

1 **Genome sequence, transcriptome, and annotation of rodent malaria parasite**

2 ***Plasmodium yoelii nigeriensis* N67**

3

4 Cui Zhang^{1*}, Cihan Oguz^{2*}, Sue Huse², Lu Xia^{1,3}, Jian Wu¹, Yu-Chih Peng¹, Margaret Smith¹,
5 Jack Chen⁴, Carole A. Long¹, Justin Lack², and Xin-zhuan Su^{1#}

6

7 ¹*Malaria Functional Genomics Section, Laboratory of Malaria and Vector Research, National*
8 *Institute of Allergy and Infectious Disease, National Institutes of Health, Bethesda, MD 20892-*
9 *8132, USA;*

10 ²*NIAID Collaborative Bioinformatics Resource (NCBR), Frederick National Laboratory for*
11 *Cancer Research, Leidos Biomedical Research, Inc., Frederick, MD 21702, USA*

12 ³*State Key Laboratory of Medical Genetics, Xiangya School of Medicine, Central South*
13 *University, Changsha, Hunan 410078, The People's Republic of China.*

14 ⁴ *the sequencing facility*

15

16

17

18

19

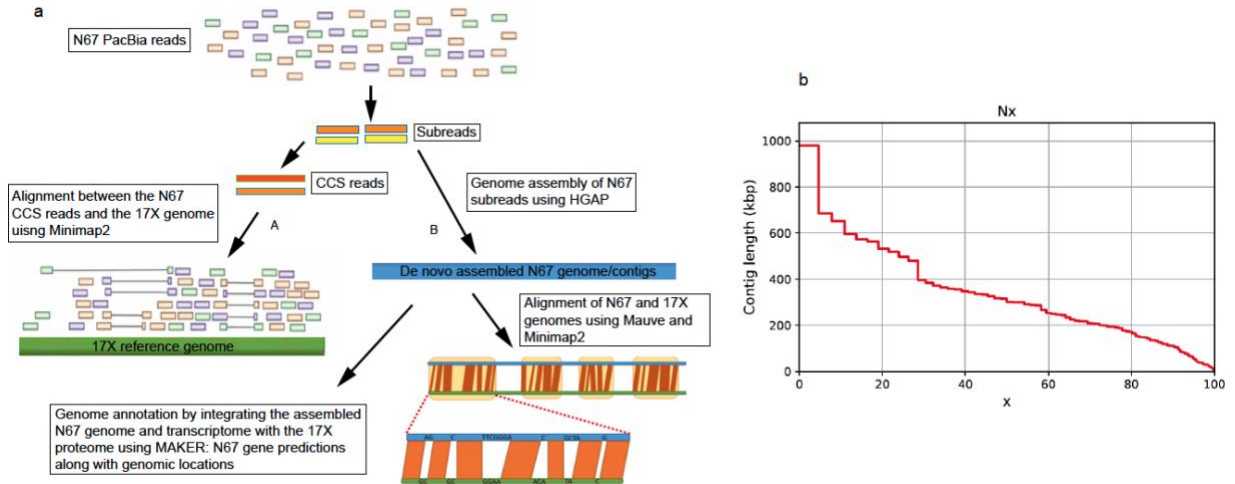
20

21

22

23 **Supplemental files**

24 **Additional file 1. Fig. S1-2.**

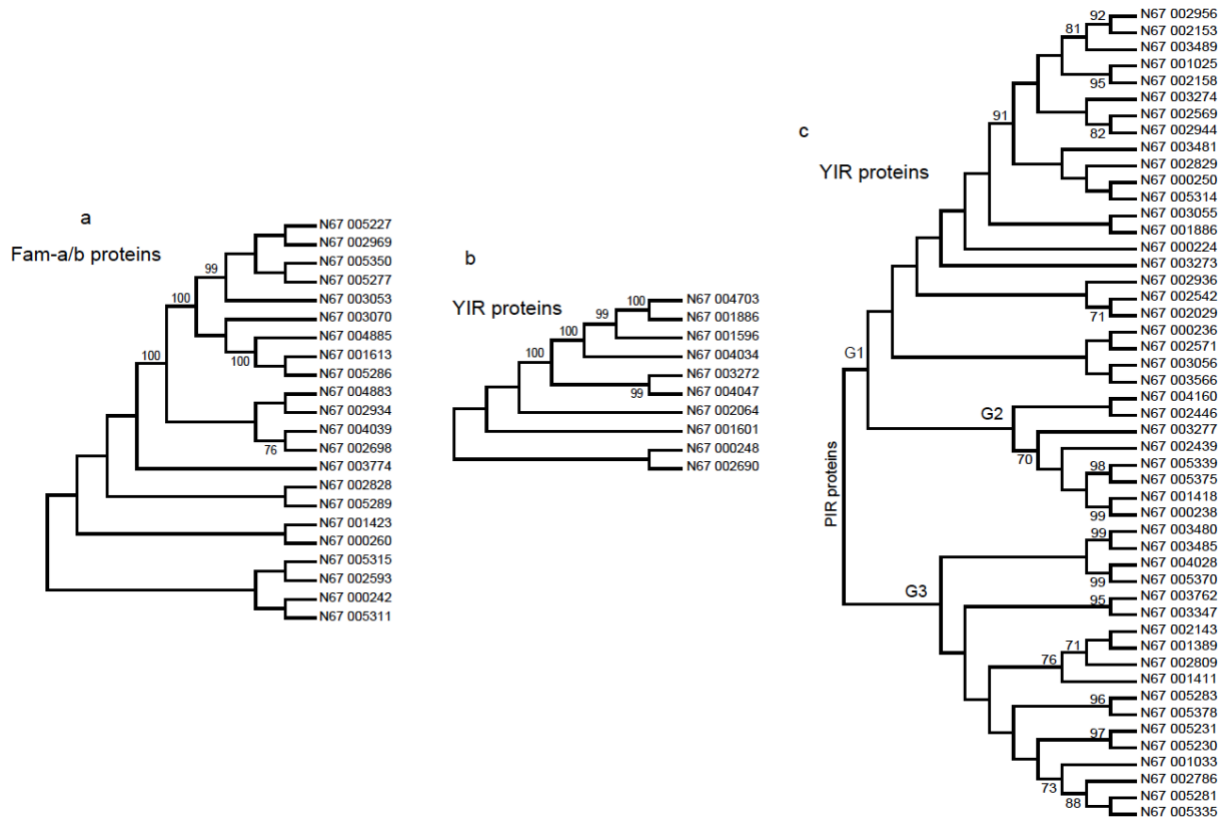


25

26 **Fig. S1.** Plots of contig length distributions of the *Plasmodium y. yoelii* N67 parasite genome

27 assembly. The X-axis is percentage of the contigs with lengths (base pair) greater than the values

28 indicated on the Y-axis.



29

30 **Fig. S2.** Clustering of protein sequences from the *Plasmodium y. nigeriensis* N67-specific
 31 orthogroups and those that are not assigned to any orthogroup. The predicted protein sequences
 32 were aligned using ClustalW algorithm and clustered using procedures described in the Methods
 33 section. **a**, Fam-A/B proteins; **b**, YIR proteins (group 1); **c**, YIR proteins (group 2).

34

35

36

37

38

39

40

41

42 **Additional file 2. Table S1**

43 **Table S1.** Statistics of *de novo* transcriptome sequence assembly of Illumina paired end reads
44 using Trinity.

45

46 **Additional file 3. Table S2**

47 **Table S2.** Predicted gene and protein sequences from *Plasmodium y. nigeriensis* N67 parasite
48 genome and transcriptome.

49

50 **Additional file 4. Table S3**

51 **Table S3.** Predicted proteins of *Plasmodium y. nigeriensis* N67 matching IntrPro domains.

52

53 **Additional file 5. Table S4**

54 **Table S4.** Predicted *Plasmodium y. nigeriensis* N67 proteins that match those in Reactome
55 pathways.

56

57 **Additional file 6. Table S5**

58 **Table S5.** Lists of *Plasmodium y. nigeriensis* N67 genes clustered into selected Reactome
59 pathways.

60

61 **Additional file 7. Table S6**

62 **Table S6.** Completeness statistics of the predicted *Plasmodium y. nigeriensis* N67 genome and
63 transcriptome in matching *Apicomplexan* and *Plasmodium* benchmarking universal single-copy
64 orthologs (BUSCO) gene sets.

65 **Additional file 8. Table S7**

66 **Table S7.** Summary statistics of genes and orthogroups from three *Plasmodium yoelii* parasites.

67

68 **Additional file 9. Table S8**

69 **Table S8.** Individual orthogroups containing genes from three *Plasmodium yoelii* parasites (17X,
70 N67, and YM).

71

72 **Additional file 10. Table S9**

73 **Table S9.** Sequences and functional annotation of genes in *Plasmodium y. nigirensis* N67
74 specific orthogroups.

75

76 **Additional file 11. Table S10**

77 **Table S10.** Gene families and copy numbers from three *Plasmodium yoelii* strains.

78

79 **Additional file 12. Table S11**

80 **Table S11.** Single nucleotide substitutions and indels between *Plasmodium y. nigeriensis* N67
81 (ALT) and *P. y. yoelii* 17XNL (REF) or *P. y. yoelii* YM (REF, second page) parasites with
82 predicted medium and high functional impacts.

83

84