The data associated with this manuscript (including mass spectrometer raw files, search results, and the associated databases and search parameters) can be downloaded from www.peptideatlas.org using the following URLS:
http://www.peptideatlas.org/PASS/PASS00095
http://www.peptideatlas.org/PASS/PASS00097
http://www.peptideatlas.org/PASS/PASS00098
http://www.peptideatlas.org/PASS/PASS00101

The data are also available for download from ProteomeCommons.org Tranch, https://proteomecommons.org/tranche/, using the hash:
7egFn9UV4wwUoNIvIF0zSs8Wwe1PcBrObDyUduwrGo83WW5+87l2ZZXuNoQhiiv9weoizb0cFXYmI1OUlbEHevk8VpYAAAAAAAKXbA==

EXTENDED METHODS

Below are extra methods details not included in the main text.

Sample Fractionation by SDS-PAGE and In-Gel Tryptic Digestion.

Purified salivary gland sporozoites (10^7 for total sporozoite proteome) or the isolated surface proteins (from 3-5x10^6 salivary gland sporozoite equivalents) were electrophoresed through a 4-20% SDS-polyacrylamide gel at 180^{o}C for 40 minutes at 22^{o}C. Gels were post-stained with Imperial Stain (Thermo Scientific) and destained in double-distilled water. In-gel tryptic digestion of proteins was automated with a TECAN Freedom Evo (Männedorf, Switzerland). Gel fractions were cut into pieces ~1-3 mm on a side and placed in 96-well plates. Gel pieces were destained by adding 50 µL of 25 mM ammonium bicarbonate in 50% methanol to each well and incubating 10 minutes at 36°C while agitating at 700 RPM. The destaining solution was removed and the destaining process was repeated two more times. The gel pieces were dehydrated by adding 50 µL of acetonitrile to each well and incubated 5 minutes at 36°C while agitating at 700 RPM. The acetonitrile was removed and the dehydrating process was repeated once more. After removal of the acetonitrile the gel pieces were dried by incubating 10 minutes at 45°C while agitating at 700 RPM. Disulfide bonds were disrupted by adding 50 µL of 10 mM dithiothreitol in 100 mM ammonium bicarbonate to each well and incubating 30 minutes at 36°C while agitating at 700 RPM. The dithiothreitol solution was removed and cysteines were alkylated by adding 50 mM iodoacetamide in 100 mM ammonium bicarbonate to each well and incubating 20 minutes at 36°C while agitating at 700 RPM. Gel pieces were then washed three times with 25 mM ammonium bicarbonate in 50% methanol and dehydrated three times with acetonitrile as above. 50 µL of 6.25 ng/mL trypsin was added to each well and incubated 4.5 hours at 36°C while agitating at 700 RPM. Tryptic peptides were extracted by adding 40 µL of 3% (v/v) acetonitrile/1% (v/v) formic acid to each well and incubating 30 minutes at 36°C while agitating at 700 RPM. The extraction solution was collected. A second extraction was performed by adding 40 µL of acetonitrile to each well and incubating 30 minutes at 36°C while agitating at 700 RPM. The extraction solution was collected and combined with the first extraction. A third extraction with 3% (v/v) acetonitrile/1% (v/v) formic acid was performed in

the same manner as the first extraction and added to the first two extractions. Samples were evaporated to dryness in a rotary vacuum and reconstituted in 20 µL of HPLC loading buffer consisting of 3% (v/v) acetonitrile/0.1% (v/v) formic acid.

Gels were cut into 1 mm long x 5 mm wide slices with a grid cutter (Gel Company, San Francisco, CA). The *P. yoelii* whole cell lysate gel was divided into a total of 26 fractions, with fractions 2 through 25 being 2 mm sections of the gel. Fraction 1, the top ~1 mm of the gel (loading portion), and fraction 26, the bottom ~2 mm of the gel (dye front) were cut free-hand with a razor blade. The HPLC loading buffer (3% (v/v) acetonitrile/0.1% (v/v) formic acid) was evaporated in a rotary vacuum and each fraction was re-constituted in 2% (v/v) acetonitrile and 0.2% (v/v) trifluoroacetic acid. 2 µL of each 20 µL fraction were injected for the first LC-MS replicate. For the next two technical replicates, fractions 1 and 2 were pooled as were fractions 20 and 21, 22 and 23, and 24 and 25. Fraction 26 (the dye front) was not observed to contain any peptides not observed in other fractions and so was discarded. 4 µL of each single fraction and 8 µL of each pooled fraction were injected for each of 2 more LC-MS replicates.

The first of two *P. falciparum* whole cell lysate gels was fractionated in a fashion identical to the *P. yoelii* gel except that fractions were kept in the original 3% acetonitrile/0.1% formic acid HPLC load buffer. Fractions 1 and 2, 20 and 21, 22 and 23, and 24 and 25 were pooled. Fraction 26 was discarded. 2 µl of the single fractions and 4 µL of the pooled fractions were injected for each of three LC-MS replicates. Due to significant contamination with mosquito protein, these data were discarded and a second attempt was made.

For the second *P. falciparum* whole cell lysate (the results of which are presented in this manuscript), the gel lane was cut into 28 fractions, fractions 1 through 26 being 2 mm sections of the gel beginning ~2mm below the loading section, fraction 27 being the top ~2 mm of the gel (loading portion) and fraction 28 being the bottom ~1 mm of the gel (dye front). The extractions of these fractions were pooled prior to drying and reconstitution in HPLC load buffer such that each 20 µL final fraction contained the contents of two wells (total 4 mm of gel). Fractions 1 & 2, 3 & 4, etc. were combined. 4 µL of each fraction were injected for each of three LC-MS replicates. The *P. yoelii* and *P. falciparum* surface protein gels were not run to completion such that the lane was ~35 mm long. The lanes were manually divided into 8 fractions each and cut free-hand with a razor blade. 6 µL of each 20 µL fraction was injected for each of 2 LC-MS replicates.

Peptide separation was performed on an Agilent 1100 Nano system with an electronically-controlled flow splitter for nanoflow rates. A separate capillary pump was used for loading the trap column. Separation and trap columns were prepared in-house by packing Picofrit (New Objective) fritted, pulled-tip capillaries (360 µm O.D., 75 µm I.D., 15 µm I.D. tip) with a 15 or 20 cm bed of Dr. Maisch ReproSil Pur C18-AQ 3µm 120Å stationary phase. Trapping columns consisted of an in-house manufactured Kasil fritted capillary (360 µm O.D., 150 µm I.D.) packed with a 1 cm bed of the same stationary phase. For each LC run, sample was injected on the trap and washed for 5 minutes at 5 µL/min with loading buffer (2% v/v acetonitrile and 0.2% v/v trifluoroacetic acid). The *P. yoelii* whole cell lysate and the first attempt with a *P. falciparum*

whole cell lysate were separated on a 15 cm column by a linear gradient changing from 95% A (0.1% v/v formic acid in water) and 5% B (0.1% v/v formic acid in acetonitrile) to 65% A and 35% B in 60 minutes at 0.6 µL/min. The second *P. falciparum* whole cell lysate, the *P. falciparum* surface proteins, and the *P. yoelii* the surface proteins were separated on a 20 cm column by a linear gradient changing from 5% B to 35% B in 90 minutes at 0.5 µL/min.

Mass Spectrometry

The *P. yoelii* salivary gland sporozoite whole cell lysate was analyzed with a Thermo LTQ-Velos Orbitrap. MS1 data were collected in the over the range of 300 - 1500 Th with resolution = 60,000. The *P. falciparum* salivary gland sporozoite whole cell lysate and the surface proteins from both species were analyzed with a Thermo LTQ-Velos Pro Orbitrap Elite. MS1 data were collected in the over the range of 300 - 1600 Th with resolution = 60,000. For all experiments, FTMS preview scan and predictive automatic gain control (pAGC) were enabled. The full scan FTMS target ion volume was $1x10^6$ with a max fill time of 500 ms. MS2 data were collected in the LTQ with a target ion volume of $1x10^4$ and a max fill time of 80 ms. The 20 most intense peaks from a preview scan of each full Orbitrap scan were selected (with a selection window of 2.0 Th) for collision-induced dissociation (CID) with wide-band activation. Dynamic exclusion was enabled to exclude an observed precursor for 30 seconds after a single observation. The dynamic exclusion list size was set at the maximum 500 and the exclusion width was set at ±5 ppm based on precursor mass. Monoisotopic precursor selection and charge state rejection were enabled to reject precursors with z = +1 or unassigned charge state.

Peak list generation

Thermo .RAW files were converted to mzXML format using MSConvert and searched with X!Tandem version 2010.10.01.1 and SEQUEST v.27 rev.0. Spectra were searched against databases comprised of the *Plasmodium* species in question, *A. gambiae* (to account for mosquito debris), common protein contaminants, and decoys. The *A. gambiae* database (vectorbase.org, version 3.6, updated 10 October 2011) contained 14,324 entries. The contaminant database was a modified version of the common Repository of Adventitious Proteins (cRAP Version 2009.05.1, last updated 18 October 2011, www.thegpm.org/crap) with the Sigma Universal Standard Proteins removed and human angiotensin II and [Glu-1] fibrinopeptide B (MS test peptides) added, for a total of 66 entries. Decoys were generated with Mimic (www.kaell.org), which randomly shuffles peptide sequences between tryptic residues but also retains peptide sequence homology in decoy entries. Decoy protein sequences were interleaved among real entries, randomly alternating between listing the real entry or the decoy entry first. *The P. falciparum* database contained a total of 39,828 entries comprised of 5,524 *P. falciparum* entries (plasmodb.org, v.8.2, updated 27 October 2011), 14,324 *A. gambiae* entries, 66 cRAP entries, and 19,914 decoys. The *P. yoelii* database contained a total of 44,746 entries comprised of 7,983 P*. yoelii* entries (based on Vaughan et al. [Bioinformatics. 2008 Jul 1;24(13):i383-9.] and manually curated to remove duplicate entries, correct spurious entries, and to add new entries), 14,324 *A. gambiae* entries, 66 cRAP entries, and 22,373 decoys. To compensate for the poor sequence coverage of the *P. yoelii* database, *P.*

yoelii data were also searched against a *P. berghei* database. The *P. berghei* database contained a total of 38,588 entries comprised of 4,904 *P.berghei* entries (plasmodb.org, v.8.2, updated 27 October 2011), 14,324 *A. gambiae* entries, 66 cRAP entries, and 19,294 decoys. Similar search criteria were used for X!Tandem and SEQUEST. A wide precursor mass tolerance of 0.1 Da was used to improve the performance of the accurate mass binning tool available in Peptide Prophet. Fragment ions were searched with a mass tolerance of ±0.4 Da in X!Tandem. Fragment ion mass tolerance is not specified in SEQUEST. Peptides were assumed to be semi-tryptic with up to 2 missed cleavages were allowed. The search parameters included a static modification of + 57.021464 Da at C for carbamidomethylation by iodoacetamide and a potential modification of +15.994915 for oxidation at M. Additionally, X!Tandem automatically searched for -17.026549 Da for deamidation at N-terminal Q and -18.010565 Da for loss of water at N-terminal E from formation of pyro-Glu as well as -17.026549 Da at N-terminal carbamidomethylated C for deamidation from formation of S-carbamoylmethylcysteine. For surface proteins, a potential modification of 339.161662 at K was added for the biotin tag. MS/MS data were analyzed using the Trans Proteomic Pipeline (Deutsch, E.W., et al., A guided tour of the Trans-Proteomic Pipeline. Proteomics, 2010. 10(6): p. 1150-9.) version 4.5 Rev.2. Peptide spectrum matches (PSM) generated by each search engine were analyzed separately with Peptide Prophet to assign each PSM a probability of being correct. Accurate mass binning was employed to promote PSMs whose theoretical mass closely matched the observed mass of the precursor ion, and to correct for any systematic mass error. Decoys and the non-parametric model option were used to improve PSM scoring. The Peptide Prophet scores for a given analysis (e.g. all X!Tandem and SEQUEST results of all the injections of all gel fractions of one sporozoite lysate) were then combined in iProphet), which assigns a probability for each unique peptide sequence based on how often it is observed at different charge states with different modifications and by different search engines, as well as whether other peptides from the same protein are also observed.

The false positive error rate (FPER) for PSMs was used as estimated by iProphet. When determining the number of peptides or PSMs identified by a certain experiment, only peptides and PSMs with iProphet probabilities corresponding to a FPER <1.00 % were counted. In this case, the estimate of FPER is actually conservative. PSMs matching to decoy entries were not removed from results files in order that the performance of the various models could be confirmed. Assuming that false positive PSMs arise randomly, the probability of a false positive PSM matching a decoy entry rather than a real entry is the same as percentage of decoy peptides in the database (0.511 for all three databases). Decoy PSMs were removed after filtering out results below the 1.00% FPER cut-off; therefore, the actual FPER for the reported peptides and PSM is more accurately estimated as 0.489%.

Protein identifications were inferred with ProteinProphet (Nesvizhskii et al., Anal. Chem. 2003, 75, 4646-4658). The false positive error rate (FPER) at a given Protein Prophet probability cut-off was calculated as the fraction of decoy protein inferences (FPER=[# of decoys]/[# of all protein entries]). Only proteins identified at Protein Prophet probabilities corresponding to a FPER less than 1.00% were reported. As with the peptides and PSMs, the decoy entries were not reported in the final dataset. If the probability of a false positive protein inference matching

to a decoy is assumed to be the database decoy ratio of 0.511, then at a decoy-estimated FPER of 1.00%, 1.00% of the protein inferences are decoys and 0.957% are false positives matching real sequences (database decoy ratio = [# decoy peptides in database]/([# decoy peptides in database] + [# of real peptides in database])). When the decoy entries are removed, the FPER is estimated as [# of false positives matching to real proteins]/[total # of proteins with decoys removed], or 0.00957*[# of proteins]/0.990*[# of proteins], or 0.967%. Therefore the percentage of decoy protein entries found above a given Protein Prophet probability cut-off is a good proxy for estimating FPER.