Rapid screening for specific glycosylation and pathogen interactions on a 78 species avian egg white glycoprotein microarray

Marta Utratna[*,1], Heidi Annuk[*,1], Jared Q. Gerlach[1,2], Yuan C. Lee[3], Marian Kane[1], Michelle Kilcoyne[4,†], Lokesh Joshi[1,†]

[1] Glycoscience Group, National Centre for Biomedical Engineering Science, National University of Ireland Galway, Galway, Ireland

[2] Regenerative Medicine Institute, National University of Ireland Galway, Galway, Ireland

[3] Department of Biology, Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218, U.S.A.

[4] Carbohydrate Signalling Group, Microbiology, School of Natural Sciences, National University of Ireland Galway, Galway, Ireland

* These authors have contributed equally.

Corresponding authors: [†] E-mail: Lokesh.Joshi@nuigalway.ie (L. Joshi).

[†] E-mail: Michelle.Kilcoyne@nuigalway.ie (M. Kilcoyne)

Running title: Glycoprofiling and antimicrobial screening of avian egg whites with microarrays

Abstract

There is an urgent need for discovery of novel antimicrobials and carbohydrate-based anti-adhesive strategies are desirable as they may not promote resistance. Discovery of novel anti-adhesive molecules from natural product libraries will require the use of a high throughput screening platform. Avian egg white (EW) provides nutrition for the embryo and protects against infection, with glycosylation responsible for binding certain pathogens. In this study, a microarray platform of 78 species of avian EWs was developed and profiled using a lectin panel with a wide range of carbohydrate specificities. The dominating linkages of sialic acid in EWs were determined for the first time using the lectins MAA and SNA-I. EW glycosylation similarity among the different orders of birds did not strictly depend on phylogenetic relationship. The interactions of five strains of bacterial pathogens, including *Escherichia coli*, *Staphylococcus aureus* and *Vibrio cholera*, identified a number of EWs as potential anti-adhesives, with some as strain- or species-specific. Of the two bacterial toxins examined, shiga-like toxin 1 subunit B bound to ten EWs with similar glycosylation more intensely than pigeon EW. This study provides a unique platform for high throughput screening of natural products for specific glycosylation and pathogen interactions. This platform may provide a useful platform in the future for discovery of anti-adhesives targeted for strain and species specificity.

Key words: avian egg white, antimicrobials, glycoproteins, glycosylation, microarray, lectin profiling, bacterial toxins, glycomics

With the rise of antimicrobial resistance in human pathogens and the serious consequences of untreatable infections for human healthcare, there is increased interest in sourcing novel antimicrobial molecules, especially those which do not promote resistance[1,2]. The food industry in particular is interested in natural antimicrobial compounds that could be effective against food spoilage and pathogenic microorganism growth[3]. Adhesion of bacteria to the host cell surface is one of the first and most critical steps in infection and anti-adhesion therapy has been shown experimentally to be effective for preventing or treating infections[4,5]. Most bacteria express cell surface adhesins that bind to particular host cell surface carbohydrates. Virus particles and bacterial toxins can also have carbohydrate-binding protein or lectin domains to bind to host cells. For example, the P-fimbriae of uropathogenic *Escherichia coli* and Shiga-like toxin 1 (Stx1) binds to the Gal-α-(1,4)-Gal (galabiose) structure found on human glycolipids in fibroblasts, erythrocytes, primary kidney cells, and intestinal tissue[6-8]. Thus binding inhibitor or anti-adhesive molecules that are structurally similar to the complex carbohydrate ligands presented by the host have been extensively investigated in recent years. Multivalent presentation of these anti-adhesives are typically more effective than the monovalent version and significantly lower concentrations of the multivalent anti-adhesives are required compared to the monovalent versions[4,5,9]. Therefore, discovery of novel anti-adhesive molecules, including those that are species- or even strain-specific, will require screening multivalently-presented carbohydrates with sufficient structural complexity to function as effective decoys or mimics of host carbohydrates and a high-throughput (HTP) screening method or platform[4,5].

Avian egg white (EW), or albumen, provides nourishment to the developing embryo and has been suggested to act as a defensive layer, shielding the embryo from pathogens. Chicken EW (CEW) comprises approximately 60% of the total egg weight and constitutes 10-12% of total protein[10]. Most EW proteins are glycoproteins, including ovalbumin (OVA), ovomucoid

(OVM) and ovotransferrin (OVT) which are modified with *N*-linked oligosaccharides[11-14]. Protein glycosylation has important biological roles in many processes including mediating host-pathogen interactions, immune response, cell development and differentiation. EW is growth restricting for pathogens as it contains multiple antimicrobial components, including lysosyme (LYZ) and OVT[10,15]. Additionally, the structures, linkages and presentation of oligosaccharides present on EW glycoproteins serve as ligands for microbial adhesion[10], presumably to arrest bacterial migration to the yolk and facilitate subsequent antimicrobial defence. The Gal-α-(1,4)-Gal terminal structure is found in EW oligosaccharides of a few species of birds distributed in closely positioned branches of the phylogenetic tree including pigeon EW (PEW), in amphibians and in a sea turtle[16,17,18]. Thus EW glycoproteins may provide an excellent source of natural antimicrobials and specificity for targeted pathogens could potentially be selected for, depending on appropriate EW glycosylation.

Enzymatic glycosylation of proteins is controlled by factors that differ greatly among cell types and species[19]. Oligosaccharide structures from a small number of purified EW glycoproteins have been previously analysed by mass spectrometric methods, the majority of which are *N*-linked oligosaccharides e.g.[11,12,20-22], with a small number of *O*-linked structures elucidated from CEW ovomucin[23,24]. A variety of crude and purified EW glycoproteins from a range of avian species have also been studied using lectin and antibody blotting[16,25]. Lectins, particularly those from plants, are used to detect and profile intact glycosylation on glycoconjugates, cells and tissues[26]. High mannose (Man) and hybrid *N*-linked structures dominate in OVA while other CEW glycoproteins contain complex *N*-linked glycans with up to five antennae[12]. Pigeon glycoproteins contain up to four Gal-α-(1,4)-Gal and α-(2,6)-linked *N*-acetylneuraminic acid (Neu5Ac) residues on the termini of the antennae of their *N*-linked oligosaccharides[20,21]. While it is important to characterize the individual EW glycoprotein oligosaccharides, the density, distribution pattern, and three-dimensional presentation of the

intact oligosaccharides on the molecule greatly impact their biological interactions and recognition[19]. Accordingly, screening for EW glycosylation and interactions with targeted pathogens should be performed with EW in its natural conformation and distribution.

While lectin and glycan microarrays are well known and widely used for profiling glycosylation and carbohydrate-based interactions, the use of microarrays comprised of other glycosylated molecules are more recent developments. A natural human milk oligosaccharide microarray was developed which was utilized for screening of biologically relevant interactions[27]. A microarray of mucins purified from the reproductive and gastrointestinal tracts of six animals and from two cell lines was an effective tool for profiling mucin glycosylation[19] and the interactions of intact bacterial pathogens[28]. No multiplexed presentation of a collection of EWs has been performed to date. The presentation of EWs in a HTP platform such as a microarray will be highly advantageous for understanding EW natural glycosylation and screening for interactions with targeted pathogens and their toxins while using minimal amounts of samples and analytes.

In this work, a natural avian EW microarray was constructed using EWs from 78 different species (Table S-1). Lectin profiling was carried out to characterise the EW glycosylation and deduced structures were correlated with previous studies and with avian phylogeny. As sialic acid linkages can be a critical factor in dictating pathogen binding to host ligands[29], the linkages of sialic acids in EW species were elucidated in this study for the first time using *Maackia amurensis* agglutinin (MAA) and *Sambucus nigra* agglutinin I (SNA-I) lectins, which bind specifically to α-(2,3)- and α-(2,6)-linked sialic acid, respectively. The novel EW microarray was also used to investigate interactions with five bacterial strains of relevance to food, including *E. coli*, *Staphylococcus aureus* and the emerging foodborne pathogen *Vibrio parahaemolyticus*, and two bacterial toxins. This platform may be useful in the future for

screening for novel species- and strain-specific potential anti-adhesives, which may be suitable for future deployment in food and food processing.

Results and discussion

Electrophoretic analysis of avian EW library. Following optimization of EW solubilisation (see Supplementary Information), the entire EW library was profiled by SDS-PAGE. The major protein components of CEW are OVA (45 kDa, 54% of total protein), OVT (77 kDa, 12%), and OVM (28 kDa, 11%)[10] and these glycoproteins were used as references for electrophoretic analysis of all species of EWs (Figure S-1A). OVA, OVT and OVM were evident in CEW and the majority of the other 77 species in the selected EW library (Table S-1 and Fig. 1), with different electrophoretic mobilities observed between species, in agreement with a previous study[16]. These mobility variations were previously attributed to differences in the amino acid sequences of individual glycoproteins and species-specific differences in the number and size of *N*-linked glycans in a comparison between CEW and pigeon EW (PEW)[21].

Similar electrophoretic patterns for OVA, OVM and OVT for species in close phylogenetic proximity were previously observed for pheasant (SVP) and ostrich (OST) EWs in addition to quail EW (QEW), duck EW (DEW) and CEW[30]. However, in the present study, which contained a larger library of species, a corresponding species relationship was not conclusively demonstrated by electrophoretic analysis. Several characteristics observed in quantitative protein distribution in the various EWs were in agreement with previous studies. For example, the major protein component in emu EW (EMU) was OVT not OVA as for CEW and the electrophoretic mobilities of OVA and OVM varied in EMU compared to CEW (Fig. 1)[15].

Interestingly, a 14 kDa band corresponding to LYZ (approximately 3.5% of total CEW protein)[10] was observed in CEW and as a very faint band in DEW and QEW (Fig.s S-1A and 1) and small buttonquail EW (BUQ). However, no other EW in the library displayed a band with similar electrophoretic mobility to LYZ. This is due to low abundance or lack of LYZ in the EW of most bird species, e.g. only 1% of DEW is LYZ, 0.5% of goose and no LYZ has been found in EMU, or dramatically different electrophoretic mobility of the LYZ in other species[30].

Total CEW and individual CEW glycoprotein glycosylation. Following optimization of microarray printing (Supplementary Fig. S1), the EW microarrays were constructed and incubated with a library of fluorescently-labelled lectins (Table 1) to generate a characteristic glycoprofile for each EW and standard. Appropriate haptenic carbohydrates or glycoproteins were also co-incubated with lectins to confirm that the lectin interactions with printed EWs were carbohydrate-mediated[26,31]. Initially individual lectin binding profiles for the purified CEW glycoproteins were examined for similarity to the glycoprofile for crude CEW (Fig. 2).

Of the purified CEW components, OVM had the highest binding intensity for the majority of the lectin panel, possibly due to its high carbohydrate content of 25% (w/w)[10]. In comparison, OVA and OVT contain 3.5% and 2.6% (w/w) carbohydrate, respectively[32,33]. Although high binding intensity with LYZ was observed for most of the lectins (Table S-3), these interactions were not inhibited by haptenic sugars and were thus considered non-carbohydrate mediated. Interestingly, LYZ hydrolyses the linkage between the *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid in the bacterial cell wall. All three CEW glycoproteins (OVA, OVT and OVM) interacted with the lectins wheat germ agglutinin (WGA, *Triticum vulgaris* agglutinin) and *Griffonia simplicifolia* lectin II (GSL-II), which indicated the presence of GlcNAc residues (Table 1). Only OVT interacted with the mannose-(Man-

)specific lectins (*Narcissus pseudonarcissus* agglutinin (NPA), *Hippeastrum hybrid* agglutinin (HHA), and *Galanthus nivalis* agglutinin (GNA)) which indicated the presence of high-mannose type *N*-linked structures[32]. NPA and HHA bind internal and terminal Man residues with preference towards high Man type *N*-linked structures containing α-(1,6)-linked Man or both α-(1,3)- and α-(1,6)-linked Man, respectively (Table 1). GNA is reported to bind most strongly to multiple terminal α-(1,3)-linked Man[34]. Interaction with Concanavalin A (Con A, *Canavalia ensiformis* agglutinin), which has high binding affinity for terminal Man and Man core structures of *N*-linked glycans, was observed for all three glycoproteins. HHA binding was low with crude CEW, but moderate with OVT. This may be due to the OVT oligosaccharides comprising a relatively low proportion of crude CEW sample. However, the binding intensity of Con A was higher for crude CEW than the individual CEW glycoproteins which may reflect the overall larger relative proportion of high mannose type *N*-linked glycosylation.

The presence of bi-, tri- and tetra-antennary complex *N*-linked oligosaccharides, terminating most often with type II *N*-acetyllactosamine (LacNAc; Gal-β-(1,4)-GlcNAc) which is often sialylated[34], on OVM was suggested by binding of both *Phaseolus vulgaris* erythroagglutinin (PHA-E) and *Phaseolus vulgaris* leucoagglutinin (PHA-L), while OVA only interacted with PHA-E and OVT did not bind with either lectin (Table 1 and Fig. 2). Furthermore, the presence of α-(2,3)-linked sialic acid on OVM was indicated by MAA binding, in agreement with previous reports[13,33]. Also, WGA binding, which recognises GlcNAc and sialic acid (Table 1), was observed only for OVM and OVT. The results for OVT and OVM correlated well with structures previously elucidated by MS and HPLC[13]. OVT *N*-linked oligosaccharides were mainly composed of GlcNAc (54%) and Man (43%) and some Gal residues (3.5%), while the structures reported for OVM were mostly complex-type, with dominating tri-, tetra- and penta-antennary structures[13]. Overall, individual lectin binding

profiles were observed for each purified CEW glycoproteins with a similar combination glycoprofile observed for crude CEW.

EW library glycosylation profiling. There was considerable variation in lectin binding patterns across the 78 avian species EWs examined (Supplementary Fig. S3). The binding intensity data for the 14 lectin library were subjected to hierarchical clustering to identify any similarities within the diverse species (Fig. 3). To facilitate the description of the groups, the clustered heat map was further divided according to the dendrogram generated by the clustering algorithm to yield five clusters of EWs (Fig. 3, clusters 1-5, C1-5). For C1-5, a minimum similarity of 50% was selected as the defining threshold (bar located at position 0.5, Fig. 3) while 75% similarity was selected for subclusters a, b, c and d.

C1 (Fig. 3) included only two species, EAW and MTQ (Supplementary Table S1) from the orders Anseriformes (A) and Galliformes (G), respectively, and was very different from all other samples, only having 33.7% minimum similarity to the rest of the EW glycoprofiles. EWs in C1 exhibited high binding with the Man-specific lectins HHA, GNA, NPA and Con A, and the GlcNAc-specific GSL-II (but not WGA) (Table 1). EWs in this cluster also exhibited an almost total lack of binding with PHA-L and very low interaction with PHA-E. This lectin profile correlates well with the high Man *N*-linked structures previously reported for birds belonging to Anseriformes and Galliformes (Supplementary Fig. S4)[35].

C2 (Fig. 3) was grouped into subclusters a-c and included EWs with high binding intensity to galactose-(Gal-)specific lectins soybean agglutinin (SBA, *Glycine max* agglutinin), *Griffonia simplicifolia* lectin I isolectin B4 (GS-I-B4) and *Pseudomonas aeruginosa* lectin I (PA-I) (Table 1) and was mostly devoid of Man-specific lectin binding signals. PHA-L and PHA-E bound intensely to EWs in C2a-c with low binding for EWs in C2d. C3 (Fig. 3) contained only one EW, which was the only representative from Piciformes in this study. CFL EW had

high binding intensity with SBA (terminal *N*-acetylgalactosamine, GalNAc), WFA (GalNAc /sulfated GalNAc), MAA, WGA and moderate interaction with Con A but not with GSL-II, which typically indicates a lower incidence of terminal GlcNAc residues, and PHA-E, which taken together indicated the presence of complex, bi-antennary *N*-linked structures with terminal GalNAc and α-(2,3)-linked sialylation (Table 1).

In contrast to C2, C4a (Fig. 3) consisted of EWs with high binding intensity to GlcNAc-specific lectins GS-II and WGA and Con A. This may be indicative of a population of hybrid structures or multi-antennary structures terminating with terminal GlcNAc polymers (polyGlcNAc residues) (Fig. 3). Interestingly, the overall lectin profiles of EWs from C1 had similarities with those of C4c, with some additional binding interactions in C4c (Fig. 3), the latter of which included a further five representatives of Galliformes and three Anseriformes. This pattern suggested that the *N*-linked EW structures in C4c were of the high-mannose and hybrid type. C4b and C4a have either much lower (C4b), or lack (C4a), interactions with HHA, GNA and NPA. Among the members of C4a and C4c which demonstrated intense GSL-II binding, representatives of Anseriformes but not Galliformes corresponded well with the abundant GlcNAc previously reported by MALDI-TOF MS in Anseriformes (Fig.s S-4 and S-5)[35]. Additionally, the five species in C4a showed very high binding with WGA and moderate binding to PHA-L and PHA-E.

High binding intensity with complex structure-specific PHA-E was observed for more than 60% of the EWs in C5. C5 EWs also had high binding intensity with PHA-L, though to a somewhat lesser extent than PHA-E, with lowest PHA-L binding intensity in C4b. The majority of EWs with high binding intensities for PHA-E and PHA-L were in C5, the largest cluster (Fig. 3). C5 subclusters could be further characterised by additional binding to (i) MAA only (C5a), (ii) Con A with high binding intensity, no interaction with PA-I and GS-I-

B4, and some moderate binding intensity with Man-specific lectins (C5b), (iii) Gal- and GalNAc-specific lectins (SBA, *Wisteria floribunda* agglutinin (WFA), GS-I-B4, PA-I) and MAA with diverse binding intensities (C5c), and (iv) WFA with high binding intensity (C5d). Only 8 out of 40 EWs in C5 interacted with SNA-I, which is specific for α-(2,6)-linked sialic acid, with no SNA-I binding at all observed in Anseriformes (A) EWs which clustered in C5. Additionally, among the EWs in C5b with high WFA binding intensity, three representatives of Anseriformes were present but none of Galliformes. WFA binding for these three Anseriformes EWs, MGP, SOS and CNG, corresponded well with the low abundance of GalNAc residues in the *N*-linked oligosaccharide structures previously described for Galliformes (Supplementary Fig. S4)[35].

Overall, most EW samples exhibited high relative binding intensity with PHA-E and PHA-L, which suggested the presence of bi-, tri- and tetra-antennary *N*-linked structures, and the majority of EWs also interacted with Con A. Con A binding to EW glycoproteins was previously demonstrated using lectin blotting which included 75 EW species overlapping with this study[16]. When comparing data from the present study with all overlapping species[16], similar trends were reported with the majority of EWs interacting with Con A (93% here compared to 84% previously reported) (Figure S-7). A similar proportion of GS-I binding to EWs was also shown (40% of the species) despite the wider binding specificity of GS-I (Gal/GalNAc specificity) compared to the more narrowly specific GS-I-B4 (exclusively terminal α-linked Gal) used here (Table 1).

Due to the potential for high yields, avian eggs are also a desirable target for producing recombinant therapeutic glycoproteins[36]. Because the major egg allergens for humans are egg glycoproteins, the possibility for inadvertent co-purification of egg allergens exists. Recent studies using glycosylated variants of chicken OVM and OVA showed that allergy response

was mainly related to carbohydrate structures and their location[37,38]. Thus, production of recombinant proteins with desired glycosylation could be achieved by selection of the avian species with desired glycosylation without having to extensively genetically re-engineer birds.

EW sialic acid linkage specificity. The overall linkage specificity of the EW library sialylation was addressed in this study using MAA and SNA-I lectins for the first time. Overall, binding to MAA was more frequent across samples (Fig. 3) and was dispersed throughout C2, C3 and C5, with all the members of C5a interacting strongly, and very low binding intensities observed in C1 and C4. SNA-I interaction was low or absent across all EWs except for two EWs from Columbiformes, PEW and RKD, grouped in C2b, with intense binding and 12 other EWs in C2 and C5 with moderate binding. This suggested that α-(2,6)-linked sialic acid is less common in the 78 EW species included in this study. The average sialic acid concentration was previously reported as four times greater in Galliformes compared to Anseriformes[35] and the relative binding intensities of MAA and SNA-I lectins demonstrated similar trends in this study (Supplementary Fig. S5).

The binding of influenza virus depends on the specificity of its surface hemagglutinin for the particular linkage of the terminal sialic acid on lung epithelial cell surface oligosaccharides. This linkage tropism dictates the species specificity of influenza strains as avian influenza virus favours binding to α-(2,3)-linked sialic acid which is found in the upper respiratory tract of birds and human influenza virus binds to α-(2,6)-linked sialic acid found in the human upper respiratory tract[29]. Many vaccines, including those against influenza, are grown in chicken eggs and it has been shown that the egg component glycosylation influences the specificity of the produced influenza virus[39]. Thus, it is beneficial to understand species-

specific EW glycosylation, including the presence or absence of certain motifs and linkages to produce a vaccine better formulated for human pathogens.

Correlation of EW glycoprofiles with avian phylogeny. Glycan composition was previously shown to be dependent on the phylogenetic relationship between the species[16,35]. According to the current classification, modern birds are divided into three taxa, Ratitae, Galloanserae, and Neoaves (Supplementary Fig. S6)[16]. Based on that classification, the expression of terminal α-linked Gal in the terminal structure Gal-α-(1,4)-Gal[20,35,40] was confined to the species from Neoaves. In a previous assessment of 148 avian species EWs, approximately 67% of EWs were shown to express Gal-α-(1,4)-Gal-containing glycoproteins based on lectin (GS-I) and Western blotting (anti-P1 monoclonal antibody) analysis[16]. In addition, difference in galactosyltransferase (GalT) expression was reported in various tissues of ostrich, chicken and pigeon from the taxa Ratitae, Galloanserae and Neoaves, respectively[41]. The β-(1,4)-GalT, which forms the type II LacNAc structure, was present in all three birds while the α-(1,4)-GalT, which forms the terminal structure Gal-α-(1,4)-Gal, was present in pigeon only[41].

In this work (Fig. 3, C2), the EWs with high binding intensity to Gal-specific lectins (SBA, GS-I-B4, PA-I) were from closely related orders of Psittaciformes, Strigiformes, Columbiformes, Passeriformes and Ciconiiformes, all members of Neoaves which represent neighbouring branches of the phylogenetic tree based on DNA relationships (Supplementary Fig. S6). However, Cuculiformes, Gruiformes and Musophagiformes, which are in close phylogenetic proximity to the aforementioned orders, showed low or no binding to PA-I and GS-I-B4. Interestingly, the EWs from Turniciformes, Piciformes and Coraciiformes, despite being the part of Neoaves but phylogenetically more distinct than Cuculiformes, Gruiformes and Musophagiformes, did not interact with PA-I and GS-I-B4 at all. In addition, no EWs

from the superorders Ratitae and Galloanserae interacted with lectins PA-I and GS-I-B4, in agreement with a previous report[16].

Overall, the hierarchically clustered lectin glycoprofiling analysis (Fig. 3) revealed that the EWs from all species included did not group strictly according to phylogeny. However, several glycosylation clusters and subclusters contained multiple members of one order, e.g. C5b was composed of seven members of Ciconiiformes, five members of Galliformes and two members of Anseriformes. While Galliformes and Anseriformes are phylogenetically close, Ciconiiformes is more phylogenetically distant (Supplementary Fig. S6). A high prevalence of EWs from these three orders, together with one representative from Turniciformes, was also found in C4c. Thus, the detailed glycosylation and hierarchical clustering analyses demonstrated that the similarity of EW glycosylation among different orders of birds is not strictly dependent on the phylogenetic relationship between the species. Previous conclusions that glycosylation is related to phylogeny was based on only one carbohydrate structure, terminal Gal-α-(1,4)-Gal[16].

HTP screening of EWs for pathogen-binding. The EW binding of five bacterial strains which can form pili or biofilm was investigated (Table 2). The bacterial strains included were two non-pathogenic *E. coli* type strains, the Type 1 fimbriae-producing ATCC 35218 and the P pili-producing ATCC 25922, which bind to Man residues and Gal-α-(1,4)-Gal, respectively, *V. parahaemolyticus* RIMD2210633, which is a seafood-borne pathogen known to express type IV pili[42], and two strains of *S. aureus*. *S. aureus* is a common food spoilage pathogen[3] but the strains selected for this study are not food isolates but were selected for their biofilm forming ability. Biofilm formation is a source of persistent contamination in the food processing industry[43] and persistent infection, for example, in lungs of cystic fibrosis patients or on implanted medical devices. *S. aureus* 8325-4 is a methicillin-susceptible isolate

which produces a polysaccharide biofilm[44] while BH1CC is methicillin-resistant and forms a proteinaceous biofilm[45]. Bacterial cells were harvested under conditions optimal for expression of pili or biofilm, fluorescently labelled and incubated on the EW microarrays. The mean binding intensity data was hierarchically clustered (Supplementary Table S3 and Fig. 4) and three main clusters with 0.5 minimum similarity were identified.

All selected bacterial strains demonstrated highest relative affinity for three EWs (Supplementary Table S3), EAW, which interacted with Con A, GS-II and Man-specific lectins, and LGK and EMU, which bound most intensely to complex *N*-linked structure-specific lectins and MAA (Fig. 3). Most EWs which exhibited high intensity interactions with all bacteria were grouped in cluster IIIa and the right- and left- hand side of cluster IIIb, with less interactions in the middle of cluster IIIb (Fig. 4). *S. aureus* strains interactions were mainly within the right-hand side portion of cluster IIIb, while the left-hand side of cluster IIIb had higher intensity interactions with the *E. coli* strains and moderate binding to the *S. aureus* strains.

Four EWs with high Man type structures indicated, EAW, DEW, QEW and BGE, were in the top 15% of most intensely binding EWs for *E. coli* ATCC 35218 (Supplementary Table S4), in agreement with the expected carbohydrate specificity of this strain. For *E. coli* ATCC 25922, the top 30% of the most intensely binding EW features included four EWs from the C2 cluster (Fig. 3), CFI, BOW, CRC and BGR, which bound intensely with Gal-specific lectins, including GS-I-B4. This affinity for terminal α-linked Gal was expected for strain ATCC 25922 but interestingly, BRC EW was also among the most intensely bound EWs for this strain. BRC was located in C5c (Fig. 3) and had moderate binding with SBA and PA-I, extremely low binding with GS-I-B4, and high binding with PHA-E and PHA-L. Therefore it is unlikely that BRC has any terminal α-linked Gal but rather terminal β-linked Gal or

GalNAc. This observation indicates that strain ATCC 25922 may have multiple adhesins with lectin functionality or that the P pili has wider specificity than Gal-α-(1,4)-Gal alone, which exemplifies the additional advantage of the use of a natural product displaying multiple complex structures as a potential anti-adhesive.

V. parahaemolyticus demonstrated high binding intensity with WCT EW, which interacted with GSL-II but not with PHA-L and PHA-E. Other intensely *V. parahaemolyticus* binding EWs, including EMU, CGC, ASC and LGK, belonged to C5 (Fig. 3) and were bound mainly by PHA-L and PHA-E. The high Man-type structure containing EWs DEW, EAW and BGE, all from Anseriformes, were also found within the top 10% of intensely binding EWs for *V. parahaemolyticus*, and binding could potentially be due to the expression of the mannose-sensitive haemagglutinin (MSHA), which has lectin functionality[42].

To our knowledge, there are no reports indicating lectin functionality of the surface proteins of the *S. aureus* strains. However, these strains bound to EWs from clusters C4 and C5, known to bind to i) Con A, GS-II and Man-specific lectins (BUQ, EAW, SCQ, SVP, RED, BGE, QEW and DEW), ii) Con A and GS-II (CPI, RIT, WCT, CPF), iii) Con A, PHA-L and PHA-E (CTP, PWD and TKY), iv) MAA, PHA-L and PHA-E (LGK and EMU), and v) WGA, PHA-L and PHA-E (MLF, CTI and GTI). Thus, these *S. aureus* strains may bind to multiple structures including high mannose and complex type with terminal GlcNAc residues and terminal α-(2,3)-linked sialic acids, indicating the presence of several lectins. Interestingly, for strain BH1CC, CFI, YKT and BOW EWs were found within the top 30% of the most intense interactions and can help distinguish BH1CC from the less intense interactions of 83252-4 with these EWs. CFI, YPT and BOW were recognized by the Gal-specific lectins SBA, GS-I-B4 and PA-I and were grouped in C2 (Fig. 3) and cluster II (Fig. 4), so BH1CC may have additional specificity for terminal α- and β-linked Gal structures.

However, it is unlikely that the Gal-α-(1,4)-Gal motif was selectively recognised by BH1CC as this strain did not similarly intensely bind to the other seven EWs recognised by subunit B of Stx1 (Stx1-B) (see below for further detail). Nevertheless, further confirmation of specific carbohydrate-mediated binding of *S. aureus* adhesins is still required.

The presence of certain glycan structures in EWs of selected avian species suggests their inhibitory role for bird pathogen adhesion. Liu, *et al.*[46] used pigeon ovalbumin with terminal Gal-α-(1,4)-Gal for detection of uropathogenic bacteria in urine samples. Extracts from edible bird's nest (regurgitated saliva of male *Collocalia swiftlets*), widely consumed by humans as a delicacy and a naturopathic food, strongly inhibit the hemagglutination of human erythrocytes by human, avian and porcine influenza viruses in a host-range-independent manner[47], possibly due to the high abundance of sialylated high-antennary *N*-linked oligosaccharides[48]. Because wild birds play an important role in the epidemiology of human-associated zoonoses[49], an understanding of the mechanisms of their immunity as carriers would be highly beneficial for future developments of disease prevention and treatments.

Binding of bacterial toxins to the EW microarray. The EW binding of two bacterial toxins was investigated (Table 2). Bacterial cholera toxin subunit B (Ctx-B) did not show any significant interaction with any feature on the EW microarray (Supplementary Table S3), except a GM1 monosialoganglioside (Gal-β-(1,3)-GalNAc-β-(1,4)-[NeuAc-α-(2,3)-]Gal-β-(1,4)-Glc-β-Cer)-containing neoglycoconjugate (GM1-HSA) which was included as a positive control[50]. The lack of Ctx-B binding indicated that the GM1 carbohydrate structure was not present in any of the arrayed EWs.

Stx1-B bound to ten EWs above the stringent five times background threshold[19,51] (Fig. 4) including i) four representatives of Passeriformes, CFI, CNS, MWS and ROB EWs, ii) two representatives of Strigiformes, BOW and SNO, iii) two representatives of Psittaciformes,

TPT and YPT, iv) one representative from Ciconiiformes, BRK; and v) one from Columbiformes, PEW. The binding of Stx1-B indicated the presence of Gal-α-(1,4)-Gal on the EWs, which is the minimum structure required for Stx1-B binding[17]. Interestingly, all ten EWs were in C2 (Fig. 3), showed high binding intensity with SBA, PA-I and GS-I-B4 and are grouped together as taxonomically close orders (Supplementary Fig. S6).

Previously, PEW was shown to be a rich source of Gal-α-(1,4)-Gal-β-(1,4)-GlcNAc (also known as the P1 antigen) terminating structures[16,21]. Pigeon ovomucoid, pigeon ovalbumin and whole PEW immobilised on Sepharose gels was used for purification of Stx1 from the crude extracts of *E. coli* SLT100[17]. Here, PEW had only moderate binding intensity with Stx1-B and ten other EWs had greater binding intensity. This could be due to a more favourable multivalent presentation of the Gal-α-(1,4)-Gal-containing determinants on the EWs with greater binding intensity for engaging the multiple binding sites of the Stx1-B pentamer. Hence, not only PEW but also the other ten EWs interacting with Stx1-B identified in this study may have potential in the development of alternative and highly efficient methods for neutralisation of toxins and other anti-adhesive strategies.

In conclusion, the complex glycosylation profiles provided in this work demonstrated that the similarity of EW glycosylation among different orders of birds is not strictly dependent on the phylogenetic relationship between the species as previously suggested[16]. The dominating linkages of sialic acid in the EW library were determined for the first time using the lectins MAA and SNA-I. The binding of five bacterial strains which form pili or biofilm was screened and a number of EWs were identified as potential anti-adhesives for these pathogens, along with the potential structures recognised by *S. aureus* strains for the first time. Two bacterial toxins were profiled and Stx1 was shown to bind to ten EWs with similar glycosylation with greater intensity than previously identified PEW. Using HTP technology

for screening pathogen interactions with the 78 EWs from different avian species could increase understanding of carbohydrate-based anti-bacterial mechanisms of egg white components and species-specific antimicrobial defence. This platform provides a valuable opportunity to promote the exploitation of natural products as anti-adhesives that could be targeted for strain and species specificity in the bid for discovery of alternative antimicrobial agents.

Methods

Materials. The lyophilised avian EW library was provided by Prof. Y. C. Lee (Johns Hopkins University, USA) (Supplementary Table S1). CEW, QEW and DEW separated from eggs were obtained from local providers (farmers market) and lyophilised. CEW OVA, OVT, OVM, LYZ, and glycoprotein standards transferrin, yeast invertase, ASF, A1AT, AGP and mouse monoclonal anti-6X His IgG-CF640R antibody were from Sigma-Aldrich (Dublin, Ireland). Neoglycoconjugate GM1-HSA was obtained from IsoSep AB (Tullinge, Sweden). TRITC-labelled lectins were from EY Laboratories (San Mateo, CA) or Vector Laboratories, Ltd. (Orton Southgate, United Kingdom). PA-I and WGA were labelled with AlexaFluor® 555 (Life Technologies, USA) according to the manufacturer's protocol. Cholera toxin subunit B (CT-B) conjugated with AlexaFluor® 555, NuPAGE® Novex® 10% and 4-12% Bis-Tris polyacrylamide gels and SeeBlue® Plus2 pre-stained molecular mass protein standards were purchased from Life Technologies (Carlsbad, CA). Shiga-like toxin B (SLT-1B) with a polyhistidine tag (6x His) attached at its N-terminus was from Nova Biotech Development (El Cajon, CA) (Table 2). The silver stain kit was from Pierce (Thermo Fisher Scientific, Dublin, Ireland). All other materials were from Sigma-Aldrich Co. unless otherwise indicated and were of the highest grade available.

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis. EW solutions (Supplementary Table S1, 5 mg/ml) were prepared in PBS, pH 7.4 and mixed by inversion (4 rpm) at room temperature for 1 h. Samples were first centrifuged for 10 min at 4,000 rpm, the floating layer of aggregate was removed and the sample was centrifuged again for 10 min at 14,000 rpm. Samples were filtered through a 0.2 μm centrifugal filter with PVDF membrane which was washed with PBS before use. The protein content of the final preparation was ascertained using the bicinchonic acid assay. All EW solutions were adjusted to protein concentrations of 1.2 mg/ml in PBS while standards were adjusted to 1 mg/ml in PBS. EWs (1-3 μg per lane) and purified chicken standards (OVA, OVM, OVT and LYZ, 0.5 μg of each) were electrophoresed on NuPAGE® Novex® 10% or 4-12% Bis-Tris polyacrylamide gels under reducing conditions using MOPS buffer at 150 V constant. Protein migration was visualised by silver staining the gels.

EW microarray construction. Microarray printing was performed essentially as previously described[19,31] with minor modifications as follows. Microarrays were printed with a SciFLEXARRAYER S3 (Scienion) piezoelectric printer equipped with a 90 μm glass nozzle with a hydrophobic coating. Crude EWs (probes) were printed onto Nexterion® Slide H microarray slides at 0.6 mg/mL in PBS containing 0.01% Tween 20 and OVA and OVM were printed at 0.5 mg/mL in PBS containing 0.015% Tween 20. Eight identical subarrays were printed per slide, with each subarray containing six replicates each of 52 different probes. The EW library was divided into twinned panels (A and B) for printing, with each panel consisting of EWs from 36 different species of birds and 17 overlapping probes per panel printed at identical locations (Supplementary Table S2). The overlapping probes consisted of six EWs (EAW, PEW, QEW, CEW, DEW, GEW), four CEW glycoproteins (OVA, OVM, OVT and LYZ), six standard glycoproteins (bovine transferrin, yeast invertase,

bovine asialofetuin (ASF), bovine fetuin, human α-1-antitrypsin (A1AT) and human α-1-acid glycoprotein (AGP) and one neoglycoconjugate (GM1-BSA).

Microarray incubation and scanning. For incubations, carried out in triplicate, fluorescently-labelled lectins (Table 1) and toxins (Table 2) were diluted in Tris buffered saline supplemented with divalent cations (TBS; 20 mM Tris-HCl, 100 mM NaCl, 1 mM $CaCl_2$, 1 mM $MgCl_2$, pH 7.2) with 0.05% Tween 20 (TBS-T). All labelled lectins and toxins were first titrated on the EW microarrays to determine the optimum concentration such that the resulting fluorescence intensity of the bound lectins was below saturation, approximately 65,500 relative fluorescence units (RFU), with minimal background (Table 1). Inhibitions were performed in parallel by diluting lectins in solutions of the haptenic sugar or glycoprotein in TBS-T (Table 1) and pre-incubated for 20 min at room temperature. Microarrays were incubated at 23 °C for 1 h, washed and dried as previously described[19,31]. After drying, microarray slides were scanned immediately in an Agilent G2505B microarray scanner (532 nm laser, 90% PMT and 5 μm resolution). Images were saved as *.tif files for data extraction. For the His-tagged Stx1-B, the slides were washed and dried after incubation with Stx1-B and then incubated immediately afterwards with AlexaFluor® 647-labelled anti-His antibody (0.5 μg/ml TBS-T). The rest of the procedure was carried out as described above, except using the 633 nm laser for scanning.

Bacterial culture. Bacterial strains were obtained from ATCC (*E. coli* type strains ATCC 35218 and 25922) or as a donation (Table 2). The strains were routinely grown in LB broth and LB agar (*E. coli* type strains) or in BHI broth and BHI agar (*V. parahaemolyticus* RIMD2210633 and *S. aureus* strains 8325-4 and BH1CC). Selected strains were first grown for 16 h (overnight) static at 37 °C in 5 mL liquid culture in a 50 mL plastic tube. Then overnight culture of each strain was inoculated into fresh medium to starting OD_{600nm} 0.05

and grown to mid-exponential phase (OD_{600nm} 0.6). Bacteria were harvested by centrifugation (5,000 x g, 5 min, 20 °C), washed twice in TBS and resuspended in TBS at half of the initial culture volume at an OD_{600nm} of 2.0 (2 x 10^9 cfu/ mL) and 0.5 ml aliquots were prepared.

Fluorescent staining of bacteria and incubation on EW microarrays. All following steps were carried out with limited light exposure. The optimum concentration of SYTO® 82 fluorescent cell-permeable nucleic acid dye was determined for each strain as previously described[52]. Bacteria aliquots were incubated at 37 °C for 1 h with rotation (200 rpm) with 10 μM SYTO® 82 and then washed seven times in 2 mL of TBS (3,000 rpm, 2 min) each wash to remove excess dye. Finally cells were resuspended in 0.5 mL of TBS with 0.05% Tween-20 (TBS-T) and adjusted to an OD_{600nm} of 1.0 for immediate use on the microarrays. Each labelled strain was titrated to determine optimal dilution and then incubated in triplicate on the EW microarrays as described above.

Data extraction and analysis. Data extraction and analysis was performed as previously described[19,31]. Local background-corrected median feature intensity data (F543median-B543) was analysed. The median of six replicate spots per subarray was handled as a single data point for graphical and statistical analysis. Data from the twinned panels and three replicates were normalised across all six microarray slides (three replicates of twinned microarrays) to the per-subarray total intensity mean of the 17 overlapping probes (Supplementary Table S3). Binding data was presented in histogram form of mean intensity with error bars of one standard deviation of all replicates. Unsupervised clustering of the data was also performed using Hierarchical Clustering Explorer v3.0 (HCE 3.0, University of Maryland, http://www.cs.umd.edu/hcil/hce/bec3.html).

References

1. Brown, E. D. & Wright, G.D. Antibacterial drug discovery in the resistance era. *Nature*, **529**, 336-343, (2016).
2. Silbergeld, E. K., Graham, J. & Price, L. B. Industrial food animal production, antimicrobial resistance, and human health. *Annu. Rev. Public Health*, **29**, 151-169, (2008).
3. Lucera, A., Costa, C., Conte, A. & Del Nobile, M.A. Food applications of natural antimicrobial compounds. *Front. Microbiol*. **3**, Article 287, (2012).
4. Ofek, I., Hasty, D. L. & Sharon, N. Anti-adhesion therapy of bacterial diseases: prospects and problems. *FEMS Immunol. Med. Microbiol.* **38**, 181-191, (2003).
5. Krachler, A. M. & Orth, K. Targeting the bacteria-host interface: Strategies in anti-adhesion therapy. *Virulence*, **4**, 284-294, (2013).
6. Källenius, G., Winberg, J., Möllby, R., Svenson, S. B. & Hultberg, H. Identification of a carbohydrate receptor recognized by uropathogenic *Escherichia coli*. *Infection*, **8**, S288-S293 (1980).
7. Jacewicz, M., Clausen, H., Nudelman, E., Donohue-Rolfe, A. & Keusch, G. T. Pathogenesis of shigella diarrhea. XI. Isolation of a shigella toxin-binding glycolipid from rabbit jejunum and HeLa cells and its identification as globotriaosylceramide. *J. Exp. Med*. **163**, 1391-1404 (1986).
8. Samuel, J. E. *et al*. Comparison of the glycolipid receptor specificities of Shiga-like toxin type II and Shiga-like toxin type II variants. *Infect. Immun*. **58**, 611-618 (1990).
9. Müller, C., Despras, G. & Lindhorst, T. K. Organising multivalency in carbohydrate recognition. *Chem. Soc. Rev*. **45**, 3275-3302, (2016).

10. Kovacs-Nolan, J., Phillips, M. & Mine, Y. Advances in the value of eggs and egg components for human health. *J. Agric. Food Chem.* **53**, 8421–8431 (2005).

11. Plasencia, M. D., Isailovic, D., Merenbloom, S. I., Mechref, Y. & Clemmer, D. E. Resolving and sssigning N-linked glycan structural isomers from ovalbumin by IMS-MS. *J. Am. Soc. Mass Spectrom.*, **19**, 1706–1715 (2008).

12. Harvey, D. J., Wing, D. R., Küster, B. & Wilson, I. B. H. Composition of N-linked carbohydrates from ovalbumin and co-purified glycoproteins. *J. Am. Soc. Mass Spectrom.*, **11**, 564–571 (2000).

13. Hwang, H. S. *et al*. Type and branched pattern of N-glycans and their structural effect on the chicken egg allergen ovotransferrin: A comparison with ovomucoid. *Glycoconj. J.*, **31**, 41–50 (2014).

14. Yet, M. G., Chin, C. C. Q. & Wold, F. The covalent structure of individual N-linked glycopeptides from ovomucoid and asialofetuin. *J. Biol. Chem.* **263**, 111–122 (1988).

15. Varon, O., Allen, K. J., Bennett, D. C., Mesak, L. R. & Scaman, C. H. Purification and characterization of tinamou egg white ovotransferrin as an antimicrobial agent against foodborne pathogenic bacteria. *Food Res. Int.* 54, 1836-1842 (2013).

16. Suzuki, N., Laskowski, M. & Lee, Y. C. Phylogenetic expression of Galalpha1-4Gal on avian glycoproteins: glycan differentiation inscribed in the early history of modern birds. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 9023–9028 (2004).

17. Tomoda, H. *et al*. Purification of Shiga-like toxin 1 by pigeon egg white glycoproteins immobilized on Sepharose gels. *Anal. Biochem.* **311**, 50–56 (2002).

18. Yagi, H. *et al*. N-Glycosylation profiling of turtle egg yolk: expression of galabiose structure. *Carbohydr. Res.* **345**, 442–448 (2010).

19. Kilcoyne, M. *et al*. Construction of a natural mucin microarray and interrogation for biologically relevant glyco-epitopes. *Anal. Chem.* **84**, 3330–3338 (2012).

20. Takahashi, N., Khoo, K. H., Suzuki, N., Johnson, J. R. & Lee, Y. C. N-Glycan structures from the major glycoproteins of pigeon egg white: Predominance of terminal Gal-α-(1-4)-Gal. *J. Biol. Chem.* **276**, 23230–23239 (2001).

21. Suzuki, N., Khoo, K. H., Chen, H. C., Johnson, J. R. & Lee, Y. C. Isolation and characterization of major glycoproteins of pigeon egg white: Ubiquitous presence of unique N-glycans containing Galalpha1-4Gal. *J. Biol. Chem.* **276**, 23221–23229 (2001).

22. Offengenden, M., Fentabil, M. A. & Wu, J. N-glycosylation of ovomucin from hen egg white. *Glycoconj. J.* **28**, 113–123 (2011).

23. Strecker, G., Wieruszeski, J., Martel, C. & Montreuil, J. Determination of the structure of sulfated tetra- and pentasaccharides obtained by alkaline borohydride degradation of hen ovomucin. A fast atom bombardment-mass spectrometric and 1H-NMR sprectroscopic study. *Glycoconj. J.* **4**, 329-337 (1987).

24. Strecker, G., Wieruszeski, J.-M., Cuvillier, O., Michalski, J. C. & Montreuil, J. 1H and 13C-NMR assignments for sialylated oligosaccharide-alditols related to mucins. Study of thirteen components from hen ovomucin and swallow nest mucin. *Biochimie*, **74**, 39-52 (1992).

25. Schaffer, S., Isci, N., Zickner, B. & Dürre, P. Differential staining of Western blots of avian egg white glycoproteins using diverse lectins. *Electrophoresis,* **23**, 8–14 (2002).

26. Gerlach, J. Q., Kilcoyne, M., Eaton, S., Bhavanandan, V. & Joshi, L. Non-carbohydrate-mediated interaction of lectins with plant proteins. *Adv. Exp. Med. Biol.* **705**, 257–269 (2011).

27. Yu, Y. *et al*. Functional glycomic analysis of human milk glycans reveals the presence of virus receptors and embryonic stem cell biomarkers. *J. Biol. Chem.* **287**, 44784–44799 (2012).

28. Naughton, J. A. *et al*. Divergent mechanisms of interaction of *Helicobacter pylori* and *Campylobacter jejuni* with mucus and mucins. *Infect. Immun.* **81**, 2838–2850 (2013).

29. Matrosovich, M. N. *et al*. Avian influenza A viruses differ from human viruses by recognition of sialyloligosaccharides and gangliosides and by a higher conservation of the HA receptor-binding site. *Virology* **233**, 224–234 (1997).

30. Miguel, M., Manso, M. A., López-Fandiño, R. & Ramos, M. Comparative study of egg white proteins from different species by chromatographic and electrophoretic methods. *Eur. Food Res. Technol.* **221**, 542–546 (2005).

31. Gerlach, J. Q., Kilcoyne, M. & Joshi, L. Microarray evaluation of the effects of lectin and glycoprotein orientation and data filtering on glycoform discrimination. *Anal. Methods* **6**, 440 (2014).

32. Gerlach, J. Q., Kilcoyne, M., Farrell, M. P., Kane, M. & Joshi, L. Differential release of high mannose structural isoforms by fungal and bacterial *endo*-β-*N*-acetylglucosaminidases. *Mol. Biosyst.* **8,** 1472-1481 (2012).

33. Yamashita, K., Ueda, I. & Kobata, A. Sulfated asparagine-linked sugar chains of hen egg albumin. *J. Biol. Chem.* **258**, 14144–14147 (1983).

34. Van Damme, E. J. M., Peumans, W. J., Barre, A. & Rougé, P. Plant lectins: A composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. *CRC. Crit. Rev. Plant Sci.* **17**, 575-692 (1998).

35. Hirose, K., Amano, M., Hashimoto, R., Lee, Y. C. & Nishimura, S. I. Insight into glycan

616 diversity and evolutionary lineage based on comparative avio-N-glycomics and sialic
617 acid analysis of 88 egg whites of galloanserae. *Biochemistry,* **50**, 4757–4774 (2011).

618 36. Ivarie, R. Competitive bioreactor hens on the horizon. *Trends Biotechnol.* **24**, 99–101
619 (2006).

620 37. Rupa, P., Nakamura, S. & Mine, Y. Genetically glycosylated ovomucoid third domain
621 can modulate Immunoglobulin E antibody production and cytokine response in BALB/c
622 mice. *Clin. Exp. Allergy* **37**, 918–928 (2007).

623 38. Hwang, H. S. *et al*. Cleavage of the terminal N-acetylglucosamine of egg-white
624 ovalbumin N-glycans significantly reduces IgE production and Th2 cytokine secretion.
625 *Biochem. Biophys. Res. Commun.* **450**, 1247–1254 (2014).

626 39. Ito, T. *et al*. Differences in sialic acid-galactose linkages in the chicken egg amnion and
627 allantois influence human influenza virus receptor specificity and variant selection. *J.*
628 *Virol.* **71**, 3357–3362 (1997).

629 40. Suzuki, N. *et al*. Structural analysist of N-glycans from gull egg white glycoproteins and
630 egg yolk IgG. *Glycobiology* **19**, 693–706 (2009).

631 41. Suzuki, N., Nawa, D. & Yamamoto, K. Distinct expression profiles of UDP-galactose:
632 beta-D-galactoside alpha1,4-galactosyltransferase and UDP-galactose: beta-D-
633 galactoside beta1,4-galactosyltransferase in pigeon, ostrich and chicken. *Glycobiology*,
634 **21**, 283–294 (2011).

635 42. O'Boyle, N., Houeix, B., Kilcoyne, M., Joshi, L. & Boyd, A. The MSHA pilus of *Vibrio*
636 *parahaemolyticus* has lectin functionality and enables TTSS-mediated pathogenicity.
637 *Int. J. Med. Microbiol.* **303**, 563–573 (2013).

638 43. Van Houdt, R. & Michiels, C.W. Biofilm formation and the food industry, a focus on

639 the bacterial outer surface. *J. Appl. Microbiol*. **109**, 1117-1131 (2010).

640 44. Horsburgh, M. J. *et al*. SigB modulates virulence determinant expression and stress
641 resistance: Characterization of a functional rsbU strain derived from *Staphylococcus*
642 *aureus* 8325-4. *J. Bacteriol.* **184**, 5457–5467 (2002).

643 45. O'Neill, E. *et al*. Association between methicillin susceptibility and biofilm regulation in
644 *Staphylococcus aureus* isolates from device-related infections. *J. Clin. Microbiol.* **45**,
645 1379–1388 (2007).

646 46. Liu, J., Tsai, P., Lee, Y. C. & Chen, Y. Affinity capture of uropathogenic *Escherichia*
647 *coli* using pigeon ovalbumin-bound Fe_3O_4@Al_2O_3 magnetic nanoparticles. .*Anal. Chem.*
648 **80**, 5425–5432 (2008).

649 47. Guo, C. T. *et al*.. Edible bird's nest extract inhibits influenza virus infection. *Antiviral*
650 *Res.* **70**, 140–146 (2006).

651 48. Yagi, H. *et al*. The expression of sialylated high-antennary N-glycans in edible bird's
652 nest. *Carbohydr. Res.* **343**, 1373–1377 (2008).

653 49. Reed, K. D., Meece, J. K., Henkel, J. S. & Shukla, S. K. Birds, migration and emerging
654 zoonoses: West Nile virus, Lyme disease and influenza A and enteropathogens. *Clin.*
655 *Med. Res.* **1**, 5–12 (2003).

656 50. Merritt, E. A. *et al*. Crystal structure of cholera toxin B-pentamer bound to receptor
657 GM1 pentasaccharide. *Protein Sci.* 3, 166–175 (1994).

658 51. Manimala, J. C., Roach, T. A., Li, Z. & Gildersleeve, J. C. High-throughput
659 carbohydrate microarray profiling of 27 antibodies demonstrates widespread specificity
660 problems. *Glycobiology*, **17**, 17C–23C (2007).

661 52. Kilcoyne, M. *et al*. *Campylobacter jejuni* strain discrimination and temperature-

dependent glycome expression profiling by lectin microarray. *Carbohydr. Res.* **389**, 123–133 (2014).

53. Laskowski, M. Jr. *et al*. Ovomucoid third domains from 100 avian species: isolation, sequences, and hypervariability of enzyme-inhibitor contact residues. *Biochem*. **26**, 202-221, (1987).

54. Laskowski, M. Jr. *et al*. Amino acid sequences of ovomucoid third domains from 25 additional species of birds. *J. Protein Chem*. **9**, 715-725, (1990).

55. Apostol, I., Giletto, A., Komiyama, T., Zhang, W. & Laskowski, M. Jr. Amino acid sequences of ovomucoid third domains from 27 additional species of birds. *J. Protein Chem*. **12**, 419-433, (1993).

Acknowledgements

This work was funded by Science Foundation Ireland (SFI) grant numbers 11/TIDA/I2044 in support of a Technology Innovation Development Award, 08/SRC/B1393 in support of the Alimentary Glycoscience Research Cluster and 07/SK/B1250 for Stokes Professor for Glycosciences (LJ), and the European Union FP7 grant number 260600 in support of the GlycoHIT consortium. The authors thank Prof. James O'Gara and Ms Andrea Flannery for the gift of *S. aureus* strains and Dr. Aoife Boyd for *V. parahaemolyticus*. Y.C.L. thanks the late Dr. M. Laskowski, Jr. for generously providing the invaluable egg white collection, which he was gifted over the course of many years by various donors including zoos, research institutes and game preserves (further detailed in [53-55]). Bird eggs should not be collected from the wild without ethical approval from the relevant national or international governing body.

Author contributions

J.Q.G., Y.C.L., M.Kane, M.K. and L.J. conceived and designed the study. M.U. and H.A. performed the majority of experiments. M.U., H.A., J.Q.G. and M.K. participated in data analysis. Y.C.L., M.Kane, M.K. and L.J. provided logistical and scientific support for the study. M.U., H.A., J.Q.G. and M.K. drafted the manuscript. All authors have read and approved the final manuscript.

Competing financial interests

The authors declare no conflict of interest or competing financial interest.

Data availability statement

The authors are willing to make data available upon request.

Legends to figures

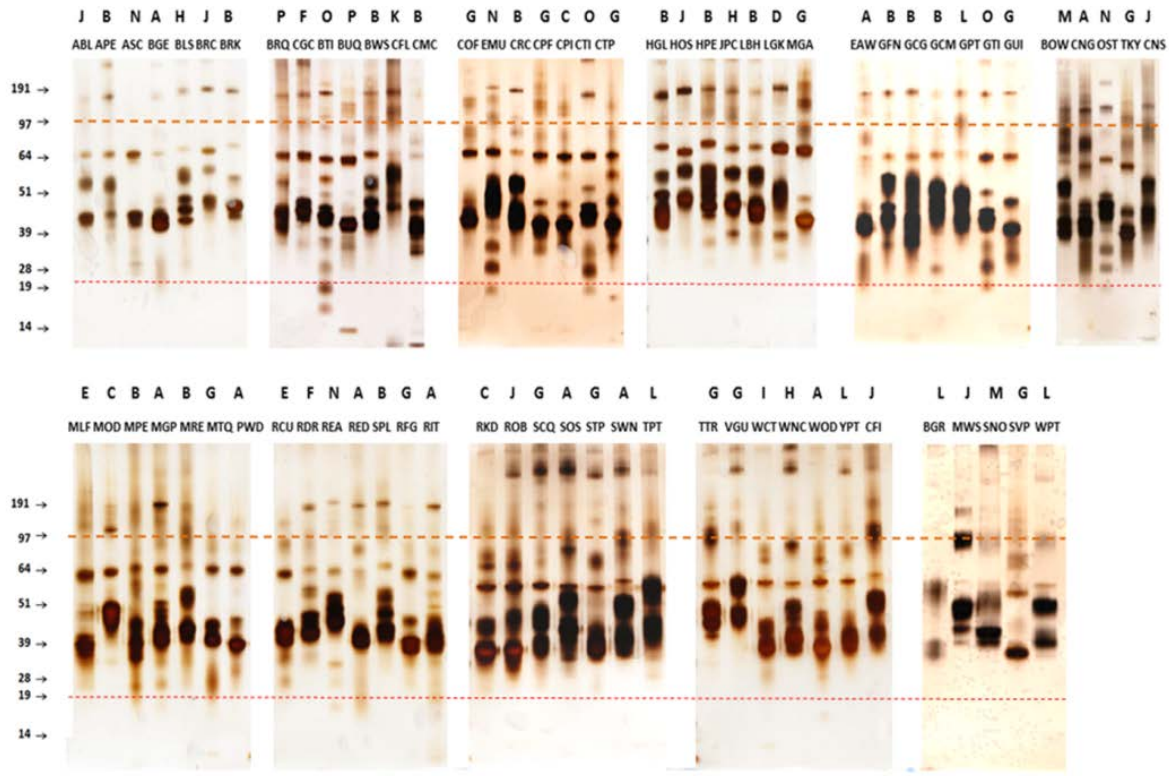
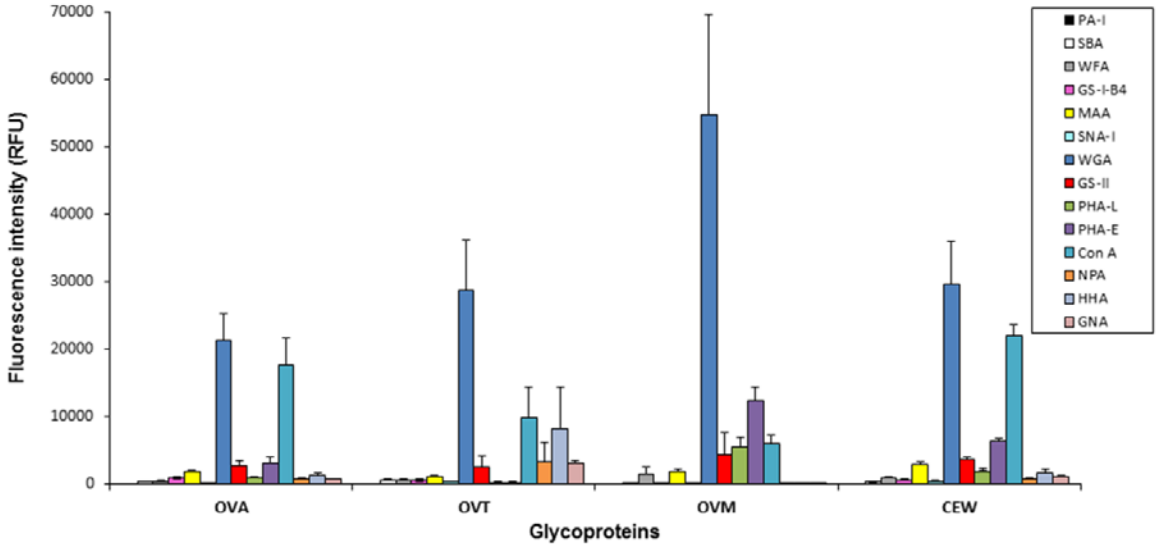


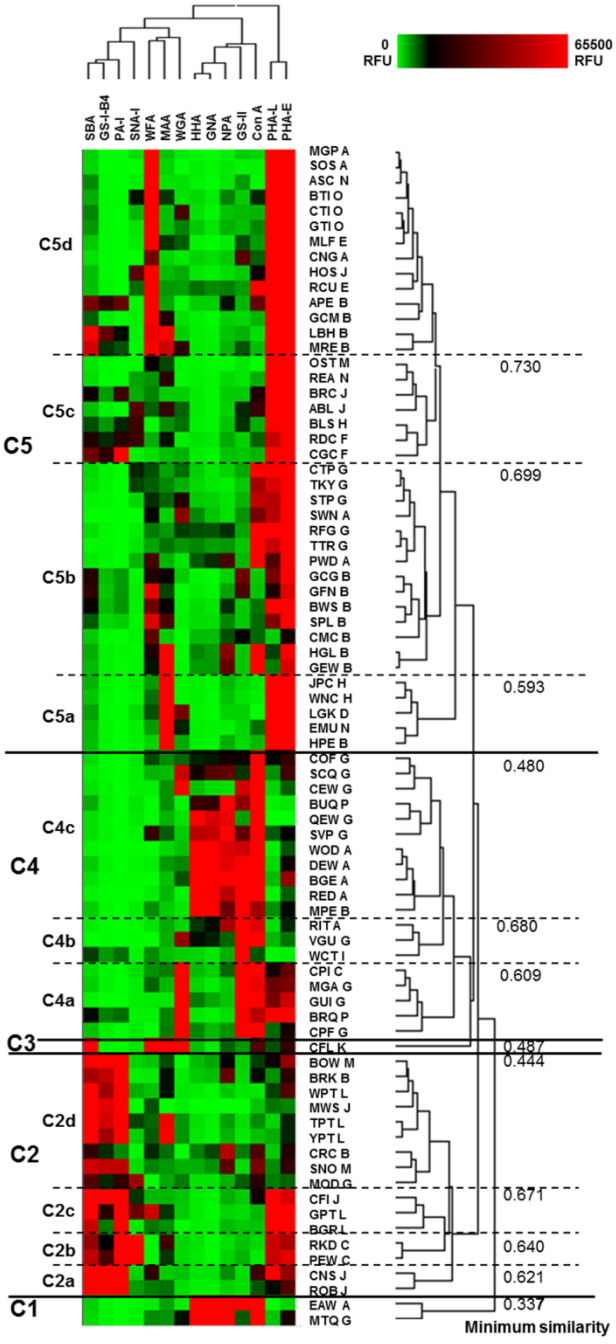
Figure 1. Electrophoretic profiles of EWs (with exception of CEW, DEW, GEW, PEW and QEW). EWs (1-3 μg per lane) were separated on NuPage 4-12% Bis-Tris gels and silver stained. The single capital letters above the gels illustrate the systematic order of birds while triple letters indicate abbreviated common names of birds (Table S-1). In each gel, 5 μL of molecular mass marker and 0.5 μg of purified chicken (CH) protein standards (OVT, OVA, OVM and LYZ) were separated (not shown for simplification). For comparison, the gels were aligned based on the migration of two marker bands: 97 (orange line) and 19 kDa (red line).



708

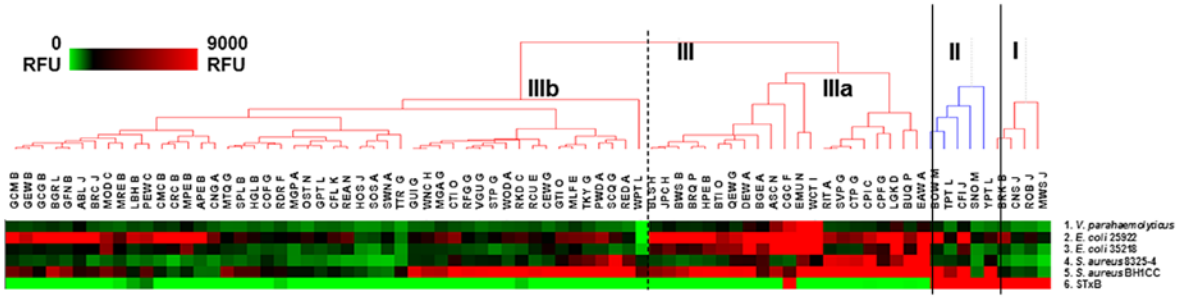
709 **Figure 2.** Bar chart representing the mean fluorescence intensity from three replicate
710 experiments of 14 individual lectins with carbohydrate-specific interactions binding to three
711 chicken EW glycoproteins (OVA, OVT and OVM) and total CEW on the EW microarray.
712 Error bars are one standard deviation for the mean of all replicates (Table S-3).

713



714

715 **Figure 3.** Clustering of the binding interactions of 14 lectins with the 78 EWs. Heat map with
716 dendrograms of hierarchical clustering generated using HCE 3.0. Fluorescence intensities
717 correspond to colour as in the legend. Species names are abbreviated and shown with a
718 capital letter denoting the systematic order as in Table S-1. Five clusters of birds (C1-C5) are
719 indicated by solid lines based on the minimum similarity by glycoprofile located at position
720 0.5 for major clusters and 0.75 for subclusters (broken lines).

721 

722 **Figure 4.** Interactions of the bacteria and Stx-1B with EWs. Heat map with dendrograms of
723 hierarchical clustering generated using HCE 3.0. Fluorescence intensities correspond to
724 colour as in the legend. Species names are abbreviated (Table S-1). Three groups of birds (I-
725 III) are indicated by solid lines based on the minimum similarity by glycoprofile located at
726 position 0.5 for major clusters and 0.75 for subclusters IIIa and IIIb (broken line).

727

728 **Table 1**. Lectins used, their abbreviation, source species, binding specificity, concentration
729 used and their haptenic sugars (100 mM) or glycoproteins (5 mg/mL).

Abbrev.	Conc. (µg/ml)	Source	Species	Specificity	Hapten
PA-I	0.1	Bacteria	*Pseudomonas aeruginosa*	Gal, Gal derivatives	Gal
AIA (Jacalin)	15	Plant	*Artocarpus integrifolia*	Gal (somewhat sialylation tolerant)	Gal
SBA	15	Plant	*Glycine max*	GalNAc	Gal
WFA	10	Plant	*Wisteria floribunda*	GalNAc/sulfated GalNAc	Gal
VVA-B4	10	Plant	*Vicia villosa*	GalNAc	Gal
PNA	15	Plant	*Arachis hypogaea*	Gal-β-(1→3)-GalNAc	Gal
VRA	20	Plant	*Vigna radiata*	Terminal α-linked Gal	Gal
GS-I-B4	4	Plant	*Griffonia simplicifolia (Bandeiraea simplicifolia)*	Terminal α-linked Gal	Gal
Con A	5	Plant	*Canavalia ensiformis*	α-linked Man, Glc or GlcNAc	Man
NPA	20	Plant	*Narcissus pseudonarcissus*	α-(1→6)-linked Man	Man
HHA	5	Plant	*Hippeastrum hybrid*	Man-α-(1→3)-Man-α-(1→6)-R	Man
GNA	4	Plant	*Galanthus nivalis*	Man-α-(1→3)-R	Man
MAA	10	Plant	*Maackia amurensis*	Neu-α-(2→3)-Gal(NAc)-R	Lac
SNA-I	2	Plant	*Sambucus nigra*	Neu-α-(2→6)-Gal(NAc)-R	Lac
WGA	0.04	Plant	*Triticum vulgaris*	NeuAc/GlcNAc	GlcNAc
GSL-II	15	Plant	*Griffonia simplicifolia*	GlcNAc	GlcNAc
PHA-L	2	Plant	*Phaseolus vulgaris*	Tri-, tetra-antennary β-Gal/Gal-β-(1→4)-GlcNAc	α-1-acid glycoprotein
PHA-E	0.7	Plant	*Phaseolus vulgaris*	Bi-antennary, bisecting GlcNAc, β-Gal/Gal-β-(1→4)-GlcNAc	Bovine IgG
UEA-I	15	Plant	*Ulex europaeus*	Fuc-α-(1→2)-R	Fuc

730

731

732 **Table 2.** Bacterial strains and toxins used and their sources.

Strain or toxin	**Source or reference**
Escherichia coli 25922	ATCC
Escherichia coli 35218	ATCC
Vibrio parahaemolyticus RIMD2210633	O'Boyle, *et al.*, 2013[42]
Staphylococcus aureus 8325-4	Horsburgh, *et al.*, 2002[44]
Staphylococcus aureus BH1CC	O'Neill, *et al.*, 2007[45]
Cholera toxin subunit B (Ctx-B)	Life Technologies (Carlsbad, CA)
*S*higa-like toxin B (Stx1-B)	Nova Biotech Development (El Cajon, CA)

733

734

735

Supplementary Information

Rapid screening for specific glycosylation and pathogen interactions on a 78 species avian egg white glycoprotein microarray

Marta Utratna[1], Heidi Annuk[1], Jared Q. Gerlach[1,2], Yuan C. Lee[3], Marian Kane[1], Michelle Kilcoyne[4], Lokesh Joshi[1]

[1]Glycoscience Group, National Centre for Biomedical Engineering Science, National University of Ireland Galway, Galway, Ireland

[2] Regenerative Medicine Institute, National University of Ireland Galway, Galway, Ireland

[3] Department of Biology, Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218, U.S.A.

[4] Carbohydrate Signalling Group, Microbiology, School of Natural Sciences, National University of Ireland Galway, Galway, Ireland

Table of Contents

LIST OF SUPPLEMENTARY FIGURES AND TABLES

RESULTS AND DISCUSSION

Optimisation of EW solubilisation

Prior to electrophoresis, solubilisation optimisation was carried out on lyophilised pigeon EW (PEW), gull EW (GEW), CEW, DEW and QEW samples to ensure maximum protein retention. Solubilisation in PBS resulted in a proportion of insoluble aggregates which were removed by centrifugation and filtration. Subsequent delipidisation and acetone precipitation resulted in significant loss of protein, e.g. the protein content of GEW and PEW decreased to 20% and 24% of the original PBS solubilised content, respectively, after delipidisation and 9% and 18%, respectively, after acetone precipitation (data not shown). Fresh CEW is approximately 12% protein by weight with only approximately 0.2% lipid and 0.7% free carbohydrate content,[3] and centrifugation and filtration was found to be sufficient prior to SDS-PAGE analysis and microarray printing.

Optimisation of EW microarray printing

The Nexterion® Slide H microarray slides were selected for optimal background and chemical compatibility as previously described.[4,5] PEW, GEW, CEW, DEW and QEW samples were used to optimise microarray printing and were printed at 0.1 to 0.6 mg/ml, either in PBS or PBS with 0.05% Tween 20 (PBS-T) using a piezoelectric dispenser capillary (PDC) with a hydrophobic coating. The printed slides were incubated with a panel of selected TRITC-labelled lectins (**Table 1**) to monitor the printed feature quality. The lectins were selected based on the glycosylation motifs expected to be present in the EW samples.[2,6] Not all interactions were inhibitable (Supplementary **Table S3** and **Figure S2**). The average inter-array coefficient of variance (%CV) for all lectins was ~16% (range 5% to 45%) for the 78 EWs and ~16% (range <1% to 40%) for the glycoprotein standards (Supplementary **Table S3**). These ranges were lower than expected based on the high viscosity of the samples and %CV reported for other protein-based microarrays.[4,7]

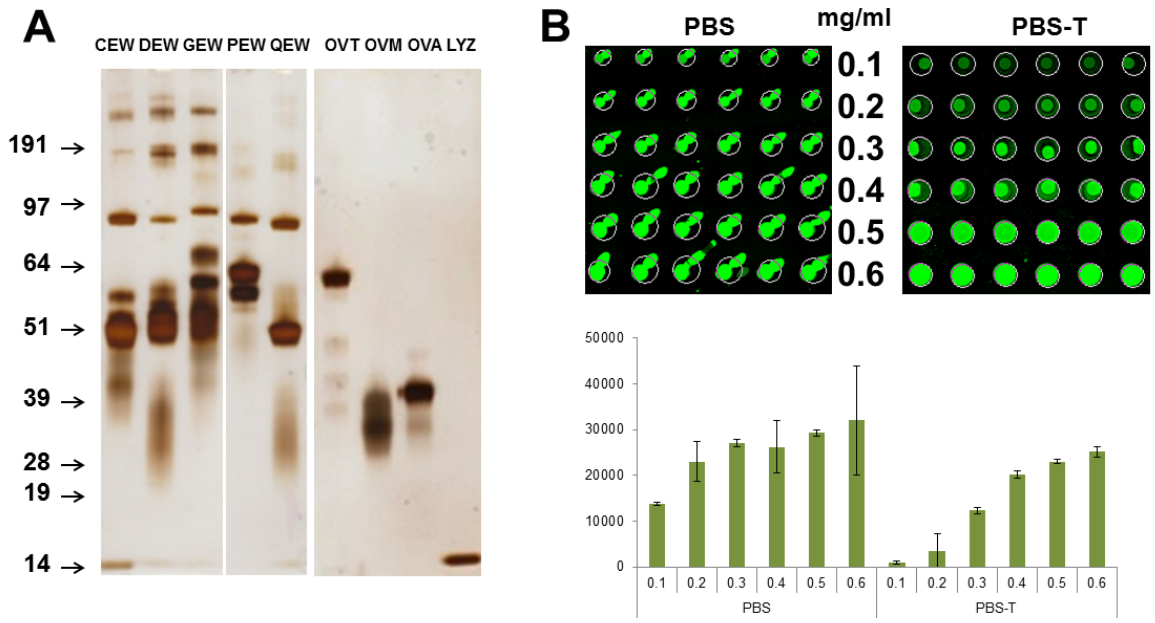


Figure S1. Optimisation of EWs for printing. A) Protein profiles of selected EWs. Samples of five EWs (1 µg per lane, dissolved in PBS pH 7.4) were separated on 10% SDS-PAGE /MOPS and silver stained; commercial (Sigma) purified chicken standards (0.5 µg per lane): ovotransferin (OVT), ovomucoid (OVM), ovalbumin (OVA) and lysosyme (LYZ) were used. B) Schematic of printing layout for the PEW sample printed in gradually increasing concentrations: 0.1-0.6 mg/mL in PBS pH 7.4 (left half) or 0.05% PBS-T (right half). Each sample was printed six times, resulting in six replicate spots. The example illustrates the binding of PEW with PHA-E incubated at a concentration of 0.7 µg/ml. An extraction grid was loaded on top of the imaged spots to indicate the challenges of data extraction, including proper feature fitting. C) Histogram illustrating the mean fluorescence intensity from part B to compare the variability within six replicates. Error bars indicate the standard deviation for the fluorescence intensity of six replicate features for each sample.

Five lectins from the initial library (VVA-B4, PNA, VRA, UEA-I and Jacalin) demonstrated very low binding (data not shown) and did not reach the threshold for inhibition (a minimum of 25% reduction of binding intensity upon co-incubation with appropriate hapten on average for all printed samples). Thus, these five lectins were excluded from the subsequent glycosylation analysis which included the remaining 14 lectins (**Table 1**).

Uniform feature morphology is required for reliable microarray data extraction and the addition of detergent resulted in round features with an average size of 150 μm (Supplementary **Figure S1B**). A concentration-dependent increase in lectin binding was observed up to 0.6 mg/ml, the maximum concentration possible to print for the majority of EWs due to their limited availability (Supplementary **Figure S1C**). Standard glycoproteins were included to monitor lectin performance and were printed at 0.5 mg/ml with 0.01% Tween 20 with the exception of OVA and OVM, which were optimally printed with 0.015% Tween 20 (Supplementary **Table S2**).

Performance of lectins with printed glycoprotein standards

Glycoprotein standards printed on the EW microarray (Supplementary **Figure S2**) demonstrated the expected associations with lectins including binding of the mannose- (Man-) specific lectins NPA and GNA to yeast invertase, which has high- and oligo-mannose type *N*-linked oligosaccharides.[7] The absence of MAA binding to asialofetuin (ASF) confirmed the loss of sialic acid when compared to fetuin and MAA also bound to the serum globulins human α-1-antitrypsin (A1AT), human α-1-acid glycoprotein (AGP) and bovine transferrin, as previously reported).[4]

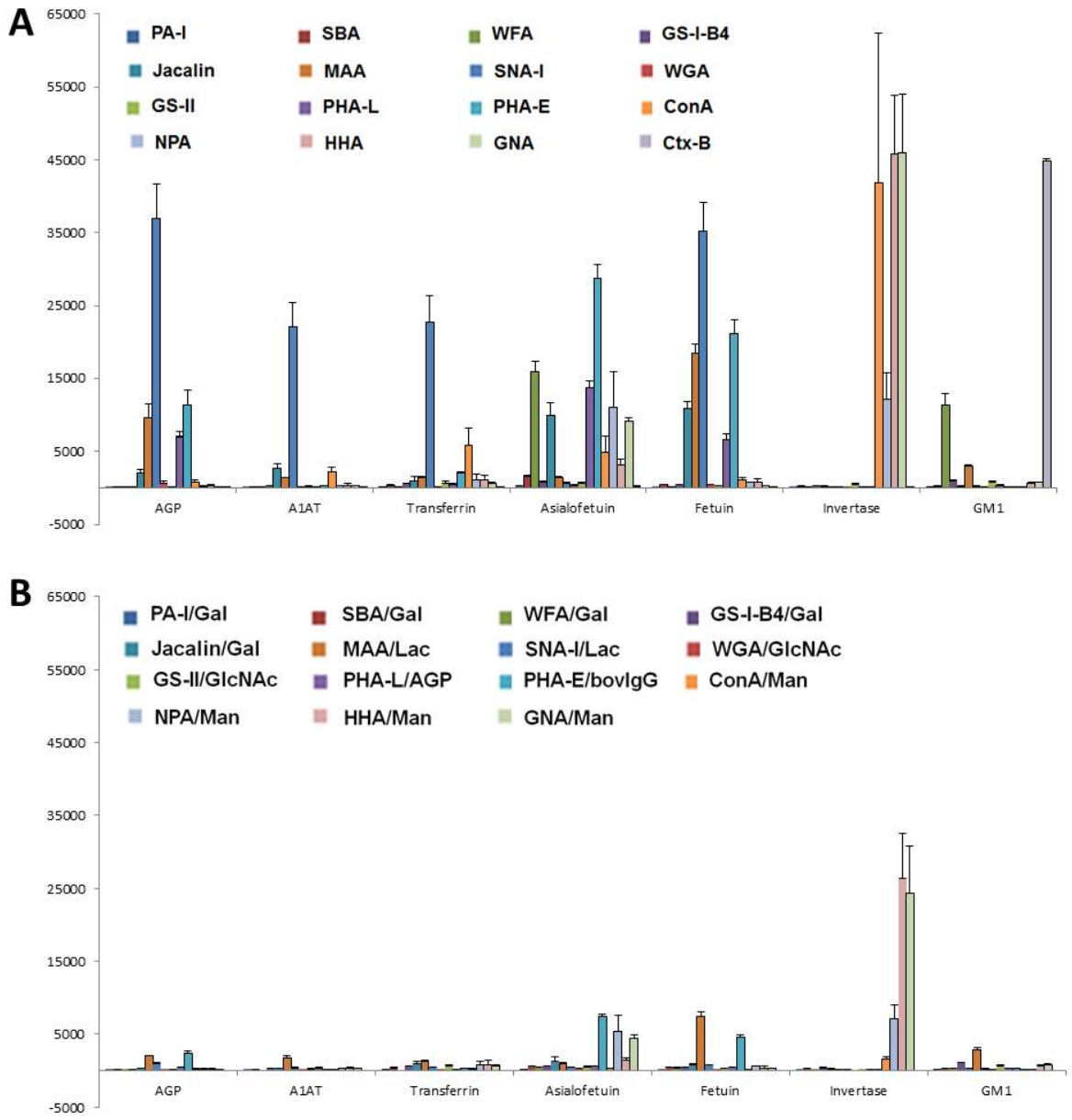


Figure S2. Analysis of glycoproteins (other than chicken EW glycoproteins) included as controls in the two panels of EW microarray. A) Lectin profiles, B) inhibition. Lectin binding profiles of fifteen selected lectins recognised as specific interaction and cholera toxin (Ctx-B), together with their concentrations listed in brackets. The lectin AIA (Jacalin) had carbohydrate inhibitable binding for the glycoprotein standards (non-EW) included on the microarray as controls (included in this figure in the carbohydrate-mediated lectin binding library). The fluorescence intensity values shown are the average of the normalised data from three replicates, with error bars indicating one standard deviation. Error bars are one standard deviation for the mean of all replicates (Supplementary **Table S3**).

Additional tables

Table S1. Species of origin of avian EW microarray. Birds grouped by systematic order indicated by an assigned capital letter (A-P).

Common name	Genus and species	Abbreviation	Order code	Order	Family
Magpie goose	*Anseranas semipolmata*	MGP	**A**	Anseriformes	Anseranatidae
Plumed whistling-duck	*Dendrocygna eytoni*	PWD	**A**	Anseriformes	Dendrocygnidae
Redhead	*Aythya americana*	RED	**A**	Anseriformes	Anatidae
Ringed teal	*Callonetta (Anas) leucophrys*	RIT	**A**	Anseriformes	Anatidae
Barrow's goldeneye	*Bucephala islandica*	BGE	**A**	Anseriformes	Anatidae
Eurasian widgeon	*Anas penelope*	EAW	**A**	Anseriformes	Anatidae
Domesticated duck	*Anas platyrhynchos*	DEW	**A**	Anseriformes	Anatidae
Wood duck	*Aix sponsa*	WOD	**A**	Anseriformes	Anatidae
Black swan	*Cygnus atratus*	SWN	**A**	Anseriformes	Anatidae
Canada goose	*Branta canadensis*	CNG	**A**	Anseriformes	Anatidae
Southern screamer	*Chauna torquata*	SOS	**A**	Anseriformes	Anhimidae
Gull	*Larus argentatus*	GEW	**B**	Ciconiiformes	Laridae
Herring gull	*Larus argentatus*	HGL	**B**	Ciconiiformes	Laridae
Black-winged stilt	*Himantopus himantopus*	BWS	**B**	Ciconiiformes	Charadriidae
Spur-winged lapwing	*Vanellus spinosus*	SPL	**B**	Ciconiiformes	Charadriidae
Adelie penguin	*Pygoscelis adeliae*	APE	**B**	Ciconiiformes	Spheniscidae
Humboldt penguin	*Spheniscus humboldti*	HPE	**B**	Ciconiiformes	Spheniscidae
Macaroni penguin	*Eudyptes chrysolophus*	MPE	**B**	Ciconiiformes	Spheniscidae
Brahminy kite	*Haliastur indus*	BRK	**B**	Ciconiiformes	Accipitridae
Crested caracara	*Polyborus plancus*	CRC	**B**	Ciconiiformes	Falconidae
Chimango caracara	*Milvago chimango*	CMC	**B**	Ciconiiformes	Falconidae
Great-crestec grebe	*Podiceps cristatus*	GCG	**B**	Ciconiiformes	Podicipedidae
Great cormorant	*Phalacrocorax carbo*	GCM	**B**	Ciconiiformes	Phalacrocoracidae
Greater flamingo	*Phoenicopterus ruber*	GFN	**B**	Ciconiiformes	Phoenicopteridae

Mascarene reef-egret	*Egretta dimorpha*	MRE	**B**	Ciconiiformes	Ardeidae
Little blue heron	*Egretta caerulae*	LBH	**B**	Ciconiiformes	Ardeidae
Domestic pigeon	*Columba liviadomestica*	PEW	**C**	Columbiformes	Columbidae
Common crowned pigeon	*Goura cristata*	CPI	**C**	Columbiformes	Columbidae
Mourning dove	*Zenaida macroura*	MOD	**C**	Columbiformes	Columbidae
Rock dove	*Columba livia*	RKD	**C**	Columbiformes	Columbidae
Laughing kookaburra	*Dacelo novaeguineae*	LGK	**D**	Coraciiformes	Halcyonidae
Malle fowl	*Leipoa ocellata*	MLF	**E**	Craciformes	Megapodiidae
Razor-billed curassow	*Crax mitu*	RCU	**E**	Craciformes	Cracidae
Coral-billedground-cuckoo	*Carpococcyx renauldi*	CGC	**F**	Cuculiformes	Cuculidae
Roadrunner	*Geococcyx californianus*	RDR	**F**	Cuculiformes	Neomorphidae
Domesticated chicken	*Gallus gallus domesticus*	CEW	**G**	Galliformes	Phasianidae
Wild turkey	*Meleagris gallopavo*	TKY	**G**	Galliformes	Phasianidae
Coqui francolin	*Francolinus coqui*	COF	**G**	Galliformes	Phasianidae
Scaled (blue) quail	*Callipepla squamata pallida*	SCQ	**G**	Galliformes	Odontophoridae
Japanese quail	*Coturnix japonica*	QEW	**G**	Galliformes	Phasianidae
Mountain quail	*Oreortyx pictus*	MTQ	**G**	Galliformes	Odontophoridae
Congo peafowl	*Afropavo congensis*	CPF	**G**	Galliformes	Phasianidae
Cabot's tragopan	*Tragopan caboti*	CTP	**G**	Galliformes	Phasianidae
White-tailed ptarmigan	*Lagopus leucurus*	TTR	**G**	Galliformes	Phasianidae
Malay great argus	*Argusianus argus*	MGA	**G**	Galliformes	Phasianidae
Ruffed grouse	*Bonasa umbellus*	RFG	**G**	Galliformes	Phasianidae
Satyr tragopan	*Tragopan Satyra*	STP	**G**	Galliformes	Phasianidae
Silver pheasant	*Lophura nycthermera*	SVP	**G**	Galliformes	Phasianidae
Helmet guineafowl	*Numida Meleagris*	GUI	**G**	Galliformes	Numididae
Vulturine guineafowl	*Acryllium vulturinum*	VGU	**G**	Galliformes	Numididae
Black-legged seriema	*Chunga burmeisteri*	BLS	**H**	Gruiformes	Cariamidae
Japanese crane	*Grus japonensis*	JPC	**H**	Gruiformes	Gruidae

White-naped crane	*Grus vipio*	WNC	**H**	Gruiformes	Gruidae
White-cheeked turaco	*Turaco leucotis*	WCT	**I**	Musophagiformes	Musophagidae
Albert's lyrebird	*Menura alberti*	ABL	**J**	Passeriformes	Menuridae
Blue-shouldered robin-chat	*Cossypha cyanocampter*	BRC	**J**	Passeriformes	Muscicapidae
Crimson finch	*Neochmia phaeton*	CFI	**J**	Passeriformes	Passeridae
House sparrow	*Passer domesticus*	HOS	**J**	Passeriformes	Passeridae
Common starling	*Sturnus vulgaris*	CNS	**J**	Passeriformes	Sturnidae
Masked woodswallow	*Artamus personatus*	MWS	**J**	Passeriformes	Corvidae
American robin	*Turdus migratorius*	ROB	**J**	Passeriformes	Muscicapidae
Common flicker	*Colaptes auratus*	CFL	**K**	Piciformes	Picidae
Budgerigar	*Melopsittacus undulatus*	BGR	**L**	Psittaciformes	Psittacidae
Grey parrot	*Psittacus erithacus*	GPT	**L**	Psittaciformes	Psittacidae
Turquois-fronted parrot	*Amazona aestiva*	TPT	**L**	Psittaciformes	Psittacidae
White-capped parrot	*Pionus senilis*	WPT	**L**	Psittaciformes	Psittacidae
Yellow-headed parrot	*Amazoa ochrocephala*	YPT	**L**	Psittaciformes	Psittacidae
Barn owl	*Tyto alba*	BOW	**M**	Strigiformes	Tytonidae
Snowy owl	*Nyctea scandiaca*	SNO	**M**	Strigiformes	Strigidae
Australian cassowary	*Casuarius casuarius*	ASC	**N**	Struthioniformes	Casuariidae
Emu	*Dromaius novaehollandiae*	EMU	**N**	Struthioniformes	Casuariidae
Ostrich	*Struthio camelus*	OST	**N**	Struthioniformes	Struthionidae
Greater rhea	*Rhea americana*	REA	**N**	Struthioniformes	Rheidae
Brushland tinamou	*Nothoprocta cinerascens*	BTI	**O**	Tinamiformes	Tinamidae
Elegant crested-tinamou	*Euotromia elegans*	CTI	**O**	Tinamiformes	Tinamidae
Great tinamou	*Tinamus major*	GTI	**O**	Tinamiformes	Tinamidae
Brown-rumped buttonquail	*Turnix nana*	BRQ	**P**	Turniciformes	Turnicidae
Small buttonquail	*Turnix sylvatica*	BUQ	**P**	Turniciformes	Turnicidae

Table S2. Two printing panels of EW microarray. The samples are listed based on their printing location.

PANEL_A				PANEL_B			
	Abbreviation	printing conc. mg/ml	final Tween20	Abbreviation	printing conc. mg/ml	final Tween20	Sample type:
1	APE (02)	0.6	0.01 %T	ABL (01)	0.6	0.01 %T	Egg white - small scale
2	BGE (05)	0.6	0.01 %T	ASC (03)	0.6	0.01 %T	Egg white - small scale
3	EAW (28)	0.6	0.01 %T	EAW (28)	0.6	0.01 %T	Egg white - small scale
4	BRK (10)	0.6	0.01 %T	BGR (06)	0.6	0.01 %T	Egg white - small scale
5	BTI (12)	0.6	0.01 %T	BLS (07)	0.6	0.01 %T	Egg white - small scale
6	OVA 0.5	0.5	0.015 %T	OVA 0.5	0.5	**0.015 %T**	glycoprotein (chicken egg white)
7	BUQ (13)	0.6	0.01 %T	BOW (08)	0.6	0.01 %T	Egg white - small scale
8	CFL (17)	0.6	0.01 %T	BRC (09)	0.6	0.01 %T	Egg white - small scale
9	OVT 0.5	0.5	0.01 %T	OVT 0.5	0.5	0.01 %T	glycoprotein (chicken egg white)
10	CNG (20)	0.6	0.01 %T	BRQ (11)	0.6	0.01 %T	Egg white - small scale
11	CNS (21)	0.6	0.01 %T	BWS (15)	0.6	0.01 %T	Egg white - small scale
12	OVM 0.5	0.5	0.015 %T	OVM 0.5	0.5	**0.015 %T**	glycoprotein (chicken egg white)
13	COF (22)	0.6	0.01 %T	CFI (16)	0.6	0.01 %T	Egg white - small scale
14	CPI (24)	0.6	0.01 %T	CGC (18)	0.6	0.01 %T	Egg white - small scale
15	Transferrin/Invertase	0.5	0.01 %T	Transferrin/Invertase	0.5	0.01 %T	glycoprotein
16	CTI (26)	0.6	0.01 %T	CMC (19)	0.6	0.01 %T	Egg white - small scale
17	CTP (27)	0.6	0.01 %T	CPF (23)	0.6	0.01 %T	Egg white - small scale
18	ASF	0.5	0.01 %T	ASF	0.5	0.01 %T	glycoprotein
19	GPT (33)	0.6	0.01 %T	CRC (25)	0.6	0.01 %T	Egg white - small scale
20	GTI (34)	0.6	0.01 %T	EMU (29)	0.6	0.01 %T	Egg white - small scale
21	Fetuin	0.5	0.01 %T	Fetuin	0.5	0.01 %T	glycoprotein
22	GUI (35)	0.6	0.01 %T	GCG (30)	0.6	0.01 %T	Egg white - small scale
23	HGL (36)	0.6	0.01 %T	GCM (31)	0.6	0.01 %T	Egg white - small scale
24	PEW	0.6	0.01 %T	PEW	0.6	0.01 %T	Egg white - large scale
25	HOS (37)	0.6	0.01 %T	GFN (32)	0.6	0.01 %T	Egg white - small scale
26	MGA (42)	0.6	0.01 %T	HPE (38)	0.6	0.01 %T	Egg white - small scale
27	GEW	0.6	0.01 %T	GEW	0.6	0.01 %T	Egg white - large scale
28	MGP (46)	0.6	0.01 %T	JPC (39)	0.6	0.01 %T	Egg white - small scale
29	MTQ (48)	0.6	0.01 %T	LBH (40)	0.6	0.01 %T	Egg white - small scale
30	PWD (51)	0.6	0.01 %T	LGK (41)	0.6	0.01 %T	Egg white - small scale
31	QEW	0.6	0.01 %T	QEW	0.6	0.01 %T	Egg white - large scale
32	RCU (52)	0.6	0.01 %T	MLF (43)	0.6	0.01 %T	Egg white - small scale
33	RDR (53)	0.6	0.01 %T	MOD (44)	0.6	0.01 %T	Egg white - small scale
34	DEW	0.6	0.01 %T	DEW	0.6	0.01 %T	Egg white - large scale
35	RED (55)	0.6	0.01 %T	MPE (45)	0.6	0.01 %T	Egg white - small scale
36	RFG (56)	0.6	0.01 %T	MRE (47)	0.6	0.01 %T	Egg white - small scale
37	CEW	0.6	0.01 %T	CEW	0.6	0.01 %T	Egg white - large scale
38	RIT (57)	0.6	0.01 %T	MWS (49)	0.6	0.01 %T	Egg white - small scale
39	RKD (58)	0.6	0.01 %T	OST (50)	0.6	0.01 %T	Egg white - small scale
40	AGP	0.5	0.01 %T	AGP	0.5	0.01 %T	glycoprotein
41	SCQ (61)	0.6	0.01 %T	REA (54)	0.6	0.01 %T	Egg white - small scale
42	SNO (62)	0.6	0.01 %T	ROB (59)	0.6	0.01 %T	Egg white - small scale
43	A1AT/Lysosyme	0.5	0.01 %T	A1AT/Lysosyme	0.5	0.01 %T	glycoprotein / glycoprotein (chicken egg white)
44	SOS (63)	0.6	0.01 %T	SPL (64)	0.6	0.01 %T	Egg white - small scale
45	STP (65)	0.6	0.01 %T	TKY (68)	0.6	0.01 %T	Egg white - small scale
46	GM1-HSA	0.5	0.01 %T	GM1-HSA	0.5	0.01 %T	neoglycoconjugate
47	SVP (66)	0.6	0.01 %T	TPT (69)	0.6	0.01 %T	Egg white - small scale
48	SWN (67)	0.6	0.01 %T	WCT (74)	0.6	0.01 %T	Egg white - small scale
49	PBS-T			PBS-T			PBS-T
50	TTR (71)	0.6	0.01 %T	WNC (75)	0.6	0.01 %T	Egg white - small scale
51	VGU (72)	0.6	0.01 %T	WPT (77)	0.6	0.01 %T	Egg white - small scale
52	WOD (76)	0.6	0.01 %T	YPT (78)	0.6	0.01 %T	Egg white - small scale
16 probes common to both panels for cross normalisation							

Additional figures

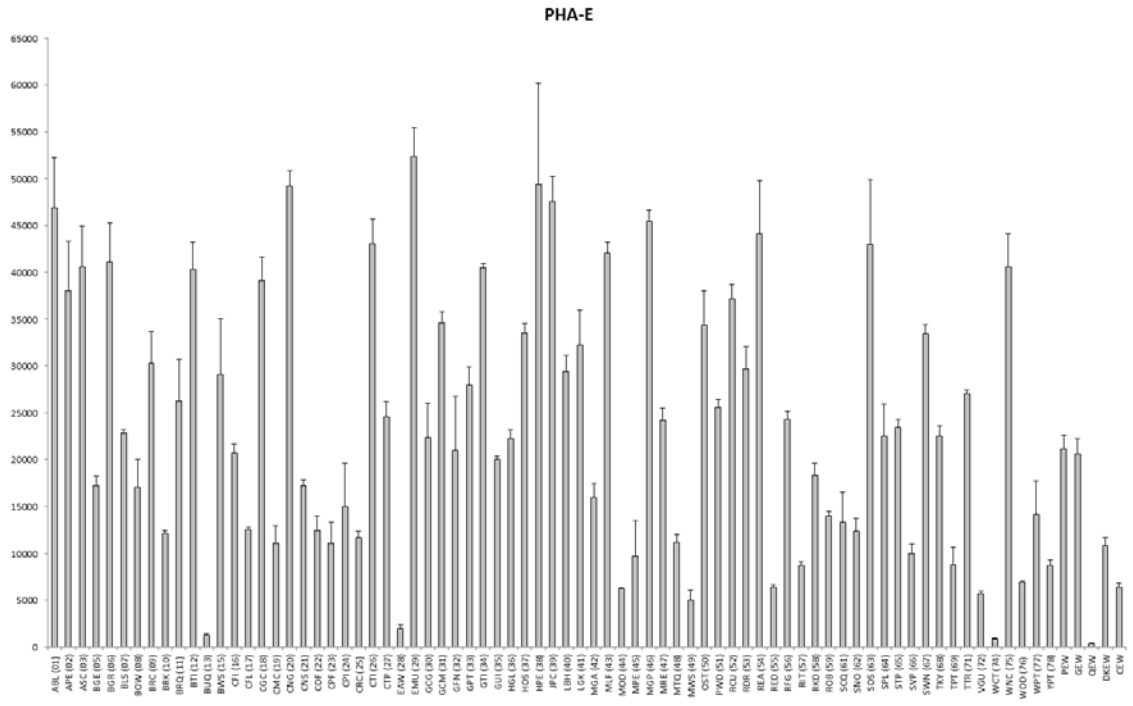


Figure S3. PHA-E lectin binding profile across all EWs. Sample bar chart demonstrating range of binding on the microarray. Error bars are one standard deviation for the mean of all replicates (Supplementary **Table S3**).

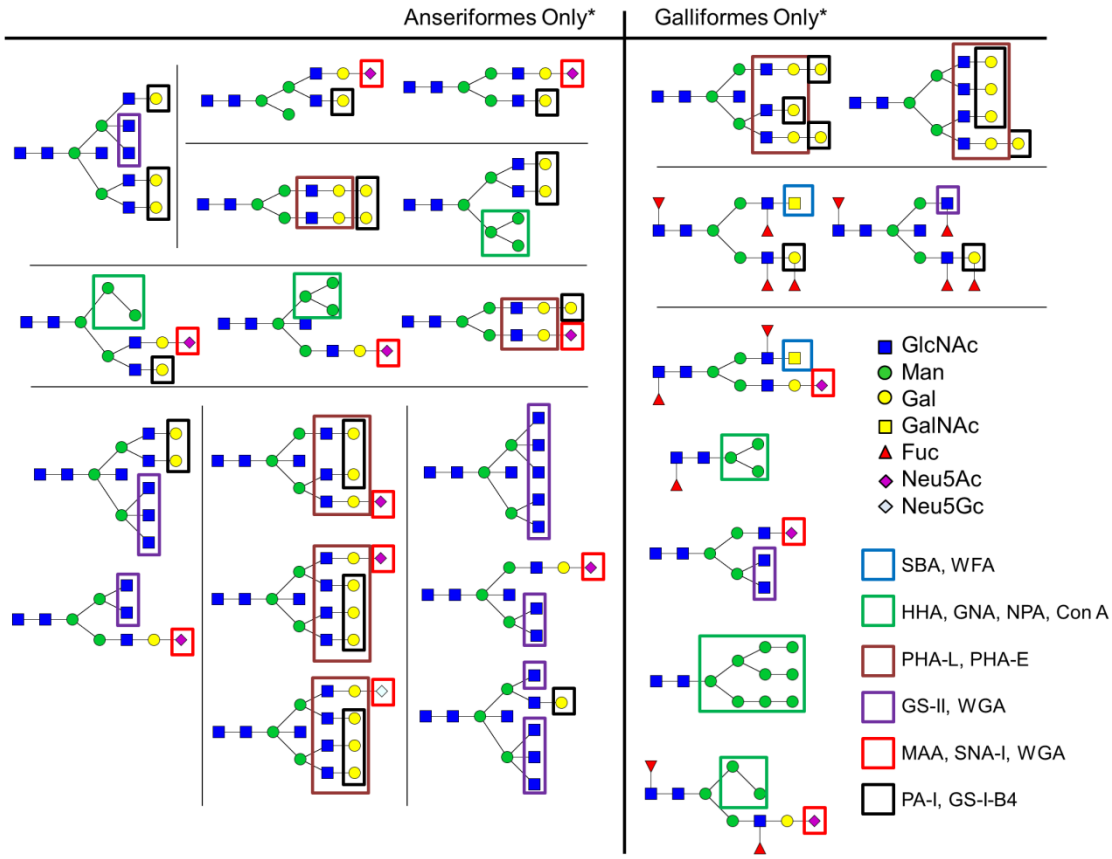


Figure S4. *N*-linked structures unique for either Anseriformes or Galliformes previously reported.[6] Coloured rectangles indicate epitopes for interactions with lectins used in this study (**Table 1**).

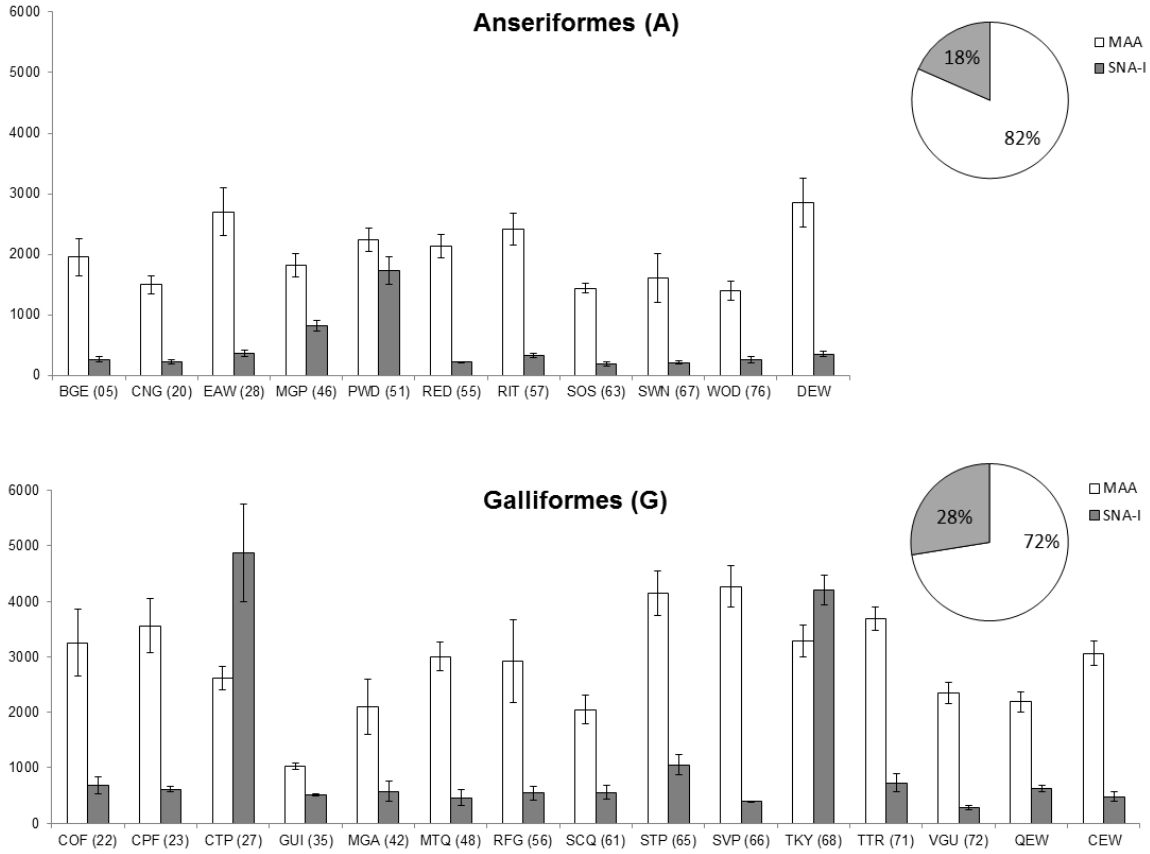
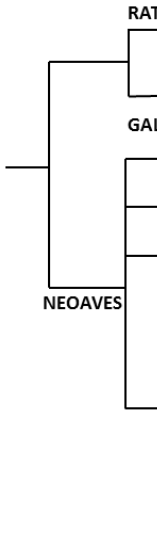


Figure S5. Bar charts for SNA-I and MAA lectin binding to the 28 species representing Anseriformes and Galliformes included in this study. Pie charts indicate proportions of MAA and SNA-I within total intensity of each order. Error bars are one standard deviation for the mean of all replicates (Supplementary **Table S3**).



			This study			Suzuki *et al*. 2004	
Avian order	Order code	Example	Con A (+)	GS-I-B4 (+)	Overlap	Con A (+)	GS-I (+)
Struthioniformes	N	ostrich	2	0	4	4	0
Tinamiformes	O	tinamou	3	0	3	3	0
Craciformes	E	currassow	2	0	2	2	0
Galliformes	G	chicken	13	0	13	13	0
Anseriformes	A	duck	9	0	10	9	*0*
Turniciformes	P	buttonquail	2	1	2	2	1
Piciformes	K	flicker	1	0	1	1	0
Coraciformes	D	kookaburra	0	0	1	0	0
Cuculiformes	F	cuckoo	2	2	2	2	2
Psittaciformes	L	parrot	5	5	5	2	5
Musophagiformes	I	turaco	1	1	1	1	1
Strigiformes	M	owl	2	2	2	2	2
Columbiformes	C	pigeon	3	2	3	2	2
Gruiformes	H	crane	2	1	3	2	1
Ciconiiformes	B	heron/gull	14	10	14	10	10
Charadriiformes	J	starling	7	5	7	7	5
Total:			68	29	73	62	29
			93%	40%		84%	40%

Figure S6. Phylogenic tree of birds used in this study according to DNA-DNA hybridisation[1] (adapted from Suzuki, *et al.*, 2004[2]). The examples of birds were taken from Supplementary **Table S-1** and listed together with a total number of species represented for each order used in this study. The overlap with birds analysed by Suzuki, *et al*., (2004)[2] is indicated, and the results for Con A and GS-I versus GS-I-B4 are summarised in comparison to this study.

REFERENCES

1. Sibley, C. G. & Monroe Jr, B. L. Taxonomy and distribution of the birds of the world. Yale University Press, New Haven, Connecticut, USA, 1990.

2. Suzuki, N., Laskowski, M. & Lee, Y. C. Phylogenetic expression of Galalpha1-4Gal on avian glycoproteins: glycan differentiation inscribed in the early history of modern birds. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 9023–9028 (2004).

3. Kovacs-Nolan, J., Phillips, M. & Mine, Y. Advances in the value of eggs and egg components for human health. *J. Agric. Food Chem*. **53**, 8421–8431 (2005).

4. Kilcoyne, M. *et al*. Construction of a natural mucin microarray and interrogation for biologically relevant glyco-epitopes. *Anal. Chem*. **84**, 3330–3338 (2012).

5. Kilcoyne, M. e*t al*. *Campylobacter jejuni* strain discrimination and temperature-dependent glycome expression profiling by lectin microarray. *Carbohydr. Res*. **389**, 123–133 (2014).

6. Plasencia, M. D., Isailovic, D., Merenbloom, S. I., Mechref, Y. & Clemmer, D. E. Resolving and sssigning N-linked glycan structural isomers from ovalbumin by IMS-MS. *J. Am. Soc. Mass Spectrom*. **19**, 1706–1715 (2008).

7. Gerlach, J. Q., Kilcoyne, M. & Joshi, L. Microarray evaluation of the effects of lectin and glycoprotein orientation and data filtering on glycoform discrimination. *Anal. Methods* **6**, 440 (2014).