scientific data

OPEN

DATA DESCRIPTOR

Borrelia PeptideAtlas: A proteome resource of common *Borrelia burgdorferi* isolates for Lyme research

Panga J. Reddy[1,6], Zhi Sun[1,6], Helisa H. Wippel[1,6,7], David H. Baxter[1], Kristian Swearingen[1], David D. Shteynberg[1], Mukul K. Midha[1], Melissa J. Caimano[2], Klemen Strle[3], Yongwook Choi[4], Agnes P. Chan[4], Nicholas J. Schork[4], Andrea S. Varela-Stokes[5] & Robert L. Moritz[1]✉

Lyme disease is caused by an infection with the spirochete *Borrelia burgdorferi*, and is the most common vector-borne disease in North America. *B. burgdorferi* isolates harbor extensive genomic and proteomic variability and further comparison of isolates is key to understanding the infectivity of the spirochetes and biological impacts of identified sequence variants. Here, we applied both transcriptome analysis and mass spectrometry-based proteomics to assemble peptide datasets of *B. burgdorferi* laboratory isolates B31, MM1, and the infective isolate B31-5A4, to provide a publicly available Borrelia PeptideAtlas. Included are total proteome, secretome, and membrane proteome identifications of the individual isolates. Proteomic data collected from 35 different experiment datasets, totaling 386 mass spectrometry runs, have identified 81,967 distinct peptides, which map to 1,113 proteins. The Borrelia PeptideAtlas covers 86% of the total B31 proteome of 1,291 protein sequences. The Borrelia PeptideAtlas is an extensible comprehensive peptide repository with proteomic information from *B. burgdorferi* isolates useful for Lyme disease research.

Background & Summary

The spirochete *Borrelia burgdorferi* is the causative agent of Lyme disease, the main vector-borne infection in North America with over 476,000 cases per year between 2010 and 2018[1,2]. *B. burgdorferi* is transmitted to humans through the bite of infected nymphal or adult ticks, and the untreated infection may cause a multisystem disorder characterized by early and later stage signs and symptoms[3,4]. Early stage symptoms occur within the first 30 days post tick bite and may include, among others, fever, joint aches and swollen lymph nodes, and a rash referred to as erythema migrans[3]. When symptoms persist for months post infection, later symptoms can manifest as facial palsy, arthritis with severe joint pain and swelling, severe headaches, and inflammation of the brain[3]. Lyme disease treatment includes the use of antibiotics (primarily doxycycline), and in most cases, early treatment can provide a cure within 2 to 4 weeks[5]. However, some patients are not diagnosed early, or may continue to present symptoms for more than 6 months after the treatment ends, a condition termed as Post-Treatment Lyme Disease Syndrome (PTLDS)[6]. For this reason, an early and correct diagnosis of Lyme disease is key to initiate the treatment during acute disease and potentially decrease the risk of PTLDS. Currently available tests, including the CDC-recommended 2-tiered testing protocol, are designed to detect antibodies against *B. burgdorferi* in patient blood, which takes several weeks to be produced and can result in a false negative diagnosis[7]. Therefore, the development of alternative diagnostic methodologies, such as next-generation

[1]Institute for Systems Biology, Seattle, Washington, USA. [2]Department of Medicine, UConn Health, Farmington, Connecticut, USA. [3]Department of Molecular Biology and Microbiology, Tufts University School of Medicine, Boston, Massachusetts, USA. [4]Translational Genomics Research Institute, Phoenix, Arizona, USA. [5]Tufts University Cummings School of Veterinary Medicine, Department of Comparative Pathobiology, Grafton, MA, 01536, USA. [6]These authors contributed equally: Panga J. Reddy, Zhi Sun, Helisa H. Wippel. [7]Deceased: Helisa H. Wippel. ✉e-mail: rmoritz@systemsbiology.org

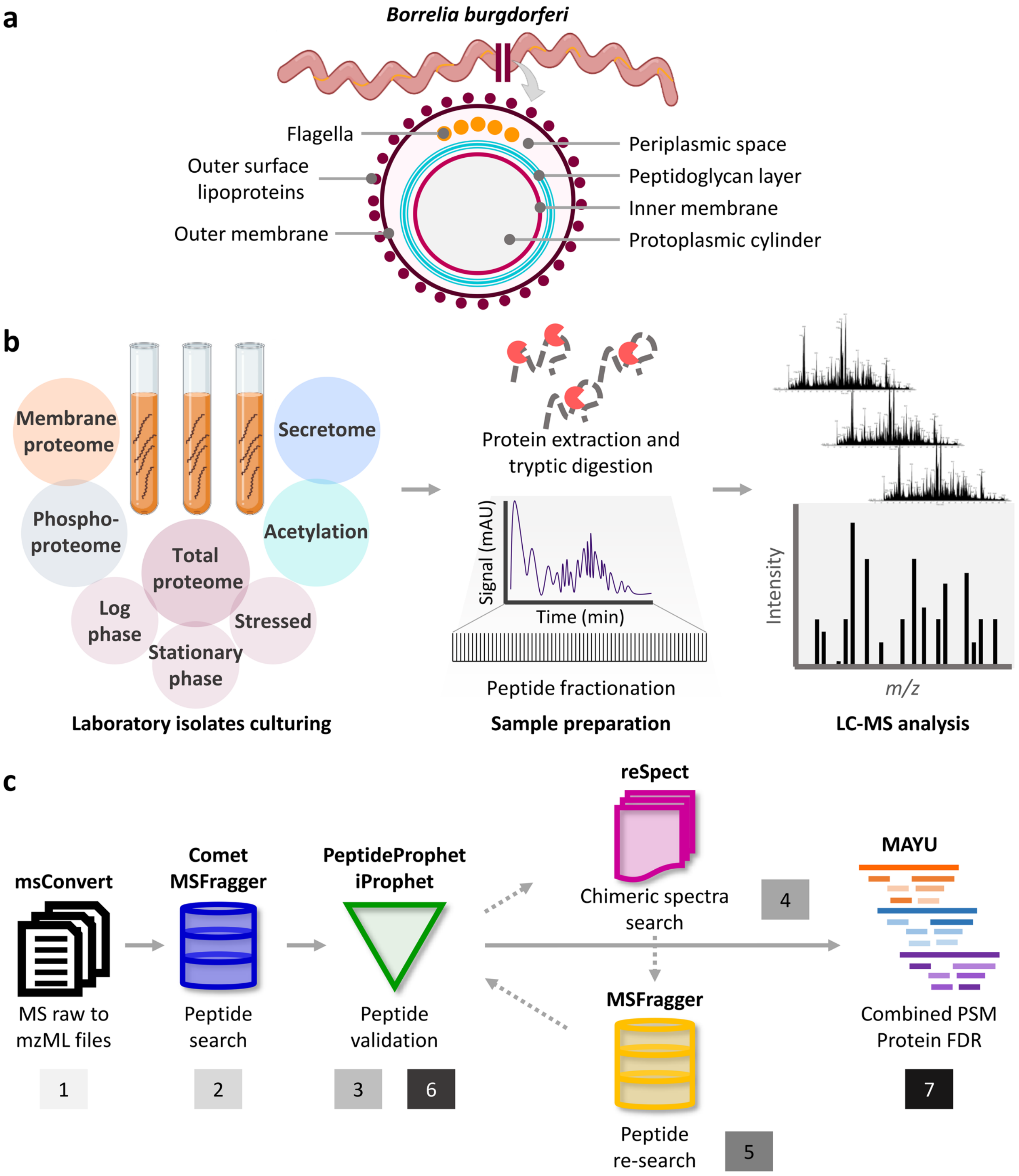


Fig. 1 Overview of experimental workflow for the development of the Borrelia PeptideAtlas. (**a**) Cartoon depiction of the *Borrelia burgdorferi* structure. (**b**) Experiment workflow. *B. burgdorferi* was cultured in different environmental conditions, including log phase, stationary phase, and stress conditions for total proteome analysis. Different enrichment assays were applied for the analysis of the secretome, the membrane proteome, phosphoproteome, and acetylation. Samples were prepared directly for LC-MS analysis, or alternatively fractionated prior to LC-MS. (**c**) Trans-Proteomic Pipeline (TPP) workflow used for the Borrelia PeptideAtlas assembly. Further details in Methods.

serologic assays[7], which include recombinant proteins and synthetic peptides targeting important factors in the *B. burgdorferi* infection, survival and proliferation mechanisms are urgently needed.

B. burgdorferi is an atypical Gram-negative bacterium due to lack of LPS in its cell wall and the presence of immuno-reactive glycolipids, a peptidoglycan layer, and lipoproteins in the outer membrane[8–11] (Fig. 1a). These lipoproteins play a key role in the infectivity and proliferation of the spirochete in ticks and in mammal hosts[12], and are mostly encoded by the spirochete linear and circular plasmids, besides the single chromosome[13].

Specifically, the *B. burgdorferi* B31 genome sequenced in 1997 revealed the presence of one linear chromosome with 843 genes, and at least 21 plasmids (12 linear and 9 circular) with 670 genes and 167 pseudogenes[9,14]. Out of a total of 1,513 potential genes, 1,291 are predicted as unique protein-coding genes[14]. B31 is the most commonly studied *B. burgdorferi* non-infective laboratory isolate, but an increasing number of infective genotypes have been isolated in North America and around the world, which are isolated from infected ticks or Lyme disease patients and display different pathogenic and infective patterns[15,16]. The genetic variability of subtypes of *B. burgdorferi* isolates – e.g., varying number of plasmids encoding for infection-related lipoproteins – may ultimately lead to revealing a (i) diverse severity of Lyme symptoms and (ii) diverse spirochetal response to the antibiotic treatment[15,16]. Hence, a proteogenomic approach combining genome sequencing data with transcriptomic and proteomic data from different isolates is a robust strategy to unveil the *B. burgdorferi* strain pathogenicity and begin to develop new strategies for more efficient diagnosis and treatment of Lyme disease.

Although numerous proteomic reports exist for *B. burgdorferi* isolates[17–24], no information is available as a comprehensive and searchable compendium of public data, instead users have to resort to obtaining mass spectrometry (MS) raw files and search any, or all, of these data individually. In this study, we performed MS-based comprehensive proteome analysis of several different laboratory *B. burgdorferi* isolates: two commercially available isolates B31[25] and MM1[26,27], and the infective isolate, B31-5A4[28]. These datasets include information on the total proteome, secretome, and membrane proteome of the *Borrelia* isolates (Fig. 1b). The uniform analysis of a total of 386 MS runs through the Trans-Proteomic Pipeline (TPP) (Fig. 1c) allowed the identification of 81,967 distinct peptide sequences at false discovery rate (FDR) levels less than 1.1%. The identified unique peptides map to 1,113 proteins among all isolates with a protein-level FDR less than 1.27%, covering 86% of the total B31 proteome. Additionally, for the comparison of protein abundance levels with mRNA levels, we performed transcriptomic analysis of isolates B31, MM1, and B31-5A4. The complex and detailed proteomic results achieved here are constructed into a searchable public repository called Borrelia PeptideAtlas adhering to FAIR (Findability, Accessibility, Interoperability, and Reusability) principles[29]. PeptideAtlas is a unique public community resource which contains large scale assembly of mass spectrometry data uniformly processed with high quality through the TPP[30,31]. This repository has data from a wide range of organisms, including human, equine, porcine, chicken, *Saccharomyces cerevisiae, Drosophila melanogaster* and *Candida albicans*[32–38] among many others. The Borrelia PeptideAtlas allows the assessment of protein content of *B. burgdorferi* isolates and compare detectable protein sequences. The continuous update of this repository with expandable data sources for many other *B. burgdorferi* isolates, including clinically relevant isolates, will enable the investigation of the dynamic proteome of this spirochete through its infection stages and their vastly different environments. The diverse proteomic information from multiple infective isolates with credible data presented by the Borrelia PeptideAtlas will be useful to understand the protein complement of each isolate, their comparisons including overall protein abundance, and assist in pinpointing potential protein targets which are common to infective isolates and may be key in the infection process. The Borrelia PeptideAtlas is readily available as an important resource for the Lyme disease research community.

Methods

B. burgdorferi isolates and spirochete culture.

Two common commercially available laboratory isolates of *B. burgdorferi* [B31 (ATCC 35210)[25] and MM1 (ATCC 51990)[26,27]], and the infective isolate B31-5A4 (a clonal isolate of 5A4 that has been passaged through rodents to maintain infectivity)[28] were cultured in BSK-H complete media with 10% rabbit serum, at 34 °C in 5.0% CO_2 incubator. B31-5A4 was cultured at a low passage to minimize loss of endogenous plasmids. The spirochetes were harvested and collected at mid-log phase (3 to 5 × 10^7) or stationary phase (3 to 5 × 10^8) for proteomic analysis. For secretome analysis, mid-log phase cells were harvested, washed and transferred to serum-free media, i.e., BSK-H media without rabbit serum, and grown for 24 h at 34 °C in 5.0% CO_2 incubator. The culture was centrifuged at 369 × *g* for 1 h and collected both the media and the bacteria were collected. The media was used for secretome analysis and bacteria were used for stress proteome analysis.

Total proteome extraction.

B. burgdorferi pellets collected from log phase, stationary phase, and stressed bacteria (grown in serum-free BSK-H media; Sigma, catalog number B8291) for 24 h were washed with PBS buffer (pH 7.4) four times to remove the media and centrifuged at 369 × *g* for 3 min at each wash. The bacterial pellets were dispersed in lysis buffer of 8 M urea in 100 mM ammonium bicarbonate (AmBic) and protease inhibitor cocktail (cOmplete, Roche, catalog number 4693132001). The bacterial cell lysis was performed using a freeze-thaw cycle followed by sonication (30 s pulse, 20% amplitude, 5 cycles). Cell lysate was centrifuged at 15,294 × *g* for 30 min and clear supernatant was collected for LC-MS analysis.

Secretome extraction.

B. burgdorferi B31 culture at mid-log phase was washed with PBS buffer to remove the media and allow transfer of the bacteria to serum-free BSK-H media for 24 h. The bacteria were collected by centrifugation at 369 × *g* for 3 min, and the media was used for the secretome analysis. To the media, four volumes of chilled acetone were added and precipitated the protein for 30 min at 4 °C. Protein pellets were collected by centrifugation and washed with acetone two more times. Protein pellet was dissolved in 8 M urea in 100 mM AmBic (pH 8.0), followed by in-solution protein digestion and LC-MS analysis.

In-solution digestion.

Protein samples of isolates B31, MM1 and B31-5A4 from log phase, stationary phase, stressed bacteria, and secretome were digested with trypsin for proteomic analysis. Briefly, 100 µg of protein from each condition were reduced with 5 mM *tris*-(2-carboxyethyl)phosphine (TCEP, Thermo Fisher Scientific, catalog number 20490) and alkylated with iodoacetamide (IAM, Merck-Sigma, catalog number 144-48-9). Sequencing grade modified trypsin (Promega, catalog number V5111) was added at a 50:1 protein-to-enzyme ratio and incubated at 37 °C overnight. Samples were acidified with 1% TFA (Thermo Fisher Scientific, catalog number

28901) and purified further using a C18 Atlas column (Tecan, USA; catalog number 30165979) and prepared for two-dimensional peptide fractionation or directly for LC-MS analysis.

Membrane proteome analysis.

B. burgdorferi B31 was cultured as above and harvested by centrifugation at 1,000 × *g* for 60 min. The bacterial pellets were washed with ice cold PBS buffer (pH 8.0) three times. The bacterial pellets were resuspended in PBS buffer with final cell number of 10^8/mL of buffer. A concentration of 10 mM Sulfo-NHS-SS-Biotin (Thermo Fisher Scientific, catalog number 21331) was prepared according to the manufacturer's guidelines. The stock solution of Sulfo-NHS-SS-Biotin was added to the bacterial pellets and mixed via pipette. The bacterial pellets were incubated at 4 °C for 60 min for the labeling reaction. Each bacterial pellet was centrifuged at 2,795 × *g* for 20 min and supernatant was discarded. Tris buffered saline (TBS, pH 7.4) was added to the bacterial pellets and incubated at room temperature for 15 min and centrifuged at 15,294 × *g* for 10 min. *B. burgdorferi* pellets were washed with PBS buffer (pH 7.4) and dispersed in 100 mM Tris-HCl buffer (pH 8.0) containing the protease inhibitor cocktail. The cell lysis was performed using freeze-thaw cycles as described above. The *B. burgdorferi* B31 lysate was centrifuged at 15,294 × *g* for 30 min and the supernatant was collected for soluble proteome analysis. The resultant protein pellet was washed with 100 mM Tris-HCl buffer (pH 8.0) and dissolved in membrane dissolving buffer (8 M urea having protease inhibitor cocktail) and incubated at 4 °C for 30 min with intermediate vortexing. The sample was centrifuged at 15,294 × *g* for 30 min and the supernatant was collected for membrane protein analysis. Alternatively, Dynabeads MyOne Streptavidin T1 (Invitrogen, catalog number 65601) were prepared by adding PBS buffer (pH 7.4). Membrane fractions were transferred to the tubes having beads and incubated for 1 h at 4 °C with end-over-end rotation. Beads were sequestered by a magnet (Invitrogen, catalog number 12321D) and sequential washing steps were performed as follows: 1 mL per wash and 8 min per wash with solution-I (2% SDS), solution-II (6 M urea, 0.1% SDS, 1 M NaCl and 50 mM Tris pH 8.0), solution-III (4 M urea, 0.1% SDS, 200 mM NaCl, 1 mM EDTA and 50 mM Tris pH 8.0) and solution-IV (0.1% SDS, 50 mM NaCl and 50 mM Tris pH 8). The bound proteins were eluted in 2 × SDS-PAGE sample buffers (Invitrogen, catalog number M8546G).

SDS-PAGE and in-gel digestion for membrane proteins.

The biotin labeled proteins eluted in 2 × SDS-PAGE sample buffers were mixed with reducing agent and bromophenol blue (BPB) and resolved on 12% SDS-PAGE gel. The gel was stained with SimplyBlue Safe Stain (Invitrogen, catalog number LC6065). Each lane of the SDS-PAGE was cut into five bands and processed for in-gel digestion. In brief, the gel pieces were washed with 50 mM AmBic and 2:1 ratio of acetonitrile:AmBic alternatively, three times for five min each to remove the stain. Gel bands were treated with DTT (56 °C for 1 h) and IAM (20 min in the dark) for reducing and alkylating the cysteine residues. Trypsin (500 ng/μL) along with sufficient 50 mM AmBic was added to each gel band and incubated at 37 °C overnight. Peptide elution was performed by adding 60% of acetonitrile in 0.1% trifluoroacetic acid (TFA) to the bands, vortexing for 10 min and collecting the solution into a fresh tube. The process was repeated two more times with acetonitrile gradient 70% and 80% in 0.1% TFA and pooled to the previous fraction. Samples were purified with a C18 Atlas column (Tecan, USA; catalog number 30165979) and prepared for peptide fractionation or directly for LC-MS analysis.

Enrichment of phosphorylated peptides.

Enrichment of phosphorylated peptides was performed as previously described[39]. Briefly, 200 μg of tryptic peptides from B31-5A4 cells were resuspended in 500 μL of loading buffer [80% acetonitrile, 5% TFA, 0.1 M glycolic acid], and incubated with 400 μg of MagReSyn Ti-IMAC HP (Resyn Biosciences, catalog number MR-THP002). Beads were washed 3 times with 500 μL of 80% acetonitrile and 1% TFA, 3 times with 500 μL of 10% acetonitrile and 0.2% TFA, and peptides were eluted with 200 μL of 2% ammonium hydroxide. A second round of enrichment was performed with MagReSyn Zr-IMAC HP beads (Resyn Biosciences, catalog number MR-ZHP005). Samples were cleaned up with AttractSPE Disks Tips C18 (Affinisep, catalog number Tips-C18.T2.200.960). In brief, acidified samples in were loaded to 200 μL-tips (30 μg binding capacity) and washed once with 200 μL of 0.5% acetic acid and 0.1% phosphoric acid, followed by elution with 100 μL of 80% acetonitrile and 0.5% acetic acid. Samples were then dried out to completion in a SpeedVac (Thermo Scientific, catalog number SPD120-115) and prepared for peptide fractionation or directly for LC-MS analysis.

High-pH fractionation.

Peptides were reconstituted in 200 mM ammonium formate (pH 10) and fractionated on an Agilent 1200 Series Gradient HPLC system. Peptides were loaded on Zorbax SB-C18 column (4.6 × 150 mm, 5 μm particle size; Agilent, catalog number 41115709) and fractionated using a linear gradient of 0-100% of B (60% acetonitrile in 20 mM ammonium formate pH 10). For the second set, tryptic peptides were fractionated with a flow rate of 100 μL/min of buffer A [0.1% (vol/vol) triethylammonium bicarbonate (TEAB, Honeywell Fluka, catalog number 17902) in water] and 1%/min gradient of buffer B [60% (vol/vol) acetonitrile, 0.1% (vol/vol) TEAB in water], with a Brownlee Aquapore RP-300 column (100 mm × 2.1 mm I.D., Perkin-Elmer; catalog number 07110060). The total 56 fractions were pooled to 14 final fractions through groupings of 3 disparate fractions to cover the range. These fractions were lyophilized and reconstituted in 0.1% formic acid (FA) and 2% acetonitrile for LC-MS/MS analysis.

LC-MS/MS analysis.

The mass spectrometry data was deposited to the ProteomeXchange Consortium via the PRIDE[40] partner repository with the dataset identifier PXD046281[41].

Q-Exactive HF.

B. burgdorferi samples, except B31-Biotin labeled samples, were analyzed on an EasyLC (Thermo Fisher Scientific) coupled with Q-Exactive HF mass spectrometer (Thermo Fisher Scientific). The purified dried peptides were dissolved in loading buffer (0.1% FA in water) and loaded on to the Acclaim PepMap 100 trap (2 cm long, 75 μm ID, C18 3 μm; Thermo Fisher Scientific, product number 164946). Analytical column

#proteins	B31	B31-5A4	MM1
RefSeq	1,359	1,354	1,159
GenBank	1,339	1,429	1,302
UniProt	1,291		
ISB		814	
Total non-redundant	1,485	1,443	1,383

Table 1. Number of protein sequences per reference database.

(PicoChip: 105 cm, 1.9 µm, REPROSIL Pur C-18-AQ, 120 Å; New Objective, USA, material number r119.aq.) with a flow rate of 300 nL/min was used for the separation of the peptides with a linear gradient of 5–35% buffer-B (90% acetonitrile in 0.1% FA) over 120 min. The data acquisition parameters include: mass range 375-1375 *m/z*, MS resolution of 30,000 (at *m/z* 200), MS2 resolution of 15,000 (at *m/z* 200), full scan target at 3 × 10^6, 40 top intense peaks with charge state > 2 were selected for fragmentation using HCD with 28% normalized collision energy, dynamic exclusion time of 25 s and profiler mode with positive polarity. Alternatively, B31-Biotin labelled peptides were analyzed using Agilent 1100 nano HPLC pump coupled to an LTQ Velos Pro-Orbitrap Elite mass spectrometer (Thermo Scientific, USA). Sample was loaded onto a trap column consisting of a fritted capillary (360 µm O.D., 150 µm I.D.). Peptides were separated with in-house packed column with a 20 cm bed of C18 (Dr. Maisch ReproSil-Pur C18-AQ, 120 Å, 3 µm; product number r13.aq.) having an integrated fritted tip (360 µm O.D.), 75 µm I.D., 15 µm I.D. tip; New Objective). Data-dependent acquisition was performed by selecting top precursor ions for fragmentation using collision-induced dissociation (CID) with a 30 sec dynamic exclusion time limit.

Orbitrap Fusion Lumos. B31, B31-5A4, and MM1 pooled fractions from high pH fractionation were individually analyzed on a Vanquish Neo nanoUHPLC coupled to an Orbitrap Fusion Lumos instrument (Thermo Scientific, USA), equipped with an EasySpray nanoelectrospray source. Peptides were loaded onto a trap column (0.5 cm × 300-µm I.D., stationary phase C18) with a flow rate of 10 µL/min of mobile phase: 98% (vol/vol) LC-MS solvent A [0.1% (vol/vol) FA in water] and 2% (vol/vol) LC-MS solvent B [0.1% (vol/vol) FA in acetonitrile]. Peptides were chromatographically separated on a 50-cm analytical column [(EASY-Spray ES803A, Thermo Scientific); 75 µm × 50 cm, PepMap RSLC C18, 2-µm I.D., 100-Å-pore-size particles] applying a 115-min linear gradient: from 3% solvent B to 8% solvent B in 10 min, to 30% solvent B in 90 min, and ramped to 80% solvent B in 5 min, at a flow rate of 250 nL/min. The column temperature was set to 45 °C. Spray voltage was set to 1.8 kV and s-lens RF levels at 30%. The mass spectrometer was set to high resolution data-dependent acquisition (DDA) of 15 topN most intense ions with charge state of + 2 to + 5. Each MS1 scan (120,000 resolving power at 200 *m/z*, automated gain control (AGC) of 125%, scan range 300 to 1,500 *m/z*, and dynamic exclusion of 30 s, with maximum fill time of 50 ms) was followed by 15 MS2 scans (30,000 resolving power at 200 *m/z*, AGC of 200%, maximum fill time of 54 ms). Higher-energy collisional dissociation (HCD) was used with 1.6 *m/z* isolation window and normalized collision energy of 30%.

Orbitrap Eclipse Tribid. B31-5A4 samples enriched for phosphorylated peptides were analyzed on a Vanquish Neo nanoUHPLC coupled to an Orbitrap Eclipse Tribid mass spectrometer (Thermo Scientific, USA), equipped with a Nanospray Flex source. Peptides were loaded onto a trap column (0.5 cm × 300-µm I.D., stationary phase C18) with a flow rate of 10 µL/min of mobile phase: 98% (vol/vol) LC-MS solvent A (0.1% FA in water) and 2% (vol/vol) LC-MS solvent B (0.1% FA in acetonitrile). Peptides were chromatographically separated on a 50-cm analytical column (Double nanoViper PepMap Neo DNV75500PN, Thermo Scientific; 75 µm × 500 mm, C18, 2-µm I.D., 100-Å-pore-size particles) applying a 135-min linear gradient: 0-35% solvent B in 120 min, and ramped to 80% solvent B in 15 min, at a flow rate of 250 nL/min. The mass spectrometer was set to high resolution data-dependent acquisition (DDA) of most intense ions with charge state of + 2 to + 5. Each MS1 scan (120,000 resolving power at 200 *m/z*%, scan range 375 to 1,550 *m/z*, maximum injection time of 118 ms) was followed by MS2 scans (30,000 resolving power at 200 *m/z*). Higher-energy collisional dissociation (HCD) was used with 1.6 *m/z* isolation window and normalized collision energy of 28%, and maximum injection time of 60 ms.

Triple-TOF. B31 and MM1 samples were analyzed using 5600 + Triple-TOF mass spectrometry (ABSciex, USA) coupled with Eksigent 400 nano-HPLC (Sciex, USA). Peptides were run separately by loading on trap column (200 µm × 0.5 mm, Chrom XP C18-CL 3 µm, 120 Å, Eksigent, AB Sciex). Peptides were separated on analytical column (75 µm × 20 cm, ChromXP C18- CL 3 µm, 120 Å, Eksigent, Sciex) with the gradient of buffer B (95% acetonitrile in 0.1% FA) and flowrate was 300 nL/min. The linear gradient profile from 3 to 40% buffer B in 103 min, increased to 80% in 105 min and continued to 113 min. Buffer B was then brought down to 3% in 115 min and continued until 140 min. Precursor mass was measured at MS1 level in high resolution mode with mass range of 400-1250 *m/z*. The TOF-MS parameters includes: nanospray ionization, curtain gas (CUR)- 25, ion source gas 1 (GS1)- 3, interface heater temperature (IHF)- 150, ion spray voltage floating (ISDF)-2300, declustering potential (DP)-100, collision energy (CE)-10, accumulation time- 50 ms, mass tolerance 100 ppm, exclude former peptide ion- 15 sec after first detection and precursors selected for each cycle top 30 intense peaks with charge state 2 to 4 having greater than or equal to 150 counts were selected for fragmentation using rolling collision energy. Similarly, at MS2 level, spectra were collected in *m/z* range of 100-1500 *m/z* with 50 ms accumulation time in high sensitivity mode.

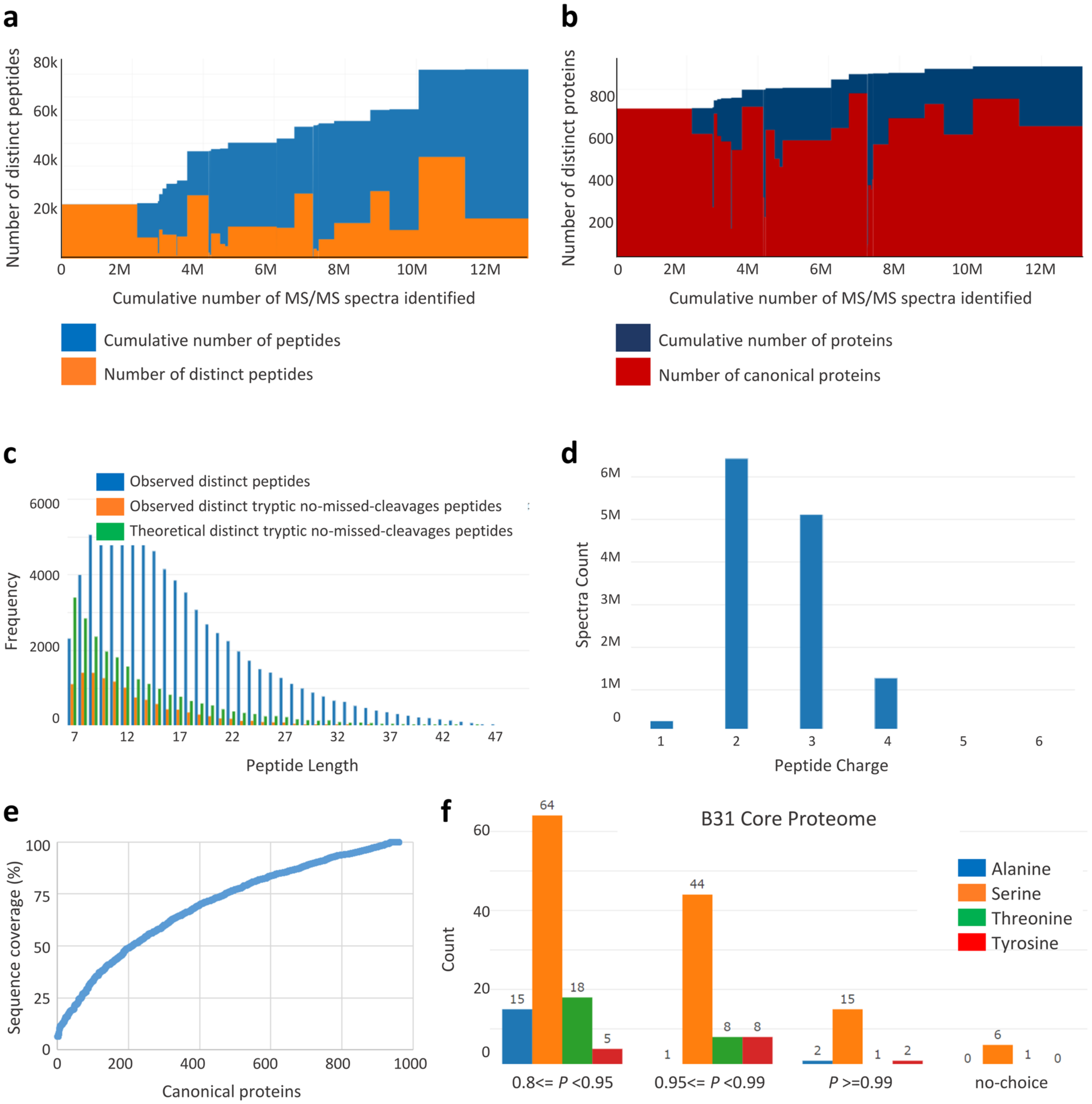


Fig. 2 Borrelia PeptideAtlas experiment contribution. (**a**) Number of peptides which contributed to each experiment, and the cumulative number of distinct peptides for the build as of that experiment. (**b**) Cumulative number of canonical proteins contributed by each experiment. Height of red bar is the number of proteins identified in experiment; height of blue bar is the cumulative number of proteins; width of the bar (x-axis) shows the number of spectra identified (PSMs), above the threshold, for each experiment. (**c**) Frequency distributions of peptide length by number of amino acids. The figure shows frequency of distinct peptides (in blue), distinct tryptic peptides with no missed cleavages (in orange), and theoretical, i.e., not observed, tryptic peptides with no missed cleavage (in green). (**d**) Frequency distributions of peptide charge. (**e**) Relative protein sequence coverage for canonical proteins based on sequence coverage, i.e., the % of amino acids of the primary sequence which were identified. (**f**) Histogram showing the frequency distribution of PSMs of phosphorylated sites (false positive-alanine, serine, threonine, and tyrosine), identified for B31 UniProt core proteome, according to PTMProphet probability (*P*). *P* ranges from 0.8 to 0.99. no-choice: shows PSMs with only one possible phosphorylation site available, hence $P = 1$. Blue, yellow, green, and red bars indicate alanine, serine, threonine, or tyrosine phosphorylated sites, respectively.

timsTOF PRO. MM1 pooled fractions from high pH fractionation were spiked in with iRT standard peptides (Biognosys AG, Schlieren, Switzerland) and subjected to mass spectrometry (MS) analysis using a timsTOF PRO mass spectrometer (Bruker), coupled to a Vanquish Neo HPLC system (ThermoFisher Scientific) in nanoflow setup for both Data-Dependent Acquisition-Parallel Accumulation-Serial Fragmentation (DDA-PASEF) and Data-Independent Acquisition (DIA) PASEF modes. Both modes were operated with buffer A (0.1% FA in water),

Protein label	#proteins*	Technical definition
Canonical	911	Proteins with at least two 9 amino acids or greater peptides with a total extent of 18 amino acids or greater that are uniquely mapping within the core reference proteome. B31 UniProt proteome is used as core proteome.
Noncore Canonical	324	Noncore canonical means that there are uniquely mapping peptides to this protein that do not map to a protein that is considered part of the core proteome of a species. A non-core canonical protein might be an isoform, contaminant, or protein missing from the core reference proteome. Contaminants are not included in the count.
Weak	62	Protein has more unique peptides than shared peptides, and only one uniquely mapping peptide 9 amino acids or greater.
Insufficient evidence	4	Protein has more unique peptides than shared peptides, but none are 9 amino acids or greater.
Marginally Distinguished	220	Protein has unique peptides, but there are not more unique peptides than shared peptides, and the extended length of unique peptides is less than 18 amino acids.
Indistinguishable Representative	49	Protein has no unique peptides, and there are several indistinguishable proteins, but this one is assigned to be an Indistinguishable Representative and the others are Indistinguishable.
Total	1,570	

Table 2. Protein identification categories in the Borrelia PeptideAtlas build. *Contaminants are not included in all protein counts.

and buffer B (0.1% formic acid in acetonitrile). Peptides were trapped on a 0.5 cm × 0.3 mm trap cartridge Chrom XP C18, 3 µm (Thermo Fisher Scientific) at 10 µL/min and separated on a C18 UHP 15 cm × 0.15 mm I.D. × 1.5 µm column (Bruker/PepSep) at either 600 nL/min or 1 µL/min for 66 and 45 minutes, respectively. The gradient elution profile for both flow rates was as follows: 3% to 25% B in 51 min (37 min for 1 µL/min), 25% to 35% B in 15 min (8 min for 1 µL/min), 35% to 80% B in 1 min, followed by an isocratic flow at 80% B for 2 min. The Captive Spray ion source was equipped with a 20 µm emitter (Bruker, catalog number 1811107) and the parameters were as follows: 1700 V Capillary voltage, 3.0 L/min dry gas, and temperature set to 180 °C. The DDA-PASEF data covered 100–1700 *m/z* range with 6 (for 45 min gradient length) or 8 (for 66 min) PASEF ramps. The TIMS settings were 100 ms ramp and accumulation time (100% duty cycle), resulting in 0.9 s (45 min) 1.1 s (66 min) of total cycle time. Active exclusion was enabled with either a 0.2 (45 min) and 0.3 (66 min) min release. The default collision energy with a base of 0.6 1/K0 [V s/cm^2] is set at 20 eV and 1.6 1/K0 [V s/cm^2] at 59 eV was used. Isolation widths were set at 2 *m/z* at < 700 *m/z* and 3 *m/z* at > 800 *m/z*. To achieve more comprehensive coverage, fractions were acquired using DIA-PASEF preformed py5 scheme (Bruker) with 32 × 25 Da windows, covering the *m/z* range of 400-1200 and 1/K0 range of 0.6 to 1.42, resulting in a total cycle time of 1.8 s.

Proteomic data analysis. The Trans-Proteomic Pipeline TPP v6.2.0 Nacreous, build 202302160135-8863 was used for the mass spectrometry data analysis for both identification and quantitation of the proteins. Mass spectrometry raw data (.raw, .d, and .wiff files) was converted into .mzML files using msConvert 3.0.5533[42] and AB_SCIEX_MS_Converter 1.3 Beta from AB SCIEX. The converted files were searched using comet version 2023.01 rev. 0[43] and MSFragger 3.7[44]. In addition, MS/MS Spectra acquired on the Thermo Fisher MS instruments were searched with comet and MSFragger with Monocle[45] modified mzML files. All files were searched against a combined reference database, which comprised the following genome assemblies and proteomes. For isolate B31, the UniProt[46] proteome (ProteomeID UP000001807[9,14]), with 1,291 protein sequences (Table 1). This database was named "core proteome" in the build. Also, the RefSeq.[47] assembly with accession GCF_000008685.2 containing 1,359 protein sequences, and the GenBank[48] assembly GCA_000008685.2 with 1,339 sequences. The total number of non-redundant protein sequences for isolate B31 is 1,485. For isolate B31-5A4, the GenBank assembly GCA_024662195.1 with 1,429 protein sequences, the RefSeq assembly GCF_024662195.1 with 1,354 sequences, and the ISB assembly with 814 sequences (unpublished). The total number of non-redundant protein sequences for isolate B31-5A4 is 1,443. For isolate MM1, the GenBank assembly GCA_003367295.1 with 1,302 protein sequences, and the RefSeq assembly GCF_003367295.1 with 1,159 sequences, and an overall total of 1,383 non-redundant protein sequences (Table 1). The final combined protein database included 116 contaminant sequences from cRAP database (http://www.thegpm.org/crap/), downloaded on July 22nd 2022, containing all 3 isolates with 2,619 unique sequences and an equal number of decoy sequences (generated using the decoy tool in Trans-Proteomic Pipeline with "randomize sequences and interleave entries" decoy algorithm). The following data analysis parameters were used: peptide mass tolerance 20 ppm, fragment ions bins tolerance of 0.02 *m/z* and monoisotopic mass offset of 0.0 *m/z* for Q-Exactive and Orbitrap Fusion Lumos; fragment ions bins tolerance of 1.0005 *m/z* and a monoisotopic mass offset of 0.4 *m/z* for LTQ Orbitrap Elite; peptide mass tolerance 20 ppm, fragment ions bins tolerance of 0.1 *m/z* and monoisotopic mass offset of 0.0 *m/z* for Triple-TOF and timsTOF. Search parameters included semi-tryptic peptides with allowed 2 missed cleavages, static modification carbamidomethylation of cysteine (+57.021464 Da), variable modifications oxidation of methionine and tryptophan (+15.994915 Da), protein N-terminal acetylation (+42.0106), peptide N-terminal Gln to pyro-Glu (−17.0265), peptide N-terminal Glu to pyro-Glu (−18.0106), phosphorylation of Ser, Thr, Tyr, and for negative control, Ala (+79.9663). PeptideProphet was used to assign the scores for peptide spectral matches (PSM) for individual files. iProphet was used to combine the search results from different search engines and assign the score for peptides[31,49,50]. UniProt proteomes are available at https://www.uniprot.org/proteomes/, and NCBI RefSeq and GenBank genome assemblies are available at https://www.ncbi.nlm.nih.gov/assembly/.

PeptideAtlas assembly. The iProphet outputs from Q-Exactive, Orbitrap Fusion Lumos, LTQ Orbitrap Elite, timsTOF, and Triple-TOF runs were further processed using two rounds of reSpect to identify chimeric spectra[51]. For the first round of reSpect, the MINPROB was set to 0 and the MINPROB was set to 0.5 for the second round of reSpect. The new set of.mzML files generated by both rounds of reSpect were searched using MSFragger with the precursor mass tolerance 3.1 and isotope_error off, and processed using the TPP as for the initial files. Using the PeptideAtlas processing pipeline, all the iProphet results from standard and reSpect were filtered at a variable probability threshold to maintain a constant peptide-spectrum match (PSM) FDR of 0.01% for each experiment. The filtered data was assessed with the MAYU software[52] to calculate decoy-based FDRs at the peptide-spectrum match (PSM), distinct peptide, and protein levels. PTMProphet[53], Build 202403260131-9175, was used to access the localization confidence of the sites with post-translational modifications (PTMs), and low resolution ITCID runs DALTONTOL = 0.6 and DENOISE parameters were applied. NIONS was set to b. Bio Tools SeqStats (https://metacpan.org/pod/Bio::Tools::SeqStats) was used to get protein molecular weight, length, pI, and GRAVY scores[54].

Label-free quantitation. StPeter was used for a label-free quantitation of the build data using spectral counting through TPP[55]. The merged protein databases were clustered using OrthoFinder[56]. The representative protein sequence from each protein cluster was extracted. The protein database of the iProphet output from each experiment was refreshed to the representative protein database mapping using the RefreshParser tool in TPP. ProteinProphet and StPeter were run on the updated iProphet file. The StPeter FDR cutoff value 0.01 and minimum probability 0.9 were used. For FTMS HCD/CID and timsTOF runs, a mass tolerance of 0.01 was used. For protein abundance ranking, the percentile of the dSIn index was used.

RNA transcript analysis. To generate RNA for sequencing, *B. burgdorferi* isolates B31, MM1, and B31-5A4 were cultured as previously described, and the cells were collected by centrifugation. Total RNA was extracted using Qiagen RNEasy Mini kits (Qiagen, USA; catalog number 74104) according to the manufacturer's instructions, including an on-column DNAse digestion step. RNA concentration was measured using a NanoDrop spectrophotometer (Thermo Fisher Scientific, USA) and quality assayed by Agilent BioAnalyzer (Agilent, USA). Prior to library construction, 1 µg of total RNA was depleted of ribosomal-RNA transcripts using MICROBExpress Bacterial mRNA Enrichment Kits (Thermo Fisher Scientific, USA; catalog number AM1905). Libraries were prepared using NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs, USA; catalog number E7770) and NEBNext Multiplex Oligos for Illumina (NEB, USA; catalog numbers E7335S, E7500S, E7710S, E7730S). The libraries were prepared according to manufacturer's instructions with insert size approximately 400 bp. Library quality was validated by Agilent Bioanalyzer and yield measured by Qubit HS DNA assay (Thermo Fisher Scientific, USA; catalog number Q32851). Libraries were run on an Illumina NextSeq500 sequencer with High Output Flowcell (Illumina, USA; catalog number 20024907) for 150 cycles. Reads were mapped to the B31 reference genome (GenBank assembly accession GCA_000008685.2) using STAR[57] with quantMode enabled. Mapped reads were visualized with Integrative Genomics Viewer[58] and counts normalized in reads per kilobase of transcript per million reads mapped (RPKM).

Data Records

Mass spectrometry data from 35 different experiments using laboratory isolates B31 and MM1, and infective B31-5A4, with a total of 358 DDA- and 28 DIA-MS runs (Thermo Scientific instrument .raw files, Bruker instruments.d files), were uniformly analyzed through the TPP pipeline (see Methods) and deposited to the ProteomeXchange Consortium via the PRIDE[40] partner repository with the dataset identifier PXD046281[41]. Supplementary Tables 1–6 are also deposited in the same PRIDE dataset. All results collated in the Borrelia PeptideAtlas are made available at http://www.peptideatlas.org/builds/borrelia/, build 2024-03.

Technical Validation

Borrelia PeptideAtlas assembly. The Borrelia PeptideAtlas repository contains information on peptides identified by mass spectrometry-based proteomics of non-infective (B31 & MM1) and infective (B31-5A4) *B. burgdorferi* laboratory isolates. The build comprises extensive proteomics analysis on the total proteome, the secretome and the membrane proteome of the isolates from 35 experiments with a total 386 MS runs (Supplementary Table S1, deposited at PXD046281[41]) generated for this study. To build the Borrelia PeptideAtlas, the dense MS-based proteomic data, which includes 60 million MS/MS spectra, was searched using combined reference databases of B31, MM1, and B31-5A4, and uniformly processed through the TPP pipeline (see Methods). This approach included the use of the post-search engine reSpect to provide peptide identification from chimeric spectra[51] and MAYU[52] to estimate decoy-based FDR levels for the Borrelia build. This strategy provided peptide identities of approximately 13 million PSMs with FDR level threshold less than 0.0001 at the PSM level, and identification of a total of 81,967 distinct peptides at 1.1% peptide FDR (Fig. 2a). These peptides mapped to a total of 1,570 proteins among all isolates with a protein-level FDR less than 1.27%, including 911 core canonical and 324 noncore canonical (not including contaminant proteins). The description of all protein categories and a summary of the proteins identified within each category in the build is shown in Table 2. The complete information on proteins identified in the build is made available in Supplementary Table S2, deposited at PXD046281[41]. Specifically, for the B31 core proteome, 1,113 non-redundant proteins to which at least one peptide was mapped were identified, covering 86% of the B31 core proteome (Table 3). Figure 2c & d shows the frequency distributions of observed and theoretical tryptic peptides by length (amino acid), and distributions of peptide charge and the number of distinct peptides per million observed in each isolate experiment, respectively. The majority of the identified peptides are at charge state $M+2H^{2+}$ or $M+3H^{3+}$ with a length of 7 to 30 amino acids, and most of the identified peptides presented at least one trypsin missed cleavage site. Figure 2e illustrates the frequency (%) of

Database	#entries	#proteins	#obs-proteins	%observed	#unObs-proteins
B31 Core Proteome	1,291	1,291	1,113	86.2	178
B31	3,989	1,485	1,240	83.5	245
MM1	2,461	1,383	1,168	84.5	215
B31-5A4	3,597	1,443	1,230	85.2	213

Table 3. Proteome coverage. Database: name of database, which is collectively from the reference database for this build. #entries: total number of entries. #proteins: total number of non-redundant entries. #obs-proteins: number of non-redundant protein sequences within the subject database to which at least one observed peptide maps. %observed: the percentage of the subject proteome covered by one or more observed peptides. #unObs-proteins: number of non-redundant protein sequences within the subject database to which no observed peptide maps.

	#pSites		#pSites
	$0.95 \le P \le 1.0$	FLR	ExpectFalse
Serine	51	6%	3
Threonine	9	20%	2
Tyrosine	9	20%	2
Alanine	2	n/a	n/a

Table 4. False Localization Rate (FLR) of phosphorylated sites. #pSites: number of phosphorylated STY sites. *P*: PTMProphet probabilities. FLR levels (0-100%) are indicated in the table. ExpectFalse: number of expected false identifications of phosphorylated sites.

the primary sequence coverage for canonical proteins, i.e., the percentage value of amino acids which were identified for each protein, which ranged from 6% to 100%. The identification of specific peptide phosphorylated sites, shown in Fig. 2f, is discussed in the next section.

Post-translational modifications - phosphorylation.

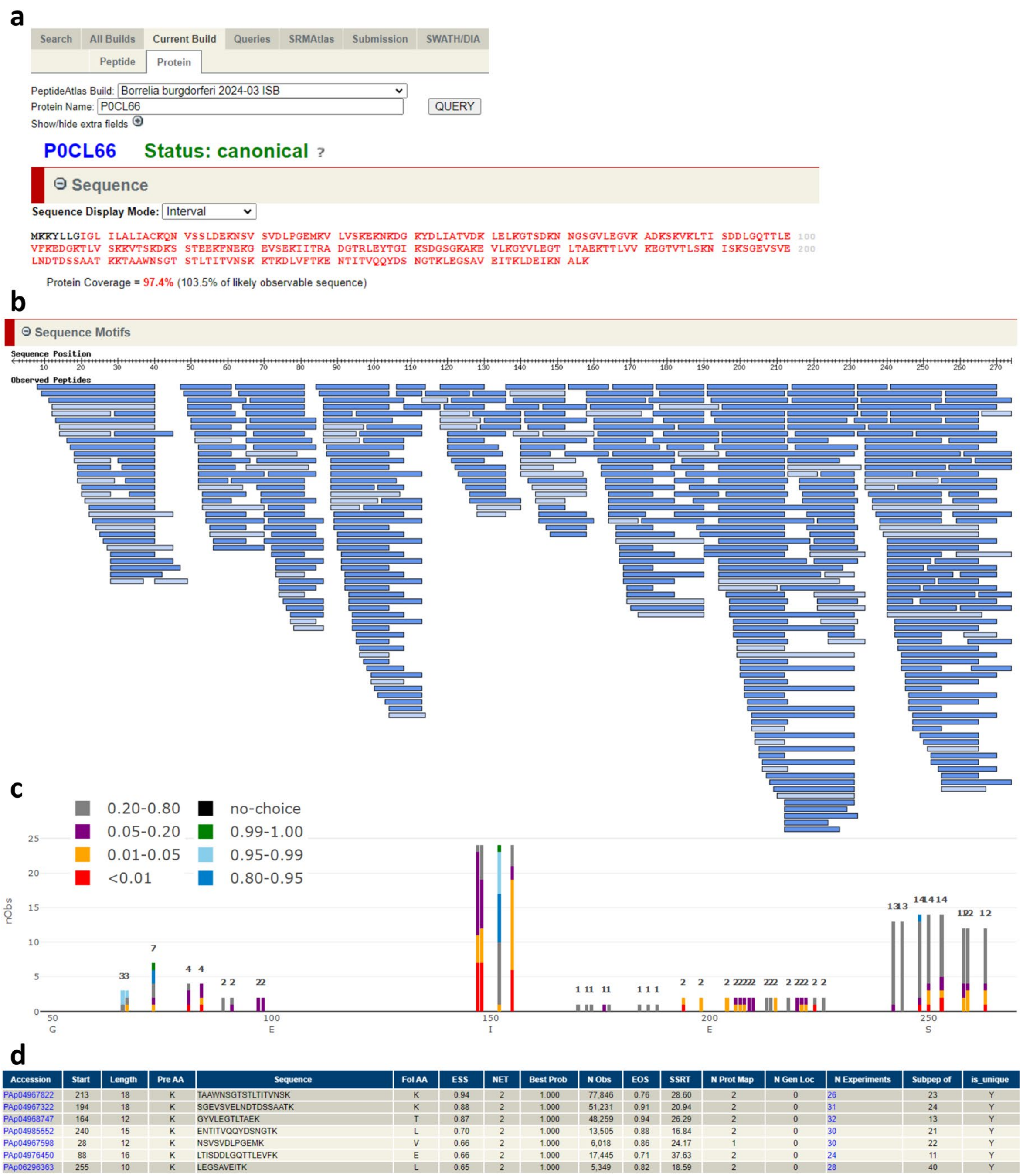
Many reports of large scale phosphorylation identification are invariably false due to poor analysis including no database-level control of false discovery rates[59]. For high-quality protein phosphorylation analysis in the Borrelia PeptideAtlas, each dataset was further analyzed by PTMProphet, embedded in the TPP, to compute localization probabilities (*P*) of phosphorylation sites, including serine (pS), threonine (pT), and tyrosine (pY) residues (Fig. 2f). We used the identification of phosphorylated alanine, which is known to be a false phosphorylated localization, as decoy to access the False Localization Rate (FLR) of phosphorylation[59,60]. PTMProphet applies Bayesian models for each passing PSM that contains a phosphorylation PTM as reported by the search engine[53]. PTMProphet probabilities for ASTY-sites present in the Borrelia build range from 0 to 1 (highest significance), with greater values indicating higher probability that a phosphate group is present at the site, based on MS/MS evidence[53]. The complete information on PTMProphet analysis for all 4 databases (B31 core proteome, B31, MM1, and B31-5A4) is made available in Supplementary Table S3, deposited at PXD046281[41].

Specifically, in the B31 core proteome, the total number of potential phosphorylated sites among the observed proteins is 15,840 for alanine, 25,711 for serine, 14,429 for threonine, and 14,834 for tyrosine. The number of potential phosphorylated sites with peptide coverage among these proteins is 311 (1.96%) for alanine, 450 (1.75%) for serine, 296 (2%) for threonine and 117 (0.79%) for tyrosine. Among these, a total of 2 phospho-alanine sites, 51 phospho-serine sites, 9 phospho-threonine sites, and 9 phospho-tyrosine sites were identified with PTMProphet probability $0.95 \le P \le 1$. For phosphorylated sites detected with $0.95 \le P \le 1.0$, FLR levels are 6% for pS, and 20% for pT and pY, which means that the expected number of false identifications for pS is 3 out of 51, and 2 out of 9 for pT and pY. Table 4 shows the FLRs for all identified phosphorylated sites with *P* probabilities ranging from 0.95 to 1.0. Considering all phosphorylation sites (ASTY) with $P \ge 0.95$ identified in all canonical proteins in the build, including the redundancy of phosphorylated sites, a total 69 phospho-sites were seen throughout 43 *Bb* proteins.

Here, we used outer surface protein A (OspA) as an example of a protein with phosphorylated peptides to illustrate the capabilities of the Borrelia PeptideAtlas interface. OspA is a canonical protein identified with 121 observations at $0.80 \le P \le 0.99$. Figure 3 shows the Borrelia PeptideAtlas interface after searching results for OspA (UniProt identifier P0CL66) in the build protein browser. Figure 3a displays OspA primary sequence coverage of 97.4%, and Fig. 3b illustrates the distribution of all observed distinct peptides for that protein. It is possible to open the peptide browser for each peptide by clicking on the individual blue bar. In the same page, it is possible to visualize phosphorylated ASTY sites distributed in the protein sequence, with the corresponding PTMProphet probabilities (Fig. 3c), and a view table with information on the distinct observed peptides, which contain the phosphorylated sites (Fig. 3d). The Borrelia PeptideAtlas PTM summary can be accessed at http://www.peptideatlas.org/builds/borrelia/, build 2024-03, in the "PTM coverage" section.

Genome coverage of *B. burgdorferi* isolates.

Due to the variability of the plasmid content in different *B. burgdorferi* isolates – which account for approximately one-third of the genome[61], combined reference databases of laboratory isolates B31, MM1, and B31-5A4 were used to search the dense proteomic data when



PeptideAtlas Build: Borrelia burgdorferi 2024-03 ISB
Protein Name: P0CL66

P0CL66 Status: canonical ?

Sequence

Sequence Display Mode: Interval

```
MKKYLLGIGL ILALIACKQN VSSLDEKNSV SVDLPGEMKV LVSKEKNKDG KYDLIATVDK LELKGTSDKN NGSGVLEGVK ADKSKVKLTI SDDLGQTTLE 100
VFKEDGKTLV SKKVTSKDKS STEEKFNEKG EVSEKIITRA DGTRLEYTGI KSDGSGKAKE VLKGYVLEGT LTAEKTTLVV KEGTVTLSKN ISKSGEVSVE 200
LNDTDSSAAT KKTAAWNSGT STLTITVNSK KTKDLVFTKE NTITVQQYDS NGTKLEGSAV EITKLDEIKN ALK
```

Protein Coverage = 97.4% (103.5% of likely observable sequence)

Accession	Start	Length	Pre AA	Sequence	Fol AA	ESS	NET	Best Prob	N Obs	EOS	SSRT	N Prot Map	N Gen Loc	N Experiments	Subpep of	is_unique
PAp04967822	213	18	K	TAAWNSGTSTLTITVNSK	K	0.94	2	1.000	77,846	0.76	28.60	2	0	26	23	Y
PAp04967322	194	18	K	SGEVSVELNDTDSSAATK	K	0.88	2	1.000	51,231	0.91	20.94	2	0	31	24	Y
PAp04968747	164	12	K	GYVLEGTLTAEK	T	0.87	2	1.000	48,259	0.94	26.29	2	0	32	13	Y
PAp04985552	240	15	K	ENTITVQQYDSNGTK	L	0.70	2	1.000	13,505	0.88	16.84	2	0	30	21	Y
PAp04967598	28	12	K	NSVSVDLPGEMK	V	0.66	2	1.000	6,018	0.86	24.17	1	0	30	22	Y
PAp04976450	88	16	K	LTISDDLGQTTLEVFK	E	0.66	2	1.000	17,445	0.71	37.63	2	0	24	11	Y
PAp06296363	255	10	K	LEGSAVEITK	L	0.65	2	1.000	5,349	0.82	18.59	2	0	28	40	Y

Fig. 3 Borrelia PeptideAtlas view of outer OspA phosphorylated sites. OspA UniProt entry P0CL66. Example of the protein PTM summary on the Borrelia PeptideAtlas. (**a**) View of the protein search tab and corresponding primary protein sequence coverage, in red. (**b**) View of the primary protein sequence display with observed peptides. (**c**) Distribution of phosphorylated sites in OspA protein sequence with PTMProphet probabilities, ranging from less than 0.01 to 1. (**d**) Information on observed peptides including empirical suitability score (ESS) empirical observability score (EOS). Accession: peptide accession; start: start position in the protein; pre AA: preceding (towards the N terminus) amino acid; sequence: amino acid sequence of detected peptide, including any mass modifications; fol AA: following (towards the C terminus) amino acid; ESS: empirical suitability score, derived from peptide probability, EOS, and the number of times observed. This is then adjusted sequence characteristics such as missed cleavage [MC] or enzyme termini [ET], or multiple genome locations [MGL]; NET: highest number of enzymatic termini for this protein; NMC: lowest number of missed cleavage for this protein; Best Prob: highest iProphet probability for this observed sequence; Best Adj Prob: highest iProphet-adjusted probability for this observed sequence; N Obs: total number of observations in all modified forms and charge states; EOS: empirical Observability Score, a measure of how many samples a particular peptide is seen in relative to other peptides from the same protein; SSRT: Sequence Specific Retention time provides a hydrophobicity measure for each peptide using the algorithm of Krohkin *et al*. Version 3.0[66]; N Prot Map: number of proteins in the reference database to which this peptide maps; N Gen Loc: number of discrete genome locations which encode this amino acid sequence; Subpep of: number of observed peptides of which this peptide is a subsequence.

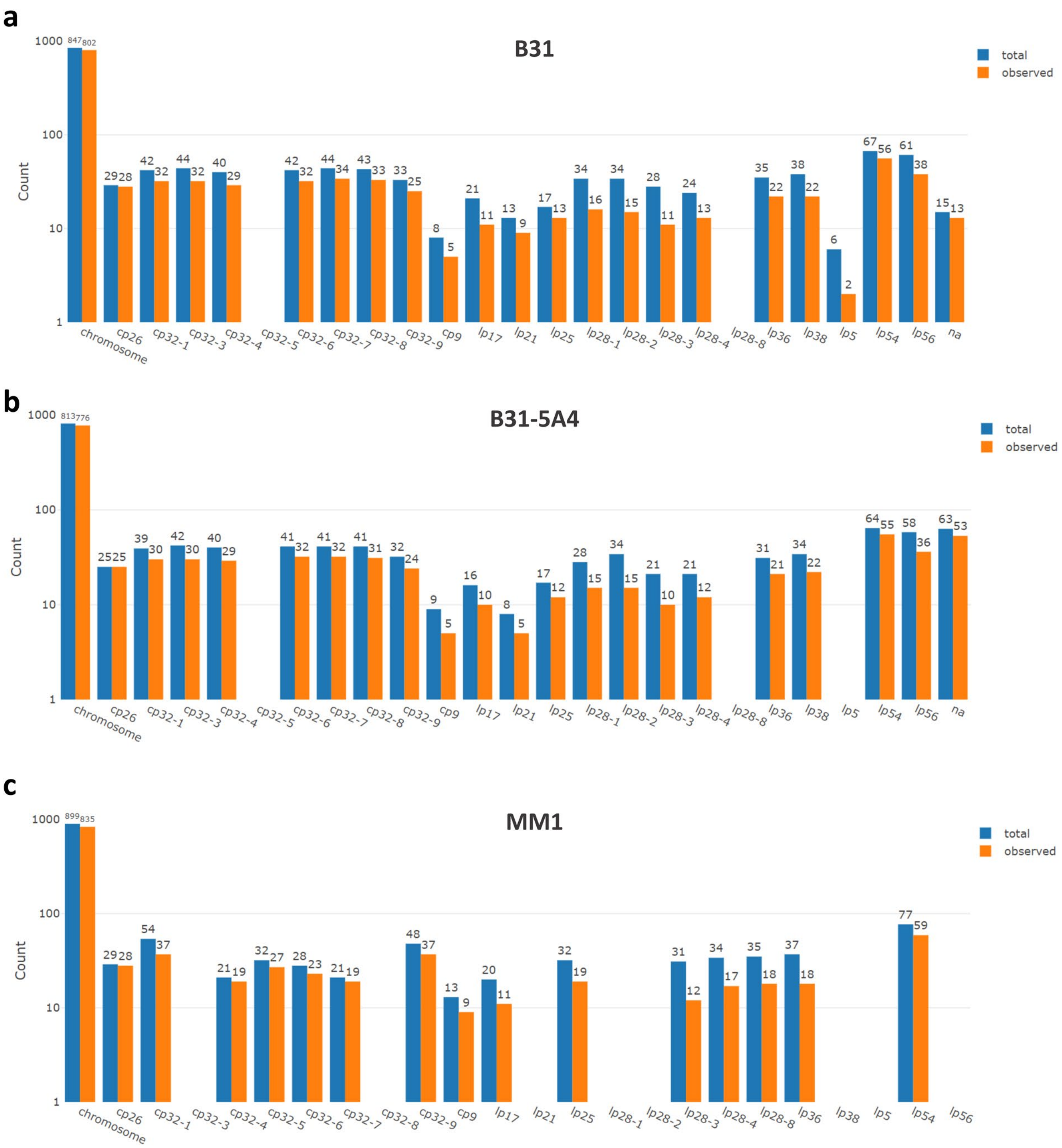


Fig. 4 Genome coverage for isolates. Histograms showing the distribution of chromosomal and plasmid coverage for the reference database of isolates B31, B31-5A4, and MM1. Blue bars indicate total number of genes expected for the chromosome or corresponding plasmid. Orange bars indicate number of genes, which correspond to proteins, observed in the chromosome or corresponding plasmid. na: not assigned.

constructing the build. These databases comprise reference genome assemblies from NCBI RefSeq, GenBank, and UniProt proteome (see Methods). As aforementioned, isolate B31 genome contains a linear chromosome (843 genes) and 21 plasmids (12 linear and 9 circular, 670 genes and 167 pseudogenes total)[14]. Of the 1,513 genes, 1,291 are predicted as unique protein-coding genes. The infective B31-5A4 genome assembly indicates the presence of, besides the linear chromosome, 11 linear plasmids and 9 circular plasmids (ISB, unpublished data). Isolate MM1 has 15 plasmids (7 linear and 8 circular), including the unique lp28-8 and the conserved chromosome[62].

The linear chromosome carries approximately 65% of all genes in *B. burgdorferi*, which encode housekeeping proteins involved in DNA replication, transcription and translation regulation, besides energy metabolism[14]. Here, more than 95% of proteins encoded by the chromosome genome were identified with FDR levels less than 1% throughout all isolates (Fig. 4 and Supplementary Table S2). Circular plasmid cp26 and linear plasmid lp54 are stable and present in all *B. burgdorferi* isolates studied to date[63], including B31, MM1, and B31-5A4, and hence considered a control for encoded proteins identified in the build. Plasmid cp26 encodes proteins which

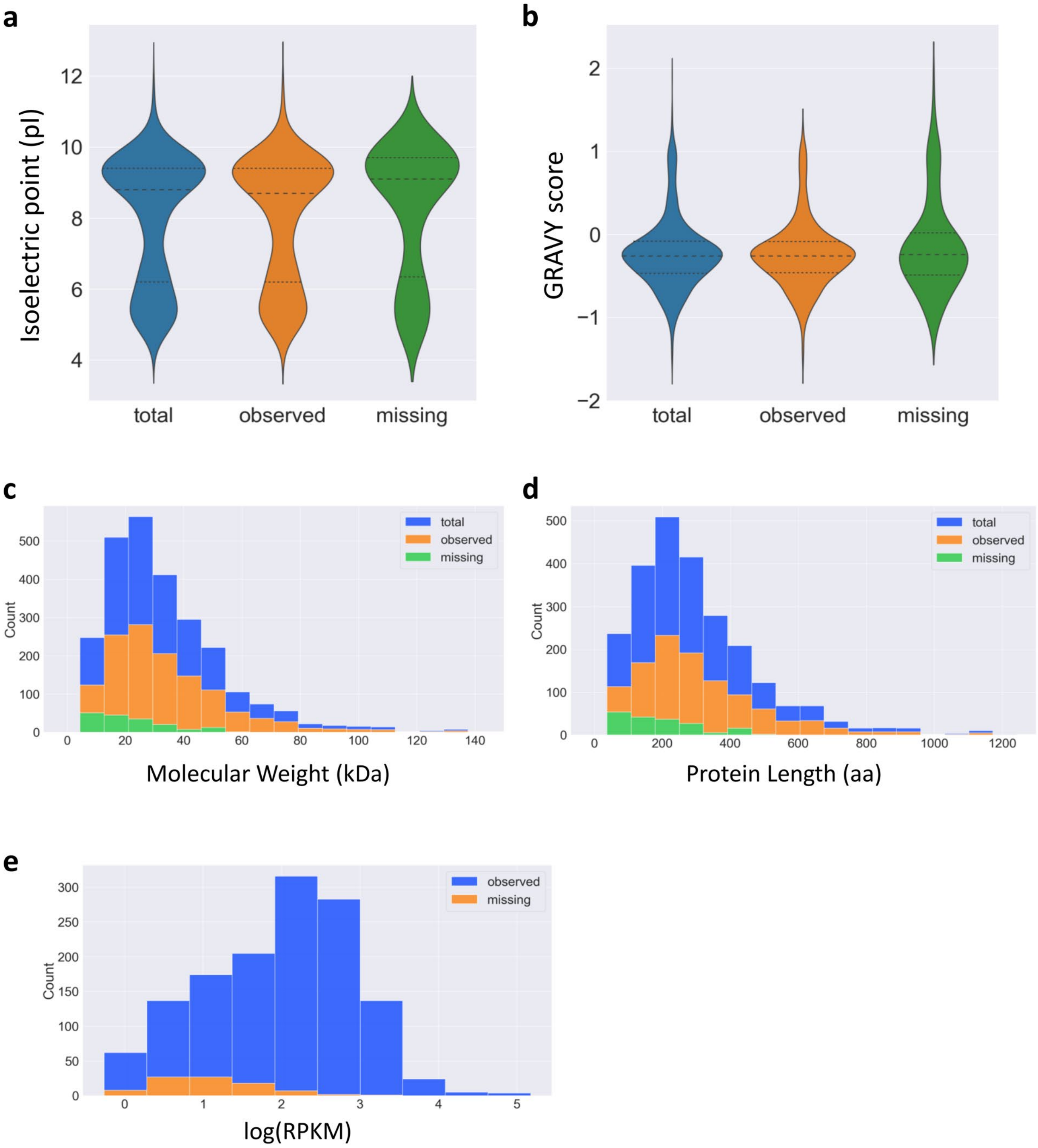
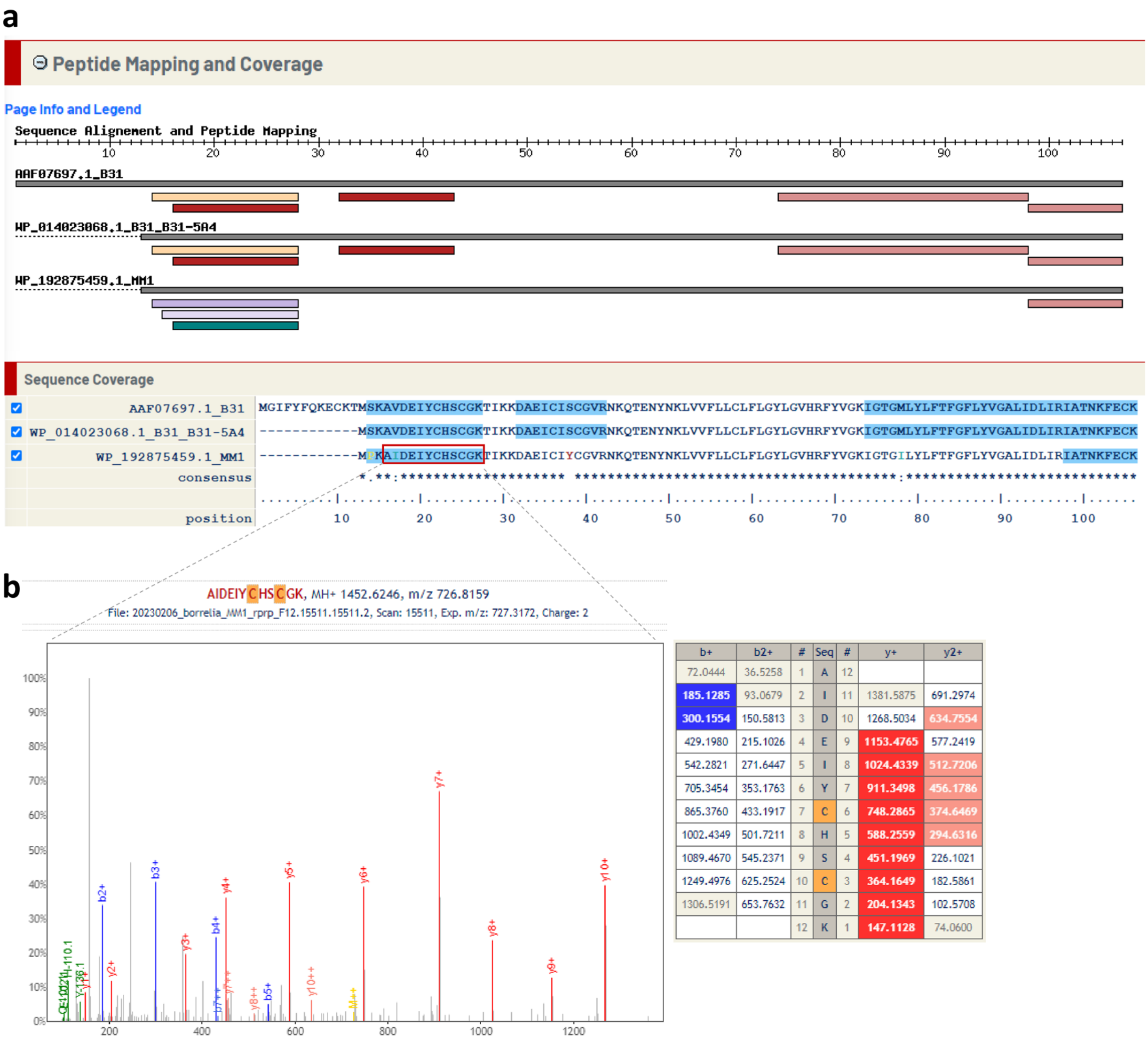


Fig. 5 Protein physicochemical properties and RNA abundance. Total: number of total proteins in the B31 UniProt reference database (core proteome). Observed: number of observed proteins in the B31 core proteome. Missing: number of proteins not observed in the B31 core proteome. (**a,b**) Frequency distributions for protein isoelectric point (pI) and GRAVY score, shown as violin plot. Protein GRAVY index score indicates average hydrophobicity and hydrophilicity. GRAVY score below 0 indicates hydrophilic protein, while scores above 0, hydrophobic[54]. (**c,d**) Frequency distribution for protein molecular weight (kDa) and protein length (number of amino acids), shown as stacked histograms. (**e**) Frequency distribution of mRNA $\log_{10}$ RPKM for observed and not observed (missing) proteins in blue and orange, respectively, shown as a histogram.

are essential for early stages of infection in mammalian hosts, e.g., outer surface protein C (OspC)[64]. Thus, it is considered an essential plasmid for the spirochete growth and survival[48]. Similarly to cp26, the linear plasmid lp54 is present in all *B. burgdorferi* genotypes and encodes critical proteins in tick colonization, e.g. surface proteins OspA and OspB, in tissue attachment and proliferation, such as Decorin-binding proteins A and B, and Crasp1, which plays a critical role in evasion of the host immune system by binding proteins of the complement system[65]. Accordingly, 96% of proteins encoded by cp26 had peptide coverage for B31, 100% for B31-5A4, and 96% for MM1; and around 84% of proteins encoded by lp54 had peptide coverage for the B31, 85% for B31-5A4,



```
☑            AAF07697.1_B31  MGIFYFQKECKTMSKAVDEIYCHSCGKTIKKDAEICISCGVRNKQTENYNKLVVFLLCLFLGYLGVHRFYVGKIGTGMLYLFTFGFLYVGALIDLIRIATNKFECK
☑ WP_014023068.1_B31_B31-5A4  -----------MSKAVDEIYCHSCGKTIKKDAEICISCGVRNKQTENYNKLVVFLLCLFLGYLGVHRFYVGKIGTGMLYLFTFGFLYVGALIDLIRIATNKFECK
☑        WP_192875459.1_MM1  -----------MPKAIDEIYCHSCGKTIKKDAEICIYCGVRNKQTENYNKLVVFLLCLFLGYLGVHRFYVGKIGTGILYLFTFGFLYVGALIDLIRIATNKFECK
                  consensus             *.**:******************* *************************************:***************************
                             .........|.........|.........|.........|.........|.........|.........|.........|.........|.........|......
                   position            10        20        30        40        50        60        70        80        90       100
```

AIDEIYCHSCGK, MH+ 1452.6246, m/z 726.8159

File: 20230206_borrelia_MM1_rprp_F12.15511.15511.2, Scan: 15511, Exp. m/z: 727.3172, Charge: 2

b+	b2+	#	Seq	#	y+	y2+
72.0444	36.5258	1	A	12		
185.1285	93.0679	2	I	11	1381.5875	691.2974
300.1554	150.5813	3	D	10	1268.5034	634.7554
429.1980	215.1026	4	E	9	1153.4765	577.2419
542.2821	271.6447	5	I	8	1024.4339	512.7206
705.3454	353.1763	6	Y	7	911.3498	456.1786
865.3760	433.1917	7	C	6	748.2865	374.6469
1002.4349	501.7211	8	H	5	588.2559	294.6316
1089.4670	545.2371	9	S	4	451.1969	226.1021
1249.4976	625.2524	10	C	3	364.1649	182.5861
1306.5191	653.7632	11	G	2	204.1343	102.5708
		12	K	1	147.1128	74.0600

Fig. 6 TM2 domain family primary protein sequence coverage in B31, B31-5A4, and MM1 databases. UniProt entry Q9S022_BORBU, gene BB_U09. (**a**) In the Peptide Mapping section, peptide highlighted with teal denotes a uniquely mapping and tryptic peptide within this set of sequences. Peptide highlighted with mauve denotes a uniquely mapping and non-tryptic peptide within this set of sequences. Peptide highlighted with red denotes a multi-mapping and tryptic peptide within this set of sequences. Peptide highlighted with orange denotes a multi-mapping and non-tryptic peptide within this set of sequences. In the Sequence Coverage section, all relevant proteins are aligned with MAFFT and all detected peptides are displayed in colors. In the consensus (bottom) row, a * indicates identity across all sequences. Other symbols denote varying degrees of similarity. Sequence highlighted with blue: PEPTIDE denotes peptides observed in specified build. (**b**) Lorikeet MS/MS spectrum view of the peptide AIDEIYCHSCGK, unique to MM1 database.

and 77% for MM1 (Fig. 4). The remaining plasmids display varying frequencies of proteins identified throughout the isolates, ranging from 37% to 85%. The complete information on non-detected proteins by LC-MS ("missing proteins") for each isolate reference database is made available in Supplementary Table S4 (deposited at PXD046281[41]), which includes the plasmid information. We note that 80% of missing proteins are described as hypothetical proteins or of unknown function in UniProt B31 core proteome, 5% are membrane proteins, and the remaining 15% have variable descriptions, including flagellar and transporter proteins.

Physicochemical protein properties. The physicochemical characteristics of expected (total proteins), observed, and missing proteins in the B31 core proteome are shown in Fig. 5. This information is included in the Borrelia PeptideAtlas. The features comprise protein isoelectric point (pI), GRAVY index score, molecular weight (kDa), and length (number of amino acids). The frequency distributions of these features indicate that missing proteins have similar characteristics as those of the observed proteins, with relatively higher frequencies of basic (pI > 10), hydrophobic (GRAVY score > 0) and small proteins (less than 20 kDa) (Fig. 5a–d). To further investigate the mRNA levels of the non-detected proteins, transcriptomic analysis of isolates B31, MM1 and

B31-5A4 was performed (Supplementary Table S5, deposited at PXD046281[41]). Transcripts were not detected by RNAseq for approximately 50% of the missing proteins. The other 50% have RPKM ranging from 1 to 1,379. A considerable number of canonical proteins detected for the B31 core proteome (around 42%) had low levels of mRNA RPKM, i.e., lower than 100 counts, and the remaining transcripts showed a range of 101-149,599 RPKM. Therefore, proteins not detected for the B31 core proteome show absence or relatively lower abundance of their corresponding transcripts. The frequency distribution of $\log_{10}$ RPKM for transcripts of observed and missing proteins is shown in Fig. 5e.

Identification of unique variant peptides across different *B. burgdorferi* isolates.

The Borrelia PeptideAtlas is an MS-based peptide repository that enables the assessment and visualization of peptides and the corresponding protein sequence coverage in different isolates using chimeric databases. The TPP-oriented analysis of the complex MS data enabled the identification, with FDR levels less than 1%, of 109 uniquely mapping peptides to 9 proteins in B31, 5,779 unique peptides in MM1, which mapped to 538 proteins, and 44 peptides unique to B31-5A4, mapping to 13 proteins (complete information in Supplementary Table S6, deposited at PXD042072). Figure 6 illustrates the capabilities of the PeptideAtlas in the comparison and visualization of protein sequences with different annotations in the isolate databases. The example used here shows a variant peptide that is uniquely detected in the isolate MM1, and which maps to the TM2 domain family protein (BB_U09). This peptide has a predicted isoleucine instead of a valine, as detected in B31 and B31-5A4. The observance of this peptide is evidence of the diverging primary protein sequence of BB_U09 in MM1.

Usage Notes

The Borrelia PeptideAtlas provides a publicly accessible resource that is important for the Lyme disease research community. Our goal is to provide an expandable data source with many other *B. burgdorferi* isolates to be added, including clinically relevant isolates, and subjected to different growth conditions, enabling the investigation of the dynamic proteome of this spirochete through its infection stages and their vastly different growth environments. The diverse proteomic information from multiple infective isolates with credible data presented by the Borrelia PeptideAtlas can be useful to pinpoint potential protein targets which are common to infective isolates and may be key in the infection process – such as outer membrane proteins. A list of membrane protein targets present in the build can be identified. With *in silico* prediction of signal peptides and secondary structures of membrane proteins, this dense proteomic data can be further investigated for host-pathogen protein interactomics with different technologies. Moreover, this resource provides access to information regarding a wide range of potential proteins and PTMs relevant to develop sensitive diagnostic assays in the Lyme disease research community. The Borrelia PeptideAtlas is a dynamic proteome resource in terms of size and complexity and will be updated to include new data periodically, as more genomic and proteomic data is made available for new clinical and laboratory isolates. The collection of the raw data, protein, and peptide information are publicly available in the Borrelia PeptideAtlas at http://www.peptideatlas.org/builds/borrelia/.

Data availability

All MS data and Supplementary Tables for the Borrelia PeptideAtlas is available and deposited at the EBI PRIDE repository at https://www.ebi.ac.uk/pride/archive/projects/PXD046281 and https://doi.org/10.6019/PXD046281.

Code availability

The authors do not have code specific to this work to disclose.

Received: 21 September 2023; Accepted: 28 October 2024;
Published online: 02 December 2024

References

1. Schwartz, A. M., Kugeler, K. J., Nelson, C. A., Marx, G. E. & Hinckley, A. F. Use of Commercial Claims Data for Evaluating Trends in Lyme Disease Diagnoses, United States, 2010-2018. *Emerging infectious diseases* **27**, 499–507, https://doi.org/10.3201/eid2702.202728 (2021).
2. Kugeler, K. J., Schwartz, A. M., Delorey, M. J., Mead, P. S. & Hinckley, A. F. Estimating the Frequency of Lyme Disease Diagnoses, United States, 2010-2018. *Emerging infectious diseases* **27**, 616–619, https://doi.org/10.3201/eid2702.202731 (2021).
3. Steere, A. C. *et al*. Erythema chronicum migrans and Lyme arthritis. The enlarging clinical spectrum. *Annals of internal medicine* **86**, 685–698 (1977).
4. Steere, A. C. *et al*. The spirochetal etiology of Lyme disease. *N Engl J Med* **308**, 733–740, https://doi.org/10.1056/NEJM198303313081301 (1983).
5. Schoen, R. T. Challenges in the Diagnosis and Treatment of Lyme Disease. *Curr Rheumatol Rep* **22**, 3, https://doi.org/10.1007/s11926-019-0857-2 (2020).
6. Maksimyan, S., Syed, M. S. & Soti, V. Post-Treatment Lyme Disease Syndrome: Need for Diagnosis and Treatment. *Cureus* **13**, e18703, https://doi.org/10.7759/cureus.18703 (2021).
7. Branda, J. A. *et al*. Advances in Serodiagnostic Testing for Lyme Disease Are at Hand. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America* **66**, 1133–1139, https://doi.org/10.1093/cid/cix943 (2018).
8. Tilly, K., Rosa, P. A. & Stewart, P. E. Biology of infection with Borrelia burgdorferi. *Infectious disease clinics of North America* **22**, 217–234, https://doi.org/10.1016/j.idc.2007.12.013 (2008). v.
9. Fraser, C. M. *et al*. Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi. *Nature* **390**, 580–586, https://doi.org/10.1038/37551 (1997).
10. DeHart, T. G., Kushelman, M. R., Hildreth, S. B., Helm, R. F. & Jutras, B. L. The unusual cell wall of the Lyme disease spirochaete Borrelia burgdorferi is shaped by a tick sugar. *Nat Microbiol* **6**, 1583–1592, https://doi.org/10.1038/s41564-021-01003-w (2021).
11. Takayama, K., Rothenberg, R. J. & Barbour, A. G. Absence of lipopolysaccharide in the Lyme disease spirochete, Borrelia burgdorferi. *Infect Immun* **55**, 2311–2313, https://doi.org/10.1128/iai.55.9.2311-2313.1987 (1987).

12. Bernard, Q. *et al*. Borrelia burgdorferi protein interactions critical for microbial persistence in mammals. *Cell Microbiol* **21**, e12885, https://doi.org/10.1111/cmi.12885 (2019).
13. Steere, A. C. Lyme disease. *N Engl J Med* **345**, 115–125, https://doi.org/10.1056/NEJM200107123450207 (2001).
14. Casjens, S. *et al*. A bacterial genome in flux: the twelve linear and nine circular extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete Borrelia burgdorferi. *Mol Microbiol* **35**, 490–516, https://doi.org/10.1046/j.1365-2958.2000.01698.x (2000).
15. Strle, K., Jones, K. L., Drouin, E. E., Li, X. & Steere, A. C. Borrelia burgdorferi RST1 (OspC type A) genotype is associated with greater inflammation and more severe Lyme disease. *Am J Pathol* **178**, 2726–2739, https://doi.org/10.1016/j.ajpath.2011.02.018 (2011).
16. Lemieux, J. E. *et al*. Whole genome sequencing of human Borrelia burgdorferi isolates reveals linked blocks of accessory genome elements located on plasmids and associated with human dissemination. *PLoS Pathog* **19**, e1011243, https://doi.org/10.1371/journal.ppat.1011243 (2023).
17. Angel, T. E. *et al*. Proteome analysis of Borrelia burgdorferi response to environmental change. *PLoS One* **5**, e13800, https://doi.org/10.1371/journal.pone.0013800 (2010).
18. Bontemps-Gallo, S. *et al*. Global Profiling of Lysine Acetylation in Borrelia burgdorferi B31 Reveals Its Role in Central Metabolism. *Front Microbiol* **9**, 2036, https://doi.org/10.3389/fmicb.2018.02036 (2018).
19. Dowdell, A. S. *et al*. Comprehensive Spatial Analysis of the Borrelia burgdorferi Lipoproteome Reveals a Compartmentalization Bias toward the Bacterial Surface. *J Bacteriol* **199** https://doi.org/10.1128/JB.00658-16 (2017).
20. Jacobs, J. M. *et al*. Proteomic analysis of Lyme disease: global protein comparison of three strains of Borrelia burgdorferi. *Proteomics* **5**, 1446–1453, https://doi.org/10.1002/pmic.200401052 (2005).
21. Schnell, G. *et al*. Proteomic analysis of three Borrelia burgdorferi sensu lato native species and disseminating clones: relevance for Lyme vaccine design. *Proteomics* **15**, 1280–1290, https://doi.org/10.1002/pmic.201400177 (2015).
22. Toledo, A., Huang, Z., Coleman, J. L., London, E. & Benach, J. L. Lipid rafts can form in the inner and outer membranes of Borrelia burgdorferi and have different properties and associated proteins. *Mol Microbiol* **108**, 63–76, https://doi.org/10.1111/mmi.13914 (2018).
23. Toledo, A., Perez, A., Coleman, J. L. & Benach, J. L. The lipid raft proteome of Borrelia burgdorferi. *Proteomics* **15**, 3662–3675, https://doi.org/10.1002/pmic.201500093 (2015).
24. Payne, S. H. *et al*. The Pacific Northwest National Laboratory library of bacterial and archaeal proteomic biodiversity. *Sci Data* **2**, 150041, https://doi.org/10.1038/sdata.2015.41 (2015).
25. Baranton, G. *et al*. Delineation of Borrelia burgdorferi sensu stricto, Borrelia garinii sp. nov., and group VS461 associated with Lyme borreliosis. *Int J Syst Bacteriol* **42**, 378–383, https://doi.org/10.1099/00207713-42-3-378 (1992).
26. Hughes, C. A. & Johnson, R. C. Methylated DNA in Borrelia species. *J Bacteriol* **172**, 6602–6604, https://doi.org/10.1128/jb.172.11.6602-6604.1990 (1990).
27. Xu, Y. & Johnson, R. C. Analysis and comparison of plasmid profiles of Borrelia burgdorferi sensu lato strains. *J Clin Microbiol* **33**, 2679–2685, https://doi.org/10.1128/jcm.33.10.2679-2685.1995 (1995).
28. Kawabata, H., Norris, S. J. & Watanabe, H. BBE02 disruption mutants of Borrelia burgdorferi B31 have a highly transformable, infectious phenotype. *Infect Immun* **72**, 7147–7154, https://doi.org/10.1128/IAI.72.12.7147-7154.2004 (2004).
29. Wilkinson, M. D. *et al*. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018, https://doi.org/10.1038/sdata.2016.18 (2016).
30. Desiere, F. *et al*. The PeptideAtlas project. *Nucleic Acids Res* **34**, D655–658, https://doi.org/10.1093/nar/gkj040 (2006).
31. Deutsch, E. W. *et al*. Trans-Proteomic Pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics. *Proteomics. Clinical applications* **9**, 745–754, https://doi.org/10.1002/prca.201400164 (2015).
32. Bundgaard, L. *et al*. The Equine PeptideAtlas: a resource for developing proteomics-based veterinary research. *Proteomics* **14**, 763–773, https://doi.org/10.1002/pmic.201300398 (2014).
33. Deutsch, E. W., Lam, H. & Aebersold, R. PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. *EMBO reports* **9**, 429–434, https://doi.org/10.1038/embor.2008.56 (2008).
34. Hesselager, M. O. *et al*. The Pig PeptideAtlas: A resource for systems biology in animal production and biomedicine. *Proteomics* **16**, 634–644, https://doi.org/10.1002/pmic.201500195 (2016).
35. McCord, J., Sun, Z., Deutsch, E. W., Moritz, R. L. & Muddiman, D. C. The PeptideAtlas of the Domestic Laying Hen. *J Proteome Res* **16**, 1352–1363, https://doi.org/10.1021/acs.jproteome.6b00952 (2017).
36. Vialas, V. *et al*. A Candida albicans PeptideAtlas. *Journal of proteomics* **97**, 62–68, https://doi.org/10.1016/j.jprot.2013.06.020 (2014).
37. Loevenich, S. N. *et al*. The Drosophila melanogaster PeptideAtlas facilitates the use of peptide data for improved fly proteomics and genome annotation. *BMC Bioinformatics* **10**, 59, https://doi.org/1471-2105-10-5910.1186/1471-2105-10-59 (2009).
38. King, N. L. *et al*. Analysis of the Saccharomyces cerevisiae proteome with PeptideAtlas. *Genome Biology* **7**, 15 https://doi.org/R10610.1186/gb-2006-7-11-r106 (2006).
39. Morrone, S. R., Hoopmann, M. R., Shteynberg, D. D., Kusebauch, U. & Moritz, R. L. [Preprint: Not Peer Reviewed] Optimization of Instrument Parameters for Efficient Phosphopeptide Identification and Localization by Data-dependent Analysis Using Orbitrap Tribrid Mass Spectrometers. *ChemRxiv* https://doi.org/10.26434/chemrxiv-2023-qklh1-v2 (2023).
40. Perez-Riverol, Y. *et al*. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res* **50**, D543–D552, https://doi.org/10.1093/nar/gkab1038 (2022).
41. Reddy, P. J. *et al*. *Borrelia PeptideAtlas: A proteome resource of common Borrelia burgdorferi isolates for Lyme community*, <EBI PRIDE repository at https://www.ebi.ac.uk/pride/archive/projects/PXD046281 and https://doi.org/10.6019/PXD046281> (2023).
42. Kessner, D., Chambers, M., Burke, R., Agus, D. & Mallick, P. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* **24**, 2534–2536, https://doi.org/10.1093/bioinformatics/btn323 (2008).
43. Eng, J. K., Jahan, T. A. & Hoopmann, M. R. Comet: an open-source MS/MS sequence database search tool. *Proteomics* **13**, 22–24, https://doi.org/10.1002/pmic.201200439 (2013).
44. Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D. & Nesvizhskii, A. I. *Nat Methods*. **14**(5), 513–520, https://doi.org/10.1038/nmeth.4256 (2017)
45. Rad, R. *et al*. Improved Monoisotopic Mass Estimation for Deeper Proteome Coverage. *J Proteome Res* **20**, 591–598, https://doi.org/10.1021/acs.jproteome.0c00563 (2021).
46. UniProt, C. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res* **51**, D523–D531, https://doi.org/10.1093/nar/gkac1052 (2023).
47. O'Leary, N. A. *et al*. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* **44**, D733–745, https://doi.org/10.1093/nar/gkv1189 (2016).
48. Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Sayers, E. W. GenBank. *Nucleic Acids Res* **44**, D67–72, https://doi.org/10.1093/nar/gkv1276 (2016).
49. Keller, A. & Shteynberg, D. Software pipeline and data analysis for MS/MS proteomics: the trans-proteomic pipeline. *Methods Mol Biol* **694**, 169–189, https://doi.org/10.1007/978-1-60761-977-2_12 (2011).
50. Shteynberg, D. *et al*. iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. *Mol Cell Proteomics* **10**, M111 007690, https://doi.org/10.1074/mcp.M111.007690 (2011).

51. Shteynberg, D. *et al*. reSpect: software for identification of high and low abundance ion species in chimeric tandem mass spectra. *J Am Soc Mass Spectrom* **26**, 1837–1847, https://doi.org/10.1007/s13361-015-1252-5 (2015).
52. Reiter, L. *et al*. Protein identification false discovery rates for very large proteomics data sets generated by tandem mass spectrometry. *Mol Cell Proteomics* **8**, 2405–2417, https://doi.org/M900317-MCP20010.1074/mcp.M900317-MCP200 (2009).
53. Shteynberg, D. D. *et al*. PTMProphet: Fast and Accurate Mass Modification Localization for the Trans-Proteomic Pipeline. *J Proteome Res* **18**, 4262–4272, https://doi.org/10.1021/acs.jproteome.9b00205 (2019).
54. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**, 105–132, https://doi.org/10.1016/0022-2836(82)90515-0 (1982).
55. Hoopmann, M. R., Winget, J. M., Mendoza, L. & Moritz, R. L. StPeter: Seamless Label-Free Quantification with the Trans-Proteomic Pipeline. *J Proteome Res* **17**, 1314–1320, https://doi.org/10.1021/acs.jproteome.7b00786 (2018).
56. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**, 238, https://doi.org/10.1186/s13059-019-1832-y (2019).
57. Dobin, A. *et al*. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21, https://doi.org/10.1093/bioinformatics/bts635 (2013).
58. Robinson, J. T. *et al*. Integrative genomics viewer. *Nat Biotechnol* **29**, 24–26, https://doi.org/10.1038/nbt.1754 (2011).
59. Ramsbottom, K. A. *et al*. Method for Independent Estimation of the False Localization Rate for Phosphoproteomics. *J Proteome Res* **21**, 1603–1615, https://doi.org/10.1021/acs.jproteome.1c00827 (2022).
60. Kalyuzhnyy, A. *et al*. Profiling the Human Phosphoproteome to Estimate the True Extent of Protein Phosphorylation. *J Proteome Res* **21**, 1510–1524, https://doi.org/10.1021/acs.jproteome.2c00131 (2022).
61. Casjens, S. R. *et al*. Primordial origin and diversification of plasmids in Lyme disease agent bacteria. *BMC genomics* **19**, 218, https://doi.org/10.1186/s12864-018-4597-x (2018).
62. Jabbari, N. *et al*. Whole genome sequence and comparative analysis of Borrelia burgdorferi MM1. *PLoS One* **13**, e0198135, https://doi.org/10.1371/journal.pone.0198135 (2018).
63. Casjens, S. R. *et al*. Plasmid diversity and phylogenetic consistency in the Lyme disease agent Borrelia burgdorferi. *BMC genomics* **18**, 165, https://doi.org/10.1186/s12864-017-3553-5 (2017).
64. Grimm, D. *et al*. Outer-surface protein C of the Lyme disease spirochete: a protein induced in ticks for infection of mammals. *Proc Natl Acad Sci USA* **101**, 3142–3147, https://doi.org/10.1073/pnas.0306845101 (2004).
65. Bestor, A. *et al*. Use of the Cre-lox recombination system to investigate the lp54 gene requirement in the infectious cycle of Borrelia burgdorferi. *Infect Immun* **78**, 2397–2407, https://doi.org/10.1128/IAI.01059-09 (2010).
66. Krokhin, O. V. Sequence-specific retention calculator. Algorithm for peptide retention prediction in ion-pair RP-HPLC: application to 300- and 100-A pore size C18 sorbents. *Anal Chem* **78**, 7785–7795, https://doi.org/10.1021/ac060777w (2006).

Acknowledgements

This work was funded in part by the National Institutes of Health grants from the National Institutes of Health, National Institute of Allergy and Infectious Diseases (NIAID) R21AI142302 (RLM), R21AI133335 (RLM), and R01AI029735 (MJC), National Institute for General Medical Sciences (NIGMS) R01GM087221 (RLM), the Office of the Director S10OD026936 (RLM), and the National Science Foundation award 1920268 (RLM). We thank the Wilke Family Foundation, and the Steven and Alexandra Cohen Foundation for their generous support (RLM).

Author contributions

Panga J. Reddy: performed *B. burgdorferi* culturing, mass spectrometry data generation experiments and data analysis, PeptideAtlas generation, manuscript generation. Zhi Sun: performed mass spectrometry data analysis, PeptideAtlas generation, and edited the manuscript. Helisa H. Wippel: performed *B. burgdorferi* culturing, mass spectrometry data generation, experiments, and data analysis, PeptideAtlas generation, manuscript writing and preparation, and finalization of the manuscript. David H. Baxter: performed *B. burgdorferi* genome and RNA sequencing data generation and sequence data analysis. Kristian E. Swearingen: performed sample preparation experiments. David D. Shteynberg: performed mass spectrometry data analysis and code generation for the Trans-Proteomic Pipeline. Mukul Midha: performed mass spectrometry experiments. Melissa J. Caimano: Provided B31-5A4, biological, and technical expertise for preparing *B. burgdorferi* and edited the manuscript. Klemen Strle: provided Clinical, biological, and technical expertise for preparing *B. burgdorferi* and edited the manuscript. Yongwook Choi: performed *B. burgdorferi* genome sequence data analysis. Agnes P. Chan: performed *B. burgdorferi* genome sequence data analysis. Nicholas J. Schork: performed *B. burgdorferi* genome sequence data analysis and edited the manuscript. Andrea S. Varela-Stokes: Provided biological, and technical expertise for preparing *B. burgdorferi* and edited the manuscript. Robert L. Moritz: conceived the project, secured funding, designed experiments, provided technical expertise, managed the project, and performed manuscript writing, preparation, and finalization. † In memory of Dr. Helisa Helena Wippel.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-024-04047-9.

Correspondence and requests for materials should be addressed to R.L.M.