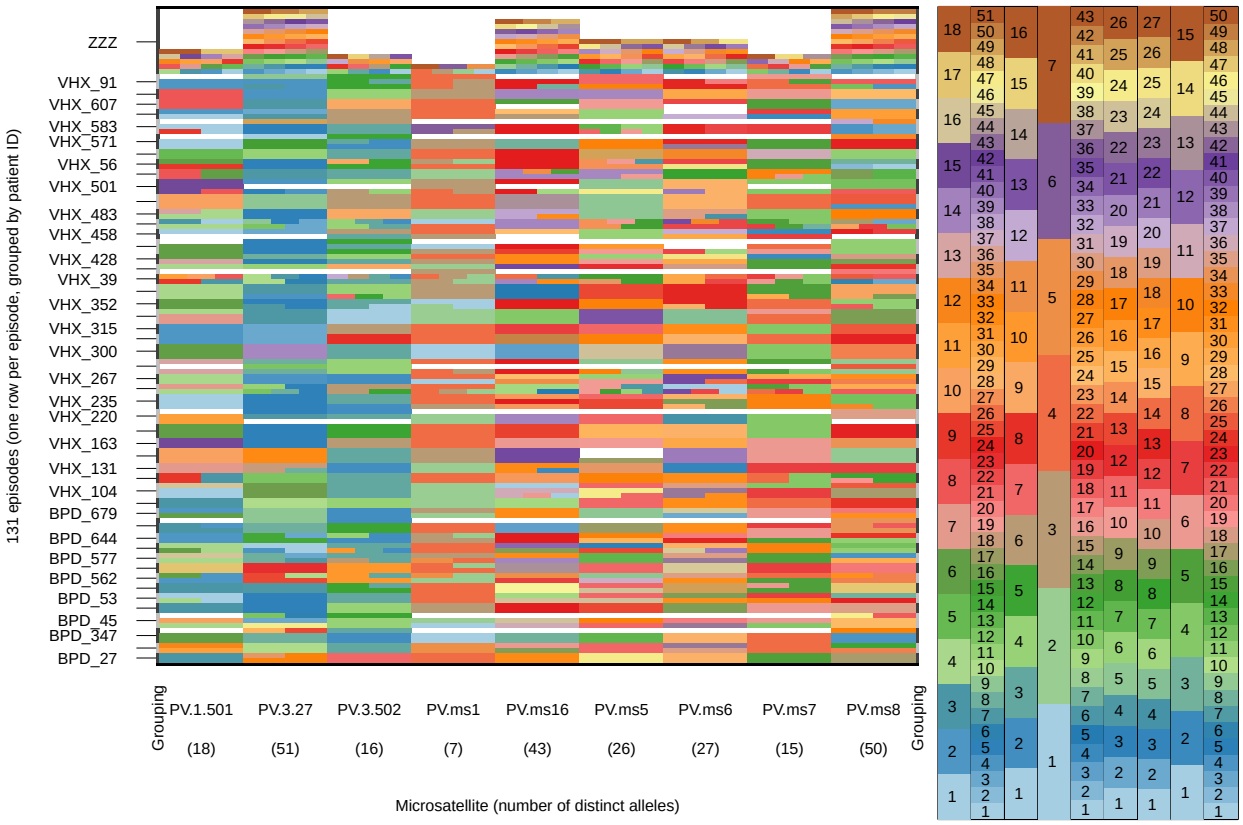
New estimates were generated for all BPD patients with paired data, allowing direct comparison between the computation of the BPD failure rate using the old and new genetic models — basically the same (see Compute_PQ_failure_rates_new.R).

The new model generates estimates for patients whose genetic data were not analysed under the old model of Taylor & Watson et al. 2019:

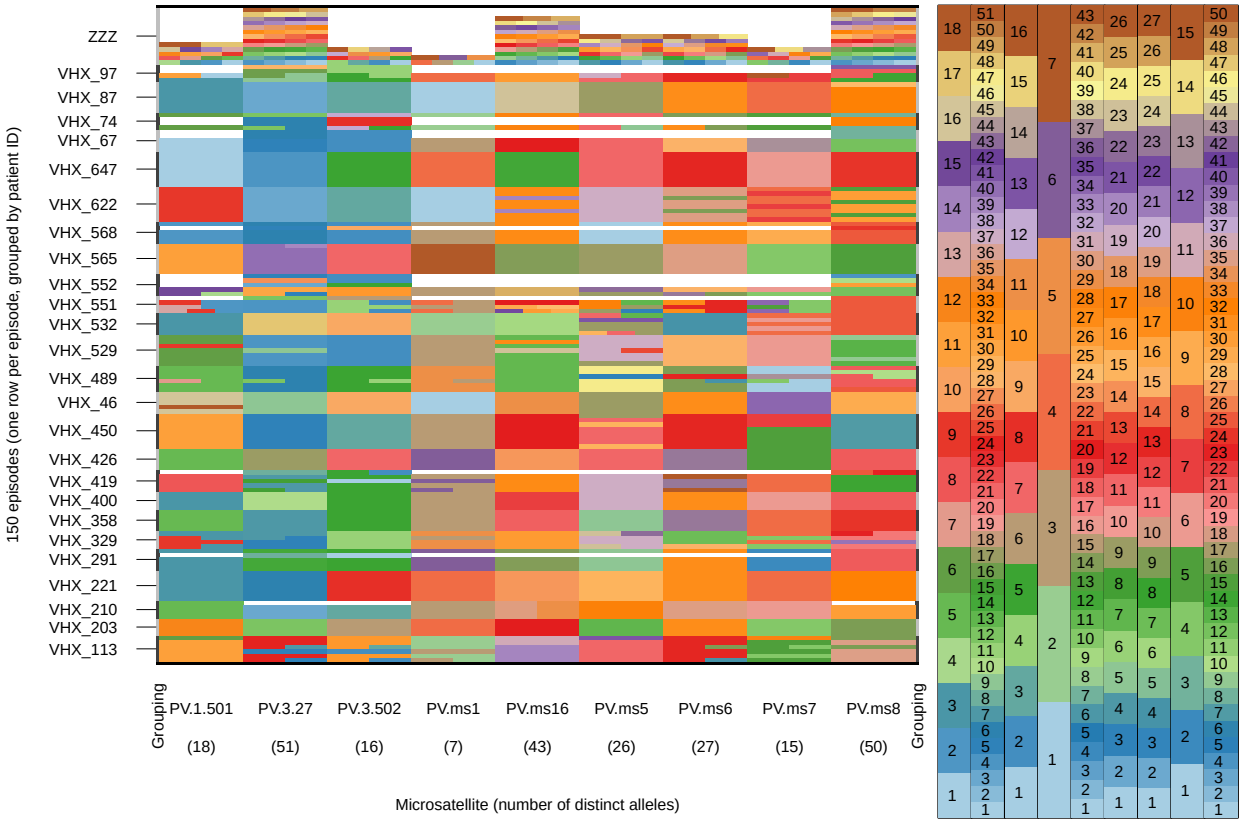


Episodes with data on nine markers look like relapse because more than three markers were typed only if the first three markers did not discriminate. . .

For patients with only one recurrence:



For patients with more than one recurrence:



Some cherry picked examples, including

- VHX_532 (an example of half sibs with a high posterior relapse probability)
- BPD_45 (a rare example of a possible reinfection with data on nine markers)

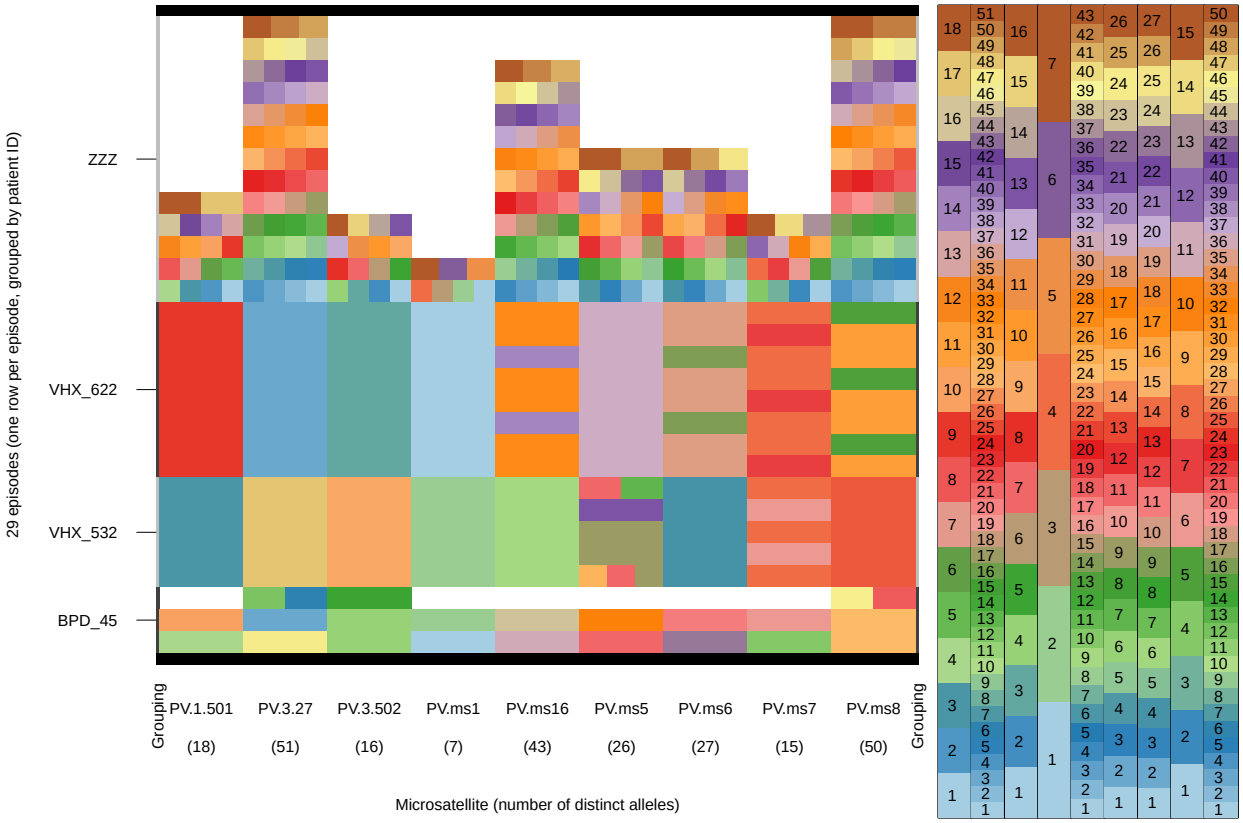


Table 1: Estimates for VHX_532_5

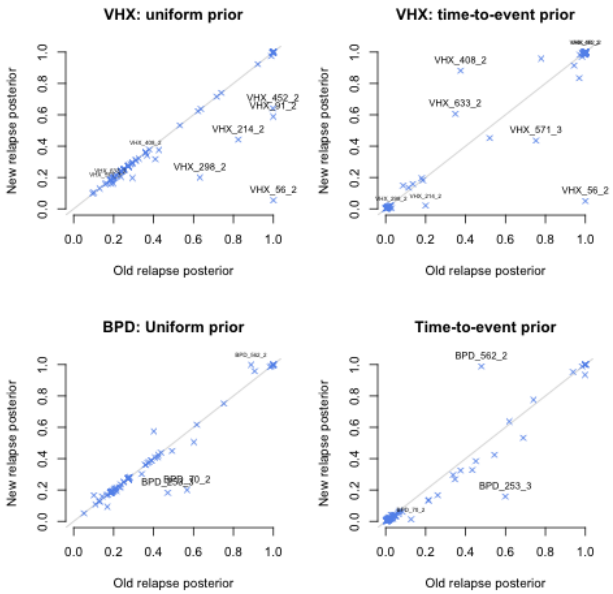
	C	L	I
prior unnorm.	0.0001391	0.9615154	0.0381671
prior norm.	0.0001392	0.9616869	0.0381739
old time-to-event	NA	NA	NA
new time-to-event	0.0000000	0.9998887	0.0001113
old uniform	NA	NA	NA
new uniform	0.0000000	0.9992839	0.0007161

Table 2: Estimates for BPD_45_2

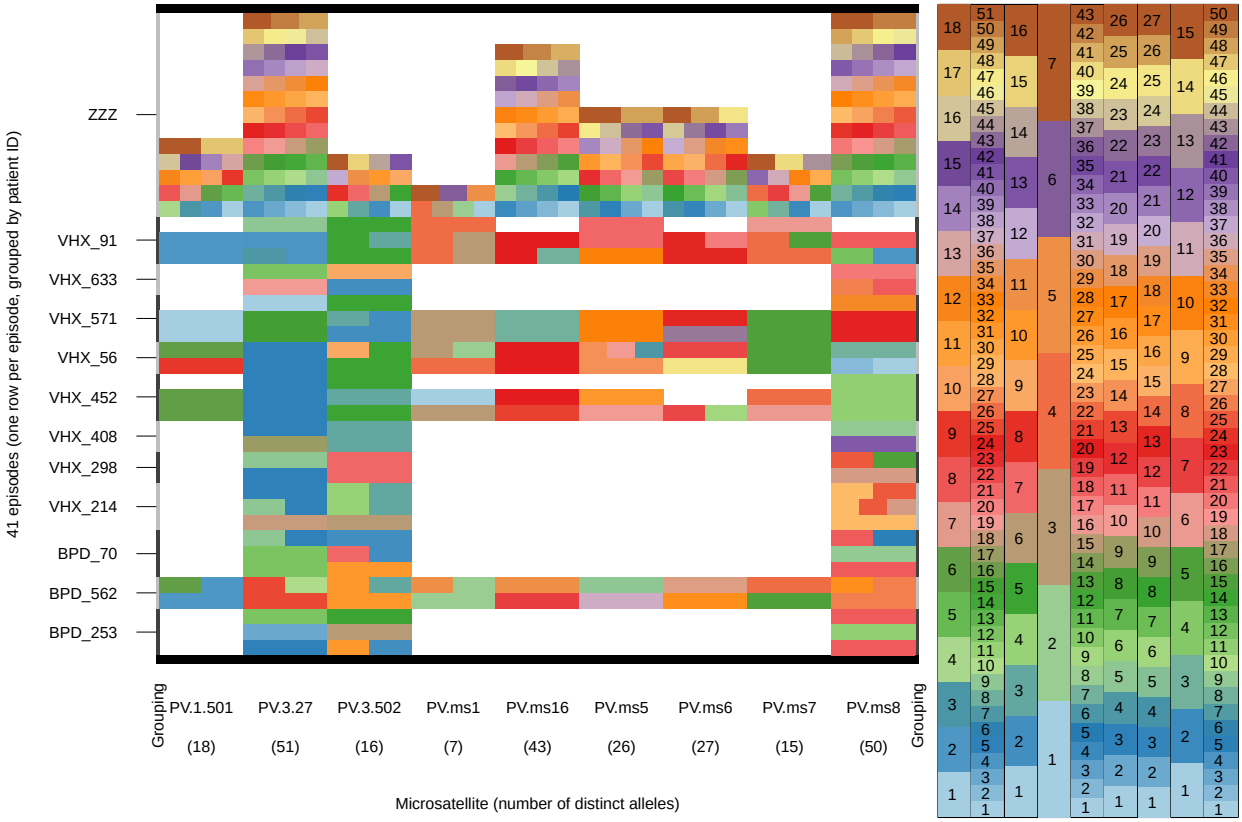
	C	L	I
prior unnorm.	2e-07	0.0692457	0.9307542
prior norm.	2e-07	0.0692457	0.9307541
old time-to-event	0e+00	0.0635692	0.9364308
new time-to-event	0e+00	0.0471660	0.9528340
old uniform	0e+00	0.4069893	0.5930107
new uniform	0e+00	0.4068454	0.5931546

Comparing the new and old model

Some new and old relapse probability estimates differ a lot:



Data for which estimates differ a lot:



Examples where change is undersirable:

Based on the genetic data VHX_56_2 and VHX_91_2 look like relapses, but with half-siblings across infections at odds with full siblings within infections:

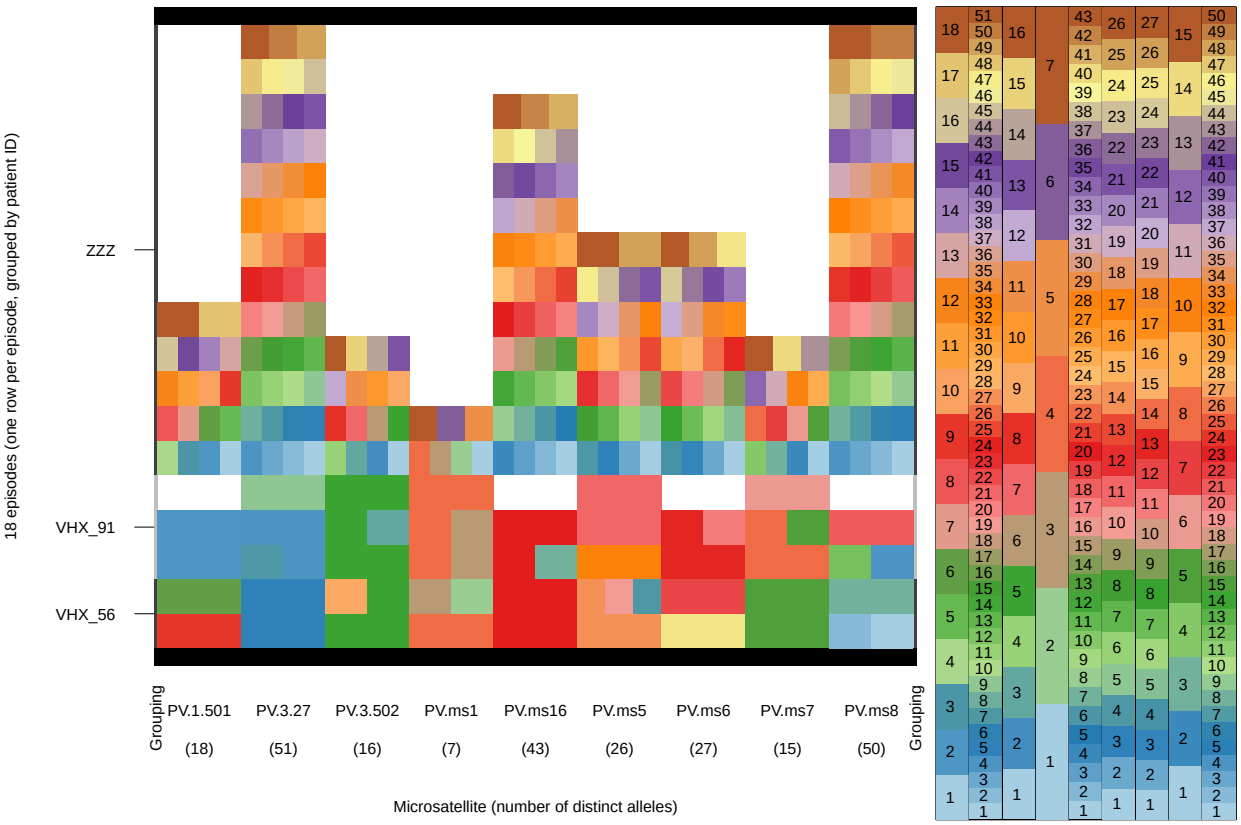


Table 3: Estimates for VHX_56_2

	C	L	I
prior unnorm.	0.0002932	0.4728243	0.5268730
prior norm.	0.0002932	0.4728288	0.5268780
old time-to-event	0.0000000	0.9999990	0.0000010
new time-to-event	0.0000000	0.0495969	0.9504031
old uniform	0.0000000	0.9999994	0.0000006
new uniform	0.0000000	0.0549548	0.9450452

Table 4: Estimates for VHX_91_2

	C	L	I
prior unnorm.	0.0013914	0.9962393	0.0023037
prior norm.	0.0013915	0.9963047	0.0023038
old time-to-event	0.0000000	0.9999995	0.0000005
new time-to-event	0.0000000	0.9990602	0.0009398
old uniform	0.0000000	0.9993305	0.0006695
new uniform	0.0000000	0.5882715	0.4117285

If state probabilities are re-estimated without using the data on PV.ms8, we recover high posterior relapse:

```
## Number of valid relationship graphs (RGs) is 30
## ================================================================================|
## Computing log p(Y|RG) for 30 RGs
## ================================================================================|
## Finding log-likelihood of each vector of recurrent states
## ================================================================================|
##
## Number of valid relationship graphs (RGs) is 250
## ================================================================================|
## Computing log p(Y|RG) for 250 RGs
## ================================================================================|
## Finding log-likelihood of each vector of recurrent states
## ================================================================================|
```

```
##          C         L           I
## VHX_56_2 0 0.9913728 0.008627174
## VHX_91_2 0 0.9988203 0.001179682
## VHX_91_3 0 0.1360378 0.863962238
```

Example where change is beneficial:

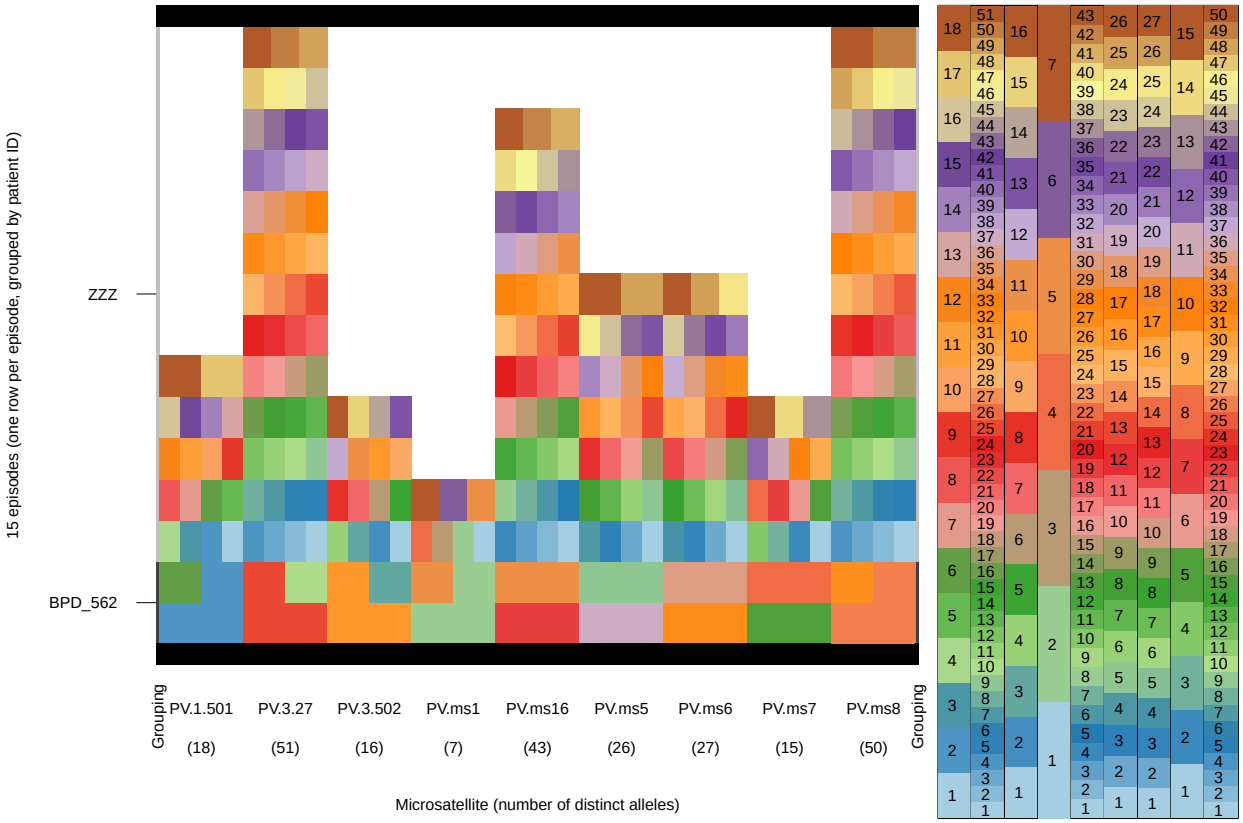


Table 5: Estimates for BPD_562_2

	C	L	I
prior unnorm.	7e-07	0.0774003	0.9225992
prior norm.	7e-07	0.0774002	0.9225990
old time-to-event	0e+00	0.4791717	0.5208283

	C	L	I
new time-to-event	0e+00	0.9881264	0.0118736
old uniform	0e+00	0.8882067	0.1117933
new uniform	0e+00	0.9989929	0.0010071

Based on the genetic data, BPD_562_2 looks like a relapse (5 of 4 markers have some homologous alleles) and has a high probability of being a relapse. However, the timing of BPD_562_2 suggests it's a reinfection. Under the new model, the posterior deviates more from the prior.