Burkholderia PglL enzymes are serine preferring oligosaccharidetransferases which target conserved proteins across the Burkholderia genus

Andrew J. Hayes¹, Jessica M. Lewis¹, Mark R. Davies¹ and Nichollas E. Scott^{1#}

¹Department of Microbiology and Immunology, University of Melbourne at the Peter Doherty Institute for Infection and Immunity, Melbourne 3000, Australia

#To whom correspondence and requests for materials should be addressed N.E.S
(Nichollas.scott@unimelb.edu.au)

Key words: Glycosylation, *Burkholderia cenocepacia*, Burkholderia, Post-translational modifications, Proteomics, Glycosyltransferases, PglL

Abstract

Glycosylation is increasingly recognised as a common protein modification within bacterial proteomes. While great strides have been made in identifying species that contain glycosylation systems, our understanding of the proteins and sites targeted by these enzymes is far more limited. Within this work we explore the conservation of glycoproteins and *O*-linked glycosylation sites across the pan-*Burkholderia* glycoproteome. Using a multi-protease glycoproteomic approach we generate high-confidence glycoproteomes and associated glycosylation sites in two widely utilized *B. cenocepacia* strains, K56-2 and H111. This resource reveals glycosylation occurs exclusively at serine residues and that glycoproteins/glycosylation sites are highly conserved across 294 publicly available *B. cenocepacia* genomes. Consistent with this we demonstrate that the substitution of Serine for Threonine residues in a model protein results in a dramatic decrease in glycosylation efficiency by the oligosaccharidetransferase pgl_{BC} even when pgl_{BC} is overexpressed. This preference for glycosylation at Serine residues is observed across at least 9 *Burkholderia* glycoproteomes supporting that Serine is the dominant residue targeted by pgl-mediated glycosylation across the *Burkholderia* genus. Using population genomics we observe that pgl targeted glycosylated proteins are common across *Burkholderia* species. Combined, this work demonstrates that Pgl enzymes of the *Burkholderia* genus are Serine-preferring oligosaccharidetransferases that target conserved and shared protein substrates across the *Burkholderia* genus.

Introduction

Glycosylation is a common class of protein modifications increasingly recognised within bacterial proteomes^{1,2}. Within bacterial proteomes, the installation of glycans has been shown to influence protein functions by enhancing protein stability³⁻⁵, stabilizing protein complexes⁶, masking antigenic sites⁷ and modulating protein enzymatic activities^{8,9}. As these events fine-tune the proteome, identifying novel glycosylation systems has been a key goal of microbial glycoproteomics. Over the last decade, with the aid of mass spectrometry-based proteomics, a range of bacterial glycosylation systems have been identified^{1,2,10,11}. Although these studies have demonstrated the commonality of bacterial glycosylation, the majority of follow up studies have focused on understanding the biology of protein glycosylation glycans, including elucidating their biosynthetic pathways¹²⁻¹⁴ and defining their diversity¹⁵⁻¹⁷. This focus has resulted in a limited understanding of the specific glycoproteins and glycosylation sites within the majority of known bacterial glycoproteomes.

To date arguably the best characterised bacterial glycoproteome is that of the *N*-linked glycosylation system of *Campylobacter jejuni*, where 134 glycosylation sites have been experimentally identified¹⁸. The success in mapping glycosylation across the *C. jejuni* glycoproteome has largely been due to two features of this system; i) glycosylation events within the *Campylobacter* genus occurring within a unique glycosylation sequon, D/E-X-N-X-S/T¹⁹, which restricts the possible glycosylation sites within a protein and ii) the chemical nature of *N*-linked glycosylation which can allow localisation information to be obtained using collision-based Mass Spectrometry (MS) fragmentation approaches²⁰. For bacterial mediated *O*-linked glycosylation systems, far less is known about the sites of glycosylation^{1,2}. This limited understanding of bacterial *O*-linked glycosylation sites is largely driven by technical limitations associated with the analysis of *O*-linked glycopeptides, where the chemical nature of *O*-linked glycosylation requires approaches such as Electron-transfer dissociation (ETD) or Electron-Transfer/Higher-Energy Collision Dissociation (ETHcD) to localise glycosylation sites²⁰. Defining the glycosylation sites are critical for enabling functional characterisation of glycoproteins as well as for understanding the conservation of sites across species. From previous studies it has

become clear that bacterial glycosylation sites can vary even within closely related strains/species^{21,22} and that understanding glycosylation site heterogeneity can provide critical insights into functionally important nuances within glycoproteins. An example of this can be seen in the pilin of *Neisseria* species where high levels of glycosylation are correlated with masking conserved antigenic regions while forms of pilin prone to amino acid sequence variation possess limited pilin glycosylation^{7,14,23,24}.

One of the most widespread families of *O*-linked glycosylating enzymes are the PglL oligosaccharidetransferases²⁵ which have been experimentally shown to *O*-glycosylate proteins within *Acinetobacter*^{26,27}, *Neisseria*^{14,17,28}, *Burkholderia*^{12,16,29}, *Francisella*³⁰, *Pseudomonas*³¹, and *Ralstonia*³² species. Within these species, PglL mediates *O*-linked glycosylation of substrates within the periplasmic space, and can be responsible for the modification of a single protein or multiple proteins depending on the enzyme^{10,11}. To date, no rigid glycosylation sequon has been observed within PglL substrates, with glycosylation predominately occurring in disordered regions rich in Alanine and Prolines^{15,29,33,34}. For PglL enzymes responsible for the glycosylation of multiple proteins, also known as general glycosylation systems, the abolishment of PglL has been shown to result in profound effects such as the attenuation of virulence^{27,29,32}. Within these systems the lack of a detailed understanding of the glycoproteomes limits our ability to understand how glycosylation influences virulence and the glycosylation sites which are critical for virulence associated processes.

Recently we began exploring the pan-glycoproteome of *Burkholderia* species focusing on the conservation of the glycans used for glycosylation^{12,16} and the properties of glycopeptides which can influence their enrichment using zwitterionic hydrophilic interaction liquid chromatography (ZIC-HILIC)³⁵. These studies have highlighted that the glycoproteomes of *Burkholderia* species appear larger than initially thought, yet, due to the glycosylation sites associated with these proteins not being defined, this has limited our ability to understand the general trends within site utilisation across the *Burkholderia* pan-glycoproteome. To improve our understanding of *Burkholderia* glycosylation we have undertaken a site focused glycoproteomic study of

103 *Burkholderia cenocepacia* within two widely utilised strains, K56-2³⁶ and H111³⁷. Leveraging this
 104 curated resource, we gain and experimentally confirm a previously unrecognised preference in
 105 the specificity of *O*-linked glycosylation as well as demonstrate that *Burkholderia* glycosylation
 106 targets conserved protein substrates across this genus. Combined this work further
 107 demonstrates a previously underappreciated aspect of *O*-linked glycosylation conservation
 108 within *Burkholderia* species, the conservation of proteins substrates.
 109

Methods

Bacterial strains and growth conditions: Strains and plasmids used in this study are listed in Tables 1 and 2, respectively. Strains of *Escherichia coli* and *B. cenocepacia* were grown at 37°C on Lysogeny Broth (LB) medium. When required, antibiotics were added to a final concentration of 50 µg/ml trimethoprim for *E. coli* and 100 µg/ml for *B. cenocepacia*, 20 µg/ml tetracycline for *E. coli* and 150 µg/ml for *B. cenocepacia* and 40 µg/ml kanamycin for *E. coli*. Ampicillin was used at 100 µg/ml and polymyxin B at 25 µg/ml for triparental mating to select against donor and helper *E. coli* strains as previously described ³⁸. Induction of tagged glycoproteins within *Burkholderia* strains was undertaken by the addition of either Rhamnose (final concentration 0.1%, for pSCrhaB2-based plasmids) or Arabinose (final concentration 0.5%, for pKM4-based plasmids) to overnight cultures. Antibiotics were purchased from Thermo Fisher Scientific while all other chemicals, unless otherwise stated, were provided by Sigma-Aldrich.

Recombinant DNA methods. Oligonucleotides used in this study are provided in Table 3. The inducible pSCrhaB2-BCAL2345-his₆ and pSCrhaB2-BCAL2466-his₆ constructs were generated using Gibson assembly ³⁹ by inserting PCR amplified fragments into *Nde*I and *Xba*I linearized pSCrhaB2 using NEBuilder® HiFi DNA master mix according to the manufacturer's instructions (New England Biolabs). pKM4 Site directed mutagenesis was undertaken using PCR-based site replacement and *Dpn*I digestion ⁴⁰. All restriction endonuclease digestions, and agarose gel electrophoresis were performed using standard molecular biology techniques ⁴⁰. All restriction enzymes were used according to the manufacturer's instructions (New England Biolabs). Chemically competent *E. coli* pir2 cells were transformed using heat shock-based transformation ⁴⁰. PCR amplifications were carried out using Phusion DNA polymerase (Thermo Fisher Scientific) according to the manufacturer's recommendations with the addition of 2.5% DMSO for the amplification of *B. cenocepacia* DNA, due to its high GC content. Genomic DNA isolations were performed using genomic DNA clean-up Kits (Zmyo research), while PCR recoveries and restriction digest purifications were performed using DNA Clean & Concentrator Kits (Zmyo research). Colony and screening PCRs were performed using GoTaq DNA polymerase (Promega;

supplemented with 10% DMSO when screening *B. cenocepacia* gDNA). DNA sequencing was undertaken at the Australian Genome Research Facility (Melbourne, Australia).

Construction of unmarked H111 Δ pglL deletion mutants. Deletions of H111 *pglL* (Gene accession: I35_RS13570) were undertaken using the approach of Flannagan *et al.* for the construction of unmarked, non-polar deletions in *B. cenocepacia*⁴¹ using the plasmid pYM8¹².

Protein Immunoblotting. Bacterial whole-cell lysates were prepared from overnight LB cultures of *Burkholderia* strains. 1ml of overnight cultures at an OD₆₀₀ of 1.0 were pelleted, resuspended in 1X Laemmli loading buffer [24.8 mM Tris, 10 mM glycerol, 0.5% (w/v) SDS, 3.6 mM β -mercaptoethanol and 0.001% (w/v) of bromophenol blue (pH 6.8)] and heated for 10 minutes at 95°C. Lysates were then subjected to SDS-PAGE using pre-cast 4-12% gels (Invitrogen) and transferred to nitrocellulose membranes. Membranes were blocked for 1 hour in 5% skim milk in TBS-T (20 mM Tris, 150 mM NaCl and 0.1% Tween 20) and then incubated for at least 16 hours at 4°C with either mouse monoclonal anti-His (1:2,000; AD1.1.10, Biorad) or mouse anti-RNA pol (1:5,000; 4RA2, Neoclone). Proteins were detected using anti-mouse IgG horseradish peroxidase (HRP)-conjugated secondary antibodies (1:3,000; catalog number NEF822001EA, Perkin-Elmer) and developed with Clarity Western ECL Substrates (BioRad). All antibodies were diluted in TBS-T with 1% bovine serum albumin (BSA; Sigma-Aldrich). Images were obtained using an Amersham imager 600 (GE life sciences) or a Biorad ChemiDoc imaging station (Biorad).

Preparation of cell lysates for proteomic analysis. *B. cenocepacia* strains were grown overnight on LB plates as previously described⁵. Plates were flooded with 5 ml of pre-chilled sterile phosphate-buffered saline (PBS) and cells collected with a cell scraper. Cells were washed 3 times in PBS and collected by centrifugation at 10,000 x *g* at 4°C then snap frozen. Frozen whole cell samples were resuspended in 4% SDS, 100 mM Tris pH 8.0, 20 mM DTT and boiled at 95°C with shaking for 10 minutes. Samples were then clarified by centrifugation at 17,000 x *g* for 10 minutes, the supernatant collected, and protein concentration determined by bicinchoninic acid assays (Thermo Fisher Scientific). For glycoproteomic analysis, 1mg of protein for each sample

(three biological replicates per strain/per enzyme) was acetone precipitated by mixing 4 volumes of ice-cold acetone with one volume of sample. For quantitative proteomic comparisons of H111 strains, 200µg of protein for each biological replicate (four biological replicates per strain) were acetone precipitated by mixing 4 volumes of ice-cold acetone with one volume of sample. Samples were precipitated overnight at -20°C and then centrifuged at 10,000 x *g* for 10 minutes at 0°C. The precipitated protein pellets were resuspended in 80% ice-cold acetone and precipitated again for an additional 4 hours at -20°C. Following incubation, samples were spun down at 17,000 x *g* for 10 minutes at 0°C to pellet precipitated protein, the supernatant discarded, and excess acetone evaporated at 65°C for 5 minutes.

Digestion of Proteome samples for glycoproteomic analysis. Dried protein pellets were resuspended in 6 M urea, 2 M thiourea, 50 mM NH₄HCO₃ and reduced / alkylated as previously described⁴². Samples were then digested with one of three different protease combinations; 1) Trypsin (Promega) and Lys-c (Wako); 2) Thermolysin (Promega) or 3) Pepsin (Promega). 1) For the Trypsin / Lys-C digests; Lys-C (1/200 w/w) was added to reduced / alkylated samples for 4 hours at room temperature before the sample was diluted with 100mM NH₄HCO₃ four-fold to reduce the urea/thiourea concentration below 2M and trypsin (1/50 w/w) added. 2) For Thermolysin digestions reduced / alkylated samples were diluted with 100mM NH₄HCO₃ four-fold to reduce the urea/thiourea concentration below 2M and Thermolysin (1/25 w/w) added. 3) For Pepsin digests reduced / alkylated samples were diluted with 0.1% TFA four-fold to reduce the urea/thiourea concentration below 2M, the pH was confirmed to be pH ~2 and Pepsin (1/25 w/w) added. Digests were allowed to proceed overnight at room temperature with shaking. Digested samples were acidified to a final concentration of 0.5% formic acid and desalted on 50 mg tC18 Sep-Pak columns (Waters corporation) according to the manufacturer's instructions. tC18 Sep-Pak columns were conditioned with 10 bed volumes of Buffer B (0.1% formic acid, 80% acetonitrile), then equilibrated with 10 bed volumes of Buffer A* (0.1% TFA, 2% acetonitrile) before use. Samples were loaded on to equilibrated columns then columns washed with at least 10 bed volumes of Buffer A* before bound peptides were eluted with Buffer B. Eluted peptides were dried by vacuum centrifugation at room temperature and stored at -20 °C.

196

197 **ZIC-HILIC enrichment of glycopeptides:** ZIC-HILIC enrichment was performed as previously
198 described with minor modifications ⁴³. ZIC-HILIC Stage-tips ⁴⁴ were created by packing 0.5cm of
199 10 µm ZIC-HILIC resin (Millipore/Sigma) into p200 tips containing a frit of C8 Empore™ (Sigma)
200 material. Prior to use, the columns were washed with Milli-Q water, followed by 95% acetonitrile
201 and then equilibrated with 80% acetonitrile and 1% TFA. Digested proteome samples were
202 resuspended in 80% acetonitrile and 1% TFA. Samples were adjusted to a concentration of 5
203 µg/µL (a total of 500 µg of peptide used for each enrichment) then loaded onto equilibrated ZIC-
204 HILIC columns. ZIC-HILIC columns were washed with 20 bed volumes of 80% acetonitrile, 1% TFA
205 to remove non-glycosylated peptides and bound peptides eluted with 10 bed volumes of Milli-Q
206 water. Eluted peptides were dried by vacuum centrifugation at room temperature and stored at
207 -20°C.

208

209 **LC-MS analysis of glycopeptide enrichments:** ZIC-HILIC enriched samples were re-suspended in
210 Buffer A* and separated using a two-column chromatography set up composed of a PepMap100
211 C18 20 mm x 75 µm trap and a PepMap C18 500 mm x 75 µm analytical column (Thermo Fisher
212 Scientific). Samples were concentrated onto the trap column at 5 µL/min for 5 minutes with
213 Buffer A (0.1% formic acid, 2% DMSO) and then infused into an Orbitrap Fusion™ Lumos™
214 Tribrid™ Mass Spectrometer (Thermo Fisher Scientific) at 300 nL/minute via the analytical column
215 using a Dionex Ultimate 3000 UPLC (Thermo Fisher Scientific). 185-minute analytical runs were
216 undertaken by altering the buffer composition from 2% Buffer B (0.1% formic acid, 77.9%
217 acetonitrile, 2% DMSO) to 28% B over 150 minutes, then from 28% B to 40% B over 10 minutes,
218 then from 40% B to 100% B over 2 minutes. The composition was held at 100% B for 3 minutes,
219 and then dropped to 2% B over 5 minutes before being held at 2% B for another 15 minutes. The
220 Lumos™ Mass Spectrometer was operated in a data-dependent mode automatically switching
221 between the acquisition of a single Orbitrap MS scan (350-1800 m/z, maximal injection time of
222 50 ms, an Automated Gain Control (AGC) set to a maximum of 1*10⁶ ions and a resolution of
223 120k) every 3 seconds and Orbitrap MS/MS HCD scans of precursors (NCE 28% with 5% Stepping,
224 maximal injection time of 60 ms, an AGC set to a maximum of 1*10⁵ ions and a resolution of 15k).

Scans containing the oxonium ions (204.0867; 138.0545 or 366.1396 m/z) triggered three additional product-dependent MS/MS scans⁴⁵ of potential glycopeptides; a Orbitrap EThcD scan (NCE 15%, maximal injection time of 250 ms, AGC set to a maximum of 2×10^5 ions with a resolution of 30k and using the extended mass range setting to improve the detection of high mass glycopeptide fragment ions⁴⁶); a ion trap CID scan (NCE 35%, maximal injection time of 40 ms, an AGC set to a maximum of 5×10^4 ions) and a stepped collision energy HCD scan (using NCE 35% with 8% Stepping, maximal injection time of 150 ms, an AGC set to a maximum of 2×10^5 ions and a resolution of 30k).

Glycopeptide identifications using Byonic: Raw data files were processed using Byonic v3.5.3 (Protein Metrics Inc.⁴⁷). Tryptic samples were searched with a n-ragged semi-tryptic specificity allowing a maximum of two missed cleavage events while Pepsin and Thermolysin samples were searched with non-specific specificity. Carbamidomethyl was set as a fixed modification of cysteine while oxidation of methionine was included as a variable modification. The Burkholderia glycans HexHexNAc₂ (elemental composition: C₂₂O₁₅H₃₆N₂, mass: 568.2115) and Suc-HexHexNAc₂ (elemental composition: C₂₆O₁₈H₄₀N₂, mass: 668.2276) were included as variable modifications at serine and threonine residues. K56-2 samples were searched against the K56-2 proteome⁴⁸ (Uniprot accession: UP000011196, 7,467 proteins) while H111 samples were searched against the H111 proteome⁴⁹ (Uniprot accession: UP000245426, 8111 proteins). A maximum mass precursor tolerance of 5 ppm was allowed while a mass tolerance of up to 10 ppm was set for HCD fragments and 20 ppm for EThcD fragments. Separate datasets from the same strain were combined using R and only glycopeptides with a Byonic score >300 used for further analysis. This score cut-off is in line with previous reports highlighting that score thresholds greater than at least 150 are required for robust tryptic glycopeptide assignments within Byonic^{50,51}. Manual inspection of glycopeptides to assess the correctness of assignments was undertaken using the guidelines of Chen *et al.*⁵² with the additional requirement that HCD glycopeptide spectra should contain evidence for glycan fragments such as oxonium ions or the presence of Y₀ to Y₂ ions. Glycosylation sites were defined as localised if EThcD scans enabled the unambiguous assignment to a specific serine / threonine residue based on c and z ions or a HCD scan contained only a single

serine / threonine residue. Partial localisation is defined as a EThcD spectra containing fragmentation information which allows the ruling out of potential serine / threonine residues yet does not provide evidence enabling the exact site of glycosylation to be assigned. All glycopeptide spectra deemed correct, as well as spectra which provided either partial or complete localisation of glycosylation sites, are provided within Supplementary Data 1 to 4.

Analysis of Glycosylation sites using O-Pair: Automated glycosylation site analysis was undertaken using O-Pair within MetaMorpheus (non-public release version MM0523⁵³). Glycopeptide enriched samples were first processed using the Freestyle Viewer (1.7 SP1, Thermo Fisher Scientific) to remove ion-trap CID scans and then searched allowing for a maximum of 4 glycans. The *Burkholderia* glycans were defined as HexHexNAc₂ (elemental composition: C₂₂O₁₅H₃₆N₂, mass: 568.2115) or Suc-HexHexNAc₂ (elemental composition: C₂₆O₁₈H₄₀N₂, mass: 668.2276). K56-2 samples were searched against the K56-2 proteome⁴⁸ (Uniprot accession: UP000011196) while H111 samples were searched against the H111 proteome⁴⁹ (Uniprot accession: UP000245426). Tryptic samples were searched using the default settings while Pepsin and Thermolysin samples were searched with non-specific specificity allowing 5 to 55 amino acids and the search partitioned into 100 partitions. Analysis of the tryptic glycoproteome samples of *B. pseudomallei* K96243; *B. multivorans* MSMB2008; *B. dolosa* AU0158; *B. humptydooensis* MSMB43; *B. ubonensis* MSMB22; *B. anthina* MSMB649; *B. diffusa* MSMB375 and *B. pseudomultivorans* MSMB2199 (PRIDE ProteomeXchange Consortium accession: PXD018587) were undertaken as above using the proteome databases as outlined in¹⁶. Only class “level 1” glycopeptides with a Q-value less than 0.01 and a site-specific probability of >0.75 were considered as localised and used for further analysis.

Digestion of whole cell samples for proteomic comparisons. Dried protein pellets of H111 strains were resuspended in 6 M urea, 2 M thiourea, 40 mM NH₄HCO₃ and reduced / alkylated prior to digestion with Lys-C then trypsin overnight as described above. Digested samples were acidified to a final concentration of 0.5% formic acid and desalted with home-made high-capacity StageTips composed of 1 mg Empore™ C18 material (3M) and 5 mg of OLIGO R3 reverse phase

resin (Thermo Fisher Scientific) as described^{44,54}. Columns were wet with Buffer B (0.1% formic acid, 80% acetonitrile) and conditioned with Buffer A* prior to use. Acidified samples were loaded onto conditioned columns, washed with 10 bed volumes of Buffer A*, and bound peptides were eluted with Buffer B, before being dried via vacuum centrifugation at room temperature and stored at -20°C.

LC-MS analysis of H111 proteome samples. Stagetip cleaned-up H111 proteome samples were resuspended in Buffer A* and separated using a two-column chromatography set up comprised of a PepMap100 C18 20 mm x 75 µm trap and a PepMap C18 500mm x 75µm analytical column (Thermo Scientific). Samples were concentrated onto the trap column at 5 µl/minute with Buffer A for 5 minutes and infused into an Orbitrap Elite™ Mass Spectrometer (Thermo Scientific) at 300 nl/minute via the analytical column using an Dionex Ultimate 3000 UPLC (Thermo Scientific). The peptides were separated using a 125-minute gradient altering the buffer composition from 1% buffer B to 23% B over 95 minutes, then from 23% B to 40% B over 10 minutes, then from 40% B to 100% B over 5 minutes, the composition was held at 100% B for 5 minutes, and then dropped to 3% B over 2 minutes and held at 3% B for another 8 minutes. The Orbitrap Elite™ was operated in a data-dependent mode automatically switching between the acquisition of a single Orbitrap MS scan (300-1650 m/z, maximal injection time of 50 ms, an AGC set to a maximum of 1×10^6 ions and a resolution of 60k) followed by 20 data-dependent CID MS-MS events (NCE 30%, maximal injection time of 80 ms, an AGC set to a maximum of 2×10^4 ions).

In-gel digestion of DsbA1_{Nm-his6} within Burkholderia lysates. Whole cell lysates of induced *Burkholderia* strains prepared as above for immunoblotting were separated on pre-cast 4-12% gels then fixed in fixative buffer (10% methanol, 7% acetic acid) for 1hr before being stained with Coomassie G-250 overnight. The region corresponding to ~25-35kDa was then excised and processed as previously described (38). Briefly, the excised gel regions were sectioned into ~2 mm² pieces then destained in a solution of 100 mM NH₄HCO₃ / 50% ethanol for 15 minutes at room temperature with shaking. Destaining was repeated twice to ensure the removal of excess

Coomassie. Destained bands were dehydrated with 100% ethanol for 5 minutes and then rehydrated in 50 mM NH_4HCO_3 containing 10 mM DTT. Samples were reduced for 60 minutes at 56 °C with shaking and washed twice in 100% ethanol for 10 minutes to remove DTT. Reduced ethanol washed samples were sequentially alkylated with 55 mM of Iodoacetamide in 50 mM NH_4HCO_3 in the dark for 45 minutes at room temperature. Alkylated samples were then washed with Milli-Q water followed by 100% ethanol twice for 5 minutes to remove residual Iodoacetamide then vacuum-dried for 10 minutes. Alkylated samples were then rehydrated with 20 ng/ μl trypsin (Promega) in 50 mM NH_4HCO_3 at 4 °C for 1 hour. Excess trypsin was removed, the gel pieces covered in 40 mM NH_4HCO_3 , and incubated overnight at 37 °C. Peptides were concentrated and desalted using C18 stage tips (64, 65) before analysis by LC-MS.

LC-MS analysis of DsbA1_{Nm}-his₆ glycopeptides. Stagetip cleaned-up samples were resuspended in Buffer A* and separated using a two-column set up as described above, coupled to a Orbitrap Lumos™ Mass Spectrometer equipped with a FAIMS Pro interface (Thermo Fisher Scientific). 125-minute gradients were run for each sample altering the buffer composition from 3% Buffer B to 28% B over 95 minutes, then from 28% B to 40% B over 10 minutes, then from 40% B to 80% B over 7 minutes, the composition was held at 80% B for 3 minutes, and then dropped to 3% B over 1 minutes and held at 3% B for another 9 minutes. The Lumos™ Mass Spectrometer was operated in a stepped FAIMS data-dependent mode at three different FAIMS CVs -25, -45 and -65 as previously described ³⁵. For each FAIMS CV a single Orbitrap MS scan (350-2000 m/z, maximal injection time of 50 ms, an AGC of maximum of 1×10^6 ions and a resolution of 120k) was acquired every 2 seconds followed by Orbitrap MS/MS HCD scans of precursors (NCE 30%, maximal injection time of 80 ms, an AGC set to a maximum of 1×10^5 ions and a resolution of 15k). HCD scans containing the oxonium ions (204.0867; 138.0545 or 366.1396 m/z) triggered an additional Orbitrap EThcD scan, an ion-trap CID scan and a Orbitrap HCD scan of potential glycopeptides with scan parameters described above. For PRM experiments, 65-minute gradients were run for each sample altering the buffer composition from 3% buffer B to 28% B over 35 minutes, then from 28% B to 40% B over 10 minutes, then from 40% B to 80% B over 7 minutes, the composition was held at 80% B for 3 minutes, and then dropped to 3% B over 1 minutes and held at 3% B for

another 9 minutes. The Lumos™ Mass Spectrometer was operated at a FAIMS CV of -25 in a data-dependent mode automatically switching between the acquisition of a single Orbitrap MS scan (350-2000 m/z, maximal injection time of 50 ms, an AGC set to a maximum of 1×10^6 ions and a resolution of 60k) every 3 seconds followed by ion-trap EThcD MS2 events (NCE 25%, maximal injection time of 200 ms, an AGC set to a maximum of 6×10^4 ions) of precursors and then a Orbitrap EThcD PRM scan (NCE 25%, maximal injection time of 450 ms, an AGC set to a maximum of 2×10^5 ions) of the m/z 1547.77, 1581.12 and 1543.10 which correspond to the +3 charge state of the HexHexNAc₂ glycosylated ²³VQTSVPADSAPAATAAAAPAGLVEGQNYTVLANPIPQQQAGK⁶⁴, the Suc-HexHexNAc₂ glycosylated ²³VQTSVPADSAPAATAAAAPAGLVEGQNYTVLANPIPQQQAGK⁶⁴ and the HexHexNAc₂ glycosylated ²³VQTSVPADSAPAASAAAAPAGLVEGQNYTVLANPIPQQQAGK⁶⁴ peptides.

Proteomic analysis: H111 Proteome and in-gel datasets were processed using MaxQuant (v1.5.5.1 or 1.6.3.4.⁵⁵). The H111 proteome dataset was searched against the H111 proteome⁴⁹ (Uniprot accession: UP000245426) and *B. cenocepacia* strain J2315 (Uniprot accession: UP000001035, 6,993 proteins) to enable the matching of J2315 gene accessions. In-gel digests were searched against either *B. cenocepacia* strain J2315 (Uniprot accession: UP000001035), *B. ubonensis* MSMB22 (Burkholderia Genome Database⁵⁶, Strain number: 3404) or *B. humptydooensis* MSMB43 (Burkholderia Genome Database⁵⁶, Strain number: 4072) depending on the sample type and a custom database of DsbA1_{Nm}-his₆ containing the desired point mutations at position 31 and 36 within Uniprot entry Q9K189. All searches were undertaken using “Trypsin” enzyme specificity with carbamidomethylation of cysteine as a fixed modification. Oxidation of methionine and acetylation of protein N-termini were included as variable modifications and a maximum of 2 missed cleavages allowed. For in-gel samples HexHexNAc₂ (elemental composition: C₂₂O₁₅H₃₆N₂, mass: 568.2115) and Suc-HexHexNAc₂ (elemental composition: C₂₆O₁₈H₄₀N₂, mass: 668.2276) were also included as variable modifications. To enhance the identification of peptides between samples, the Match between Runs option was enabled with a precursor match window set to 2 minutes and an alignment window of 10 minutes. For label free quantitation (LFQ) the MaxLFQ option in Maxquant⁵⁷ was enabled. The

resulting outputs were processed within the Perseus (v1.5.0.9)⁵⁸ analysis environment to remove reverse matches and common protein contaminants prior to further analysis. For LFQ comparisons missing values were imputed based on the observed total peptide intensities with a range of 0.3σ and a downshift of 2.5σ using Perseus. Enrichment analysis was undertaken using Fisher exact tests in Perseus with PglL altered proteins defined as those proteins which were previously reported by Oppy *et al.* as differentially altered in K56-2 Δ pglL compared to K56-2 Wt⁵. To compare the relative abundance of glycosylated and non-glycosylated peptides from DsbA1_{Nm}-his₆ point mutants the area under the curve of these peptides were extracted using the FreeStyle viewer and the resulting data provided within the supplementary document. Statistically analysis of the area under the curve was undertaken in Prism (version 7.0e) using a two-side t-test.

Visualization of glycoproteome and proteome datasets: Data visualization was undertaken using ggplot2⁵⁹ within R with all scripts included in the PRIDE uploaded datasets. To aid in the analysis of the MS/MS data the Interactive Peptide Spectral Annotator⁶⁰ (<http://www.interactivepeptidespectralannotator.com/PeptideAnnotator.html>) was used. Mass spectrometry data (Raw data files, Byonic/Maxquant/O-pair search outputs, R Scripts and output tables) have been deposited into the PRIDE ProteomeXchange Consortium repository^{61,62}. The glycoproteomic datasets are available with the identifier: PXD024090 and can be accessed using the Username: reviewer_pxd024090@ebi.ac.uk and Password: HDbdkj6; the H111 proteome analysis is available with the identifier: PXD023755 and can be accessed using the username: reviewer_pxd023755@ebi.ac.uk, Password: QMdQyTe2; the DsbA1_{Nm}-his₆ (*B. cenocepacia* K56-2) associated analysis is available with the identifier: PXD023955 and can be accessed using the Username: reviewer_pxd023955@ebi.ac.uk Password: vAQRDw5X; the DsbA1_{Nm}-his₆ (*B. humptydooensis* MSMB43 and *B. ubonensis* MSMB22) associated analysis is available with the identifier: PXD024056 and can be accessed using the Username: reviewer_pxd024056@ebi.ac.uk Password: Bj3dwihm.

Bioinformatic analysis of glycoproteins and glycosylation sites: Draft and complete genome sequences of *B. cenocepacia* (n=294) and complete genome sequences of other *Burkholderia* species, subsetting to contain a maximum of 20 of any individual species (n=174), were obtained from the *Burkholderia* database ⁵⁶. Isolates were screened for the presence of the glycosylated motif containing genes from *B. cenocepacia* J2315 using an 80% identity and length BlastN thresholds using screen_assembly3.py v1.2.7 ⁶³. Gene hits were translated to protein sequences and the translated hits screened for the presence of the motif using seqkit locate ⁶⁴, allowing for up to 4 mismatched amino acids (> 80% conservation of the 21 amino acid motif). The identity and coverage of the genes at 80% identity as well as the motif coverage at 80% identity were visualised using ggplot2. Sequence logos were generated using ggseqlogo ⁶⁵. To compare glycoprotein homologues and pglL sequences between *Burkholderia* species the *Burkholderia* Orthologous Groups information provided by the *Burkholderia* database was used ⁵⁶.

Results

O-linked glycosylation targets similar proteins and glycosylation disruption impacts the proteome in a similar manner across *B. cenocepacia* strains.

Recently we noted that the physiochemical properties of peptides heavily influences the ability of ZIC-HILIC enrichment to isolate bacterial glycopeptides³⁵. This observation suggests numerous glycoproteins have likely been overlooked using trypsin alone to assess the glycoproteome of *B. cenocepacia* in previous studies^{16,29}. To increase the coverage of the *B. cenocepacia* glycoproteome we undertook glycoproteomic analysis using multiple proteases^{66,67} and two widely used *B. cenocepacia* strains, K56-2 and H111. Glycopeptide enrichments of Trypsin, Thermolysin and Pepsin digested samples enabled the identification of 584 and 666 unique glycopeptides from K56-2 and H111 strains (Figure 1 A, Supplementary Tables 1 and 2). Although these glycopeptide datasets identified more glycoproteins than identified within previously published studies (Supplementary Figure 1A^{16,29}) the majority of identified glycoproteins were unique to a single strain (Supplementary Figure 1B). This high degree of heterogeneity suggested the presence of erroneous assignments within the datasets, a known issue associated with the assignment of a relatively limited population of modified peptides within large database searches⁶⁸. Consistent with this, examination of the glycopeptide scores reveal a bias toward lower scores (Supplementary Figure 2A and B) supporting a higher than ideal false discovery rate despite stringent filtering⁶⁹. To improve the quality of these glycopeptide datasets we manually curated the Byonic assigned glycopeptides, identifying 256 and 328 high-quality unique glycopeptides in K56-2 and H111 (Figure 1A; Supplementary Data 1 and 2). Of these unique glycopeptides, 167 and 175 glycopeptides within K56-2 and H111 respectively provided partial or complete site localisation (Figure 1A; Supplementary Data 3 and 4). Consistent with improving the data quality, the curated glycopeptides revealed a gaussian distribution within the scores, removing predominately low scoring identifications (Supplementary Figure 2C-F). Of these manually assessed glycopeptides, we noted >65% of glycoproteins were identified within both strains (Figure 1B) supporting a conserved glycoproteome across *B. cenocepacia* strains.

To validate the accuracy of these manually filtered assignments two small novel glycoproteins, BCAL2466 and BCAL2345, were His₆ epitope tagged and expressed within K56-2 WT and K56-2 $\Delta pgII$ (Figure 1C). Consistent with a single glycosylation event on BCAL2466 and multiple glycosylation events on BCAL2345, the expression of these proteins within K56-2 $\Delta pgII$ resulted in reduced gel mobility supporting their glycosylation status. To functionally support the observed similarities within the glycoproteomes of *B. cenocepacia* strains, we generated two independent H111 $\Delta pgII$ mutants and assessed the proteome changes observed in the absence of glycosylation within H111, compared to our previously published K56-2 findings (Figure 1D⁵). Although the proteomic alterations show congruent behaviour, it should be noted strain differences are observed within some proteins such as BCAL1086, BCAL0282 and BCAL1270 (Figure 1D; Supplementary Table 3). Regardless of these differences between K56-2 and H111 the proteins observed to undergo statistically significant alterations within H111 $\Delta pgII$ were enriched for proteins observed to be altered within K56-2 $\Delta pgII$ (Fisher exact test p-value = 0.0006 Supplementary table 4). Independent H111 $\Delta pgII$ mutants also showed similar alterations with the proteome and were highly similar (Fisher exact test p-value = 5.67×10^{-15} , Supplementary table 3 and 4). Taken together these findings support that the O-linked glycoproteome of *B. cenocepacia* strains are similar in substrates and functional consequences.

Glycosylation site analysis reveals that *B. cenocepacia* glycosylation occurs solely on serine residues

Our manually curated glycoproteomic data revealed ~50% of unique glycopeptides provided at least partial site localisation information (Figure 1A, Supplementary Data 3 and 4) of which 88 sites could be precisely localised across 70 glycoproteins (Supplementary Table 5). Examination of the sites of glycosylation revealed 69 out of 70 sites (>98%) were observed on Serine residues and that glycosylation favoured Alanine at the -1 position, yet this was not a strict requirement (Figure 2A). As only a single Threonine modification was identified on a single glycopeptide derived from the protein BCAM0996, we sought to confirm the accuracy of this assignment.

Examination of glycopeptides from BCAM0996 revealed multiple serine residues are modified within the same peptide observed to be modified at T¹⁵⁹ (Supplementary Data 3 and 4). The close proximity of this sole threonine modification to multiple serine modification events further raised concerns of miss-localisation. Manual annotation of the assigned glycopeptide supported the incorrect localisation of the glycosylation site due to the incorrect assignment of the glycan composition and peptide caused by a secondary modification within the peptide sequence revealing S¹⁶⁷ to be the correct localisation site (Supplementary Figure 3). This finding suggested all localised glycosylation sites within *B. cenocepacia* were observed on Serine residues. To independently validate this preference for Serine, we re-analysed these glycoproteomic datasets with O-Pair, a glycosylation site localisation focused glycoproteomic tool⁵³, supporting that the majority of high quality localisable glycosylation events occurred on Serine residues (Supplementary Figure 4; Supplementary Table 6). Taken together, these findings support that glycosylation within the *B. cenocepacia* glycoproteome localises exclusively to Serine residues.

Threonine residues undergo poor glycosylation in *B. cenocepacia*

To experimentally explore this preference for Serine, we utilised the glycoprotein DsbA1_{Nm}-his₆ that we have previously observed is predominately glycosylated by a single glycosylation event when expressed in *B. cenocepacia*^{5,29}. Past work showed that glycosylation occurs at an unknown site within the peptide ²³VQTSVPADSAPAASAAAAPAGLVEGQNYTVLANPIPQQQAGK⁶⁴, suggesting that one of the three Serine residues (S²⁶, S³¹ and S³⁶) was the site of glycosylation within DsbA1_{Nm}-his₆. Of these sites, S³¹ and S³⁶ are flanked by Alanine residues, consistent with the preferred glycosylation sites within native substrates (Figure 2A), yet only substitution of S³⁶ with either Alanine or Threonine resulted in gel mobility shifts consistent with the loss of glycosylation (Figure 2B). Glycopeptide analysis further confirmed S³⁶ was the sole residue modified within this peptide (Supplementary Figure 5 and Supplementary Table 7). As MS analysis provides greater dynamic range than western blotting, we manually investigated if glycosylation still occurred at T³⁶ with analysis supporting the presence of a glycosylated form of the ²³VQTSVPADSAPAATAAAAAPAGLVEGQNYTVLANPIPQQQAGK⁶⁴ at <10% of the abundance of the

unmodified form (Figure 2C, Supplementary Figure 6A). Targeted MS analysis supported the identity of this glycopeptide yet could not provide definitive site localisation to T³⁶ (Supplementary Figure 6B). To further assess if Threonine could be modified, we introduced the DsbA1_{Nm}-his₆ point mutants T³⁶ and A³⁶ into K56-2 ΔpglL AmrAB::S7-pglL-his, an overexpressing Pgl_{BC} strain ⁵, to assess if increasing Pgl_{BC} levels enhanced Threonine glycosylation. Western analysis demonstrates the majority of DsbA1_{Nm}-his₆ T³⁶ remained unglycosylated within K56-2 ΔpglL AmrAB::S7-pglL-his. Yet, in contrast to K56-2 WT, a faint band potentially corresponding to glycosylated DsbA1_{Nm}-his₆ T³⁶ was observed (Figure 3A). Targeted MS analysis confirms the glycosylation of T³⁶ in DsbA1_{Nm}-his₆ T³⁶ when expressed in K56-2 ΔpglL AmrAB::S7-pglL-his (Figure 3B and C). Combined this data supports that it is possible to glycosylate Threonine in *B. cenocepacia*, but the preferred residue for glycosylation is Serine, even when Pgl_{BC} is overexpressed.

Glycoproteins / glycosylation sites are highly conserved across *B. cenocepacia* strains

With our proteome and site-directed analysis supporting Serine as the favoured residue for glycosylation, we next addressed whether these glycosylation proteins/sites are conserved across a diverse set of *B. cenocepacia* strains. Leveraging our curated glycoprotein dataset (Supplementary Table 5) we assessed the conservation of these 70 glycoproteins across 294 publicly available *B. cenocepacia* genomes ⁵⁶. Across strains, the frequency of observation, defined as the presence of proteins with a minimum blast identity of >80%, revealed 65 glycoproteins are present in >80% of available genomes and that these glycoproteins are highly similar, with an average amino acid sequence identity >95% (Figure 4A, Supplementary Table 8). This high level of conservation is also reflected at the glycosylation site (±10 amino acids) where the majority of glycosylation sites are highly conserved at the sequence level (Figure 4B, Supplementary Table 8). It should be noted that although the majority of glycosylation sites are conserved (Supplementary Figure 7 to 12), variations are observed; such as in S¹⁴⁰ of BCAL1746, which is present in only a fraction of strains (Figure 4C,); S⁴⁰⁴ of BCAL1674, where multiple alterations within the flanking sequences were noted (Figure 4E); and S²⁶² of BCAL0163, where the serine required for glycosylation is lost across multiple strains (Figure 4F). This data supports

that both *B. cenocepacia* glycoproteins and glycosylation sites are highly conserved across *B. cenocepacia* strains.

Across *Burkholderia* species similar glycoproteins are targeted for glycosylation at Serine residues.

In light of the high sequence identity of PgIL across *Burkholderia* species (Supplementary Figure 13) we sought to assess if the preference for serine glycosylation was a common feature across *Burkholderia* glycoproteomes. To investigate this, we utilized our recently published glycoproteomic datasets of eight *Burkholderia* species (*B. pseudomallei* K96243; *B. multivorans* MSMB2008; *B. dolosa* AU0158; *B. humptydooensis* MSMB43; *B. ubonensis* MSMB22; *B. anthina* MSMB649; *B. diffusa* MSMB375 and *B. pseudomultivorans* MSMB2199¹⁶) and re-analysed these datasets using O-Pair. As with *B. cenocepacia*, Serine was the dominant residue targeted for glycosylation with only 11 out of 440 high confidence sites localised to Threonine (Figure 5A; Supplementary table 9). To experimentally support this preference, we introduced DsbA1_{Nm}-his₆ variants into two strains, *B. humptydooensis* MSMB43 and *B. ubonensis* MSMB22, to compare the preference for glycosylation at Serine over Threonine residues. Similar to *B. cenocepacia* we find DsbA1_{Nm}-his₆ T³⁶ is predominantly unable to be glycosylated within both *B. humptydooensis* MSMB43 and *B. ubonensis* MSMB22 (Figure 5B and C). As with *B. cenocepacia*, the glycosylated form of the peptide ²³VQTSVPADSAPAATAAAAPAGLVEGQNYTVLANPIPQQQAGK⁶⁴ was observed at low levels within both *B. humptydooensis* MSMB43 and *B. ubonensis* MSMB22 (Supplementary Figure 14; Supplementary Table 10). Finally, to survey the similarities between *Burkholderia* glycoproteomes we assessed the sequence identity of glycoproteins revealing the majority of glycoproteins are homologues of confirmed glycoproteins within *B. cenocepacia* (Figure 5D, Supplementary Table 11). Enrichment analysis supports that for *B. anthina* MSMB649; *B. ubonensis* MSMB22; *B. multivorans* MSMB2008; *B. pseudomultivorans* MSMB2199, and *B. dolosa* AU0158, this overlap represents a statistically significant enrichment (Figure 5E Supplementary Table 11). Combined these results support that Serine is the preferred target of glycosylation

555 across *Burkholderia* glycoproteomes and that similar glycoproteins were targeted across the
556 *Burkholderia* genus.
557

Discussion

The conservation of glycosylation systems across bacterial species is increasingly being recognised as a common phenomenon in bacterial genera^{17,70,71}. Although previous studies have sought to confirm the presence of protein glycosylation as well as highlight differences within the glycans utilised across species, little attention has been given to the protein's substrates and glycosylation sites themselves. Within this work we undertook a glycosylation-site-focused analysis of *Burkholderia* species, revealing both experimentally (Figure 1B and 5D) and bioinformatically (Figure 4A) that proteins subjected to glycosylation are conserved across *Burkholderia* strains / species and that *O*-linked glycosylation within this genus is overwhelmingly restricted to Serine residues (Figure 2A and 5A). Our findings here highlight that protein glycosylation targets nearly 100 protein substrates, the majority of which appear conserved and targeted for glycosylation across multiple *Burkholderia* species, supporting that these glycosylation events are responsible for shared, yet still largely unknown, functions across the *Burkholderia* genus. Consistent with this concept, phenotypic studies have previously shown the loss of *O*-linked glycosylation leads to similar impacts on *Burkholderia* species, such as in the case of glycosylation null *B. cenocepacia*, *B. pseudomallei* and *B. thailandensis* strains, which all possess biofilm defects^{5,72}. Thus, taken together this highlights that *Burkholderia* glycoproteomes are far more similar and conserved than previously noted.

Similar to the optimal motif proposed for the *N. meningitidis* PglL, WPAAASAG⁷³, our proteomic analysis supports that Serine residues flanked by Alanine are predominately targeted for glycosylation within *B. cenocepacia*, yet glycosylation can still occur within divergent sequences lacking consensus residues (Figure 2A). An unexpected finding within this study is that Serine is strongly targeted for glycosylation across *Burkholderia* glycoproteomes, and that a dramatic difference in the extent by which threonine/serine residues can be glycosylated exists. In fact, although multiple studies have examined the glycan promiscuity and general ability of PglL enzymes to glycosylate proteins⁷⁴⁻⁷⁶, no previous reports have suggested that threonine/serine residues are not equivalent targets for glycosylation. This difference in preference between these

residues is striking with semi-quantitative comparisons of the unglycosylated and glycosylated peptides suggesting Serine is glycosylated ~500 times more efficiently than Threonine within *B. cenocepacia* (Figure 2B and C), with a similar trend also noted within *B. humptydooensis* MSMB43 and *B. ubonensis* MSMB22 (Supplementary Figure 14). Although we confirmed that Threonine can be glycosylated at a low occupancy, this dramatic difference in preference supports that Threonine glycosylation events are likely rare in *Burkholderia* glycoproteomes, or if observed, are likely found at modest levels of occupancy. Interestingly, within other PglL glycosylation systems where only a limited number of glycosylation sites have been confirmed to date, such as in *N. gonorrhoeae*^{33,34}, *A. baumannii*¹⁵ and *F. tularensis*³⁰, all confirmed glycosylation sites appear to occur on serine residues. These trends suggest that Serine specificity may be a more general feature of the PglL oligosaccharidetransferases, yet further work is required to confirm this preference across known PglL enzymes.

At the genus level our analysis supports that across *Burkholderia* species glycoproteins are conserved (Figure 5D and Supplementary Figure 15). Although the limited number of known glycosylation sites within other bacterial glycosylation systems has hampered the analysis of glycosylation site conservation, the identification of varying numbers of glycosylation sites in proteins such as pilin of *Neisseria* species^{7,14,23,24} has suggested that glycosylation sites maybe highly variable across strains/species. In contrast, our analysis highlights that the majority of glycosylation sites, within *B. cenocepacia* at least, appear conserved and that loss of glycosylation sites appear rare. This said, we did note examples of glycosylation sites which were only observed within a subset of genomes, such as S¹⁴⁰ of BCAL1746, yet this level of variability was the exception, not the rule, across *B. cenocepacia*. Although our glycoproteome of *B. cenocepacia* significantly expands the known number of glycosylation sites from six sites in 23 glycoproteins²⁹ to 88 sites in 70 glycoproteins, it is important to note these glycosylation events likely still only represent a subset of the complete *B. cenocepacia* glycoproteome. It is well known that growth under different conditions, such as nutrient limitation⁷⁷ and or in the presence of chemical queues^{78,79} can dramatically alter the observable proteome of *B. cenocepacia*. As the growth of *B. cenocepacia* strains were undertaken under laboratory conditions with growth on rich media,

it is possible multiple glycoproteins not expressed under these conditions have been missed. Thus, to better understand glycoproteins associated with virulence or survival under specific environmental conditions further studies may be required.

In summary, this work furthers our understanding of both *B. cenocepacia* and the *Burkholderia* pan-glycoproteome as a whole and represents the first step in understanding the roles and specific functions of glycoproteins across the *Burkholderia* genus. The insights into the preference for glycosylation at Serine residues improves our ability to predict the specific residues likely to be modified within a glycopeptide, while the identification that glycoproteins identified within one species are likely glycosylated within another provides a new opportunity to dissect the conserved roles of glycoproteins across this genus. These insights will aid in future studies to understand why glycosylation events have been maintained within different *Burkholderia* species.

ACKNOWLEDGEMENTS

N.E.S is supported by an Australian Research Council Future Fellowship (FT200100270) and an ARC Discovery Project Grant (DP210100362). M.R.D is supported by a University of Melbourne CR Roper fellowship. We thank the Melbourne Mass Spectrometry and Proteomics Facility of The Bio21 Molecular Science and Biotechnology Institute for access to MS instrumentation and Byonic. We thank Lei Lu, Nick Riley and Lloyd M. Smith for providing the custom version of O-Pair for the analysis of *Burkholderia* glycopeptides. pSCrhaB2 was a gift from Miguel Valvano (Addgene plasmid # 113634; <http://n2t.net/addgene:113634> ; RRID:Addgene_113634)

Tables

Table 1: Strains list

Strain name	Description	Source/Reference
<i>E. coli</i> strain		
<i>E. coli</i> pir2	F ⁻ Δ lac169 <i>rpoS</i> (Am) <i>robA1</i> <i>creC510</i> <i>hsdR514</i> <i>endA</i> <i>recA1</i> <i>uidA</i> (Δ Mlul)::pir-116	Thermo Scientific
<i>B. cenocepacia</i> strains		
<i>B. cenocepacia</i> K56-2	Clinical isolate of the ET12 lineage (Darling P 1998, Mahenthiralingam E <i>et al</i> 2005)	Canadian <i>B. cepacia</i> research and referral repository ³⁶
<i>B. cenocepacia</i> K56-2 Δ pgl	Δ pgl (BCAL0960) derivative of K56-2 created using pYM8	5
<i>B. cenocepacia</i> K56-2 Δ pgl <i>amrAB::S7-pgl-his₁₀</i>	<i>amrAB::S7-pgl-his₁₀</i> chromosomal complement derivative of Δ pgl (BCAL0960) overexpressing pgl from the S7 promoter	5
<i>B. cenocepacia</i> H111	Clinically non-ET12 lineage isolate	37
<i>B. cenocepacia</i> H111 Δ pgl candidate 1	Δ pgl (I35_RS13570) derivative of H111 created using pYM8 candidate	This study
<i>B. cenocepacia</i> H111 Δ pgl candidate 2	Δ pgl (I35_RS13570) derivative of H111 created using pYM8 candidate	This study

Table 2: Plasmid list

Plasmid	Description	Source/ Reference
pRK2013	ori _{colE1} , RK2 derivative, Kan ^R mob ⁺ tra ⁺	80
pDAI-SceI-SacB	ori _{pBBR1} , Tet ^R , P _{dhfr} , mob ⁺ , expressing ISce-I and the negative selection marker SacB	41,81
pYM8	pGPI-SceI with fragments flanking <i>pglL</i> (BCAL0960)	12
pSCrhaB2	ori _{pBBR1} , <i>rhaR</i> , <i>rhaS</i> , P _{rhaB} Tp ^R mob ⁺ (Addgene: #113634)	82
pSCrhaB2- BCAL2466-his ₆	Tp ^R pSCrhaB2 Rhamnose inducible plasmid containing C-terminal his ₆ -tagged BCAL2466 (Ecotin)	This study
pSCrhaB2- BCAL2345-his ₆	Tp ^R pSCrhaB2 Rhamnose inducible plasmid containing C-terminal his ₆ -tagged BCAL2345 (SecG)	This study
pKM4 (DsbA1 _{Nm} -his ₆)	Tp ^R pMLBad-based plasmid containing C-terminal his ₆ -tagged DsbA1 from <i>N. meningitidis</i> MC58	74
pKM4 ^{S31A} (DsbA1 _{Nm} -his ₆ A ³¹)	Tp ^R pMLBad-based plasmid containing C-terminal his ₆ -tagged DsbA1-his ₆ from <i>N. meningitidis</i> MC58 site directed mutant of S ³¹ to A ³¹	This study
pKM4 ^{S31T} (DsbA1 _{Nm} -his ₆ T ³¹)	Tp ^R pMLBad-based plasmid containing C-terminal his ₆ -tagged DsbA1-his ₆ from <i>N. meningitidis</i> MC58 site directed mutant of S ³¹ to T ³¹	This study
pKM4 ^{S36A} (DsbA1 _{Nm} -his ₆ A ³⁶)	Tp ^R pMLBad-based plasmid containing C-terminal his ₆ -tagged DsbA1-his ₆ from <i>N. meningitidis</i> MC58 site directed mutant of S ³⁶ to A ³⁶	This study
pKM4 ^{S36T} (DsbA1 _{Nm} -his ₆ T ³⁶)	Tp ^R pMLBad-based plasmid containing C-terminal his ₆ -tagged DsbA1-his ₆ from <i>N. meningitidis</i> MC58 site directed mutant of S ³⁶ to T ³⁶	This study

Supplementary Data and Tables

Supplementary Data 1: Manually curated *B. cenocepacia* H111 glycopeptides (Best scoring unique glycopeptides). For each of the best scoring unique glycopeptide, the Byonic assigned spectra is provided in addition to the assigned J2315 gene number, protein name, peptide assignment, glycan assignment, if the spectra enables localisation of the glycan within the sequence, assignment associated metrics (m/z; mass error, score, delta score, delta mod score, scan number within the LC-MS run and scan time), enzymatic digest for which the best scoring glycopeptide was observed in, replicate for which the best scoring glycopeptide was observed in and page within the pdf the spectra can be found.

Supplementary Data 2: Manually curated *B. cenocepacia* K56-2 glycopeptides (Best scoring unique glycopeptides). For each of the best scoring unique glycopeptide, the Byonic assigned spectra is provided in addition to the assigned J2315 gene number, protein name, peptide assignment, glycan assignment, if the spectra enables localisation of the glycan within the sequence, assignment associated metrics (m/z; mass error, score, delta score, delta mod score, scan number within the LC-MS run and scan time), enzymatic digest for which the best scoring glycopeptide was observed in, replicate for which the best scoring glycopeptide was observed in and page within the pdf the spectra can be found.

Supplementary Data 3: Manually curated *B. cenocepacia* H111 glycopeptides (Best localised unique glycopeptides). For each of the best localising unique glycopeptide, the Byonic assigned spectra is provided in addition to the assigned J2315 gene number, protein name, peptide assignment, glycan assignment, if the spectra enables localisation of the glycan within the sequence, assignment associated metrics (m/z; mass error, score, delta score, delta mod score, scan number within the LC-MS run and scan time), enzymatic digest for which the best scoring glycopeptide was observed in, replicate for which the best scoring glycopeptide was observed in and page within the pdf the spectra can be found.

Supplementary Data 4: Manually curated *B. cenocepacia* K56-2 glycopeptides (Best localised unique glycopeptides). For each of the best localising unique glycopeptide, the Byonic assigned spectra is provided in addition to the assigned J2315 gene number, protein name, peptide assignment, glycan assignment, if the spectra enables localisation of the glycan within the sequence, assignment associated metrics (m/z; mass error, score, delta score, delta mod score, scan number within the LC-MS run and scan time), enzymatic digest for which the best scoring glycopeptide was observed in, replicate for which the best scoring glycopeptide was observed in and page within the pdf the spectra can be found.

Supplementary Table 1. Combined Byonic search results of *B. cenocepacia* H111 glycopeptide enrichments. The combined Byonic search results from all proteases across each of the three biological replicates of *B. cenocepacia* H111 are provided. The complete list of all glycopeptides identified, the best unique glycopeptides filtered for glycopeptides with a Byonic score over 300 and the curated list of glycopeptides associated with Supplementary data 1 are provided.

Supplementary Table 2. Combined Byonic search results of *B. cenocepacia* K56-2 glycopeptide enrichments. The combined Byonic search results from all proteases across each of the three biological replicates of *B. cenocepacia* K56-2 are provided. The complete list of all glycopeptides identified, the best unique glycopeptides filtered for glycopeptides with a Byonic score over 300 and the curated list of glycopeptides associated with supplementary data 2 are provided.

Supplementary Table 3. LFQ proteome analysis of *B. cenocepacia* strains H111 WT vs H111 Δ pglL candidate 1 vs H111 Δ pglL candidate 8. The Perseus processed MaxQuant search results for the protein analysis of four biological replicates of strains H111 WT, H111 Δ pglL candidate 1 and H111 Δ pglL candidate 8 are provided. For each identified protein, the log2 LFQ protein values, the T-test information including the $-\log_{10}(p\text{-value})$, difference in the mean between the groups and if the resulting p -values is below the multiple hypothesis corrected p -value are provided. Categorical information used for enrichment analysis including GO terms and if proteins were considered to be altered within K56-2 Δ pglL within Oppy *et al.* 2019⁵ are provided. For each protein the protein score, number of MS/MS for the corresponding protein, expected molecular weight, number of peptides identified, peptide sequence coverage and corresponding J2315 gene number for the protein are provided.

Supplementary Table 4: Enrichment analysis of global proteome changes in response to loss of glycosylation in *B. cenocepacia* H111. Using Perseus, proteins determined to undergo statistically significant changes were assessed for co-occurrence of statistically significant changes across strains as well as enrichment for proteins considered to be altered within K56-2 Δ pglL within Oppy *et al.* 2019⁵.

Supplementary Table 5. Curated *B. cenocepacia* glycosylation sites. The identified glycosylation sites from *B. cenocepacia* K56-2 and H111 are provided. For each site the J315 Gene name, position of the site within the J2315 version of the protein, residue the glycan was localised to, amino acid sequence +/- 10 amino acid either side of the localised site and which strain these sites were observed in are provided.

Supplementary Table 6. O-Pair analysis of *B. cenocepacia* glycosylation sites. O-pair site localisation information is provided with only class "level 1" glycopeptides with a Q-value less than 0.01 and a site-specific probability of >0.75 considered as localised and used for data visualisation.

Supplementary Table 7. Proteome analysis of the 25-35kDa of *B. cenocepacia* lysates expressing DsbA1_{Nm}-his₆ variants. MaxQuant peptide search results of induced K56-2 expressing DsbA1_{Nm}-his₆ WT or T³⁶. For each peptide the mass, Protein Group, Modification status, identification type (by MS/MS or matching), Retention time, Charge, PEP, MS/MS scan number for the best scoring peptide spectral match, Raw file of the best scoring peptide spectral match, Score, Delta Score, Intensity and number of MS/MS are provided.

Supplementary Table 8. Glycoprotein and glycosylation site conservation across Burkholderia genomes. Summary of sequence identity percentages of glycoproteins and glycosylation motifs identified across *B. cenocepacia* strains and representative Burkholderia species.

Supplementary Table 9. O-Pair analysis of multiple Burkholderia species. O-pair site localisation information is provided with only class “level 1” glycopeptides with a Q-value less than 0.01 and a site-specific probability of >0.75 considered as localised and used for data visualisation.

Supplementary Table 10. Proteome analysis of the 25-35kDa of *B. ubonensis* MSMB22 and *B. humptydooensis* MSMB43 lysates expressing DsbA1_{Nm}-his₆ variants. MaxQuant peptide search results of induced K56-2 expressing DsbA1_{Nm}-his₆ WT or T³⁶. For each peptide the mass, Protein Group, Modification status, identification type (by MS/MS or matching), Retention time, Charge, PEP, MS/MS scan number for the best scoring peptide spectral match, Raw file of the best scoring peptide spectral match, Score, Delta Score, Intensity and number of MS/MS are provided.

Supplementary Table 11. Glycoprotein homologues identified across Burkholderia species. Summary of glycoprotein homologues identified, and enrichment analysis of glycoproteins identified across Burkholderia species.

Figures

Figure 1. *B. cenocepacia* glycoproteomic analysis reveals similar substrates and functional impacts on *B. cenocepacia* strains. A) Within *B. cenocepacia* strains K56-2 and H111 584 and 666 unique glycopeptides were identified respectively (Byonic scores >300). Within these glycopeptides manual curation identified 256 and 328 glycopeptides as high-quality unique glycopeptides of which 167 and 175 provided partial or complete glycosylation site information. **B)** Across the manually curated glycopeptides a total of 98 unique glycoproteins were identified of which 47 were identified within both strains. **C)** Western blotting of two potential glycoproteins, His₆ tagged BCAL2466 and BCAL2345, reveals alterations in protein banding when expressed within the glycosylation null strain K56-2 Δ *pglL* compared to the K56-2 WT. **D)** LFQ analysis of the most altered proteins observed between H111 Δ *pglL* compared to H111 WT reveals similar proteome alterations as seen previously within K56-2⁵.

Figure 2. O-linked glycosylation predominately occurs on Serine residues across the *B. cenocepacia* glycoproteome. A) Sequence analysis of localised glycosylation sites across *B. cenocepacia* strains reveal the majority of assigned sites are Serine. **B)** Western analysis of *DsbA1_{Nm}-his₆* variants reveals substitution of only S³⁶ to Alanine or Threonine results in reduced glycosylation. **C)** Relative amounts of the *DsbA1_{Nm}-his₆* glycosylated and non-glycosylated peptides ²³VQTSVPADSAPAASAAAAPAGLVEGQNYTVLANPIPPQQAGK⁶⁴ and ²³VQTSVPADSAPAATAAAAAPAGLVEGQNYTVLANPIPPQQAGK⁶⁴ observed within K56-2 WT reveals the alteration of S³⁶ to T³⁶ dramatically reduces glycosylation.

Figure 3. Threonine residues undergo poor glycosylation in *B. cenocepacia* even with overexpression of *pglL_{BC}*. A) Western analysis of *DsbA1_{Nm}-his₆* variants reveals a minor product, indicated with an arrow, is observed within *DsbA1_{Nm}-his₆* T³⁶ when expressed within K56-2 Δ *pglL* AmrAB::S7-*pglL*-his. **B and C)** Targeted MS analysis using EThcD fragmentation enabled the confirmation of glycosylation at residue T³⁶ within ²³VQTSVPADSAPAATAAAAAPAGLVEGQNYTVLANPIPPQQAGK⁶⁴.

Figure 4. Glycoproteins and glycosylation sites are conserved across *B. cenocepacia*. A) Violin plot showing the amino acid sequence variation and frequency of 69 glycoproteins across 294 *B. cenocepacia* genomes. Relative frequency is shown in red (right axis) and protein sequence identity (>80% minimum) is shown in black (violin plot, left axis). **B)** Density plot representing the conservation of the glycosylation site sequence across 69 glycoproteins within 294 *B. cenocepacia* genomes **C)** Sequence logos of glycoproteins denoted within the density plot; BCAL1746 S¹⁴⁰; BCAL1674 S⁴⁰⁴ and BCAL0163 S²⁶².

Figure 5. Serine glycosylation and the conservation of glycoproteins across *Burkholderia* species. A) The re-analysis of *Burkholderia* glycopeptide enrichment datasets reveals the majority of assigned glycosylation sites are localised to Serine residues. **B)** and **C)** Western analysis of *DsbA1_{Nm}-his₆* variants expressed within *B. ubonensis* MSMB22 (B) and *B. humptydooensis* MSMB43 (C) supports that Threonine is disfavoured for glycosylation similar to *B. cenocepacia*. **D)** Analysis of glycoproteins observed across *Burkholderia* species supports multiple B.

799 *cenocepacia* glycoprotein homologues are observed glycosylated within other species. **E)**
800 Enrichment analysis of glycoproteins shows an enrichment of shared glycoproteins in *B. anthina*
801 MSMB649; *B. ubonensis* MSMB22; *B. multivorans* MSMB2008; *B. pseudomultivorans* MSMB2199
802 and *B. dolosa* AU0158 compared to *B. cenocepacia*
803

References

- 1 Koomey, M. O-linked protein glycosylation in bacteria: snapshots and current perspectives. *Curr Opin Struct Biol* **56**, 198-203, doi:10.1016/j.sbi.2019.03.020 (2019).
- 2 Nothaft, H. & Szymanski, C. M. New discoveries in bacterial N-glycosylation to expand the synthetic biology toolbox. *Curr Opin Chem Biol* **53**, 16-24, doi:10.1016/j.cbpa.2019.05.032 (2019).
- 3 Cain, J. A. *et al.* Proteomics reveals multiple phenotypes associated with N-linked glycosylation in *Campylobacter jejuni*. *Mol Cell Proteomics*, **8**, 715-734. doi:10.1074/mcp.RA118.001199 (2019).
- 4 Abouelhadid, S. *et al.* Quantitative Analyses Reveal Novel Roles for N-Glycosylation in a Major Enteric Bacterial Pathogen. *MBio* **10**, doi:10.1128/mBio.00297-19 (2019).
- 5 Oppy, C. C. *et al.* Loss of O-linked protein glycosylation in *Burkholderia cenocepacia* impairs biofilm formation, siderophore activity and alters transcriptional regulators. *mSphere* **4**, e00660-19. doi: 10.1128/mSphere.00660-19. (2019)
- 6 Abouelhadid, S., Raynes, J., Bui, T., Cuccui, J. & Wren, B. W. Characterization of Posttranslationally Modified Multidrug Efflux Pumps Reveals an Unexpected Link between Glycosylation and Antimicrobial Resistance. *mBio* **11**, doi:10.1128/mBio.02604-20 (2020).
- 7 Gault, J. *et al.* Neisseria meningitidis Type IV Pili Composed of Sequence Invariable Pilins Are Masked by Multisite Glycosylation. *PLoS Pathog* **11**, e1005162, doi:10.1371/journal.ppat.1005162 (2015).
- 8 Lassak, J. *et al.* Arginine-rhamnosylation as new strategy to activate translation elongation factor P. *Nat Chem Biol* **11**, 266-270, doi:10.1038/nchembio.1751 (2015).
- 9 El Qaidi, S. *et al.* An intra-bacterial activity for a T3SS effector. *Sci Rep* **10**, 1073, doi:10.1038/s41598-020-58062-y (2020).
- 10 Iwashkiw, J. A., Vozza, N. F., Kinsella, R. L. & Feldman, M. F. Pour some sugar on it: the expanding world of bacterial protein O-linked glycosylation. *Mol Microbiol* **89**, 14-28, doi:10.1111/mmi.12265 (2013).
- 11 Kightlinger, W., Warfel, K. F., DeLisa, M. P. & Jewett, M. C. Synthetic Glycobiology: Parts, Systems, and Applications. *ACS Synth Biol* **9**, 1534-1562, doi:10.1021/acssynbio.0c00210 (2020).
- 12 Fathy Mohamed, Y. *et al.* A general protein O-glycosylation machinery conserved in *Burkholderia* species improves bacterial fitness and elicits glycan immunogenicity in humans. *J Biol Chem*, **294**, 13248-13268, doi:10.1074/jbc.RA119.009671 (2019).
- 13 Lees-Miller, R. G. *et al.* A common pathway for O-linked protein-glycosylation and synthesis of capsule in *Acinetobacter baumannii*. *Mol Microbiol* **89**, 816-830, doi:10.1111/mmi.12300 (2013).
- 14 Anonsen, J. H. *et al.* Characterization of a Unique Tetrasaccharide and Distinct Glycoproteome in the O-Linked Protein Glycosylation System of *Neisseria elongata* subsp. *glycolytica*. *J Bacteriol* **198**, 256-267, doi:10.1128/JB.00620-15 (2016).
- 15 Scott, N. E. *et al.* Diversity within the O-linked protein glycosylation systems of *acinetobacter* species. *Mol Cell Proteomics* **13**, 2354-2370, doi:10.1074/mcp.M114.038315 (2014).

- 16 Ahmad Izaham, A. R. & Scott, N. E. Open database searching enables the identification and comparison of bacterial glycoproteomes without defining glycan compositions prior to searching. *Mol Cell Proteomics* **19**, 1561-1574. doi:10.1074/mcp.TIR120.002100 (2020).
- 17 Hadjineophytou, C. *et al.* Genetic determinants of genus-Level glycan diversity in a bacterial protein glycosylation system. *PLoS Genet* **15**, e1008532, doi:10.1371/journal.pgen.1008532 (2019).
- 18 Cain, J. A., Dale, A. L., Sumer-Bayraktar, Z., Solis, N. & Cordwell, S. J. Identifying the targets and functions of N-linked protein glycosylation in *Campylobacter jejuni*. *Mol Omics* **16**, 287-304, doi:10.1039/d0mo00032a (2020).
- 19 Kowarik, M. *et al.* Definition of the bacterial N-glycosylation site consensus sequence. *EMBO J* **25**, 1957-1966, doi:10.1038/sj.emboj.7601087 (2006).
- 20 Riley, N. M., Malaker, S. A., Driessen, M. & Bertozzi, C. R. Optimal Dissociation Methods Differ for N- and O-glycopeptides. *J Proteome Res*, **19**, 3286-3301. doi:10.1021/acs.jproteome.0c00218 (2020).
- 21 Scott, N. E. *et al.* Mass spectrometric characterization of the surface-associated 42 kDa lipoprotein JlpA as a glycosylated antigen in strains of *Campylobacter jejuni*. *J Proteome Res* **8**, 4654-4664, doi:10.1021/pr900544x (2009).
- 22 Zampronio, C. G., Blackwell, G., Penn, C. W. & Cooper, H. J. Novel glycosylation sites localized in *Campylobacter jejuni* flagellin FlaA by liquid chromatography electron capture dissociation tandem mass spectrometry. *J Proteome Res* **10**, 1238-1245, doi:10.1021/pr101021c (2011).
- 23 Stimson, E. *et al.* Meningococcal pilin: a glycoprotein substituted with digalactosyl 2,4-diacetamido-2,4,6-trideoxyhexose. *Mol Microbiol* **17**, 1201-1214, doi:10.1111/j.1365-2958.1995.mmi_17061201.x (1995).
- 24 Gault, J., Malosse, C., Dumenil, G. & Chamot-Rooke, J. A combined mass spectrometry strategy for complete posttranslational modification mapping of *Neisseria meningitidis* major pilin. *J Mass Spectrom* **48**, 1199-1206, doi:10.1002/jms.3262 (2013).
- 25 Schulz, B. L. *et al.* Identification of bacterial protein O-oligosaccharyltransferases and their glycoprotein substrates. *PLoS One* **8**, e62768, doi:10.1371/journal.pone.0062768 (2013).
- 26 Harding, C. M. *et al.* *Acinetobacter* strains carry two functional oligosaccharyltransferases, one devoted exclusively to type IV pilin, and the other one dedicated to O-glycosylation of multiple proteins. *Mol Microbiol* **96**, 1023-1041, doi:10.1111/mmi.12986 (2015).
- 27 Iwashkiw, J. A. *et al.* Identification of a general O-linked protein glycosylation system in *Acinetobacter baumannii* and its role in virulence and biofilm formation. *PLoS Pathog* **8**, e1002758, doi:10.1371/journal.ppat.1002758 (2012).
- 28 Power, P. M. *et al.* Genetic characterization of pilin glycosylation in *Neisseria meningitidis*. *Microbiology* **146** (Pt 4), 967-979, doi:10.1099/00221287-146-4-967 (2000).
- 29 Lithgow, K. V. *et al.* A general protein O-glycosylation system within the *Burkholderia cepacia* complex is involved in motility and virulence. *Mol Microbiol* **92**, 116-137, doi:10.1111/mmi.12540 (2014).
- 30 Egge-Jacobsen, W. *et al.* O-linked glycosylation of the PilA pilin protein of *Francisella tularensis*: identification of the endogenous protein-targeting oligosaccharyltransferase and characterization of the native oligosaccharide. *J Bacteriol* **193**, 5487-5497, doi:10.1128/JB.00383-11 (2011).

- 31 Castric, P. pilO, a gene required for glycosylation of Pseudomonas aeruginosa 1244 pilin. *Microbiology* **141** (Pt 5), 1247-1254, doi:10.1099/13500872-141-5-1247 (1995).
- 32 Elhenawy, W. *et al.* Protein O-linked glycosylation in the plant pathogen Ralstonia solanacearum. *Glycobiology* **26**, 301-311, doi:10.1093/glycob/cwv098 (2016).
- 33 Anonsen, J. H., Vik, A., Egge-Jacobsen, W. & Koomey, M. An extended spectrum of target proteins and modification sites in the general O-linked protein glycosylation system in Neisseria gonorrhoeae. *J Proteome Res* **11**, 5781-5793, doi:10.1021/pr300584x (2012).
- 34 Vik, A. *et al.* Broad spectrum O-linked protein glycosylation in the human pathogen Neisseria gonorrhoeae. *Proc Natl Acad Sci U S A* **106**, 4447-4452, doi:10.1073/pnas.0809504106 (2009).
- 35 Ahmad Izaham, A. R. *et al.* What Are We Missing by Using Hydrophilic Enrichment? Improving Bacterial Glycoproteome Coverage Using Total Proteome and FAIMS Analyses. *J Proteome Res*, **20**, 599-612, doi:10.1021/acs.jproteome.0c00565 (2021).
- 36 Darling, P., Chan, M., Cox, A. D. & Sokol, P. A. Siderophore production by cystic fibrosis isolates of Burkholderia cepacia. *Infect Immun* **66**, 874-877 (1998).
- 37 Romling, U. *et al.* Epidemiology of chronic Pseudomonas aeruginosa infections in cystic fibrosis. *J Infect Dis* **170**, 1616-1621, doi:10.1093/infdis/170.6.1616 (1994).
- 38 Aubert, D. F., Hamad, M. A. & Valvano, M. A. A markerless deletion method for genetic manipulation of Burkholderia cenocepacia and other multidrug-resistant gram-negative bacteria. *Methods Mol Biol* **1197**, 311-327, doi:10.1007/978-1-4939-1261-2_18 (2014).
- 39 Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods* **6**, 343-345, doi:10.1038/nmeth.1318 (2009).
- 40 Sambrook, J., E. F. Fritsch, and T. Maniatis. *Molecular cloning: a laboratory manual*. (Cold Spring Harbor Laboratory, Cold Spring Harbor, 1990).
- 41 Flannagan, R. S., Linn, T. & Valvano, M. A. A system for the construction of targeted unmarked gene deletions in the genus Burkholderia. *Environ Microbiol* **10**, 1652-1660, doi:10.1111/j.1462-2920.2008.01576.x (2008).
- 42 Scott, N. E. *et al.* Simultaneous Glycan-Peptide Characterization Using Hydrophilic Interaction Chromatography and Parallel Fragmentation by CID, Higher Energy Collisional Dissociation, and Electron Transfer Dissociation MS Applied to the N-Linked Glycoproteome of Campylobacter jejuni. *Mol Cell Proteomics* **10**, M000031MCP000201, doi:M000031-MCP201 [pii] 10.1074/mcp.M000031-MCP201 (2011).
- 43 Mysling, S., Palmisano, G., Hojrup, P. & Thaysen-Andersen, M. Utilizing ion-pairing hydrophilic interaction chromatography solid phase extraction for efficient glycopeptide enrichment in glycoproteomics. *Anal Chem* **82**, 5598-5609, doi:10.1021/ac100530w (2010).
- 44 Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature protocols* **2**, 1896-1906, doi:10.1038/nprot.2007.261 (2007).
- 45 Saba, J., Dutta, S., Hemenway, E. & Viner, R. Increasing the productivity of glycopeptides analysis by using higher-energy collision dissociation-accurate mass-product-dependent electron transfer dissociation. *Int J Proteomics* **2012**, 560391, doi:10.1155/2012/560391 (2012).

- 46 Caval, T., Zhu, J. & Heck, A. J. R. Simply Extending the Mass Range in Electron Transfer
Higher Energy Collisional Dissociation Increases Confidence in N-Glycopeptide
Identification. *Anal Chem* **91**, 10401-10406, doi:10.1021/acs.analchem.9b02125 (2019).
- 47 Bern, M., Kil, Y. J. & Becker, C. Byonic: advanced peptide and protein identification
software. *Curr Protoc Bioinformatics* **Chapter 13**, Unit13 20,
doi:10.1002/0471250953.bi1320s40 (2012).
- 48 Varga, J. J. *et al.* Draft Genome Sequences of Burkholderia cenocepacia ET12 Lineage
Strains K56-2 and BC7. *Genome Announc* **1**, doi:10.1128/genomeA.00841-13 (2013).
- 49 Carlier, A. *et al.* Genome Sequence of Burkholderia cenocepacia H111, a Cystic Fibrosis
Airway Isolate. *Genome Announc* **2**, doi:10.1128/genomeA.00298-14 (2014).
- 50 Lee, L. Y. *et al.* Toward Automated N-Glycopeptide Identification in Glycoproteomics.
Journal of proteome research **15**, 3904-3915, doi:10.1021/acs.jproteome.6b00438
(2016).
- 51 Lenco, J., Khalikova, M. A. & Svec, F. Dissolving Peptides in 0.1% Formic Acid Brings Risk
of Artificial Formylation. *Journal of proteome research* **19**, 993-999,
doi:10.1021/acs.jproteome.9b00823 (2020).
- 52 Chen, Y., Kwon, S. W., Kim, S. C. & Zhao, Y. Integrated approach for manual evaluation of
peptides identified by searching protein sequence databases with tandem mass spectra.
J Proteome Res **4**, 998-1005, doi:10.1021/pr049754t (2005).
- 53 Lu, L., Riley, N. M., Shortreed, M. R., Bertozzi, C. R. & Smith, L. M. O-Pair Search with
MetaMorpheus for O-glycopeptide characterization. *Nat Methods* **17**, 1133-1138,
doi:10.1038/s41592-020-00985-5 (2020).
- 54 Ishihama, Y., Rappsilber, J. & Mann, M. Modular stop and go extraction tips with stacked
disks for parallel and multidimensional Peptide fractionation in proteomics. *J Proteome
Res* **5**, 988-994, doi:10.1021/pr050385q (2006).
- 55 Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized
p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol*
26, 1367-1372, doi:10.1038/nbt.1511 (2008).
- 56 Winsor, G. L. *et al.* The Burkholderia Genome Database: facilitating flexible queries and
comparative analyses. *Bioinformatics* **24**, 2803-2804, doi:10.1093/bioinformatics/btn524
(2008).
- 57 Cox, J. *et al.* Accurate proteome-wide label-free quantification by delayed normalization
and maximal peptide ratio extraction, termed MaxLFQ. *Molecular & cellular proteomics :
MCP* **13**, 2513-2526, doi:10.1074/mcp.M113.031591 (2014).
- 58 Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of
(prote)omics data. *Nature methods* **13**, 731-740, doi:10.1038/nmeth.3901 (2016).
- 59 Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York,
2016).
- 60 Brademan, D. R., Riley, N. M., Kwiecien, N. W. & Coon, J. J. Interactive Peptide Spectral
Annotator: A Versatile Web-based Tool for Proteomic Applications. *Mol Cell Proteomics*
18, S193-S201, doi:10.1074/mcp.TIR118.001209 (2019).
- 61 Perez-Riverol, Y. *et al.* The PRIDE database and related tools and resources in 2019:
improving support for quantification data. *Nucleic Acids Res* **47**, D442-D450,
doi:10.1093/nar/gky1106 (2019).

979 62 Vizcaino, J. A. *et al.* 2016 update of the PRIDE database and its related tools. *Nucleic Acids*
980 *Res* **44**, D447-456, doi:10.1093/nar/gkv1145 (2016).

981 63 Davies, M. R. *et al.* Atlas of group A streptococcal vaccine candidates compiled using large-
982 scale comparative genomics. *Nat Genet* **51**, 1035-1043, doi:10.1038/s41588-019-0417-8
983 (2019).

984 64 Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q
985 File Manipulation. *PLoS One* **11**, e0163962, doi:10.1371/journal.pone.0163962 (2016).

986 65 Wagih, O. ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics*
987 **33**, 3645-3647, doi:10.1093/bioinformatics/btx469 (2017).

988 66 Scott, N. E. *et al.* Comparative proteomics and glycoproteomics reveal increased N-linked
989 glycosylation and relaxed sequon specificity in *Campylobacter jejuni* NCTC11168 O. *J*
990 *Proteome Res* **13**, 5136-5150, doi:10.1021/pr5005554 (2014).

991 67 Scott, N. E. *et al.* Modification of the *Campylobacter jejuni* N-linked glycan by EptC
992 protein-mediated addition of phosphoethanolamine. *J Biol Chem* **287**, 29384-29396,
993 doi:10.1074/jbc.M112.380212 (2012).

994 68 Fu, Y. & Qian, X. Transferred subgroup false discovery rate for rare post-translational
995 modifications detected by mass spectrometry. *Mol Cell Proteomics* **13**, 1359-1368,
996 doi:10.1074/mcp.O113.030189 (2014).

997 69 Krug, K. *et al.* Deep coverage of the *Escherichia coli* proteome enables the assessment of
998 false discovery rates in simple proteogenomic experiments. *Mol Cell Proteomics* **12**, 3420-
999 3430, doi:10.1074/mcp.M113.029165 (2013).

1000 70 Coyne, M. J. *et al.* Phylum-wide general protein O-glycosylation system of the
1001 Bacteroidetes. *Mol Microbiol* **88**, 772-783, doi:10.1111/mmi.12220 (2013).

1002 71 Nothaft, H. *et al.* Diversity in the protein N-glycosylation pathways within the
1003 *Campylobacter* genus. *Mol Cell Proteomics* **11**, 1203-1219,
1004 doi:10.1074/mcp.M112.021519 (2012).

1005 72 Willcocks, S. J. *et al.* Virulence of the emerging pathogen, *Burkholderia pseudomallei*,
1006 depends upon the O-linked oligosaccharyltransferase, PglL. *Future Microbiol* **15**, 241-257,
1007 doi:10.2217/fmb-2019-0165 (2020).

1008 73 Pan, C. *et al.* Biosynthesis of Conjugate Vaccines Using an O-Linked Glycosylation System.
1009 *mBio* **7**, e00443-00416, doi:10.1128/mBio.00443-16 (2016).

1010 74 Gebhart, C. *et al.* Characterization of exogenous bacterial oligosaccharyltransferases in
1011 *Escherichia coli* reveals the potential for O-linked protein glycosylation in *Vibrio cholerae*
1012 and *Burkholderia thailandensis*. *Glycobiology* **22**, 962-974, doi:10.1093/glycob/cws059
1013 (2012).

1014 75 Faridmoayer, A. *et al.* Extreme substrate promiscuity of the *Neisseria* oligosaccharyl
1015 transferase involved in protein O-glycosylation. *J Biol Chem* **283**, 34596-34604,
1016 doi:10.1074/jbc.M807113200 (2008).

1017 76 Faridmoayer, A., Fentabil, M. A., Mills, D. C., Klassen, J. S. & Feldman, M. F. Functional
1018 characterization of bacterial oligosaccharyltransferases involved in O-linked protein
1019 glycosylation. *J Bacteriol* **189**, 8088-8098, doi:10.1128/JB.01318-07 (2007).

1020 77 Shropshire, H., Jones, R. A., Aguilo-Ferretjans, M. M., Scanlan, D. J. & Chen, Y. Proteomics
1021 insights into the *Burkholderia cenocepacia* phosphorus stress response. *Environ*
1022 *Microbiol*, doi:10.1111/1462-2920.15451 (2021).

1023 78 Inhulsen, S. *et al.* Identification of functions linking quorum sensing with biofilm formation
1024 in *Burkholderia cenocepacia* H111. *Microbiologyopen* **1**, 225-242, doi:10.1002/mbo3.24
1025 (2012).
1026 79 Riedel, K. *et al.* Analysis of the quorum-sensing regulon of the opportunistic pathogen
1027 *Burkholderia cepacia* H111 by proteomics. *Electrophoresis* **24**, 740-750,
1028 doi:10.1002/elps.200390089 (2003).
1029 80 Figurski, D. H. & Helinski, D. R. Replication of an origin-containing derivative of plasmid
1030 RK2 dependent on a plasmid function provided in trans. *Proc Natl Acad Sci U S A* **76**, 1648-
1031 1652, doi:10.1073/pnas.76.4.1648 (1979).
1032 81 Hamad, M. A., Di Lorenzo, F., Molinaro, A. & Valvano, M. A. Aminoarabinose is essential
1033 for lipopolysaccharide export and intrinsic antimicrobial peptide resistance in
1034 *Burkholderia cenocepacia*(dagger). *Mol Microbiol* **85**, 962-974, doi:10.1111/j.1365-
1035 2958.2012.08154.x (2012).
1036 82 Cardona, S. T. & Valvano, M. A. An expression vector containing a rhamnose-inducible
1037 promoter provides tightly regulated gene expression in *Burkholderia cenocepacia*.
1038 *Plasmid* **54**, 219-228, doi:10.1016/j.plasmid.2005.03.004 (2005).
1039

A)

K56-2 Glycopeptides

H111 Glycopeptides

Figure legend

B)

Glycoproteins overlap

Figure legend

C)

BCAL2466 - Ecotin

BCAL2345 - SecG

D)

A)

B)

C)

A)

B)

C)

B) **Conservation of Glycosylation sites (+/-10 amino acids)**

C)

chemistry ■ Acidic ■ Basic ■ Hydrophobic ■ Neutral ■ Polar

A)

Class I localised glycosylation sites

Strains

B)

B. ubonensis MSMB22

C)

B. humptydooensis MSMB43

D)

Glycosylation across homologues

Strains

E)

Enrichment of shared glycoproteins