

## *Burkholderia* PglL enzymes are Serine preferring oligosaccharyltransferases which target conserved proteins across the *Burkholderia* genus

Andrew J. Hayes<sup>1</sup>, Jessica M. Lewis<sup>1</sup>, Mark R. Davies <sup>1</sup> & Nichollas E. Scott <sup>1</sup>✉

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 systems is far more limited. Within this work we explore the conservation of glycoproteins and glycosylation sites across the pan-*Burkholderia* glycoproteome. Using a multi-protease glycoproteomic approach, we generate high-confidence glycoproteomes 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 *B. cenocepacia* isolates. This preference for glycosylation at Serine residues is observed across at least 9 *Burkholderia* glycoproteomes, supporting that Serine is the dominant residue targeted by PglL-mediated glycosylation across the *Burkholderia* genus. Combined, this work demonstrates that PglL enzymes of the *Burkholderia* genus are Serine-preferring oligosaccharyltransferases that target conserved and shared protein substrates.

<sup>1</sup>Department of Microbiology and Immunology, University of Melbourne at the Peter Doherty Institute for Infection and Immunity, Melbourne, Australia.  
✉email: [Nichollas.scott@unimelb.edu.au](mailto:Nichollas.scott@unimelb.edu.au)

**G**lycosylation is a common class of protein modifications increasingly recognised within bacterial proteomes<sup>1,2</sup>. Within bacterial proteomes, the attachment of glycans has been shown to influence protein functions by enhancing protein stability<sup>3–5</sup>, stabilising protein complexes<sup>6</sup>, masking antigenic sites<sup>7</sup>, and modulating protein enzymatic activities<sup>8,9</sup>. 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 with many of these now demonstrated to be conserved across genera<sup>1,2,10,11</sup>. Within these systems follow up studies have mostly focused on understanding the glycans used for protein glycosylation, defining their diversity<sup>12–14</sup> and elucidating their biosynthetic pathways<sup>15–17</sup>. This focus has resulted in a limited understanding of the specific glycoproteins and glycosylation sites within the majority of known bacterial glycoproteomes.

One of the best characterised bacterial glycoproteomes to date is that of the *N*-linked glycosylation system of *Campylobacter jejuni*, where 134 glycosylation sites have been experimentally identified<sup>18</sup>. 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 the glycosylation sequon, D/E-X<sub>1</sub>-N-X<sub>2</sub>-S/T (where neither X<sub>1</sub> or X<sub>2</sub> can be proline)<sup>19</sup>, which restricts the possible glycosylation sites within a protein, and ii) the chemical nature of *N*-linked glycosylation that can allow localisation information to be obtained using collision-based Mass Spectrometry (MS) fragmentation approaches<sup>20</sup>. For bacterial mediated *O*-linked glycosylation systems, far less is known about the sites of glycosylation<sup>1,2</sup>. 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 sites<sup>20</sup>. Defining 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<sup>21,22</sup>, 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 of conserved antigenic regions, conversely, forms of pilin prone to amino acid sequence variation possess limited pilin glycosylation<sup>7,17,23,24</sup>.

One of the most widespread families of *O*-linked glycosylating enzymes are the PglL oligosaccharyltransferases<sup>25</sup>, which have been experimentally shown to *O*-glycosylate proteins within *Acinetobacter*<sup>26,27</sup>, *Neisseria*<sup>14,17,28</sup>, *Burkholderia*<sup>13,15,29</sup>, *Francisella*<sup>30</sup>, *Pseudomonas*<sup>31</sup>, and *Ralstonia*<sup>32</sup> 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<sup>10,11</sup>. To date, no rigid glycosylation sequon has been observed within PglL substrates, with glycosylation predominantly occurring in disordered regions rich in Alanine and Prolines<sup>12,29,33,34</sup>. For PglL enzymes responsible for the glycosylation of multiple proteins, known as general glycosylation systems, the abolishment of PglL has been shown to result in profound effects on virulence<sup>27,29,32</sup>. Despite this, the lack of a detailed understanding of most *O*-linked glycoproteomes has limited our ability to understand how glycosylation influences virulence as well as which glycosylation sites are critical for

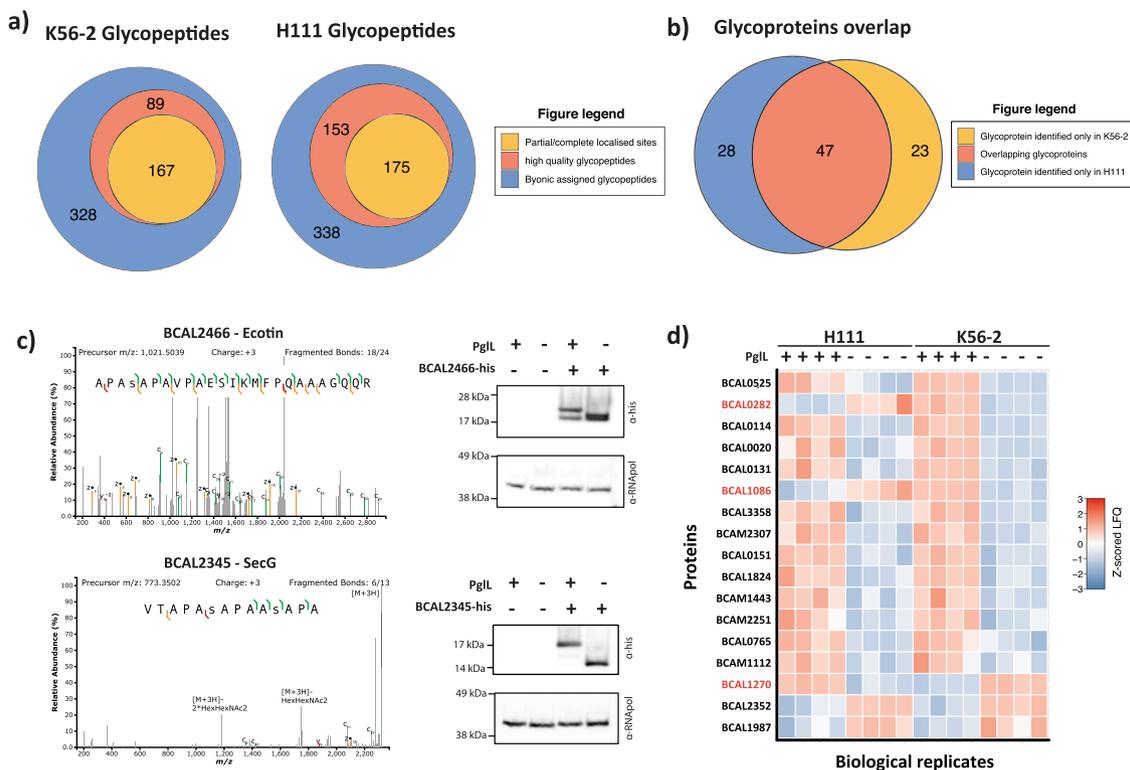
virulence-associated processes. Understanding the proteins targeted for glycosylation thus could provide an important first step to improve our understanding of virulence within poorly understood pathogens such as *Burkholderia cenocepacia*.

The opportunistic pathogen *B. cenocepacia* is associated with life-threatening infections in people with cystic fibrosis (CF)<sup>35</sup>. While ubiquitous in the environment not all *B. cenocepacia* strains are associated with human disease with the phylogenetic grouping known as the IIIA genomovar overrepresented in CF infections<sup>36–38</sup>. Within the IIIA genomovar, the K56-2<sup>39</sup> and H111<sup>40</sup> strains have been established as the most widely used models of *B. cenocepacia* pathogenesis. While both strains were isolated from CF patients, only K56-2 is a member of the ET12 lineage, a highly transmissible lineage associated with high mortality rates<sup>41,42</sup>. The absence of key genetic elements in H111, such as the low-oxygen-activated locus (*lxa*)<sup>43</sup> and the *cciIR* quorum-sensing system<sup>44,45</sup>, has been suggested to account for the notable differences in virulence traits<sup>46</sup> and quorum-sensing<sup>47,48</sup> observed between K56-2 and H111. This genetic diversity between IIIA genomovar isolates makes understanding multiple strains essential for identifying core features shared across the majority of pathogenic *B. cenocepacia* isolates.

Recently we began exploring the pan-glycoproteome of *Burkholderia* species focusing on the conservation of the glycans used for glycosylation<sup>13,15</sup> and the properties of glycopeptides that can influence their enrichment using zwitterionic hydrophilic interaction liquid chromatography (ZIC-HILIC)<sup>49</sup>. These studies have highlighted that the glycoproteomes of *Burkholderia* species appear larger than initially thought, yet, the lack of defined glycosylation sites associated with these proteins 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 *Burkholderia cenocepacia* within K56-2 and H111. Leveraging this curated resource, we gain and experimentally confirm a previously unrecognised preference in the specificity of *O*-linked glycosylation, as well as demonstrate that *Burkholderia* glycosylation targets conserved protein substrates across this genus. Combined, this work demonstrates the high levels of *O*-linked glycosylation conservation across the *Burkholderia* genus.

## 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 physicochemical properties of peptides heavily influences the ability of ZIC-HILIC enrichment to isolate bacterial glycopeptides<sup>49</sup>. This observation suggests numerous glycoproteins have likely been overlooked when previous studies used Trypsin alone to assess the glycoproteome of *B. cenocepacia*<sup>13,29</sup>. To increase the coverage of the *B. cenocepacia* glycoproteome we undertook glycoproteomic analysis using multiple proteases<sup>50,51</sup> 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 respectively (Fig. 1a, Supplementary Data 1 and 2). Although these glycopeptide datasets identified more glycoproteins than identified within previously published studies using the ET12 *B. cenocepacia* strain J2315 (Supplementary Fig. 1a<sup>13,29</sup>), the majority of identified glycoproteins were unique to a single strain (Supplementary Fig. 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



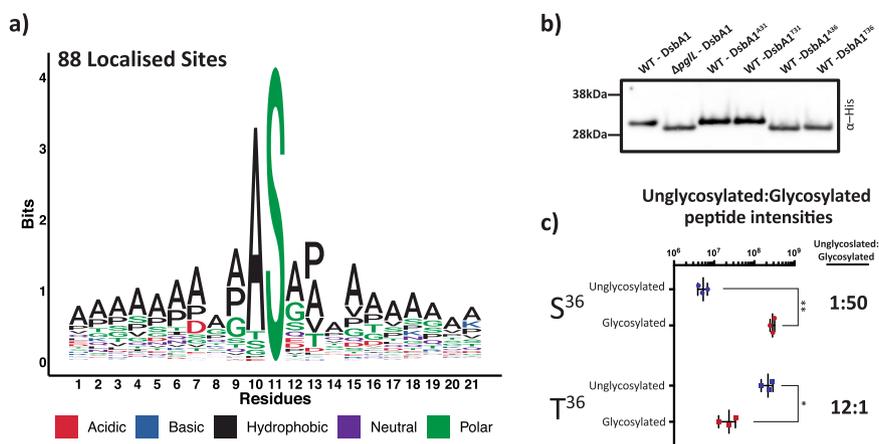
**Fig. 1 B. cenocepacia glycoproteomic analysis reveals similar substrates and functional impacts on B. cenocepacia strains.** **a** Within *B. cenocepacia* strains K56-2 and H111 584 and 666 unique glycopeptides were identified respectively (Byonic scores >300). Within these glycopeptides manual curation 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, His<sub>6</sub> tagged BCAL2466 and BCAL2345, reveals alterations in protein banding when expressed within the glycosylation null strain K56-2  $\Delta$ pglL compared to K56-2 WT with anti-RNA pol westerns included as loading controls. **d** LFQ analysis of the most altered proteins observed between H111  $\Delta$ pglL compared to H111 WT reveals similar proteome alterations as seen previously within K56-2<sup>5</sup>.

of modified peptides within large database searches<sup>52</sup>. Consistent with this hypothesis, examination of the glycopeptide scores reveal a bias toward lower scores (Supplementary Fig. 2a, b) supporting a higher than ideal false discovery rate despite stringent filtering<sup>53</sup>. 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 respectively (Fig. 1a; Supplementary Data 3). Of these unique glycopeptides, 167 and 175 glycopeptides within K56-2 and H111 respectively provided partial or complete site localisation (Fig. 1a; Supplementary Data 3). Consistent with improving the data quality, the curated glycopeptides revealed a Gaussian distribution within the scores, removing predominantly low scoring identifications (Supplementary Fig. 2c–f). Comparing the observed score distributions for each proteases revealed that for both H111 and K56-2 Thermolysin and Pepsin glycopeptides typically possessed lower and narrower score distributions, yet still enabled the localisation of glycosylation sites at a similar frequency as Trypsin (>50% of unique glycopeptides localised, Supplementary Fig. 3a–d). Of these manually assessed glycopeptides, we noted ~65% of glycoproteins were identified within both K56-2 and H111 strains (Fig. 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 His<sub>6</sub> epitope tagged and expressed within K56-2 WT and K56-2  $\Delta$ pglL strains (Fig. 1c). Consistent with a single glycosylation event on BCAL2466 and multiple glycosylation events on BCAL2345, the expression of these proteins within K56-2  $\Delta$ pglL resulted in increased gel mobility supporting their glycosylation

status. Within the wildtype strains, differences in the occupancy of these proteins were also noted, with no non-glycosylated protein observed for BCAL2345, contrasting BCAL2466 where both glycosylated and non-glycosylated forms were detected (Fig. 1c). Interestingly it should be noted that a subtle difference in the mobility of the loading control RpoA was also noted between K56-2 WT and K56-2  $\Delta$ pglL strains yet the cause of this is unknown (Fig. 1c). To functionally support the observed similarities within the glycoproteomes of *B. cenocepacia* strains, we generated two independent H111  $\Delta$ pglL mutants and assessed the proteome changes observed in the absence of glycosylation within H111, compared to our previously published K56-2 findings (Fig. 1d<sup>5</sup>). Although the proteomic alterations show congruent behaviour, it should be noted strain differences are observed within some proteins such as BCAL1086, BCAL0282, and BCAL1270 (Fig. 1d; Supplementary Data 4). Regardless of these differences between K56-2 and H111, the proteins observed to undergo statistically significant alterations within H111  $\Delta$ pglL were enriched for proteins observed to be altered within K56-2  $\Delta$ pglL (Fisher exact test p-value = 0.0006 Supplementary Data 5). Independent H111  $\Delta$ pglL mutants also showed similar alterations with the proteome and were highly similar (Fisher exact test p-value =  $5.67 \times 10^{-15}$ , Supplementary Data 4 and 5). 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



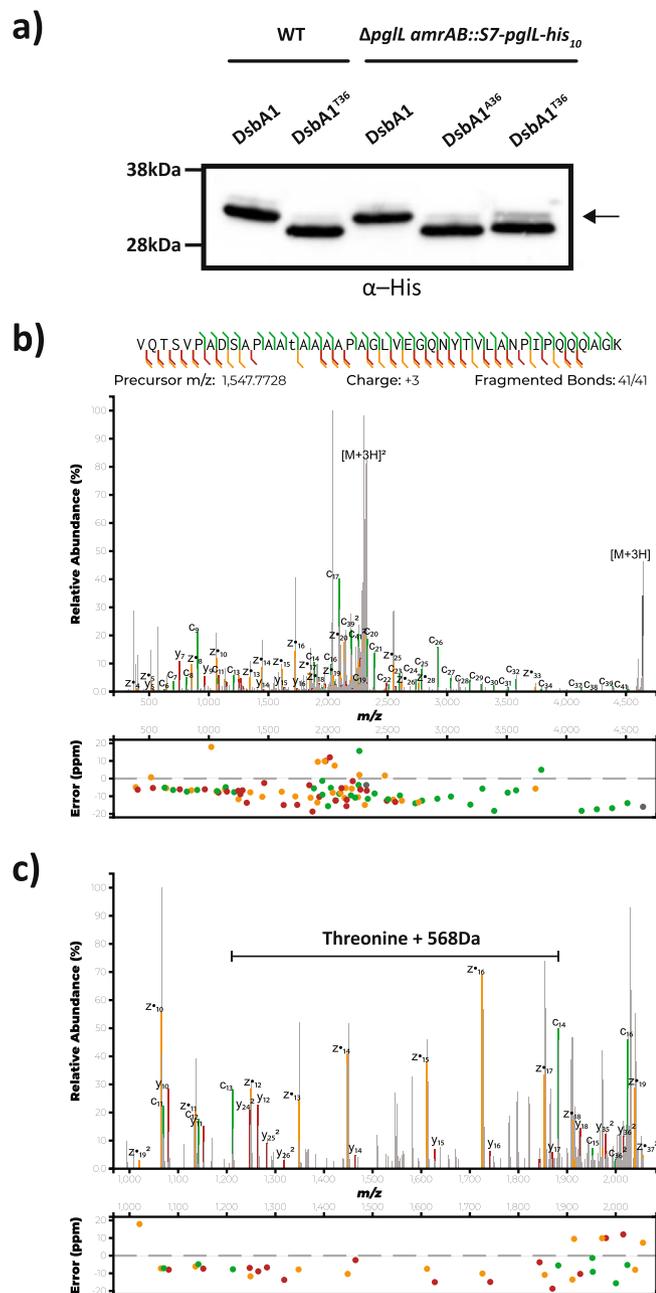
**Fig. 2 O-linked glycosylation predominantly occurs on Serine residues across the *B. cenocepacia* glycoproteome.** **a** Sequence analysis of localised glycosylation sites across *B. cenocepacia* strains reveals all assigned sites are Serine. **b** Western analysis of DsbA1<sub>Nm</sub>-his<sub>6</sub> variants reveals substitution of only S<sup>36</sup> to Alanine or Threonine results in reduced glycosylation. **c** Relative amounts of the DsbA1<sub>Nm</sub>-his<sub>6</sub> glycosylated and non-glycosylated peptides <sup>23</sup>VQTSVPADSAPAASAAAAPAGLVEGQNYTVLANPIPQQQAGK<sup>64</sup> and <sup>23</sup>VQTSVPADSAPAATAAAAAPAGLVEGQNYTVLANPIPQQQAGK<sup>64</sup> observed within K56-2 WT reveals the alteration of S<sup>36</sup> to T<sup>36</sup> dramatically reduces glycosylation. Glycosylation site within peptides denoted by underlining. Differences in the abundance of glycosylated and non-glycosylated peptides containing S<sup>36</sup> to T<sup>36</sup> correspond to (*N* = 3) biologically independent samples assessed using two-sided t-tests resulting in *p*-values of 0.0054 and 0.036 respectively.

glycoproteomic data revealed >50% of unique glycopeptides provided at least partial site localisation information (Fig. 1a, Supplementary Fig. 3, Supplementary Data 4) of which 88 sites could be precisely localised across 70 glycoproteins (Supplementary Data 6). Surprisingly this analysis suggested a single Threonine residue, T<sup>159</sup> within BCAM0996, was modified within H111. Due to the discordance of this assignment with respect to all other sites (Supplementary Fig. 4), we sought to confirm the accuracy of BCAM0996 T<sup>159</sup>. Examination of glycopeptides from BCAM0996 revealed that multiple Serine residues are modified within the same peptide observed to be modified at T<sup>159</sup> (Supplementary Data 3). 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 and a secondary modification within the peptide sequence, ultimately revealing S<sup>167</sup> to be the correct localisation site (Supplementary Fig. 5). This finding suggested all localised glycosylation sites within both *B. cenocepacia* strains were observed on Serine residues. Examination of these 88 sites demonstrates that glycosylation favoured Alanine at the -1 position, yet this was not a strict requirement (Fig. 2a). To independently validate this preference for Serine, we re-analysed these glycoproteomic datasets with O-Pair, a glycosylation site localisation focused glycoproteomic tool<sup>54</sup>, supporting that the majority of high-quality localisable glycosylation events occurred on Serine residues (Supplementary Fig. 6; Supplementary Data 7). 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<sub>Nm</sub>-his<sub>6</sub> that we have previously observed is predominantly glycosylated by a single glycosylation event when expressed in *B. cenocepacia*<sup>5,29</sup>. Past work showed that glycosylation occurs at an unknown site within the peptide <sup>23</sup>VQTSVPADSAPAASAAAAPAGLVEGQNYTVLANPIPQQQAGK<sup>64</sup>, suggesting that one of the three Serine residues (S<sup>26</sup>, S<sup>31</sup> and S<sup>36</sup>) was the site of glycosylation within DsbA1<sub>Nm</sub>-his<sub>6</sub>. Of these sites, S<sup>31</sup> and S<sup>36</sup> are flanked by Alanine residues, consistent

with the preferred glycosylation sites within native substrates (Fig. 2a), yet only substitution of S<sup>36</sup> with either Alanine or Threonine resulted in gel mobility shifts consistent with the loss of glycosylation (Fig. 2b). Curiously DsbA1<sub>Nm</sub>-his<sub>6</sub> A<sup>36</sup> leads to a slight but reproducible increased mobility compared to other point mutants yet our proteomic and sequencing result support the correctness of this construct, and as such the cause of this shift remains unknown. Glycopeptide analysis further confirmed S<sup>36</sup> was the sole residue modified within this peptide (Supplementary Fig. 7 and Supplementary Data 8). As MS analysis provides greater dynamic range than western blotting, we investigated if glycosylation still occurred at T<sup>36</sup> with analysis supporting the presence of a glycosylated form of the <sup>23</sup>VQTSVPADSAPAATAAAAAPAGLVEGQNYTVLANPIPQQQAGK<sup>64</sup> at <10% of the abundance of the unmodified form (Fig. 2c, Supplementary Fig. 8a, Student's *t* test *p*-values = 0.036). Targeted MS analysis supported the identity of this glycopeptide, yet could not provide definitive site localisation to T<sup>36</sup> (Supplementary Fig. 8b). To further assess if Threonine could be modified, we introduced the DsbA1<sub>Nm</sub>-his<sub>6</sub> point mutants T<sup>36</sup> and A<sup>36</sup> into K56-2  $\Delta$ pglL *amrAB::S7-pglL-his10*, an overexpressing PglL<sub>BC</sub> strain<sup>5</sup>, to assess if increasing PglL<sub>BC</sub> levels enhanced Threonine glycosylation. Western analysis demonstrates the majority of DsbA1<sub>Nm</sub>-his<sub>6</sub> T<sup>36</sup> remained unglycosylated within K56-2  $\Delta$ pglL *amrAB::S7-pglL-his10*. Yet, in contrast to K56-2 WT an additional faint band is observable within DsbA1<sub>Nm</sub>-his<sub>6</sub> T<sup>36</sup> (Fig. 3a), consistent with the presence of low abundance glycosylated DsbA1<sub>Nm</sub>-his<sub>6</sub> T<sup>36</sup>. Targeted MS analysis confirms the glycosylation of T<sup>36</sup> in DsbA1<sub>Nm</sub>-his<sub>6</sub> T<sup>36</sup> when expressed in K56-2  $\Delta$ pglL *amrAB::S7-pglL-his10* (Fig. 3b, 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 PglL<sub>BC</sub> is overexpressed.

**Glycoproteins and 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 Data 6) we assessed the conservation of these 70



**Fig. 3** Threonine residues undergo poor glycosylation in *B. cenocepacia* even with overexpression of *pgll<sub>BC</sub>*. **a** Western analysis of DsbA1<sub>Nm</sub>-his<sub>6</sub> variants reveals a minor product, indicated with an arrow, is observed within DsbA1<sub>Nm</sub>-his<sub>6</sub> T<sup>36</sup> when expressed within K56-2  $\Delta$ *pgll amrAB::S7-pgll-his<sub>10</sub>*. **b, c** Targeted MS analysis using EThcD fragmentation enabled the confirmation of glycosylation at residue T<sup>36</sup> within <sup>23</sup>VQTSVPADSAPAATAAAAAPAGLVEGQNYTVLANPIPIQQQAGK<sup>64</sup>. Glycosylation site within peptide denoted by underlining.

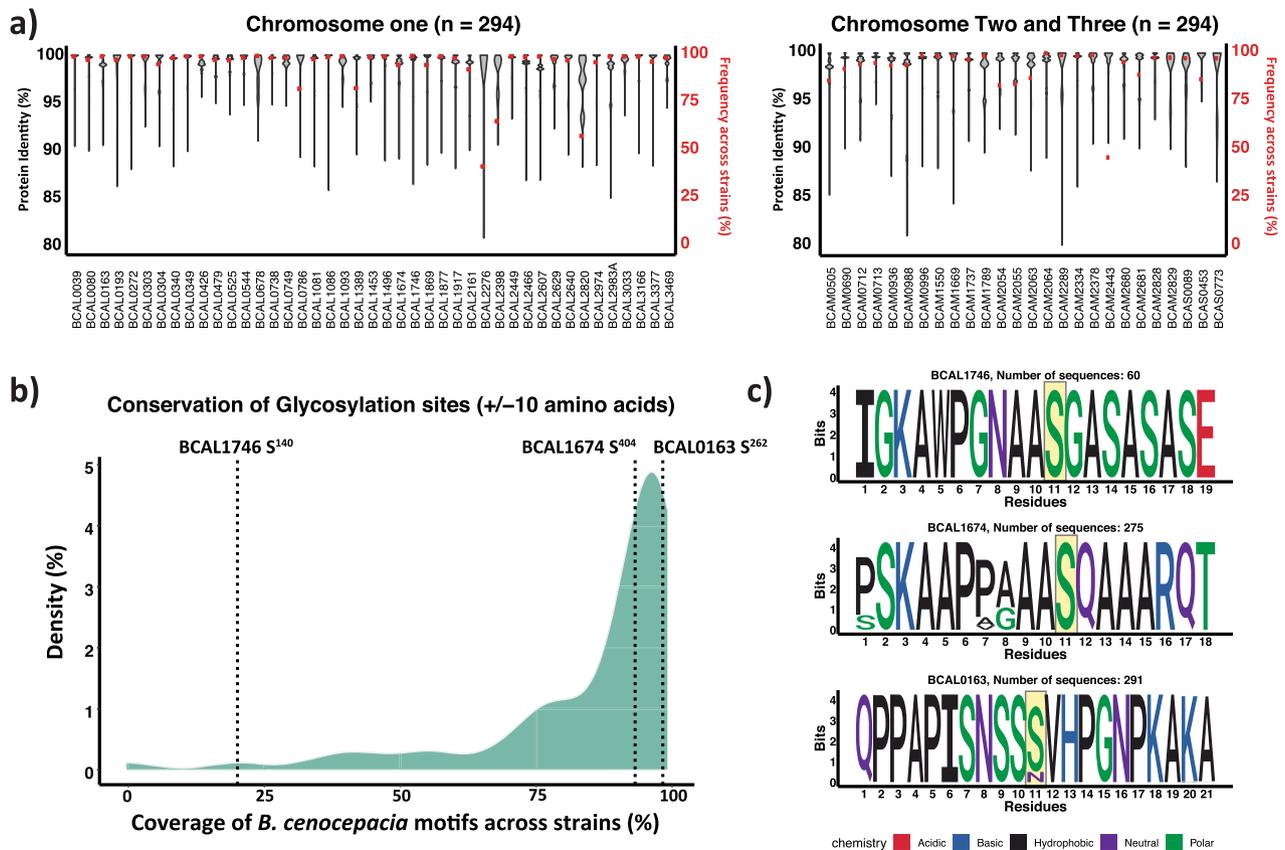
glycoproteins across 294 publicly available *B. cenocepacia* genomes<sup>55</sup>. Across strains, the frequency of observation, defined as the presence of proteins with a minimum blast identity of >80%, revealed 65 glycoproteins are present in >80% of available genomes and that these glycoproteins are highly similar, with an average amino acid sequence identity >95% (Fig. 4a, Supplementary Data 9). This high level of conservation is also reflected at the glycosylation site ( $\pm 10$  amino acids) where the majority of glycosylation sites are highly conserved at the sequence level (Fig. 4b, Supplementary Data 9). It should be noted that although

the majority of glycosylation sites are conserved (Supplementary Figs. 9–14), variations are observed; such as in S<sup>140</sup> of BCAL1746, which is present in only a fraction of strains (Fig. 4c); S<sup>404</sup> of BCAL1674, where multiple alterations within the flanking sequences were noted (Fig. 4e); and S<sup>262</sup> of BCAL0163, where the serine required for glycosylation is lost across multiple strains (Fig. 4f). These data support 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 PglL across *Burkholderia* species (Supplementary Fig. 15) we sought to assess if the preference for Serine glycosylation was a common feature across *Burkholderia* glycoproteomes. To investigate this, we utilised 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* MSMB2199<sup>13</sup>) 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 (Fig. 5a; Supplementary Data 10). To experimentally support this preference, we introduced DsbA1<sub>Nm</sub>-his<sub>6</sub> 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<sub>Nm</sub>-his<sub>6</sub> T<sup>36</sup> is predominantly unable to be glycosylated within both *B. humptydooensis* MSMB43 and *B. ubonensis* MSMB22 (Fig. 5b, c). As with *B. cenocepacia*, the glycosylated form of the peptide <sup>23</sup>VQTSVPADSAPAATAAAAAPAGLVEGQNYTVLANPIPIQQQAGK<sup>64</sup> was observed at low levels within both *B. humptydooensis* MSMB43 and *B. ubonensis* MSMB22 (Supplementary Fig. 16; Supplementary Data 11). 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* (Fig. 5d, Supplementary Data 12). 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 (Fig. 5e Supplementary Data 12). Combined these results support that Serine is the preferred target of glycosylation across *Burkholderia* glycoproteomes and that similar glycoproteins are targeted across the *Burkholderia* genus.

## Discussion

The conservation of glycosylation systems across bacterial species is increasingly being recognised as a common phenomenon in bacterial genera<sup>14,56,57</sup>. 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 (Figs. 1b, 5d) and bioinformatically (Fig. 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 (Figs. 2a, 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



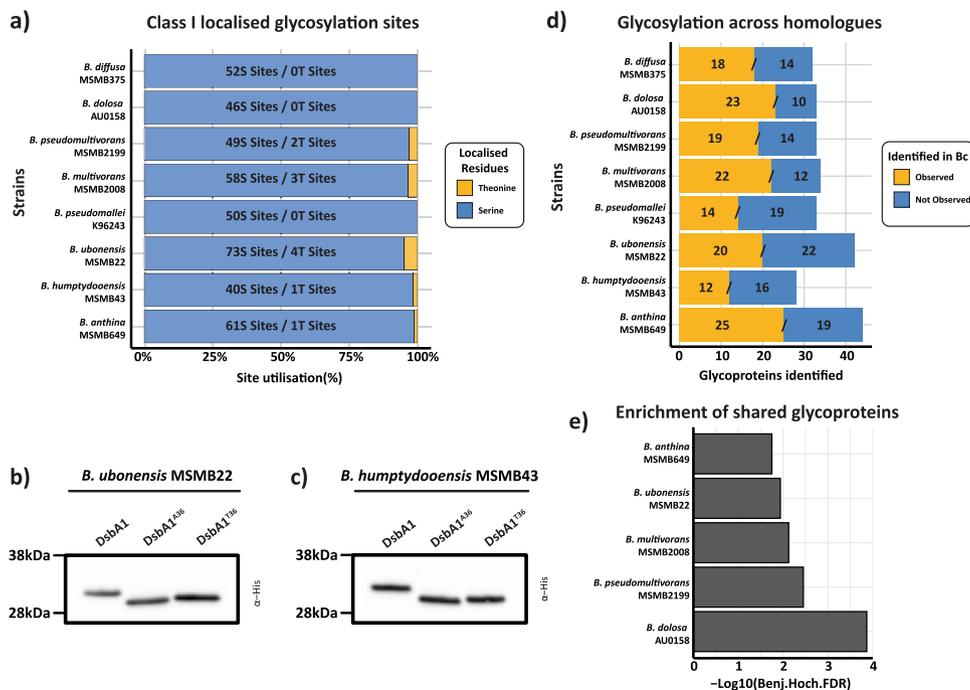
**Fig. 4 Glycoproteins and glycosylation sites are conserved across *B. cenocepacia*.** **a** Violin plot showing the amino acid sequence variation and frequency of 69 glycoproteins across 294 *B. cenocepacia* genomes. Relative frequency is shown in red (right axis) and protein sequence identity (>80% 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 S<sup>140</sup>; BCAL1674 S<sup>404</sup> and BCAL0163 S<sup>262</sup>.

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<sup>5,58</sup>. 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* PgL, WPAAASAG<sup>59</sup>, our proteomic analysis supports that Serine residues flanked by Alanine are predominantly targeted for glycosylation within *B. cenocepacia*, yet glycosylation can still occur within divergent sequences lacking consensus residues (Fig. 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 PgL enzymes to glycosylate proteins<sup>60–62</sup>, 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* (Fig. 2b, c), with a similar trend also noted within *B. humptydoensis* MSMB43 and *B. ubonensis* MSMB22 (Supplementary Fig. 16). 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 PgL glycosylation systems where only a limited number of glycosylation sites have been confirmed to date, such as in *N. gonorrhoeae*<sup>33,34</sup>, *A. baumannii*<sup>12</sup> and *F. tularensis*<sup>30</sup>, all confirmed glycosylation sites appear to occur on Serine residues. These trends suggest that Serine specificity may be a more general feature of the PgL oligosaccharyltransferases, yet further work is required to confirm this preference across known PgL enzymes. Although Serines are the preferred glycosylated residue across *Burkholderia* glycoproteins, it is unclear if all glycosylated Serine's are modified to a high level of occupancy. Our western blot analysis of His<sub>6</sub> tagged BCAL2466 and BCAL2345 (Fig. 1c) supports that differences in glycosylation occupancy do exist, yet whether specific sequence or protein characteristics predict this efficiency are unclear. Together, this suggests further studies are required to understand the properties which promote occupancy at specific sites.

At the genus level our analysis supports that across *Burkholderia* species glycoproteins are conserved (Fig. 5d and Supplementary Fig. 17). 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<sup>7,17,23,24</sup> has suggested that glycosylation sites may be 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



**Fig. 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, c** Western analysis of DsbA1<sub>Nm-his<sub>6</sub></sub> variants expressed within *B. ubonensis* MSMB22 (**b**) and *B. humptydoensis* MSMB43 (**c**) supports that Threonine is disfavoured for glycosylation similar to *B. cenocepacia*. **d** Analysis of glycoproteins observed across *Burkholderia* species supports multiple *B. cenocepacia* glycoprotein homologues are observed glycosylated within other species. **e** Enrichment analysis of glycoproteins shows an enrichment of shared glycoproteins in *B. anthina* MSMB649; *B. ubonensis* MSMB22; *B. multivorans* MSMB2008; *B. pseudomultivorans* MSMB2199 and *B. dolosa* AU0158 compared to *B. cenocepacia*.

subset of genomes, such as S<sup>140</sup> of BCAL1746, yet this level of variability was the exception, not the rule, across *B. cenocepacia*. Although highly conserved, it is important to note at the glycoproteome level we observed only 47 of the 98 glycoproteins identified within both strains (Fig. 1b). Examination of the proteomes of H111 and K56-2 supports that this modest overlap is not due to the lack of expression of these glycoproteins, as at least 66 of these glycoproteins are expressed in both strains (Supplementary Fig. 18). Rather, this overlap supports that the differences in the observable glycoproteome are driven by either under-sampling of the glycoproteome due to the low abundance of these glycoproteins, or differences in glycosylation occupancy between strains. Nonetheless our glycoproteome analysis of *B. cenocepacia* significantly expands the known number of glycosylation sites from six sites in 23 glycoproteins<sup>29</sup> 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, as it is well known that growth under different conditions, such as nutrient limitation<sup>63</sup> and or in the presence of chemical queues<sup>64,65</sup> 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 the breadth of the *B. cenocepacia* glycoproteome and the general features of glycosylation across members of the *Burkholderia* genus. The identification that the *B. cenocepacia* glycoproteome is far larger than initially thought, containing at least 98 proteins between strains, supports that glycosylation plays a multifaceted and pleiotropic role within *Burkholderia* species. The curation of a high-quality list of known glycoproteins and sites provides a

unique resource to facilitate studies that work towards understanding the roles of glycosylation within glycoproteins. From a mechanistic and technical standpoint, the insights into the preference for glycosylation at Serine residues improves our ability to predict the specific sites likely to be modified in glycoproteins, as well as improves our ability to assign glycosylation within proteomic datasets. Finally, the demonstration that glycoproteins are highly conserved across *Burkholderia* species also provides a new opportunity to use comparative glycoproteomics to dissect the conserved roles of glycoproteins across this genus. Combined these insights will aid in future studies to understand why glycosylation events are common and have been maintained across *Burkholderia* species.

## Methods

**Bacterial strains and growth conditions.** Strains and plasmids used in this study are listed in Supplementary 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<sup>66</sup>. 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 Supplementary Table 3. The inducible pSCrhaB2-BCAL2345-his<sub>6</sub> and pSCrhaB2-BCAL2466-his<sub>6</sub> constructs were generated using Gibson assembly<sup>67</sup> by inserting PCR amplified fragments into *Nde*I and *Xba*I linearised pSCrhaB2 using NEB-uilider® 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<sup>68</sup>. All restriction endonuclease digestions, and agarose gel electrophoresis were performed using standard molecular

biology techniques<sup>68</sup>. 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<sup>68</sup>. 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 (Zymo research), while PCR recoveries and restriction digest purifications were performed using DNA Clean & Concentrator Kits (Zymo 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  $\Delta$ pgl deletion mutants.** Deletions of H111 *pglL* (Gene accession: I35\_RS13570) were undertaken using the approach of Flanagan et al. for the construction of unmarked, non-polar deletions in *B. cenocepacia*<sup>69</sup> using the plasmid pYM8<sup>15</sup>.

**Protein Immunoblotting.** Bacterial whole-cell lysates were prepared from overnight LB cultures of *Burkholderia* strains. 1 ml of overnight cultures at an OD<sub>600</sub> 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  $\beta$ -mercaptoethanol and 0.001% (w/v) of bromophenol blue (pH 6.8)] and heated for 10 min 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 h 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 h 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 described<sup>5</sup>. 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  $\times$  g at 4 °C then snap frozen. Frozen whole-cell samples were resuspended in 4% SDS, 100 mM Tris pH 8.0, 20 mM Dithiothreitol (DTT) and boiled at 95 °C with shaking for 10 min. Samples were then clarified by centrifugation at 17,000  $\times$  g for 10 min, the supernatant collected, and protein concentration determined by bicinchoninic acid assays (Thermo Fisher Scientific). For glycoproteomic analysis, 1 mg of protein for each sample (three biological replicates per strain/enzyme) was acetone precipitated by mixing 4 volumes of ice-cold acetone with one volume of sample. For quantitative proteomic comparisons of H111 strains, 200  $\mu$ g 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  $\times$  g for 10 min at 0 °C. The precipitated protein pellets were resuspended in 80% ice-cold acetone and precipitated again for an additional 4 h at –20 °C. Following incubation, samples were spun down at 17,000  $\times$  g for 10 min at 0 °C to pellet precipitated protein, the supernatant discarded, and excess acetone evaporated at 65 °C for 5 min.

**Digestion of proteome samples for glycoproteomic analysis.** Dried protein pellets were resuspended in 6 M urea, 2 M thiourea, 50 mM NH<sub>4</sub>HCO<sub>3</sub> and reduced with 20 mM DTT for 1 h followed by alkylation with 40 mM Iodoacetamide in the dark for 1 h<sup>70</sup>. 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 h at room temperature before the sample was diluted with 100 mM NH<sub>4</sub>HCO<sub>3</sub> four-fold to reduce the urea/thiourea concentration below 2 M and trypsin (1/50 w/w) added. (2) For Thermolysin digestions reduced/alkylated samples were diluted with 100 mM NH<sub>4</sub>HCO<sub>3</sub> four-fold to reduce the urea/thiourea concentration below 2 M 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 2 M, 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.

**ZIC-HILIC enrichment of glycopeptides.** ZIC-HILIC enrichments were performed as according to the protocol of Mysling et al. with minor modifications<sup>71</sup>. ZIC-HILIC Stage tips<sup>72</sup> were created by packing 0.5 cm of 10  $\mu$ m ZIC-HILIC resin (Millipore/Sigma) into p200 tips containing a frit of C8 Empore™ (Sigma) material. Prior to use, the columns were washed with Milli-Q water, followed by 95% acetonitrile and then equilibrated with 80% acetonitrile and 1% TFA. Digested proteome samples were resuspended in 80% acetonitrile and 1% TFA. Samples were adjusted to a concentration of 5  $\mu$ g/ $\mu$ l (a total of 500  $\mu$ g of peptide used for each enrichment) then loaded onto equilibrated ZIC-HILIC columns. ZIC-HILIC columns were washed with 20 bed volumes of 80% acetonitrile, 1% TFA to remove non-glycosylated peptides and bound peptides eluted with 10 bed volumes of Milli-Q water. Eluted peptides were dried by vacuum centrifugation at room temperature and stored at –20 °C.

**LC-MS analysis of glycopeptide enrichments.** ZIC-HILIC enriched samples were resuspended in Buffer A\* and separated using a two-column chromatography set up composed of a PepMap100 C18 20 mm $\times$ 75  $\mu$ m trap and a PepMap C18 500 mm $\times$ 75  $\mu$ m analytical column (Thermo Fisher Scientific). Samples were concentrated onto the trap column at 5  $\mu$ l/min for 5 min with Buffer A (0.1% formic acid, 2% DMSO) and then infused into an Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer (Thermo Fisher Scientific) at 300 nl/min via the analytical column using a Dionex Ultimate 3000 UPLC (Thermo Fisher Scientific). 185-min analytical runs were undertaken by altering the buffer composition from 2% Buffer B (0.1% formic acid, 77.9% acetonitrile, 2% DMSO) to 28% B over 150 min, then from 28% B to 40% B over 10 min, then from 40% B to 100% B over 2 min. The composition was held at 100% B for 3 min, and then dropped to 2% B over 5 min before being held at 2% B for another 15 min. The Lumos™ Mass Spectrometer was operated in a data-dependent mode automatically switching between the acquisition of a single Orbitrap MS scan (350–1800 *m/z*, maximal injection time of 50 ms, an Automated Gain Control (AGC) set to a maximum of 1 $\times$ 10<sup>6</sup> ions and a resolution of 120k) every 3 s and Orbitrap MS/MS HCD scans of precursors (NCE 28% with 5% Stepping, maximal injection time of 60 ms, an AGC set to a maximum of 1 $\times$ 10<sup>5</sup> 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 scans<sup>73</sup> of potential glycopeptides; an Orbitrap EThcD scan (NCE 15%, maximal injection time of 250 ms, AGC set to a maximum of 2 $\times$ 10<sup>5</sup> ions with a resolution of 30k and using the extended mass range setting to improve the detection of high mass glycopeptide fragment ions<sup>74</sup>); an ion trap CID scan (NCE 35%, maximal injection time of 40 ms, an AGC set to a maximum of 5 $\times$ 10<sup>4</sup> 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 $\times$ 10<sup>5</sup> ions and a resolution of 30k).

**Glycopeptide identifications using Byonic.** Raw data files were processed using Byonic v3.5.3 (Protein Metrics Inc.<sup>75</sup>). 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 HexHexNAc<sub>2</sub> (elemental composition: C<sub>22</sub>O<sub>15</sub>H<sub>36</sub>N<sub>2</sub>, mass: 568.2115) and Suc-HexHexNAc<sub>2</sub> (elemental composition: C<sub>26</sub>O<sub>18</sub>H<sub>40</sub>N<sub>2</sub>, mass: 668.2276) were included as variable modifications at Serine and Threonine residues. K56-2 samples were searched against the K56-2 proteome<sup>76</sup> (Uniprot accession: UP000011196, 7467 proteins) while H111 samples were searched against the H111 proteome<sup>41</sup> (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<sup>77,78</sup>. Manual inspection of glycopeptides to assess the correctness of assignments was undertaken using the guidelines of Chen et al.<sup>79</sup> with the additional requirement that HCD glycopeptide spectra should contain evidence for glycan fragments such as oxonium ions or the presence of Y<sub>0</sub> to Y<sub>2</sub> 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 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 satisfying the correctness of assignments guidelines outlined above, as well as spectra which provided either partial or complete localisation of glycosylation sites, are provided within Supplementary Data 3.

**Analysis of glycosylation sites using O-Pair.** Automated glycosylation site analysis was undertaken using O-Pair within MetaMorpheus (non-public release version MM0523<sup>54</sup>). 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 HexHexNAc<sub>2</sub> (elemental composition: C<sub>22</sub>O<sub>15</sub>H<sub>36</sub>N<sub>2</sub>, mass: 568.2115) or Suc-HexHexNAc<sub>2</sub> (elemental composition: C<sub>26</sub>O<sub>18</sub>H<sub>40</sub>N<sub>2</sub>, mass: 668.2276). K56-2 samples were searched against the K56-2 proteome<sup>76</sup> (Uniprot accession: UP000011196) while H111 samples were searched against the H111 proteome<sup>41</sup> (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. humptydoensis* MSMB43; *B. ubonensis* MSMB22; *B. anthina* MSMB649; *B. diffusa* MSMB375 and *B. pseudomultivorans* MSMB2199 (PRIDE ProteomeXchange Consortium accession: PXD018587) was undertaken as above using the proteome databases as outlined in<sup>13</sup>. 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 NH<sub>4</sub>HCO<sub>3</sub> 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 (3 M) and 5 mg of OLIGO R3 reverse phase resin (Thermo Fisher Scientific) as described<sup>72,80</sup>. 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×75 μm trap and a PepMap C18 500 mm×75 μm analytical column (Thermo Scientific). Samples were concentrated onto the trap column at 5 μl/min with Buffer A for 5 min and infused into an Orbitrap Elite™ Mass Spectrometer (Thermo Scientific) at 300 nl/min via the analytical column using a Dionex Ultimate 3000 UPLC (Thermo Scientific). The peptides were separated using a 125-min gradient altering the buffer composition from 1% buffer B to 23% B over 95 min, then from 23% B to 40% B over 10 min, then from 40% B to 100% B over 5 min, the composition was held at 100% B for 5 min, and then dropped to 3% B over 2 min and held at 3% B for another 8 min. 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<sup>6</sup> ions and a resolution of 60k) followed by 20 data-dependent ion-trap CID MS-MS events (NCE 30%, maximal injection time of 80 ms, an AGC set to a maximum of 2×10<sup>4</sup> ions).

**In-gel digestion of DsbA1<sub>Nm</sub>-his<sub>6</sub> 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 1 h before being stained with Coomassie G-250 overnight. The region corresponding to ~25–35 kDa was then excised and processed as previously described (38). Briefly, the excised gel regions were sectioned into ~2 mm<sup>2</sup> pieces then destained in a solution of 100 mM NH<sub>4</sub>HCO<sub>3</sub>/50% ethanol for 15 min 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 min and then rehydrated in 50 mM NH<sub>4</sub>HCO<sub>3</sub> containing 10 mM DTT. Samples were reduced for 60 min at 56 °C with shaking and washed twice in 100% ethanol for 10 min to remove DTT. Reduced ethanol washed samples were sequentially alkylated with 55 mM of iodoacetamide in 50 mM NH<sub>4</sub>HCO<sub>3</sub> in the dark for 45 min at room temperature. Alkylated samples were then washed with Milli-Q water followed by 100% ethanol twice for 5 min to remove residual iodoacetamide then vacuum-dried for 10 min. Alkylated samples were then rehydrated with 20 ng/μl trypsin (Promega) in 50 mM NH<sub>4</sub>HCO<sub>3</sub> at 4 °C for 1 h. Excess trypsin was removed, the gel pieces covered in 40 mM NH<sub>4</sub>HCO<sub>3</sub>, 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<sub>Nm</sub>-his<sub>6</sub> glycopeptides.** StageTip cleaned-up samples were resuspended in Buffer A\* and separated using a two-column set up as described above, coupled to an Orbitrap Lumos™ Mass Spectrometer equipped with a FAIMS Pro interface (Thermo Fisher Scientific). 125-min gradients were run for each sample altering the buffer composition from 3% Buffer B to 28% B over 95 min, then from 28% B to 40% B over 10 min, then from 40% B to 80% B over 7 min, the composition was held at 80% B for 3 min, and then dropped to 3% B over 1 min and held at 3% B for another 9 min. 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<sup>49</sup>. 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<sup>6</sup> ions and a resolution of 120k) was acquired every 2 s 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<sup>5</sup> 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 an Orbitrap HCD scan of potential glycopeptides with scan parameters described above. For parallel reaction monitoring (PRM) experiments, 65-min gradients were run for each sample altering the buffer composition from 3% buffer B to 28% B over 35 min, then from 28% B to 40% B over 10 min, then from 40% B to 80% B over 7 min, the composition was held at 80% B for 3 min, and then dropped to 3% B over 1 min and held at 3% B for another 9 min. 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<sup>6</sup> ions and a resolution of 60k) every 3 s followed by ion-trap EThcD MS2 events (NCE 25%, maximal injection time of 200 ms, an AGC set to a maximum of 6×10<sup>4</sup> 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<sup>5</sup> ions) of the *m/z* 1547.77, 1581.12 and 1543.10 which corresponds to the +3 charge states of the HexHexNAc<sub>2</sub> glycosylated <sup>23</sup>VQTSVP ADSAPAAATAAAAPAGLVEGQNYTVLANPIPPQQAGK<sup>64</sup>; the Suc-HexHexNAc<sub>2</sub> glycosylated <sup>23</sup>VQTSVPADSAPAAATAAAAPAGLVEGQNYTVLANPIPPQQAGK<sup>64</sup> and the HexHexNAc<sub>2</sub> glycosylated <sup>23</sup>VQTSVPADSAPAAASAAAAPAGLVEGQNYTVLANPIPPQQAGK<sup>64</sup> peptides with the site of glycosylation within these peptides underlined.

**Proteomic analysis.** H111 Proteome and in-gel datasets were processed using MaxQuant (v1.5.5.1 or 1.6.3.4<sup>81</sup>). The H111 proteome dataset was searched against the H111 proteome<sup>41</sup> (Uniprot accession: UP000245426) and *B. cenocepacia* strain J2315 (Uniprot accession: UP000001035, 6993 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 Database<sup>55</sup>, Strain number: 3404) or *B. humptydoensis* MSMB43 (Burkholderia Genome Database<sup>55</sup>, Strain number: 4072) depending on the sample type and a custom database of DsbA1<sub>Nm</sub>-his<sub>6</sub> 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 HexHexNAc<sub>2</sub> (elemental composition: C<sub>22</sub>O<sub>15</sub>H<sub>36</sub>N<sub>2</sub>, mass: 568.2115) and Suc-HexHexNAc<sub>2</sub> (elemental composition: C<sub>26</sub>O<sub>18</sub>H<sub>40</sub>N<sub>2</sub>, 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 min and an alignment window of 10 min. For label free quantitation (LFQ) the MaxLFQ option in Maxquant<sup>82</sup> was enabled. The resulting outputs were processed within the Perseus (v1.5.0.9)<sup>83</sup> 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 PgL 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 WT<sup>5</sup>. Fisher exact tests in Perseus was undertaken allow a 5% FDR. To compare the relative abundance of glycosylated and non-glycosylated peptides from DsbA1<sub>Nm</sub>-his<sub>6</sub> point mutants the area under the curve of peptides were extracted using the FreeStyle viewer and the resulting data provided within the Supplementary document. Statistical analysis of the area under the curve was undertaken in Prism (version 7.0e) using a two-sided t-test.

**Visualisation of glycoproteome and proteome datasets.** Data visualisation was undertaken using ggplot2<sup>84</sup> 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 Annotator<sup>85</sup> (<http://www.interactivepeptidespectralannotator.com/PeptideAnnotator.html>) was used.

**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, subsetted to contain a maximum of 20 of any individual species (*n* = 174), were obtained from the *Burkholderia* database<sup>55</sup>. 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<sup>86</sup>. Gene hits were translated to protein sequences and the translated hits screened for the presence of the motif using seqkit locate<sup>87</sup>, 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<sup>88</sup>. To compare glycoprotein homologues and pglL

sequences between *Burkholderia* species the *Burkholderia* Orthologous Groups information provided by the *Burkholderia* database was used<sup>55</sup>.

**Statistics and Reproducibility.** Statistical analyses of biological samples were undertaken on a minimum of three biological replicates for Western blotting or Proteomic analysis. A biological replicate is defined as separately grown cultures treated with a given induction or growth condition. All raw uncropped and protein marker associated Western blotting images are provided within Supplementary Fig. 19.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### Data availability

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<sup>89,90</sup> (<https://www.ebi.ac.uk/pride/archive/>). The glycoproteomic datasets are available with the identifier: PXD024090; the H111 proteome analysis is available with the identifier: PXD023755; the DsbA1<sub>Nm</sub>-his<sub>6</sub> (*B. cenocepacia* K56-2) associated analysis is available with the identifier: PXD023955; the DsbA1<sub>Nm</sub>-his<sub>6</sub> (*B. humptdooensis* MSMB43 and *B. ubonensis* MSMB22) associated analysis is available with the identifier: PXD024056.

Received: 20 April 2021; Accepted: 23 August 2021;

Published online: 07 September 2021

### References

- Koomey, M. O-linked protein glycosylation in bacteria: snapshots and current perspectives. *Curr. Opin. Struct. Biol.* **56**, 198–203 (2019).
- Nothhaft, H. & Szymanski, C. M. New discoveries in bacterial N-glycosylation to expand the synthetic biology toolbox. *Curr. Opin. Chem. Biol.* **53**, 16–24 (2019).
- Cain, J. A. et al. Proteomics reveals multiple phenotypes associated with N-linked glycosylation in *Campylobacter jejuni*. *Mol. Cell Proteomics* <https://doi.org/10.1074/mcp.RA118.001199> (2019).
- Abouelhadid, S. et al. Quantitative analyses reveal novel roles for N-glycosylation in a major enteric bacterial pathogen. *MBio* **10**, <https://doi.org/10.1128/mBio.00297-19> (2019).
- 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 (2019).
- 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**, <https://doi.org/10.1128/mBio.02604-20> (2020).
- Gault, J. et al. *Neisseria meningitidis* type IV Pili composed of sequence invariable pilins are masked by multisite glycosylation. *PLoS Pathog.* **11**, e1005162 (2015).
- Lassak, J. et al. Arginine-rhamnosylation as new strategy to activate translation elongation factor P. *Nat. Chem. Biol.* **11**, 266–270 (2015).
- El Qaidi, S. et al. An intra-bacterial activity for a T3SS effector. *Sci. Rep.* **10**, 1073 (2020).
- Iwashkiw, J. A., Voza, 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 (2013).
- Kightlinger, W., Warfel, K. F., DeLisa, M. P. & Jewett, M. C. Synthetic glycobiology: parts, systems, and applications. *ACS Synth. Biol.* **9**, 1534–1562 (2020).
- Scott, N. E. et al. Diversity within the O-linked protein glycosylation systems of acinetobacter species. *Mol. Cell Proteom.* **13**, 2354–2370 (2014).
- 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 Proteom.* **19**, 1561–1574 (2020).
- Hadjineophytou, C. et al. Genetic determinants of genus-Level glycan diversity in a bacterial protein glycosylation system. *PLoS Genet.* **15**, e1008532 (2019).
- 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 (2019).
- 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 (2013).
- 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 (2016).
- 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 (2020).
- Kowarik, M. et al. Definition of the bacterial N-glycosylation site consensus sequence. *EMBO J.* **25**, 1957–1966 (2006).
- 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 (2020).
- 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 (2009).
- 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 (2011).
- Stimson, E. et al. Meningococcal pilin: a glycoprotein substituted with digalactosyl 2,4-diacetamido-2,4,6-trideoxyhexose. *Mol. Microbiol.* **17**, 1201–1214 (1995).
- 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 (2013).
- Schulz, B. L. et al. Identification of bacterial protein O-oligosaccharyltransferases and their glycoprotein substrates. *PLoS ONE* **8**, e62768 (2013).
- 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 (2015).
- 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 (2012).
- Power, P. M. et al. Genetic characterization of pilin glycosylation in *Neisseria meningitidis*. *Microbiology* **146**, 967–979 (2000).
- 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 (2014).
- 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 (2011).
- Castric, P. pilO, a gene required for glycosylation of *Pseudomonas aeruginosa* 1244 pilin. *Microbiology* **141**, 1247–1254 (1995).
- Elhenawy, W. et al. Protein O-linked glycosylation in the plant pathogen *Ralstonia solanacearum*. *Glycobiology* **26**, 301–311 (2016).
- 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 (2012).
- Vik, A. et al. Broad spectrum O-linked protein glycosylation in the human pathogen *Neisseria gonorrhoeae*. *Proc. Natl Acad. Sci. USA* **106**, 4447–4452 (2009).
- Mahenthalingam, E., Urban, T. A. & Goldberg, J. B. The multifarious, multiplexed *Burkholderia cepacia* complex. *Nat. Rev. Microbiol.* **3**, 144–156 (2005).
- Manno, G. et al. Epidemiology and clinical course of *Burkholderia cepacia* complex infections, particularly those caused by different *Burkholderia cenocepacia* strains, among patients attending an Italian Cystic Fibrosis Center. *J. Clin. Microbiol.* **42**, 1491–1497 (2004).
- De Souza, A. et al. Prevalence and clonality of *Burkholderia cepacia* complex genomovars in UK patients with cystic fibrosis referred for lung transplantation. *Thorax* **59**, 526–528 (2004).
- Drevinek, P. & Mahenthalingam, E. *Burkholderia cenocepacia* in cystic fibrosis: epidemiology and molecular mechanisms of virulence. *Clin. Microbiol. Infect.* **16**, 821–830 (2010).
- 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).
- Romling, U. et al. Epidemiology of chronic *Pseudomonas aeruginosa* infections in cystic fibrosis. *J. Infect. Dis.* **170**, 1616–1621 (1994).
- Carlier, A. et al. Genome sequence of *Burkholderia cenocepacia* H111, a cystic fibrosis airway isolate. *Genome Announc.* **2**, <https://doi.org/10.1128/genomeA.00298-14> (2014).
- Bodilis, J. et al. Comparative genomics of environmental and clinical *Burkholderia cenocepacia* strains closely related to the highly transmissible epidemic ET12 lineage. *Front Microbiol.* **9**, 383 (2018).
- Sass, A. M. et al. The unexpected discovery of a novel low-oxygen-activated locus for the anoxic persistence of *Burkholderia cenocepacia*. *ISME J.* **7**, 1568–1581 (2013).

44. Malott, R. J., Baldwin, A., Mahenthiralingam, E. & Sokol, P. A. Characterization of the *cciIR* quorum-sensing system in *Burkholderia cenocepacia*. *Infect. Immun.* **73**, 4982–4992 (2005).
45. Baldwin, A., Sokol, P. A., Parkhill, J. & Mahenthiralingam, E. The *Burkholderia cenocepacia* epidemic strain marker is part of a novel genomic island encoding both virulence and metabolism-associated genes in *Burkholderia cenocepacia*. *Infect. Immun.* **72**, 1537–1547 (2004).
46. Uehlinger, S. et al. Identification of specific and universal virulence factors in *Burkholderia cenocepacia* strains by using multiple infection hosts. *Infect. Immun.* **77**, 4102–4110 (2009).
47. Malott, R. J. et al. A *Burkholderia cenocepacia* orphan LuxR homolog is involved in quorum-sensing regulation. *J. Bacteriol.* **191**, 2447–2460 (2009).
48. Gomes, M. C. et al. The *afc* antifungal activity cluster, which is under tight regulatory control of ShvR, is essential for transition from intracellular persistence of *Burkholderia cenocepacia* to acute pro-inflammatory infection. *PLoS Pathog.* **14**, e1007473 (2018).
49. 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 (2021).
50. Scott, N. E. et al. Comparative proteomics and glycoproteomics reveal increased N-linked glycosylation and relaxed sequon specificity in *Campylobacter jejuni* NCTC11168 O. *J. Proteome Res.* **13**, 5136–5150 (2014).
51. Scott, N. E. et al. Modification of the *Campylobacter jejuni* N-linked glycan by EptC protein-mediated addition of phosphoethanolamine. *J. Biol. Chem.* **287**, 29384–29396 (2012).
52. Fu, Y. & Qian, X. Transferred subgroup false discovery rate for rare post-translational modifications detected by mass spectrometry. *Mol. Cell Proteom.* **13**, 1359–1368 (2014).
53. Krug, K. et al. Deep coverage of the *Escherichia coli* proteome enables the assessment of false discovery rates in simple proteogenomic experiments. *Mol. Cell Proteom.* **12**, 3420–3430 (2013).
54. 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 (2020).
55. Winsor, G. L. et al. The *Burkholderia* Genome Database: facilitating flexible queries and comparative analyses. *Bioinformatics* **24**, 2803–2804 (2008).
56. Coyne, M. J. et al. Phylum-wide general protein O-glycosylation system of the Bacteroidetes. *Mol. Microbiol.* **88**, 772–783 (2013).
57. Nothhaft, H. et al. Diversity in the protein N-glycosylation pathways within the *Campylobacter* genus. *Mol. Cell Proteom.* **11**, 1203–1219 (2012).
58. Willcocks, S. J. et al. Virulence of the emerging pathogen, *Burkholderia pseudomallei*, depends upon the O-linked oligosaccharyltransferase, PglL. *Future Microbiol.* **15**, 241–257 (2020).
59. Pan, C. et al. Biosynthesis of conjugate vaccines using an O-linked glycosylation system. *mBio* **7**, e00443–00416 (2016).
60. Gebhart, C. et al. Characterization of exogenous bacterial oligosaccharyltransferases in *Escherichia coli* reveals the potential for O-linked protein glycosylation in *Vibrio cholerae* and *Burkholderia thailandensis*. *Glycobiology* **22**, 962–974 (2012).
61. Faridmoayer, A. et al. Extreme substrate promiscuity of the *Neisseria* oligosaccharyl transferase involved in protein O-glycosylation. *J. Biol. Chem.* **283**, 34596–34604 (2008).
62. Faridmoayer, A., Fentabil, M. A., Mills, D. C., Klassen, J. S. & Feldman, M. F. Functional characterization of bacterial oligosaccharyltransferases involved in O-linked protein glycosylation. *J. Bacteriol.* **189**, 8088–8098 (2007).
63. Shropshire, H., Jones, R. A., Aguilo-Ferretjans, M. M., Scanlan, D. J. & Chen, Y. Proteomics insights into the *Burkholderia cenocepacia* phosphorus stress response. *Environ Microbiol.* <https://doi.org/10.1111/1462-2920.15451> (2021).
64. Inhulsen, S. et al. Identification of functions linking quorum sensing with biofilm formation in *Burkholderia cenocepacia* H111. *Microbiologyopen* **1**, 225–242 (2012).
65. Riedel, K. et al. Analysis of the quorum-sensing regulon of the opportunistic pathogen *Burkholderia cenocepacia* H111 by proteomics. *Electrophoresis* **24**, 740–750 (2003).
66. 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 (2014).
67. Gibson, D. G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
68. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: a Laboratory Manual* (Cold Spring Harbor Laboratory, 1990).
69. 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 (2008).
70. 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 Proteom.* **10**, M000031MCP000201 (2011).
71. Mysling, S., Palmisano, G., Højrup, P. & Thaysen-Andersen, M. Utilizing ion-pairing hydrophilic interaction chromatography solid phase extraction for efficient glycopeptide enrichment in glycoproteomics. *Anal. Chem.* **82**, 5598–5609 (2010).
72. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2**, 1896–1906 (2007).
73. 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. Proteom.* **2012**, 560391 (2012).
74. 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 (2019).
75. Bern, M., Kil, Y. J. & Becker, C. Byonic: advanced peptide and protein identification software. *Curr. Protoc. Bioinform.* **Chapter 13**, Unit13.20, <https://doi.org/10.1002/0471250953.bil320s40> (2012).
76. Varga, J. J. et al. Draft genome sequences of *Burkholderia cenocepacia* ET12 lineage strains K56-2 and BC7. *Genome Announc.* **1**, <https://doi.org/10.1128/genomeA.00841-13> (2013).
77. Lee, L. Y. et al. Toward automated N-glycopeptide identification in glycoproteomics. *J. Proteome Res.* **15**, 3904–3915 (2016).
78. Lenco, J., Khalikova, M. A. & Svec, F. Dissolving peptides in 0.1% formic acid brings risk of artificial formylation. *J. Proteome Res.* **19**, 993–999 (2020).
79. 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 (2005).
80. 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 (2006).
81. 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 (2008).
82. Cox, J. et al. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteom.* **MCP 13**, 2513–2526 (2014).
83. Tyanova, S. et al. The Perseus computational platform for comprehensive analysis of (pro)teomics data. *Nat. Methods* **13**, 731–740 (2016).
84. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag, 2016).
85. Brademan, D. R., Riley, N. M., Kwicien, N. W. & Coon, J. J. Interactive peptide spectral annotator: a versatile web-based tool for proteomic applications. *Mol. Cell Proteom.* **18**, S193–S201 (2019).
86. Davies, M. R. et al. Atlas of group A streptococcal vaccine candidates compiled using large-scale comparative genomics. *Nat. Genet.* **51**, 1035–1043 (2019).
87. Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS ONE* **11**, e0163962 (2016).
88. Wagih, O. *ggseqlogo*: a versatile R package for drawing sequence logos. *Bioinformatics* **33**, 3645–3647 (2017).
89. 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 (2019).
90. Vizcaino, J. A. et al. 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* **44**, D447–D456 (2016).

## 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)

## Author contributions

A.J.H. undertook the bioinformatic analysis of glycoproteins and glycosylation sites conserved across *Burkholderia* species. J.M.L. aided in the preparation of the manuscript and its figures. M.R.D. oversaw bioinformatic analysis and aided in the preparation of the manuscript. N.E.S. undertook all proteomic and molecular biology studies as well as wrote the initial manuscript draft.

## 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/s42003-021-02588-y>.

**Correspondence** and requests for materials should be addressed to N.E.S.

**Peer review information** *Communications Biology* thanks Benjamin Schulz, James Linton and the other, anonymous, reviewers for their contribution to the peer review of this work. Primary Handling Editors: Calvin Henard and Christina Karlsson Rosenthal. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021