

Supplemental Materials of the Deconvolution Method Validation

S1 Method validation on lab controlled strains

The *P. falciparum* genetic crosses project (Miles et al., 2015) finds that due to sequencing error or applying different variant calling methods, genotype calls vary at the same position given the same strain of *P. falciparum*. Thus we apply inference methods to mutiple samples that contains the same parasite strains, and infer the genotypes of a reference strain.

S1.1 Use inference method to reconstruct the reference strains

1. Mixtures of strains 3D7 and Dd2 Since 3D7 is reference strain, we can assume that strain Dd2 is the only source of ‘ALT’ reads in samples PG0389-C, PG0390-C, PG0391-C, PG0392-C, PG0393-C and PG0394-C. Assume markers are independent from each other, let y be the read count for ‘ALT’ allele and x be the weighted coverage, of which the weight are the proportions that are used during the mixing (see Table ??), we use the following regression model to infer the Dd2 variant calling,

$$y = \beta_0 + \beta_{Dd2}x,$$

from which significant coefficient β_{Dd2} implies a Dd2 variant (Fig. S1.1b).

2. Mixtures of strains HB3 and 7G8. Similarly, for sample from PG0398-C to PG0415-C, we let variables x_1, x_2 be the weighted coverages, of which the weights are the mixing proportions for strains HB3 and 7G8 respectively. We use regression model $y = \beta_0 + \beta_{Hb3}x_1 + \beta_{7G8}x_2$ to investigate the relationships between the total allele count and weighted coverage of HB3 and 7G8. Hb3 variant is inferred as coefficients β_{Hb3} is significant (Fig. S1.2a and S1.2b), so is 7G8 (Fig. S1.2a and S1.2c).

S1.2 Validation performance

S1.2.1 Assessing quality of the proportion inference

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

References

Miles, A., Z. Iqbal, P. Vauterin, R. Pearson, S. Campino, M. Theron, K. Gould, D. Mead, E. Drury, J. O’Brien, V. Ruano Rubio, B. MacInnis, J. Mwangi, U. Samarakoon, L. Ranford-Cartwright, M. Ferdig,

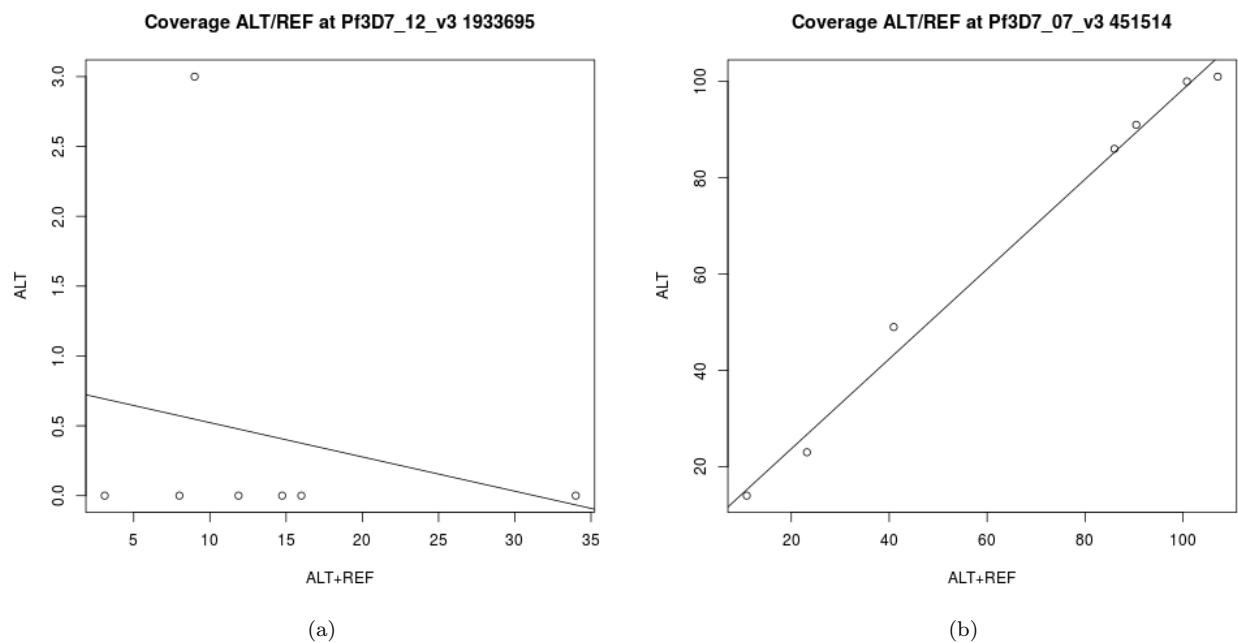


Figure S1.1: XXXXXXXXXXXXXXXX

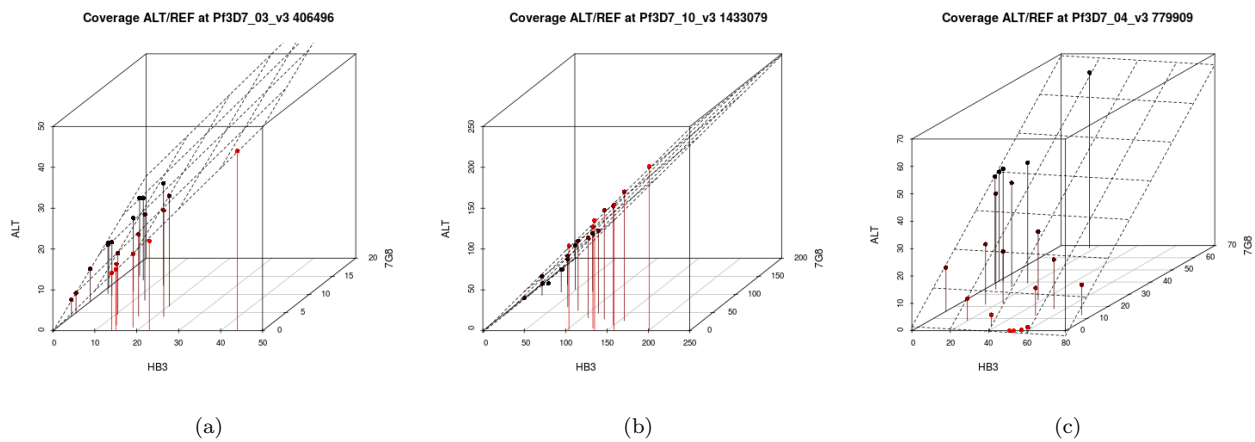
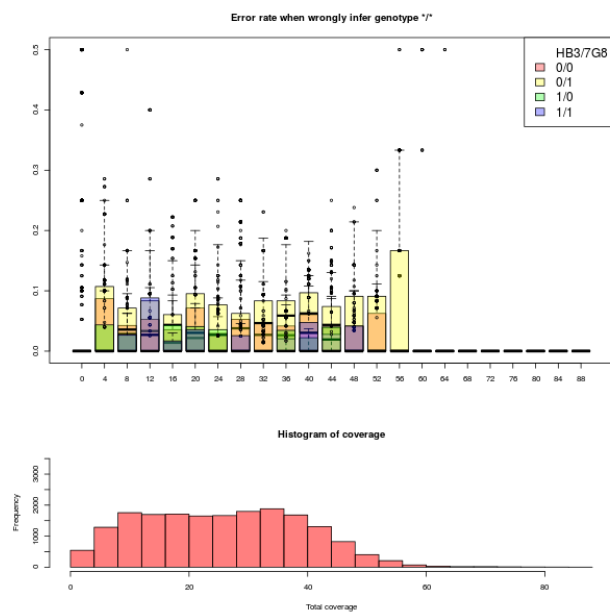
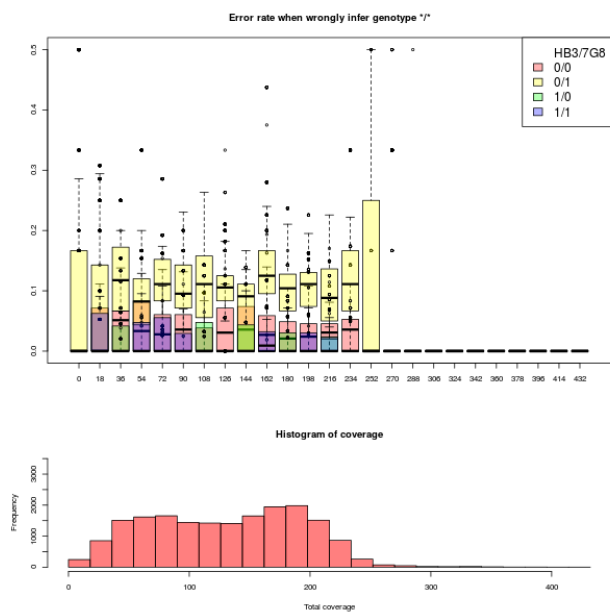


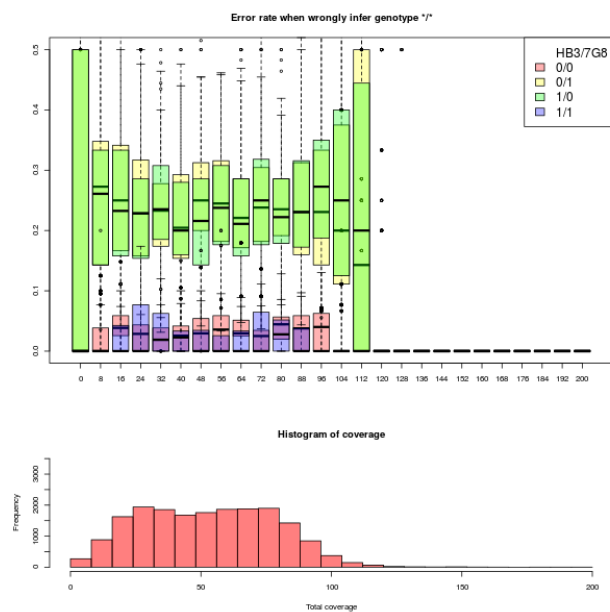
Figure S1.2: XXXXXXXXXXXXXXXX



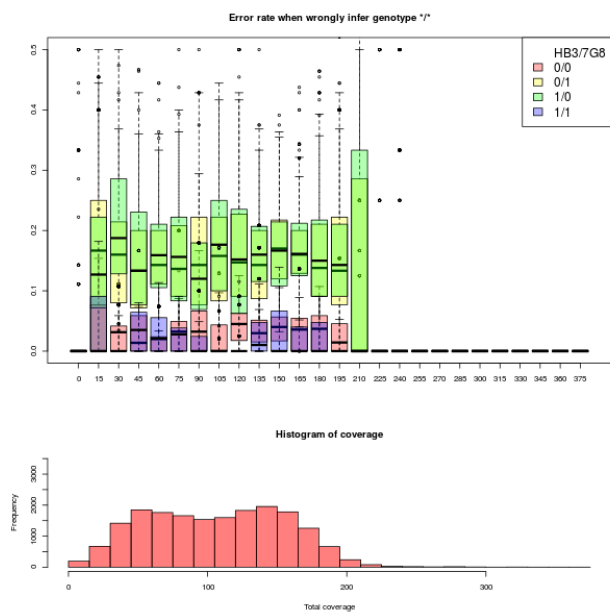
(a) PG402-C low coverage data deconvolution with panel V.



(b) PG402-C high coverage data deconvolution with panel V.



(c)



(d)

Figure S1.3: to be done

K. Hayton, X. Su, T. Wellems, J. Rayner, G. McVean, and D. Kwiatkowski (2015). Genome variation and meiotic recombination in *Plasmodium falciparum*: insights from deep sequencing of genetic crosses. *bioRxiv*.

Wendler, J. (2015). *Accessing complex genomic variation in Plasmodium falciparum natural infection*. Ph.D. thesis, University of Oxford.

Supplemental Materials of DEploid

S2 DEploid

Our program *DEploid* is freely available at <https://github.com/mcveanlab/DEploid> under the conditions of the GPLv3 license. A detailed document can be found at <http://deploid.readthedocs.io/en/latest/>.

- (a)
 1. alt vs ref
 2. wsaf hist
 3. wsaf vs plaf
 4. proportion
 5. wsaf obs vs est
 6. llk
- (b)
- (c) (d) and (e)

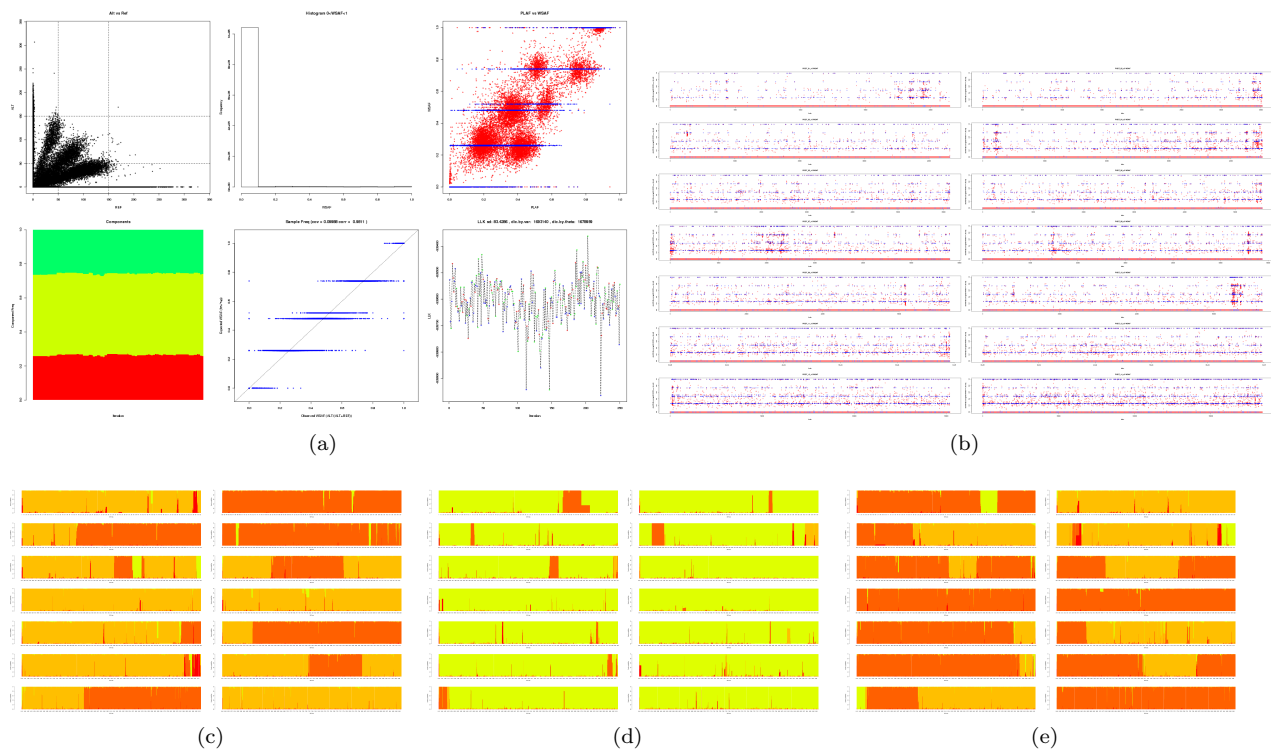


Figure S2.1

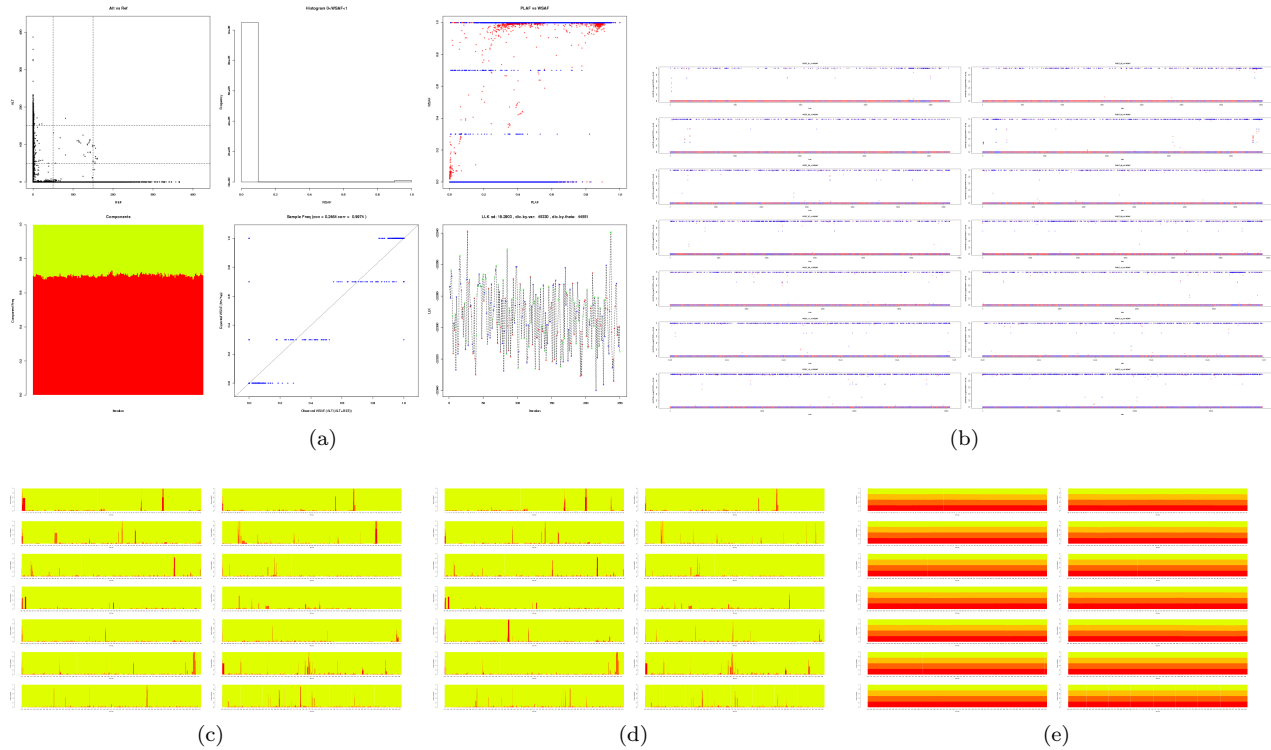


Figure S2.2