

A Host-Targeting Signal in Virulence Proteins Reveals a Secretome in Malarial Infection

N. Luisa Hiller, Souvik Bhattacharjee, Christiaan van Ooij, Konstantinos Liolios, Travis Harrison, Carlos Lopez-Estraño,* Kasturi Haldar†

Malaria parasites secrete proteins across the vacuolar membrane into the erythrocyte, inducing modifications linked to disease and parasite survival. We identified an 11-amino acid signal required for the secretion of proteins from the *Plasmodium falciparum* vacuole to the human erythrocyte. Bioinformatics predicted a secretome of >320 proteins and conservation of the signal across parasite species. Functional studies indicated the predictive value of the signal and its role in targeting virulence proteins to the erythrocyte and implicated its recognition by a receptor/transporter. Erythrocyte modification by the parasite may involve plasmoidal heat shock proteins and be vastly more complex than hitherto realized.

P. falciparum causes the most virulent form of human malaria. Blood stage parasites that infect mature erythrocytes are responsible for all the symptoms and pathologies of the disease (1). Proteins secreted from the parasite to the host erythrocyte are responsible for many disease pathologies and death (1). They are also thought to underlie structural and transport changes in the erythrocyte required for parasite survival (2, 3). Thus, secreted protein families such as subtelomeric variable open reading frame (STEVAR), repetitive interspersed family (RIFIN), and *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) are responsible for antigenic and adhesive changes in the infected erythrocyte (4). PfEMP1s are the major virulence determinants in cerebral and placental malaria commonly seen in young children and pregnant women [the groups most vulnerable to malaria (4, 5)]. Despite advances in proteomics (6, 7), parasite proteins that underlie multiple phenotypic modifications in the erythrocyte membrane, as well as the number of exported proteins, remain enigmatic (2, 3). We investigated whether critical, conserved transport signals target proteins from the parasite to the erythrocyte to establish the presence of a major host-targeting pathway in malarial infection and to enable recognition of a wide range of proteins (a “secretome”) that present high-value candidate effectors of disease and infection.

Departments of Pathology and Microbiology-Immunology, Feinberg School of Medicine, Northwestern University, 303 East Chicago Avenue, Chicago, IL 60611, USA.

*Present address: Department of Microbiology and Molecular Cell Sciences, University of Memphis, 3774 Walker Avenue, 409B Life Sciences Building, Memphis, TN 38152, USA.

†To whom correspondence should be addressed. E-mail: k-haldar@northwestern.edu

Within the erythrocyte, *P. falciparum* resides and develops surrounded by a parasitophorous vacuolar membrane. Several studies have established that a cleavable endoplasmic reticulum (ER)-type signal sequence (SS) is sufficient for protein recruitment into the secretory pathway within the parasite, as well as release at the parasite plasma membrane and into the lumen of the

parasitophorous vacuole (8–10). We have recently demonstrated that for two histidine-rich proteins [the knob-associated histidine-rich protein or histidine-rich protein I (HRPI) as well as histidine-rich protein II (HRPII)], cleavage of the ER-type SS reveals a vacuolar transport signal (VTS) that resides within the next 40 amino acids of each protein (11). The VTS is required to export a reporter such as green fluorescent protein (GFP) from the lumen of the parasitophorous vacuole to the erythrocyte cytoplasm and must be exposed at the N terminus (11). We now show (Fig. 1A) VTSs located within 60 amino acids downstream from the SS cleavage site of non-histidine-rich proteins [such as PfEMP2 (12) and glycophorin-binding protein 130 (GBP 130) (13) that are known to be exported to the erythrocyte].

Multiple Expectation Maximization for Motif Elicitation (MEME) analysis revealed the presence of a primary pattern common in five experimentally validated VTS sequences (fig. S1, A and B) (14). Reiterative use of MEME and of the Motif Alignment and Search Tool (MAST) (fig. S1) (14) in the *P. falciparum* database yielded an optimized MEME motif of RxSRILAExxx (15) (blue bar in Fig. 1B; the positional value indicated by x is elaborated in fig. S1A). This motif was detected in 259 parasite proteins, 8 of which were removed by hand curation, leav-

Table 1. Summary of predicted secretome based on table S1.

Category	No. of proteins	Notable characteristics	Annotation source
Unknowns	91	No annotation	None
Protein families	8	Range between two to four sequences from predicted secretome per family	This report
Protein rich in internal repeats	10	Repeats	This report
Proteins highly enriched in one or more amino acids	11	At least one residue is highly enriched over expected frequencies	This report
RIFINs	119	Variant antigens	PlasmoDB
STEVARs	22	Variant antigens	PlasmoDB
Candidate phosphatases	3	Aminopeptidase/ α/β hydrolase	PlasmoDB, this report
		α/β hydrolase	PlasmoDB, this report
Candidate serine-threonine kinases	3	Calcineurin-like phosphoesterase	This paper
		3.8 protein	PlasmoDB
		Kinase	This report
		Kinase	This report
Predicted heat shock proteins	3	hsp40 homolog	PlasmoDB, this report
		DNAJ	PlasmoDB, this report
		DNAJ domain	PlasmoDB, this report
Predicted ARF	1	Missing one domain critical for guanosine 5'-triphosphate binding	PlasmoDB
Presence of glycophorin-binding repeats	3	GBP 130	PlasmoDB
		Five repeats	PlasmoDB
		Seven repeats	PlasmoDB
Putative ABC transporter	1	Transporter	PlasmoDB
Proteins in the input	5	GBP 130	PlasmoDB
		HRP I	PlasmoDB
		HRP II	NCBI
		Erythrocyte membrane protein 2	PlasmoDB
		Erythrocyte membrane protein 3	PlasmoDB
PfEMP1	67	Absence of N-terminal SS	PlasmoDB

ing 251 parasite proteins (fig. S1A). These included our original input of five soluble proteins, a surprisingly large number of hypothetical proteins (Table 1 and table S1), and known exported antigens (16–18) not in our original input (thereby validating the MAST). Prominent among the known proteins were 119 and 22 membrane-bound RIFINs and STEVORs, respectively, even though our original input contained only soluble exported proteins. Consistent with the location of STEVORs in erythrocytic “cleft” structures (16), expression of a full-length STEVOR fused at its C terminus to GFP resulted in detection of green fluorescence in punctate domains in the erythrocyte (Fig. 1Ca). Notably, the first 60 amino acids downstream of STEVOR SS (but lacking the transmembrane domains) constitute a VTS and export the GFP reporter to the erythrocyte cytosol (Fig. 1Cb), suggesting a shared function for the motif in both soluble and membrane antigens. Additional signals must localize full-length STEVORs to punctate domains in the erythrocyte.

A sequence logo (19) of the MEME motif and its surrounding region is presented in Fig. 2A. It reveals the high information content of the 11 amino acids in the motif. It indicates that Arg in position 4 and Leu in position 6 are the most highly conserved residues in the motif. It also shows the lower but finite positional value of the three C-terminal residues represented as “xxx” in the linearized motif RxSRILAExxx. Placing an Ala in position 4 or in position 6 in the PfHRPII VTS blocks export of GFP to the erythrocyte (Fig. 2B), indicating that the motif provides a signal for protein export from the vacuole to the erythrocyte. Inhibition of export upon single amino acid replacement implicates recognition of the signal by a receptor/transporter.

To independently test the predictive power of the motif in identifying unknown exported proteins, we investigated whether a hypothetical protein in the MAST output (fig. S1A) contained a functional export signal. We selected a protein (PFE1615c) expressed during blood stage infection (as determined by transcriptome analysis) (20, 21) with no National Center for Biotechnology Information (NCBI) homologs or Pfam pattern recognized in silico. Furthermore, it is only present in the second MAST output and not in the first (fig. S1A). Although still included in the matrix, its motif (THSRILKQLEF) differs from that of PfHRPII (LNKRLLEHETQA) (and all of the initial experimental data set). PFE1615c contains a vacuolar translocation signal (Fig. 2Ca) that can export GFP to the erythrocyte, but this export is blocked by replacement of RILKQLE in its motif with LNAKALA (Fig. 2Cb). Thus, the motif

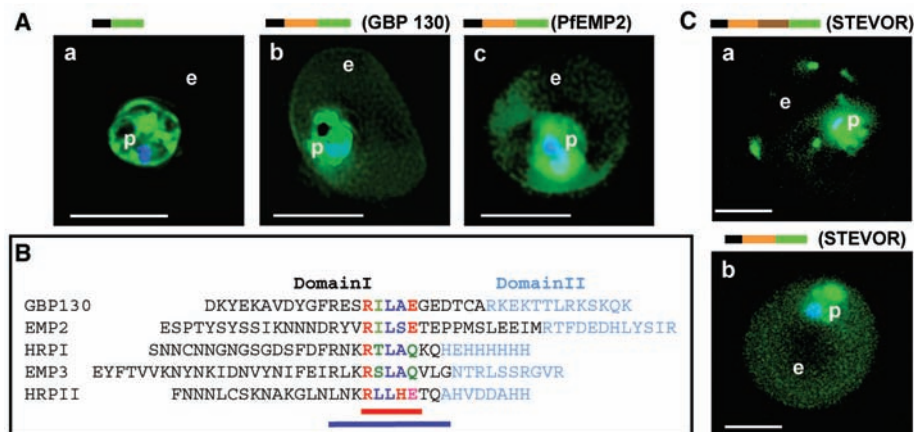


Fig. 1. Identification of a conserved motif of 11 amino acids in VTSs of parasite proteins exported to the erythrocyte. (A) VTS of proteins PfGBP 130 and PfEMP2 target GFP to the erythrocyte. Projections (0°) of live cells expressing GFP chimeras of a SS alone (a), SSVTSGBP130 (b), and SSVTSEMP2 (c). (B) Alignment of MEME motifs in VTSs of indicated five exported soluble proteins. Red bar underlines 5-amino acid MEME motif 1. Blue bar underlines 11-amino acid MEME motif 2. (C) Projections (0°) of live cells expressing GFP chimeras of full-length STEVOR (a) and SSVTSSTEVOR (b). (A) and (C) were detected by digitized fluorescence microscopy. Parasite (p) nucleus is Hoechst stained (blue). Green, GFP; e, erythrocyte. Schematic above panels indicate constructs containing SS (black), VTS (orange) of the indicated proteins, and GFP (green). STEVOR sequences downstream of the VTS are indicated in brown in (C). Scale bars, 5 μm.

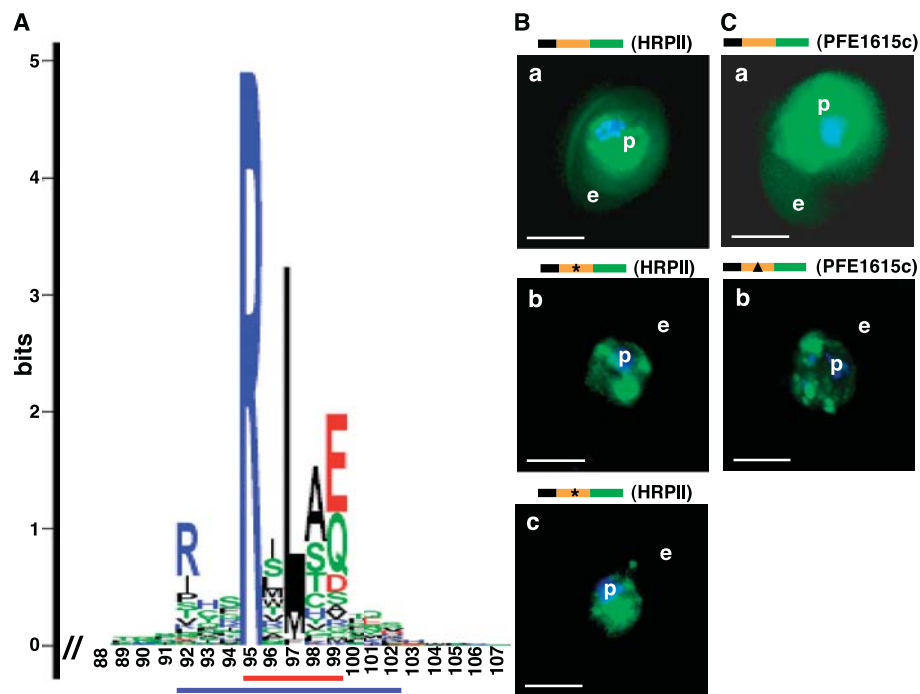


Fig. 2. The VTS motif is a signal for vacuolar protein export and shows high value in identification of unknown parasite proteins exported to the erythrocyte. (A) Sequence logo derived from the predicted secretome. Amino acids are represented by one-letter abbreviations and color coded as follows: blue, basic; red, acidic; black, hydrophobic; green, polar. Height of amino acids is proportional to the fraction of the observed frequency relative to the expected frequency in *P. falciparum* proteins. (B) Projections (0°) of live cells expressing GFP chimeras of SSVHRPIIVTS with no change (a) or motif point mutants R4A (b) or L6A (c), where 4 and 6 indicate position in the signal. Export of green fluorescence (GFP) to erythrocyte (e) is abrogated in both point mutants. (C) Projections (0°) of live cells expressing GFP chimeras of SSVTSPFE1615c with intact motif (a), or replacement of motif sequences RILKQLE with LNAKALA (b). Replacement of motif residues blocks export of green fluorescence to the erythrocyte. (B) and (C) were detected by digitized fluorescence microscopy. Parasite (p) nucleus is Hoechst stained (blue). Schematics above panels indicate constructs containing SS (black), VTS (orange) of the indicated proteins, and GFP (green). Stars indicate single amino acid substitutions and the triangle indicates seven-residue replacement. Scale bars, 5 μm.

provides a vacuolar export signal in unknown proteins, shows high predictive value in identifying hitherto unknown parasite proteins exported to the erythrocyte, and is conserved across parasite species (fig. S4) (22).

Arguably, the most important of known parasite protein families exported to the erythrocyte is the PfEMP1 family (1). However, PfEMP1s lack a leader SS and are presumably recruited into the parasite secretory pathway by means of an internal SS that serves as membrane anchor for these type I membrane proteins (23). Thus, PfEMP1s were not included in our initial database subjected to MAST (fig. S1 series). We therefore investigated the presence of the MEME motif in a data set that contained *P. falciparum* proteins that lack a predicted N-terminal ER-type SS (fig. S2). Three predicted PfEMP1 sequences containing QFFRFWFSEWSE or IGKRVHAQVQN were detected in the MAST (fig. S2A). IGKRVHAQVQN was present in only two proteins, whereas QFFRFWFSEWSE is a highly conserved sequence in all PfEMP1s (Fig. 3A and fig. S2B). We reasoned that an export motif should be preserved in the family. Unfortunately, because of their size, full-length PfEMP1s cannot be cloned and

expressed as transgenes. Nonetheless, we successfully synthesized a “minitransgene” containing the motif that is present in full-length PfEMP1s beyond amino acid 200 but before the cysteine-rich interdomain region α (Fig. 3, A and B). The synthetic PfEMP1 also contained the conserved transmembrane and C-terminal cytoplasmic domains (amino acids 1521 to 2042 of *P. falciparum* PfEMP1, NCBI accession number AAB09769.1), but most of the PfEMP1 adhesive domains were replaced by GFP (Fig. 3B). Expression of this transgene resulted in the export of green fluorescence primarily to “spots” (possibly reflecting clefts) in the erythrocyte (Fig. 3Ca). When FFRWFSEWS in the motif was replaced by AASTDIAST (Fig. 3Cb) or the N-terminal fragment was deleted (Fig. 3Cc), green fluorescence remained with the parasite. This provides the first identification of a conserved sequence in the virulence membrane protein PfEMP1 that can signal protein export to the erythrocyte. This is likely to occur without exporting the membrane protein into the erythrocyte cytosol. Although the exact step of vacuolar export catalyzed by the signal in transmembrane proteins is not yet delineated, the motif clearly provides a host/erythrocyte-

targeting signal in PfEMP1 transport. The experimentally validated PfEMP1 signal QFFRFWFSEWSE can be incorporated into the MEME and MAST analysis in fig. S1A to optimize the plasmodial host-targeting (PlasmoHT) pattern (22).

The presence of the host-targeting signal in putative parasite phosphatases and kinases expressed during blood stage infection (Table 1 and table S1) suggests that these activities modify the erythrocyte, consistent with previous data suggesting export of these parasite functions to the erythrocyte (24). A putative parasite adenosine 5'-diphosphate-ribosylation factor (ARF) or ARF-like protein is also predicted to be exported, the functional evaluation of which may reveal whether the parasite exports both the machinery and cargo underlying erythrocyte modification. Notably, our study provides the first prediction for the export of parasite-encoded heat shock proteins to the erythrocyte. Moreover, evidence that a GFP fusion of a soluble heat shock protein 40 (HSP40) (with motif TSLRSLAEFNS) is exported to the erythrocyte (fig. S3) strongly supports a role for these parasite chaperones in the process of protein export and/or erythrocyte modification.

The vacuolar export/host-targeting signal is also present in 91 unknown proteins (Table 1). Conservative estimates project that at least 50 of these unknowns are expressed during blood stage infection (table S1), suggesting the possibility that the parasite induces substantial complex molecular changes in the erythrocyte, of which we remain largely ignorant. Importantly, they provide high-value candidate effectors, the functional analyses of which may yield further insights into mechanisms underlying virulence and disease associated with blood stage malarial infection as well as the ability of the parasite to survive within the erythrocyte.

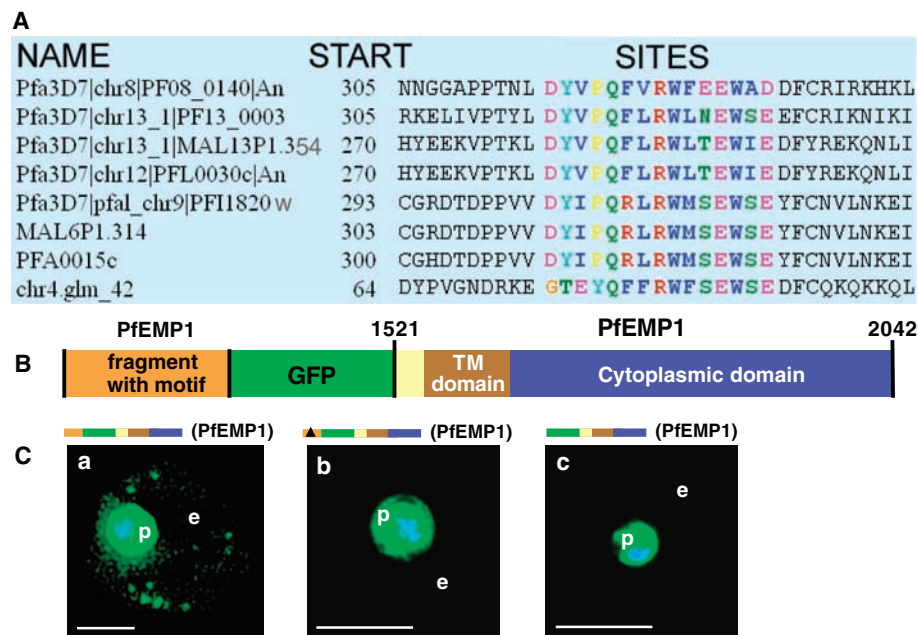


Fig. 3. A host-targeting signal is detected in the virulence membrane protein PfEMP1. (A) Alignment of multiple predicted PfEMP1 sequences with colored amino acids highlighting the MEME of the conserved region (14), in which the last 11 colored amino acids correspond to a predicted export signal. (B) Schematic representation of a PfEMP1 minigene-GFP chimera containing PfEMP1 domain (orange) with the MEME motif as well as transmembrane and cytosolic domain (amino acids 1521 to 2042 from PfEMP1 AAB09769.1) (yellow, brown, and blue). (C) Projections (0°) of live cells expressing the PfEMP1 minigene-GFP chimera with no change (a), nine-residue replacement (triangle) of FFRWFSEWS in signal with AASTDIAST (b), and deletion of the 203-to-321 fragment (c), as detected by digitized fluorescence microscopy. Export of green fluorescence is primarily to “spots” (possibly reflecting clefts) in the erythrocyte in (a). This export is blocked by either replacement of indicated residues in the predicted signal or loss of the fragment (containing the motif). p, parasite; e, erythrocyte. Scale bars, 5 μm. Identification codes shown are from PlasmoDB.

References and Notes

1. L. H. Miller, D. I. Baruch, K. Marsh, O. K. Doumbo, *Nature* **415**, 673 (2002).
2. K. Haldar *et al.*, *Cell. Microbiol.* **4**, 383 (2002).
3. L. Aravind, L. M. Iyer, T. E. Wellems, L. H. Miller, *Cell* **115**, 771 (2003).
4. N. Rasti, M. Wahlgren, Q. Chen, *FEMS Immunol. Med. Microbiol.* **41**, 9 (2004).
5. M. Fried, P. E. Duffy, *Science* **272**, 1502 (1996).
6. L. Florens *et al.*, *Nature* **419**, 520 (2002).
7. N. Nirmalan, P. F. Sims, J. E. Hyde, *Mol. Microbiol.* **52**, 1187 (2004).
8. M. E. Wickham *et al.*, *EMBO J.* **20**, 5636 (2001).
9. P. Cheresch, T. Harrison, H. Fujioka, K. Haldar, *J. Biol. Chem.* **277**, 16265 (2002).
10. R. F. Waller, M. B. Reed, A. F. Cowman, G. I. McFadden, *EMBO J.* **19**, 1794 (2000).
11. C. Lopez-Estraño, S. Bhattacharjee, T. Harrison, K. Haldar, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 12402 (2003).
12. S. Lustigman, R. Anders, G. Brown, R. L. Coppel, *Mol. Biochem. Parasitol.* **38**, 261 (1990).
13. M. Perkins, *Exp. Parasitol.* **65**, 61 (1988).
14. Materials and methods are available as supporting material on Science Online.
15. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu;

- F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
16. M. Kaviratne, S. M. Khan, W. Jarra, P. R. Preiser, *Eukaryot. Cell* **1**, 926 (2002).
17. M. Haeggstrom et al., *Mol. Biochem. Parasitol.* **133**, 1 (2004).
18. T. Y. Sam-Yellowe et al., *Genome Res.* **14**, 1052 (2004).
19. J. Gorodkin, L. J. Heyer, S. Brunak, G. D. Stormo, *Comput. Appl. Biosci.* **13**, 583 (1997).
20. Z. Bozdech et al., *PLoS Biol.* **1**, E5 (2003).
21. K. G. Le Roch et al., *Science* **301**, 1503 (2003).
22. A search engine to identify proteins containing the PlasmoHT motif is available at www.haldarlab.northwestern.edu.
23. X.-Z. Su et al., *Cell* **82**, 89 (1995).
24. J. F. Kun et al., *Mol. Biochem. Parasitol.* **85**, 41 (1997).
25. We thank W. Kibbe, L. Zhu, V. Haztimanikatis, A. Vania Apkarian, and A. Chenn for helpful discussion. Supported by American Heart Association fellowship (0215246z to N.L.H.) and the NIH (HL69630, AI39071 to K.H.). PlasmoDB and GenBank identification codes, respectively: PFE1615c: NP_703661; PfHSP40: PFE0055c and NP_703357; PfEMP1 fragment chr4.glm_42. The PfEMP1 used for transmembrane

domain and cytoplasmic tail has NCBI identification code AAB09769.1.

Supporting Online Material
www.sciencemag.org/cgi/content/full/306/5703/1934/DC1

Materials and Methods
Figs. S1 to S4
Table S1
Bioinformatic Data

13 July 2004; accepted 19 October 2004
10.1126/science.1102737

A Draft Sequence for the Genome of the Domesticated Silkworm (*Bombyx mori*)

Biology analysis group: Qingyou Xia,^{1*†} Zeyang Zhou,^{1*} Cheng Lu,^{1*} Daojun Cheng,¹ Fangyin Dai,¹ Bin Li,¹ Ping Zhao,¹ Xingfu Zha,¹ Tingcai Cheng,¹ Chunli Chai,¹ Guoqing Pan,¹ Jinshan Xu,¹ Chun Liu,¹ Ying Lin,¹ Jifeng Qian,¹ Yong Hou,¹ Zhengli Wu,¹ Guanrong Li,¹ Minhui Pan,¹ Chunfeng Li,¹ Yihong Shen,¹ Xiqian Lan,¹ Lianwei Yuan,¹ Tian Li,¹ Hanfu Xu,¹ Guangwei Yang,¹ Yongji Wan,¹ Yong Zhu,¹ Maode Yu,¹ Weide Shen,¹ Dayang Wu,¹ Zhonghuai Xiang^{1†}

Genome analysis group: Jun Yu,^{2,3*†} Jun Wang,^{2,3*} Ruiqiang Li,^{2*} Jianping Shi,² Heng Li,² Guangyuan Li,² Jianning Su,² Xiaoling Wang,² Guoqing Li,² Zengjin Zhang,² Qingfa Wu,² Jun Li,² Qingpeng Zhang,² Ning Wei,² Jianzhe Xu,² Haibo Sun,² Le Dong,² Dongyuan Liu,² Shengli Zhao,² Xiaolan Zhao,² Qingshun Meng,² Fengdi Lan,² Xiangang Huang,² Yuanzhe Li,² Lin Fang,² Changfeng Li,² Dawei Li,² Yongqiao Sun,² Zhenpeng Zhang,² Zheng Yang,² Yanqing Huang,² Yan Xi,² Qiuhui Qi,² Dandan He,² Haiyan Huang,² Xiaowei Zhang,² Zhiqiang Wang,² Wenjie Li,² Yuzhu Cao,² Yingpu Yu,³ Hong Yu,³ Jinhong Li,³ Jiehua Ye,³ Huan Chen,³ Yan Zhou,³ Bin Liu,² Jing Wang,² Jia Ye,³ Hai Ji,² Shengting Li,² Peixiang Ni,² Jianguo Zhang,² Yong Zhang,² Hongkun Zheng,² Bingyu Mao,² Wen Wang,² Chen Ye,² Songgang Li,² Jian Wang,^{2,3} Gane Ka-Shu Wong,^{2,3,4†} Huanming Yang^{2,3†}

We report a draft sequence for the genome of the domesticated silkworm (*Bombyx mori*), covering 90.9% of all known silkworm genes. Our estimated gene count is 18,510, which exceeds the 13,379 genes reported for *Drosophila melanogaster*. Comparative analyses to fruitfly, mosquito, spider, and butterfly reveal both similarities and differences in gene content.

Silk fibers are derived from the cocoon of the silkworm *Bombyx mori*, which was domesticated over the past 5000 years from the wild progenitor *Bombyx mandarina* (1). Silkworms are second only to fruitfly as a model for insect genetics, owing to their ease of rearing, the availability of mutants from genetically homogeneous inbred lines, and the existence of a large body of information on their biology (2). There are about 400 visible phenotypes, and ~200 of these are assigned to linkage groups (3). Silkworms

can also be used as a bioreactor for proteinaceous drugs and as a source of biomaterials. Here, we present a draft sequence of the silkworm genome with 5.9× coverage.

B. mori has 28 chromosomes. More than 1000 genetic markers have been mapped at an average spacing of 2 cM (~500 kb) (4). A physical map is being constructed through the fingerprinting and end sequencing of bacterial artificial chromosome (BAC) clones (5). Many expressed sequence tags (ESTs) have been produced (6), and a 3×

draft sequence has just been announced by the International Lepidopteran Genome Project (7). Our project is independent of, but complementary to, that of the consortium. Our sequence has been submitted to the DNA Data Bank of Japan/European Molecular Biology Laboratory/GenBank (project accession number AADK00000000, version AADK01000000) and is also accessible from our Web site (<http://silkworm.genomics.org.cn>) (8). ESTs discussed in this Report can be found at GenBank (accession numbers CK484630 to CK565104).

DNA for genome sequencing is derived from an inbred domesticated variety, *Dazao* (posterior silk gland, fifth-instar day 3, on a mix of 1225 males). A whole-genome shotgun (9) technique was used, and our coverage is 5.9×. Including the unassembled reads, the total estimated genome size is 428.7 Mb, or 3.6 and 1.54 times larger than that of fruitfly (10) and mosquito (11). The N50 contig and scaffold sizes are 12.5 kb and 26.9 kb. Our assembly contains 90.9% of the 212 known silkworm genes (with full-length cDNA sequence), 90.9% of ~16,425 EST clusters, and 82.7% of the 554 known genes from other Lepidoptera. Additional details of our quality analyses are given in the supporting online material (fig. S1 and tables S1 to S6).

We developed a gene-finder algorithm *BGF* (BGI GeneFinder) (fig. S2), based on *GenScan* and *FgeneSH*. To determine a gene count for silkworm, one must correct for erroneous and partial predictions (Table 1). The final corrected gene count for silkworm is 18,510 genes, which far exceeds the official gene count of 13,379 for fruitfly

¹Southwest Agricultural University, Chongqing Beibei, 400716, China. ²Beijing Institute of Genomics of Chinese Academy of Sciences, Beijing Genomics Institute, Beijing Proteomics Institute, Beijing 101300, China. ³James D. Watson Institute of Genome Sciences of Zhejiang University, Hangzhou Genomics Institute, Key Laboratory of Genomic Bioinformatics of Zhejiang Province, Hangzhou 310008, China. ⁴University of Washington Genome Center, Department of Medicine, University of Washington, Seattle, WA 98195, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed.
E-mail: xiaqy@swau.cq.cn (Q.X.), xzh@swau.cq.cn (Z.X.), junyu@genomics.org.cn (J.Y.), gks@genomics.org.cn (G.K.-S.W.), yanghm@genomics.org.cn (H.Y.)