The landscape of antibody binding to SARS-CoV-2

Anna S. Heffron[1], Sean J. McIlwain[2,3], David A. Baker[1], Maya F. Amjadi[4], Saniya Khullar[2], Ajay K. Sethi[5], Miriam A. Shelef[4,6], David H. O'Connor[1,7], Irene M. Ong[2,3,8]*

[1]Department of Pathology and Laboratory Medicine, University of Wisconsin-Madison, Madison, WI, United States of America
[2]Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI, United States of America
[3]University of Wisconsin Carbone Comprehensive Cancer Center, University of Wisconsin-Madison, Madison, WI, United States of America
[4]Department of Medicine, University of Wisconsin-Madison, Madison, WI, United States of America
[5]Department of Population Health Sciences, University of Wisconsin-Madison, Madison, WI, United States of America
[6]William S. Middleton Memorial Veterans Hospital, Madison, WI, United States of America
[7]Wisconsin National Primate Research Center, University of Wisconsin-Madison, Madison, Wisconsin, United States of America
[8]Department of Obstetrics and Gynecology, University of Wisconsin-Madison, Madison, WI, United States of America

*Corresponding author

Abstract

The search for potential antibody-based diagnostics, vaccines, and therapeutics for pandemic severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has focused almost exclusively on the spike (S) and nucleocapsid (N) proteins[1–8]. Coronavirus membrane (M), orf3a, and orf8 proteins are also humoral immunogens in other coronaviruses (CoVs)[8–11] but remain largely uninvestigated for SARS-CoV-2. Here we show that SARS-CoV-2 infection induces robust antibody responses to epitopes throughout the SARS-CoV-2 proteome, particularly in M, in which one epitope achieved near-perfect diagnostic accuracy. We map 79 B cell epitopes throughout the SARS-CoV-2 proteome and demonstrate that anti-SARS-CoV-2 antibodies appear to bind homologous peptide sequences in the 6 known human CoVs. Our results demonstrate previously unknown, highly reactive B cell epitopes throughout the full proteome of SARS-CoV-2 and other CoV proteins, especially M, which should be considered in diagnostic, vaccine, and therapeutic development.

Introduction

Antibodies mediate protection from coronaviruses (CoVs) including SARS-CoV-2[1–8], severe acute respiratory syndrome coronavirus (SARS-CoV)[8,12–15] and Middle Eastern respiratory syndrome coronavirus (MERS-CoV)[8,16–19]. All CoVs encode 4 main structural proteins, spike (S), envelope (E), membrane (M), and nucleocapsid (N), as well as multiple non-structural proteins and accessory proteins[20]. In SARS-CoV-2, anti-S and anti-N antibodies have received the most attention to date[1–8], including in serology-based diagnostic tests[1–5] and vaccine candidates[6–8]. However, anti-S antibodies have been linked to antibody dependent enhancement for SARS-CoV-2 and other CoVs[7,21–26], and prior reports observed that not all individuals infected with SARS-CoV-2 produce detectable antibodies against S or N[1–5], indicating a need for

expanded antibody-based options. Much less is known about antibody responses to other SARS-CoV-2 proteins, though data from other CoVs suggest they may be important. Antibodies against SARS-CoV M can be more potent than antibodies against SARS-CoV S[9–11], and some experimental SARS-CoV and MERS-CoV vaccines elicit responses to M, E, and orf8[8]. Additionally, previous work has demonstrated humoral cross-reactivity between CoVs[7,14,27–30] and suggested it could be protective[24,30], although full-proteome cross-reactivity has not been investigated. We designed a peptide microarray tiling the proteome of SARS-CoV-2 and 8 other human and animal CoVs in order to assess antibody epitope specificity and potential cross-reactivity with other CoVs, and we used this microarray to profile IgG antibody responses in 40 COVID-19 convalescent patients and 20 SARS-CoV-2-naive controls.

CoV reactivity in uninfected controls

Greater than 90% of adult humans are seropositive for the "common cold" CoVs (CCCoVs: HCoV-HKU1, HCoV-OC43, HCoV-NL63, and HCoV-229E)[31,32], but it is unknown how these pre-existing antibodies might affect reactivity to SARS-CoV-2 or other CoVs. We measured IgG reactivity in sera from 20 SARS-CoV-2-naïve control subjects to CoV linear peptides, considering reactivity that was >3 standard deviations above the mean for the $\log_2$-quantile normalized array data to be indicative of antibody binding. All sera exhibited binding in known epitopes of at least 1 of the control non-CoV strains (poliovirus vaccine and rhinovirus; Fig. 1, Extended data 1, Extended data 2) and were collected in Wisconsin, USA, where exposure to SARS-CoV or MERS-CoV was extremely unlikely. We found that at least one epitope in structural or accessory proteins had binding in 100% of controls for HCoV-HKU1, 85% of controls for HCoV-OC43, 65% for HCoV-NL63, and 55% for HCoV-229E (Fig. 2, Extended data 2). Apparent cross-reactive binding was observed in 45% of controls for MERS-CoV, 50% for SARS-CoV, and 50% for SARS-CoV-2.

SARS-CoV-2 proteome humoral profiling

We aimed to map the full extent of binding of antibodies induced by SARS-CoV-2 infection and to rank the identified epitopes in terms of likelihood of importance and immunodominance. We defined epitope recognition as antibody binding to contiguous peptides in which the average $\log_2$-normalized intensity for patients was at least 2-fold greater than for controls with *t*-test statistics yielding adjusted *p*-values <0.1. We chose these criteria, rather than the 3 standard deviation cut-off, in order to ensure that binding detected would be greater than background binding seen in controls (2-fold greater) and to remove regions of binding that were not at least weakly significantly different from controls (adjusted $p<0.1$). These criteria identified 79 B cell epitopes (Fig. 3, Extended data 3) in S, M, N, orf1ab, orf3a, orf6, and orf8. We ranked these epitopes by minimum adjusted *p*-value for any 16-mer in the epitope in order to determine the greatest likelihood of difference from controls as a proxy for immunodominance. The highest-ranking epitope occurred in the N-terminus of M (1-M-24). Patient sera showed high-magnitude reactivity (up to 6.7 fluorescence intensity units) in other epitopes in S, M, N, and orf3a, with lower-magnitude reactivity (<3.3 fluorescence intensity units) epitopes in other proteins. The epitopes with the greatest reactivity in S occurred in the fusion peptide (residues 788-806), with less reactivity in the receptor binding domain (residues 319-541)[6] (Fig. 3). Four detected epitopes (553-S-26, 624-S-23, 807-S-26, and 1140-S-25) have previously been shown to be potently neutralizing[33–35], and all 4 of these ranked within the top 10 epitopes. Forty-two of our detected epitopes (including 1-M-24, 553-S-26, 624-S-23, 807-S-26, and 1140-S-25; Extended

data 3) confirm bioinformatic predictions of antigenicity based on SARS-CoV and MERS-CoV[7,8,36–38], with all top-ranking epitopes confirming bioinformatic predictions.

The highest specificity (100%) and sensitivity (98%), determined by linear discriminant analysis leave-one-out cross-validation, for any individual peptide was observed for a 16-mer within the 1-M-24 epitope: ITVEELKKLLEQWNLV (Extended data 4). Fifteen additional individual peptides in M, S, and N had 100% measured specificity and at least 80% sensitivity. Combinations of 1-M-24 with 1 of 5 other epitopes (384-N-33, 807-S-26, 6057-orf1ab-17, 227-N-17, 4451-orf1b-16) yielded an area under the curve receiver operating characteristic of 1 (Extended data 5) based on linear discriminant analysis leave-one-out-cross-validation.

Human, animal CoV cross-reactivity

We defined cross-reactivity as binding by antibodies in COVID-19 convalescent sera to non-SARS-CoV-2 peptides at an average $\log_2$-normalized intensity at least 2-fold greater than in controls with *t*-test statistics yielding adjusted *p*-values <0.1. Antibodies in COVID-19-convalsecent sera appeared to be cross-reactive with homologous epitopes in S, M, N, orf1ab, orf3, orf6, and orf8 in other CoVs (Fig. 4, Extended data 6, Extended data 7, Extended data 8). The greatest number of cross-reactive epitopes (70) were in the RaTG13 bat betacoronavirus (β-CoV), the closest known relative of SARS-CoV-2 (96% nucleotide identity)[39,40], then the pangolin CoV (51 epitopes, 85% nucleotide identity with SARS-CoV-2)[41], then SARS-CoV (40 epitopes, 78% identity[39]). One region, corresponding to SARS-CoV-2 epitope 807-S-26, was cross-reactive across all CoVs, and one, corresponding to SARS-CoV-2 epitope 1140-S-25, was cross-reactive across all β-CoVs (Fig. 4). Epitope 807-S-26 includes the CoV S fusion peptide, and 1140-S-25 is immediately adjacent to the heptad repeat region 2, both of which are involved in membrane fusion[42].

Discussion

M proteins are the most abundant proteins in CoV virions[20]. The N-terminus of M is known in other CoVs to be a small, glycosylated ectodomain that protrudes outside the virion and interacts with S, N, and E[20], while the rest of M resides within the viral particle. Full-length SARS-CoV M has been shown to induce protective antibodies[11,43], and patterns of antibodies binding to SARS-CoV M are similar to those we found in SARS-CoV-2[36]. SARS-CoV anti-M antibodies can synergize with anti-S and anti-N antibodies[11,43], and M has been used in protective SARS-CoV and MERS-CoV vaccines[8]. However, the mechanism of protection of anti-M antibodies remains unknown, and this protein remains largely understudied and underutilized as an antigen. Other groups have not previously identified the high magnitude binding we observed in M, though that may be due to using earlier sample timepoints or different techniques or algorithms[44,45]. Our results, in concert with prior knowledge of anti-SARS-CoV antibodies, strongly suggest that M, particularly the 1-M-24 epitope, as well as other novel epitopes that we identified should be investigated further as potential targets in SARS-CoV-2 diagnostics, vaccines, and therapeutics.

We also found that antibodies produced in response to SARS-CoV-2 infection appeared to cross-react with homologous epitopes throughout the proteomes of other human and non-human CoVs. Hundreds of CoVs have been discovered in bats and other species[24,39–41,46,47], making future spillovers inevitable. The broad cross-reactivity we observed in some homologous peptide

sequences may help guide the development of pan-CoV vaccines[18], especially given that antibodies binding to 807-S-26 and 1140-S-25, epitope motifs cross-reactive across all CoVs and all β-CoVs, respectively, are known to be potently neutralizing[33,34]. We cannot determine whether the increased IgG binding to CCCoVs in COVID-19 convalescent sera is due to newly developed cross-reactive antibodies or the stimulation of a memory response against the original CCCoV antigens. However, cross-reactivity of anti-SARS-CoV-2 antibodies with SARS-CoV or MERS-CoV is likely real, since our population was very unlikely to have been exposed to those viruses. A more stringent assessment of cross-reactivity as well as functional investigations into these cross-reactive antibodies will be vital in determining their capacity for cross-protection. Further, our methods efficiently detect antibody binding to linear epitopes[48], but their sensitivity for detecting parts of conformational epitopes is unknown, and additional analyses will be required to determine whether epitopes identified induce neutralizing or otherwise protective antibodies.

Many questions remain regarding the biology and immunology of SARS-CoV-2. Our extensive profiling of epitope-level resolution antibody reactivity in COVID-19 convalescent subjects provides new epitopes that could serve as important targets in the development of improved diagnostics, vaccines, and therapeutics against SARS-CoV-2 and dangerous human CoVs that may emerge in the future.

Figures

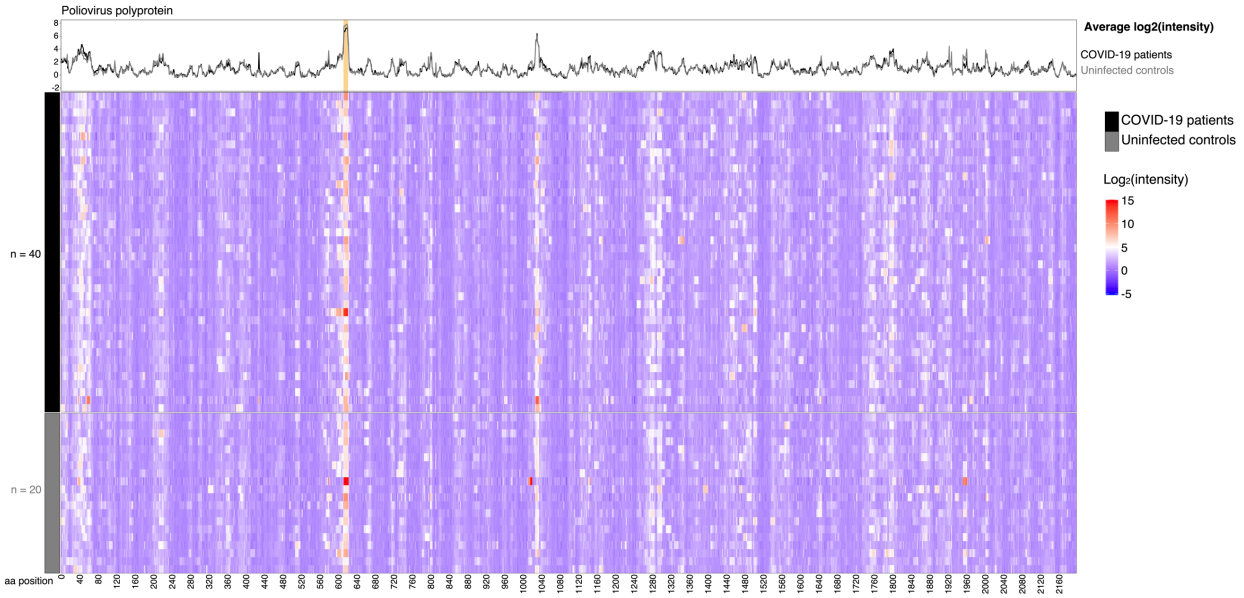

Figure 1. Patients and control subjects show reactivity to a poliovirus control. Sera from 20 control subjects collected before 2019 were assayed for IgG binding to the full proteome of human poliovirus 1 on a peptide microarray. Binding was measured as reactivity that was >3 standard deviations above the mean for the $\log_2$-quantile normalized array data. Patients and controls alike showed reactivity to a well-documented linear poliovirus epitope (start position 613 [IEDB.org]; orange shading in line plot).

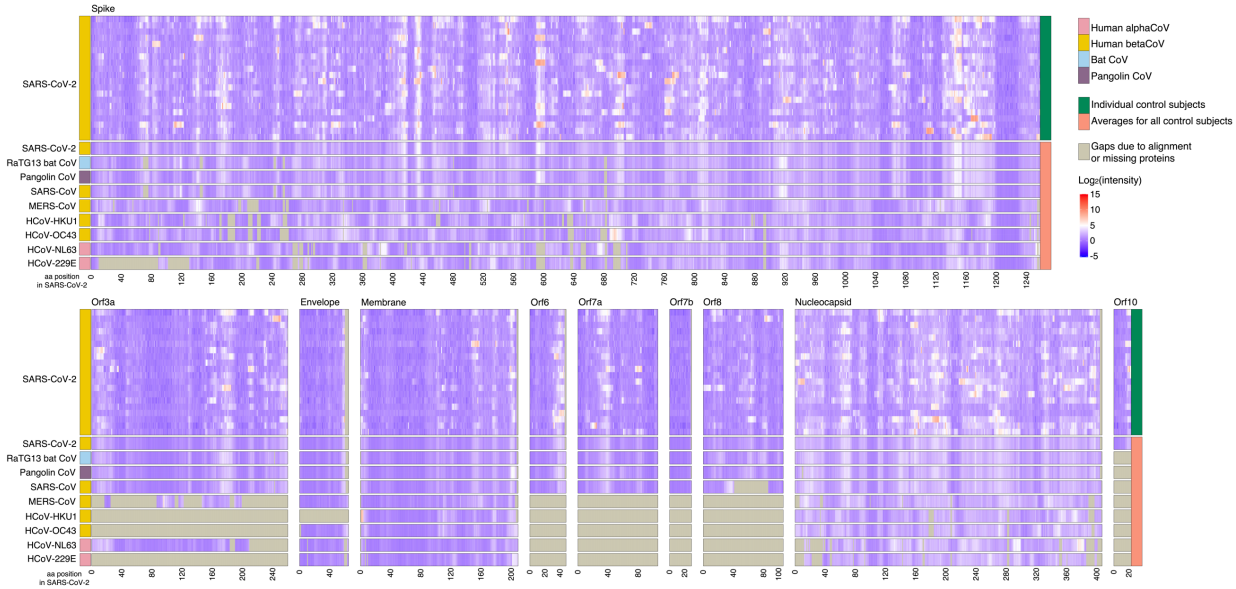

Figure 2. Control sera show reactivity frequently to CCCoVs and rarely to SARS-CoV, MERS-CoV, and SARS-CoV-2. Sera from 20 control subjects collected before 2019 were assayed for IgG binding to the full proteomes of 9 CoVs on a peptide microarray. Viral proteins are shown aligned to the SARS-CoV-2 proteome with each virus having an individual panel; SARS-CoV-2 amino acid (aa) position is represented on the x-axis. Binding was measured as

reactivity that was >3 standard deviations above the mean for the log_2-quantile normalized array data.

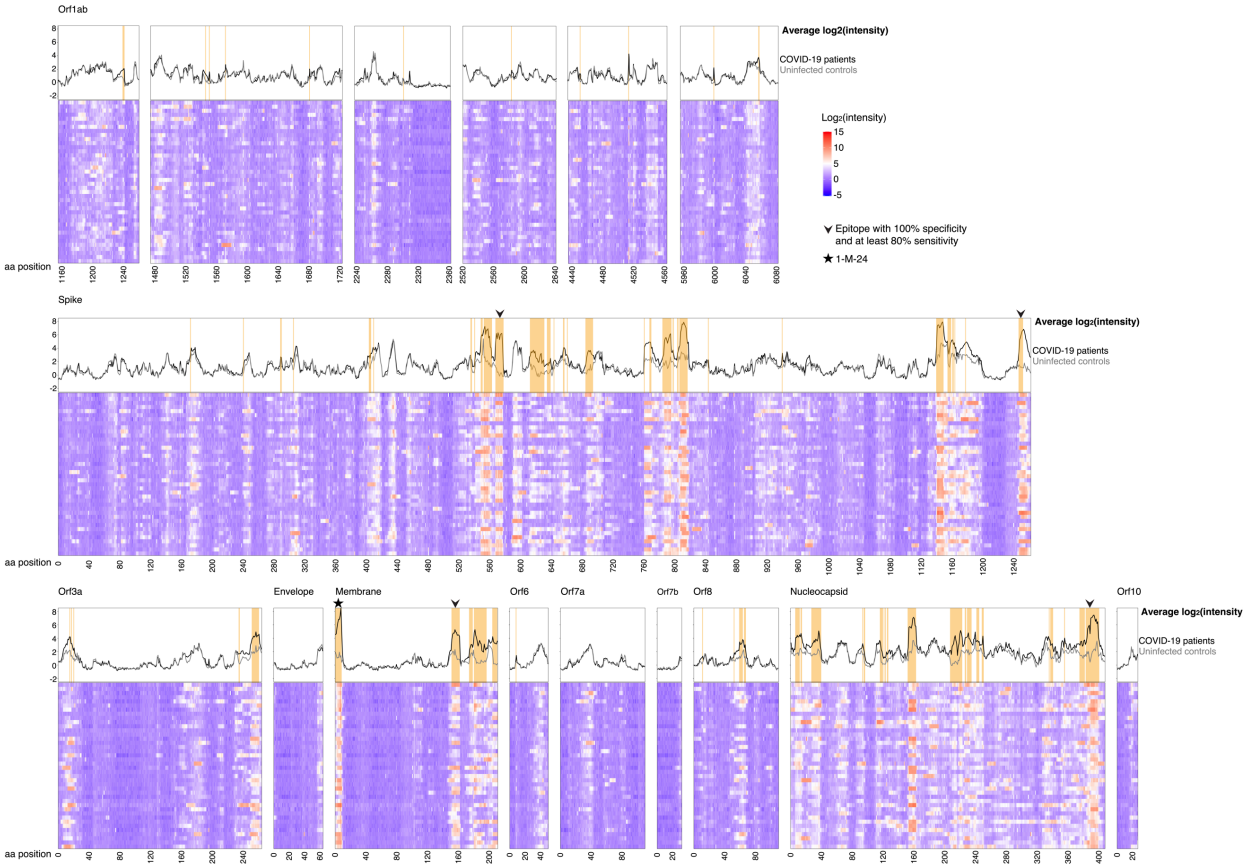

Figure 3. Anti-SARS-CoV-2 antibodies bind throughout the viral proteome. Sera from 40 COVID-19 convalescent subjects were assayed for IgG binding to the full SARS-CoV-2 proteome on a peptide microarray. B cell epitopes were defined as peptides in which patients' average log_2-normalized intensity (black lines in line plots) is 2-fold greater than controls' (gray lines in line plots) and *t*-test statistics yield adjusted *p*-values < 0.1; epitopes are identified by orange shading in the line plots.

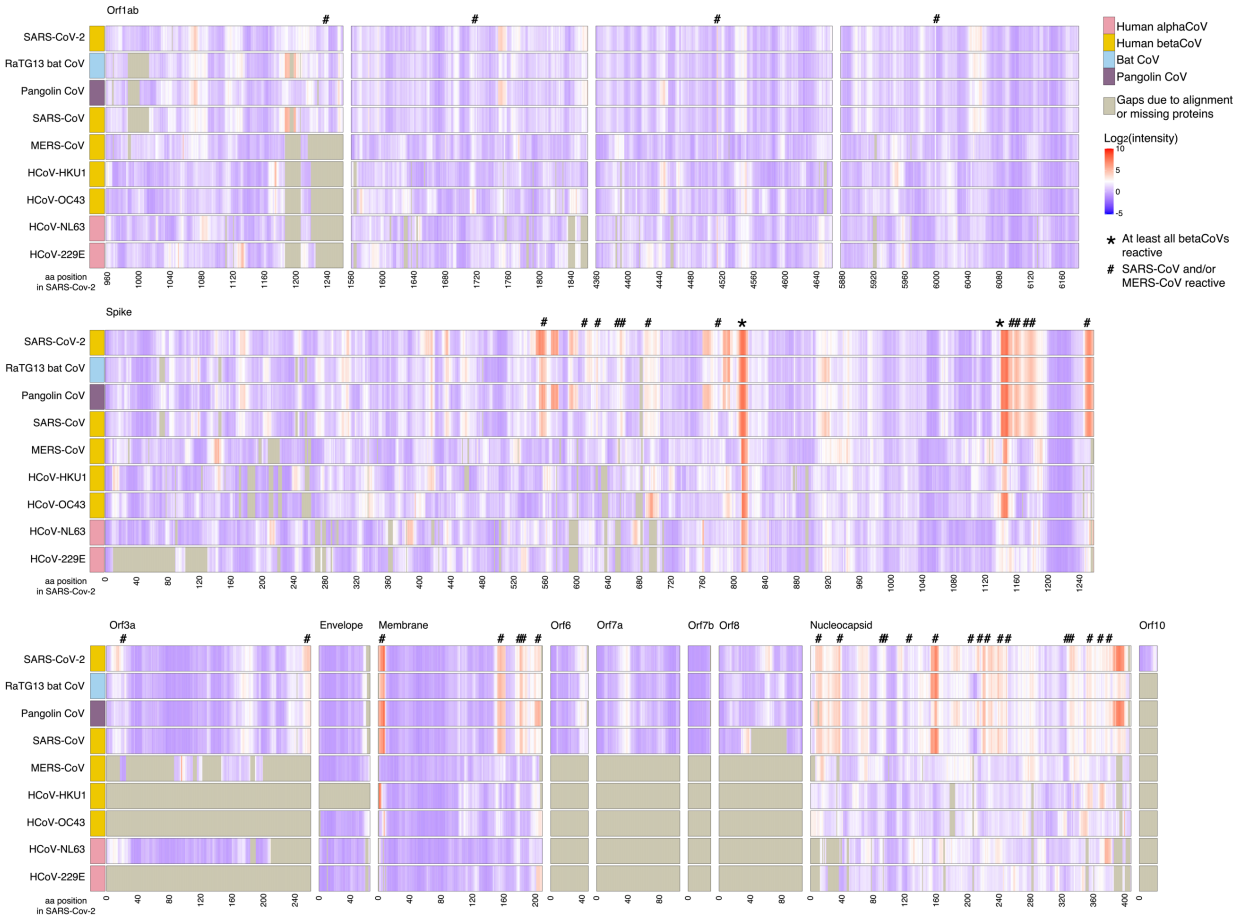

Figure 4. Anti-SARS-CoV-2 antibodies cross-react with other CoVs. Sera from 40 COVID-19 convalescent patients were assayed for IgG binding to 9 CoVs on a peptide microarray; averages for all 40 are shown. Viral proteins are aligned to the SARS-CoV-2 proteome; SARS-CoV-2 amino acid (aa) position is represented on the x-axis. Regions cross-reactive across all β-CoVs (*) or cross-reactive for SARS-CoV or MERS-CoV (#) are indicated. Gray shading indicates gaps due to alignment or lacking homologous proteins. Cross-reactive binding is defined as peptides in which patients' average $\log_2$-normalized intensity is 2-fold greater than controls' and *t*-test statistics yield adjusted *p*-values < 0.1.

Extended data

Extended data 1. All 40 COVID-19 convalescent patients and all 20 naïve controls reacted to known epitopes in at least one control virus (rhinovirus and poliovirus strains).

Extended data 2. Percentages of the 40 COVID-19 convalescent patients and 20 naïve controls reacted to known epitopes in at least one control virus (rhinovirus and poliovirus strains).

Extended data 3. B cell epitopes in the SARS-CoV-2 proteome identified by antibody binding in 40 COVID-19 convalescent patients compared to 20 naïve controls.

Extended data 4. Specificity and sensitivity for past SARS-CoV-2 infection in 40 COVID-19 convalescent patients compared to 20 naïve controls of individual 16-mer peptides comprising epitopes throughout the full SARS-CoV-2 proteome.

Extended data 5. Epitopes paired with the 1-M-24 epitope obtained an area under the receiver operating characteristic curve (AUC-ROC) of 1.0 for SARS-CoV-2 infection in 40 COVID-19 convalescent patients and 20 naïve controls using leave-one-out cross validation with linear discriminant analysis.

Extended data 6. Alignment of epitopes in human and animal CoVs for which antibodies in sera from 40 COVID-19 convalescent patients showed apparent cross-reactive binding. Alignments were performed in Geneious Prime 2020.1.2 (Auckland, New Zealand).

Extended data 7. Cross-reactive binding of antibodies against other CoVs in 40 COVID-19 convalescent patients compared to 20 naïve controls.

Extended data 8. Cross-reactive binding of antibodies in 40 COVID-19 convalescent patients compared to 20 naïve controls in protein motifs in other CoVs aligned to SARS-CoV-2.

Extended data 9. B cell epitopes in the SARS-CoV-2 proteome identified by antibody binding in 40 COVID-19 convalescent patients compared to 20 naïve controls were differentiated using a cut-off of at least a 2-fold greater magnitude reactivity in patients vs controls and *t*-test statistics yielding adjusted *p*-values <0.1. Degrees of freedom, fold change, and standard deviation for each peptide are given.

Methods

Peptide microarray design and synthesis

Viral protein sequences were selected and submitted to Nimble Therapeutics (Madison, WI, USA) for development into a peptide microarray[48]. Sequences represented include proteomes of all seven coronaviruses known to infect humans, proteomes of closely related coronaviruses found in bats and pangolins, and spike proteins from other coronaviruses (accession numbers and replicates per peptide shown in **Supplementary Table 1**). A number of proteins were included as controls, including poliovirus, seven strains of human rhinovirus, and human cytomegalovirus 65kDa phosphoprotein. We chose these controls given that we expect most human adults will have antibody reactivity to at least one of these proteins and proteomes. Accession numbers used to represent each viral protein are listed in the supplemental material (accession numbers and replicates per peptide shown in **Supplementary Table 1**). All proteins were tiled as 16 amino acid peptides overlapping by 15 amino acids. All unique peptides were tiled in a lawn of thousands of copies, with each unique peptide represented in at least 3 and up to 5 replicates (**Supplementary Table 1**). The peptide sequences were synthesized in situ with a Nimble Therapeutics Maskless Array Synthesizer (MAS) by light-directed solid-phase peptide synthesis using an amino-functionalized support (Geiner Bio-One) coupled with a 6-aminohexanoic acid linker and amino acid derivatives carrying a photosensitive 2-(2-nitrophenyl) propyloxycarbonyl (NPPOC) protection group (Orgentis Chemicals). Unique peptides were synthesized in random

positions on the array to minimize impact of positional bias. Each array consists of twelve subarrays, where each subarray can process one sample and each subarray contains up to 389,000 unique peptide sequences.

Supplementary Table 1. Proteins represented on the peptide microarray

	Protein(s)	**GenBank accession number(s)**	**Number of replicates of each unique peptide**
Coronavirus proteins	Severe acute respiratory syndrome coronavirus 2 proteome	NC_045512.2	4-5
	Severe acute respiratory syndrome coronavirus proteome	NC_004718.3	3
	Middle Eastern respiratory syndrome coronavirus proteome	NC_019843.3	3
	Human coronavirus HKU1 proteome	NC_006577.2	3
	Human coronavirus OC43 proteome	NC_006213.1	3
	Human coronavirus 229E proteome	NC_002645.1	3
	Human coronavirus NL63 proteome	NC_005831.2	3
	Bat coronavirus (RaTG13 isolate) proteome	MN996532.1	3
	Pangolin coronavirus proteome	MT072864.1	3
Control proteins	Human rhinovirus A1 polyprotein	NC_038311.1	3
	Human rhinovirus A7 polyprotein	DQ473503.1	3
	Human rhinovirus A16 polyprotein	L24917.1	3
	Human rhinovirus A36 polyprotein	JX074050.1	3

	Human rhinovirus C2 polyprotein	EF077280.1	3
	Human rhinovirus C15 polyprotein	GU219984.1	3
	Human rhinovirus C41 polyprotein	KY189321.1	3
	Human poliovirus 1 polyprotein	ANA67904.1	3
	Human cytomegalovirus 65 kDa phosphoprotein	P06725.2	3

Human subjects

The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of the University of Wisconsin-Madison. Clinical data and sera from subjects infected with SARS-CoV-2 were obtained from the University of Wisconsin (UW) COVID-19 Convalescent Biobank and from control subjects (sera collected prior to 2019) from the UW Rheumatology Biobank[49]. All subjects were 18 years of age or older at the time of recruitment and provided informed consent. COVID-19 convalescent subjects had a positive SARS-COV-2 PCR test at UW Health with sera collected 5-6 weeks after self-reported COVID-19 symptom resolution except blood was collected for one subject after 9 weeks. Age, sex, medications, and medical problems were abstracted from UW Health's electronic medical record (EMR). Race and ethnicity were self-reported. Hospitalization and intubation for COVID-19 and smoking status at the time of blood collection (controls) or COVID-19 were obtained by EMR abstraction and self-report and were in complete agreement. Two thirds of COVID-19 convalescent subjects and all controls had a primary care appointment at UW Health within 2 years of the blood draw as an indicator of the completeness of the medical information. Subjects were considered to have an immunocompromising condition if they met any of the following criteria: immunosuppressing medications, systemic inflammatory or autoimmune disease, cancer not in remission, uncontrolled diabetes (secondary manifestations or hemoglobin A1c $\geq$7.0%), or congenital or acquired immunodeficiency. Control and COVID-19 subjects were similar in regard to demographics and health (**Supplementary Table 2**). No subjects were current smokers.

Supplementary Table 2. Characteristics of COVID-19 Convalescent and Control Subjects

	COVID-19 (n=40)	Control (n=20)	*p*
Age, median (IQR) years	51 (19-83)	55 (22-83)	0.378
Sex, number female (%)	17 (42.5)	11 (55.0)	0.360

Race, number (%)			0.866
White	34 (85.0)	18 (90.0)	
Black	3 (7.5)	1 (5.0)	
Asian	3 (7.5)	1 (5.0)	
Native American	0 (0.0)	0 (0.0)	
Pacific Islander	0 (0.0)	0 (0.0)	
Ethnicity, number Hispanic (%)	5 (12.5)	1 (5.0)	0.361
Charlson comorbidity score, median (IQR)	2 (0, 3)	2 (0.5, 4)	0.551
Immunocompromised, number (%)	9 (22.5)	7 (35.0)	0.302
COVID-19 disease severity, number (%)			
Hospitalized and intubated	8 (20.0)	-	-
Hospitalized without intubation	7 (17.5)	-	-
Not hospitalized	25 (62.5)	-	-

Peptide array sample binding

Samples were diluted 1:100 in binding buffer (0.01M Tris-Cl, pH 7.4, 1% alkali-soluble casein, 0.05% Tween-20) and bound to arrays overnight at 4°C. After sample binding, the arrays were washed 3× in wash buffer (1× TBS, 0.05% Tween-20), 10 min per wash. Primary sample binding was detected via Alexa Fluor® 647-conjugated goat anti-human IgG secondary antibody (Jackson ImmunoResearch). The secondary antibody was diluted 1:10,000 (final concentration 0.1 ng/µl) in secondary binding buffer (1x TBS, 1% alkali-soluble casein, 0.05% Tween-20). Arrays were incubated with secondary antibody for 3 h at room temperature, then washed 3× in wash buffer (10 min per wash), washed for 30 sec in reagent-grade water, and then dried by spinning in a microcentrifuge equipped with an array holder. Fluorescent signal of the secondary antibody was detected by scanning at 635 nm at 2 µm resolution using an Innopsys 910AL

299 microarray scanner. Scanned array images were analyzed with proprietary Nimble Therapeutics
300 software to extract fluorescence intensity values for each peptide.
301
302 **Peptide microarray findings validation**
303 We included sequences on the array of viruses which we expected all adult humans to be likely
304 to have been exposed to as positive controls: one poliovirus strain (measuring vaccine exposure),
305 and seven rhinovirus strains. Any subject whose sera did not react to at least one positive control
306 would be considered a failed run and removed from the analysis. All subjects in this analysis
307 reacted to epitopes in at least one control strain (Fig. 1, Extended data 1, Extended data 2).
308
309 **Peptide microarray data analysis and data availability**
310 The raw fluorescence signal intensity values were $\log_2$ transformed. Clusters of fluorescence
311 intensity of statistically unlikely magnitude, indicating array defects, were identified and
312 removed. Local and large area spatial corrections were applied, and the median transformed
313 intensity of the peptide replicates was determined. The resulting median data was cross-
314 normalized using quantile normalization. All peptide microarray datasets and code used in these
315 analyses can be downloaded from https://github.com/Ong-Research/Ong_UW_Adult_Covid-
316 19.git.
317
318 **Statistical analysis**
319 Statistical analyses were performed in R (v 4.0.2) using in-house scripts. For each peptide, a p-
320 value from a two-sided *t*-test with unequal variance between sets of patient and control
321 responses were calculated and adjusted using the Benjamini-Hochberg (BH) algorithm. To
322 determine whether the peptide was in an epitope (in SARS-CoV-2 proteins) or cross-reactive for
323 anti-SARS-CoV-2 antibodies (in non-SARS-CoV-2 proteins), we used an adjusted *p*-value cutoff
324 of <0.1 (based on multiple hypothesis testing correction for all 119,487 unique sequences on the
325 array) and a fold-change of greater than or equal to 2 and grouped consecutive peptides as a
326 represented epitope. Linear discriminant analysis leave-one-out cross validation was used to
327 determine specificity and sensitivity on each peptide and from each epitope using the average
328 signal of the component peptides.
329
330 To identify cross reactive epitopes, we used each SARS-CoV-2 epitope sequence as a query,
331 searched the database of proteins from the sequences in the peptide array using blastp (-word-
332 size 2, num-targets 4000) to find homologous sequences in the bat, pangolin, and other human
333 CoV strains, then determined whether the average $\log_2$-normalized intensity for these sequences
334 in patients was at least 2-fold greater than in controls with *t*-test statistics yielding adjusted *p*-
335 values <0.1. Each blast hit was then mapped back to the corresponding probe ranges.
336
337 The clinical and demographic characteristics of convalescent subjects were compared to those of
338 the controls using χ^2 tests for categorical variables and Wilcoxon rank-sum tests for non-
339 normally distributed continuous measures.
340
341 Heatmaps were created using the gridtext[50] and complexheatmap[51] packages in R. Alignments
342 for heatmaps were created using MUSCLE[52].
343
344

Funding

I.M.O. acknowledges support by the Clinical and Translational Science Award (CTSA) program, through the NIH National Center for Advancing Translational Sciences (NCATS), grants UL1TR002373 and KL2TR002374. This research was also supported by 2U19AI104317-06 (to I.M.O) and R24OD017850 (to D.H.O.) from the National Institute of Allergy and Infectious Diseases of the National Institutes of Health (www.niaid.nih.gov). A.S.H. has been supported by NRSA award T32 AI007414 and M.F.A. by T32 AG000213. S.J.M. acknowledges support by the National Cancer Institute, NIH and UW Carbone Comprehensive Cancer Center's Cancer Informatics Shared Resource (grant P30-CA-14520). This project was also funded through a COVID-19 Response Grant from the Wisconsin Partnership Program and the University of Wisconsin School of Medicine and Public Health (to M.A.S.), startup funds through the University of Wisconsin Department of Obstetrics and Gynecology (I.M.O.), and the Data Science Initiative grant from the University of Wisconsin-Madison Office of the Chancellor and the Vice Chancellor for Research and Graduate Education (with funding from the Wisconsin Alumni Research Foundation) (I.M.O.).

Acknowledgments

The authors are grateful to Dr. Christina Newman, Dr. Nathan Sherer, Dr. Thomas Friedrich, Dr. Amelia Haj, Dr. James Gern, Dr. Christine Seroogy, and Gage Moreno for their thoughtful comments and helpful discussions in preparing this manuscript.

Author contributions

A.S.H., S.J.M., D.A.B., M.F.A., M.A.S., D.H.O. and I.M.O. conceptualized this study. A.S.H., D.A.B., and I.M.O. created the array design. M.F.A. and M.A.S. selected patient and control samples. A.S.H., M.F.A., and M.A.S. collected patient demographics and characteristics by medical record chart review. A.S.H., S.J.M., D.A.B., S.K., and I.M.O. performed data normalizations, analyses, and validations, and created graphical data visualizations. A.S.H., S.J.M., D.A.B., A.K.S., and I.M.O. performed formal statistical analyses. S.J.M., D.A.B., and I.M.O. wrote the custom software scripts used. A.S.H. wrote the original draft of the manuscript with input from D.H.O. and I.M.O.. A.S.H., S.J.M., D.A.B., M.F.A., and I.M.O. wrote sections of the methods. All authors contributed to review and editing.

The authors declare the following competing interests: A.S.H., S.J.M., D.A.B., M.F.A., S.K., M.A.S., D.H.O., and I.M.O are listed as the inventors on a patent filed that is related to findings in this study. Application: 63/080568, 63/083671. Title: IDENTIFICATION OF SARS-COV-2 EPITOPES DISCRIMINATING COVID-19 INFECTION FROM CONTROL AND METHODS OF USE. Application type: Provisional. Status: Filed. Country: United States. Filing date: September 18, 2020, September 25, 2020.

References:

1. Deeks, J. J. et al. Antibody tests for identification of current and past infection with SARS-CoV-2. *Cochrane Database Syst Rev* **6**, CD013652 (2020).

2. Liu, W. et al. Evaluation of Nucleocapsid and Spike Protein-Based Enzyme-Linked Immunosorbent Assays for Detecting Antibodies against SARS-CoV-2. *J Clin Microbiol* **58**, (2020).
3. Tré-Hardy, M. et al. Analytical and clinical validation of an ELISA for specific SARS-CoV-2 IgG, IgA, and IgM antibodies. *J Med Virol* (2020).
4. Lisboa Bastos, M. et al. Diagnostic accuracy of serological tests for covid-19: systematic review and meta-analysis. *BMJ* **370**, m2516 (2020).
5. Ayouba, A. et al. Multiplex detection and dynamics of IgG antibodies to SARS-CoV2 and the highly pathogenic human coronaviruses SARS-CoV and MERS-CoV. *J Clin Virol* **129**, 104521 (2020).
6. Huang, Y., Yang, C., Xu, X. F., Xu, W. & Liu, S. W. Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. *Acta Pharmacol Sin* **41**, 1141-1149 (2020).
7. Chen, W. Promise and challenges in the development of COVID-19 vaccines. *Hum Vaccin Immunother* 1-5 (2020).
8. Ong, E., Wong, M. U., Huffman, A. & He, Y. COVID-19 Coronavirus Vaccine Design Using Reverse Vaccinology and Machine Learning. *Front Immunol* **11**, 1581 (2020).
9. Chow, S. C. et al. Specific epitopes of the structural and hypothetical proteins elicit variable humoral responses in SARS patients. *J Clin Pathol* **59**, 468-476 (2006).
10. He, Y., Zhou, Y., Siddiqui, P., Niu, J. & Jiang, S. Identification of immunodominant epitopes on the membrane protein of the severe acute respiratory syndrome-associated coronavirus. *J Clin Microbiol* **43**, 3718-3726 (2005).
11. Pang, H. et al. Protective humoral responses to severe acute respiratory syndrome-associated coronavirus: implications for the design of an effective protein-based vaccine. *J Gen Virol* **85**, 3109-3113 (2004).
12. Sui, J. et al. Potent neutralization of severe acute respiratory syndrome (SARS) coronavirus by a human mAb to S1 protein that blocks receptor association. *Proc Natl Acad Sci U S A* **101**, 2536-2541 (2004).
13. ter Meulen, J. et al. Human monoclonal antibody as prophylaxis for SARS coronavirus infection in ferrets. *Lancet* **363**, 2139-2141 (2004).
14. Rockx, B. et al. Structural basis for potent cross-neutralizing human monoclonal antibody protection against lethal human and zoonotic severe acute respiratory syndrome coronavirus challenge. *J Virol* **82**, 3220-3235 (2008).
15. Martin, J. E. et al. A SARS DNA vaccine induces neutralizing antibody and cellular immune responses in healthy adults in a Phase I clinical trial. *Vaccine* **26**, 6338-6343 (2008).
16. Jiang, L. et al. Potent neutralization of MERS-CoV by human neutralizing monoclonal antibodies to the viral spike glycoprotein. *Sci Transl Med* **6**, 234ra59 (2014).
17. Chen, Z. et al. Human Neutralizing Monoclonal Antibody Inhibition of Middle East Respiratory Syndrome Coronavirus Replication in the Common Marmoset. *J Infect Dis* **215**, 1807-1815 (2017).
18. Burton, D. R. & Walker, L. M. Rational Vaccine Design in the Time of COVID-19. *Cell Host Microbe* **27**, 695-698 (2020).
19. Zhou, Y., Jiang, S. & Du, L. Prospects for a MERS-CoV spike vaccine. *Expert Rev Vaccines* **17**, 677-686 (2018).

20. Maier, H. J., Bickerton, E. & Britton, P. *Coronaviruses : methods and protocols* (Humana Press; Springer, New York, 2015).
21. Liu, L. et al. Anti-spike IgG causes severe acute lung injury by skewing macrophage responses during acute SARS-CoV infection. *JCI Insight* **4**, (2019).
22. Lee, W. S., Wheatley, A. K., Kent, S. J. & DeKosky, B. J. Antibody-dependent enhancement and SARS-CoV-2 vaccines and therapies. *Nat Microbiol* (2020).
23. Takano, T., Kawakami, C., Yamada, S., Satoh, R. & Hohdatsu, T. Antibody-dependent enhancement occurs upon re-infection with the identical serotype virus in feline infectious peritonitis virus infection. *J Vet Med Sci* **70**, 1315-1321 (2008).
24. Morens, D. M. & Fauci, A. S. Emerging Pandemic Diseases: How We Got to COVID-19. *Cell* **182**, 1077-1092 (2020).
25. Tseng, C. T. et al. Immunization with SARS coronavirus vaccines leads to pulmonary immunopathology on challenge with the SARS virus. *PLoS One* **7**, e35421 (2012).
26. Jaume, M. et al. SARS CoV subunit vaccine: antibody-mediated neutralisation and enhancement. *Hong Kong Med J* **18 Suppl 2**, 31-36 (2012).
27. Tian, X. et al. Potent binding of 2019 novel coronavirus spike protein by a SARS coronavirus-specific human monoclonal antibody. *Emerg Microbes Infect* **9**, 382-385 (2020).
28. Lv, H. et al. Cross-reactive Antibody Response between SARS-CoV-2 and SARS-CoV Infections. *Cell Rep* **31**, 107725 (2020).
29. Pinto, D. et al. Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature* **583**, 290-295 (2020).
30. Nickbakhsh, S. et al. Epidemiology of Seasonal Coronaviruses: Establishing the Context for the Emergence of Coronavirus Disease 2019. *J Infect Dis* **222**, 17-25 (2020).
31. Gorse, G. J., Patel, G. B., Vitale, J. N. & O'Connor, T. Z. Prevalence of antibodies to four human coronaviruses is lower in nasal secretions than in serum. *Clin Vaccine Immunol* **17**, 1875-1880 (2010).
32. Premkumar, L. et al. The receptor binding domain of the viral spike protein is an immunodominant and highly specific target of antibodies in SARS-CoV-2 patients. *Sci Immunol* **5**, (2020).
33. Poh, C. M. et al. Two linear epitopes on the SARS-CoV-2 spike protein that elicit neutralising antibodies in COVID-19 patients. *Nat Commun* **11**, 2806 (2020).
34. Zhang, B. Z. et al. Mining of epitopes on spike protein of SARS-CoV-2 from COVID-19 patients. *Cell Res* **30**, 702-704 (2020).
35. Li, Y. et al. Linear epitopes of SARS-CoV-2 spike protein elicit neutralizing antibodies in COVID-19 patients. *Cell Mol Immunol* **17**, 1095-1097 (2020).
36. Grifoni, A. et al. A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2. *Cell Host Microbe* **27**, 671-680.e2 (2020).
37. Ahmed, S. F., Quadeer, A. A. & McKay, M. R. Preliminary Identification of Potential Vaccine Targets for the COVID-19 Coronavirus (SARS-CoV-2) Based on SARS-CoV Immunological Studies. *Viruses* **12**, (2020).
38. Crooke, S. N., Ovsyannikova, I. G., Kennedy, R. B. & Poland, G. A. Immunoinformatic identification of B cell and T cell epitopes in the SARS-CoV-2 proteome. *Sci Rep* **10**, 14179 (2020).

39. Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270-273 (2020).
40. Ge, X. Y. et al. Coexistence of multiple coronaviruses in several bat colonies in an abandoned mineshaft. *Virol Sin* **31**, 31-40 (2016).
41. Xiao, K. et al. Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. *Nature* **583**, 286-289 (2020).
42. Tang, T., Bidon, M., Jaimes, J. A., Whittaker, G. R. & Daniel, S. Coronavirus membrane fusion mechanism offers a potential target for antiviral development. *Antiviral Res* **178**, 104792 (2020).
43. Shi, S. Q. et al. The expression of membrane protein augments the specific responses induced by SARS-CoV nucleocapsid DNA immunization. *Mol Immunol* **43**, 1791-1798 (2006).
44. Shrock, E. et al. Viral epitope profiling of COVID-19 patients reveals cross-reactivity and correlates of severity. *Science* (2020).
45. Mishra, N. et al. Immunoreactive peptide maps of SARS-CoV-2 and other human coronaviruses. *bioRxiv* https://www.biorxiv.org/content/10.1101/2020.08.13.249953v1.full.pdf (2020).
46. Woo, P. C., Lau, S. K., Huang, Y. & Yuen, K. Y. Coronavirus diversity, phylogeny and interspecies jumping. *Exp Biol Med (Maywood)* **234**, 1117-1127 (2009).
47. Anthony, S. J. et al. Global patterns in coronavirus diversity. *Virus Evol* **3**, vex012 (2017).
48. Heffron, A. S. et al. Antibody responses to Zika virus proteins in pregnant and non-pregnant macaques. *PLoS Negl Trop Dis* **12**, e0006903 (2018).
49. Holmes, C. L. et al. Reduced IgG titers against pertussis in rheumatoid arthritis: Evidence for a citrulline-biased immune response and medication effects. *PLoS One* **14**, e0217221 (2019).
50. Wilke, C. O. gridtext: Improved Text Rendering Support for 'Grid' Graphics. R package version 0.1.1. (2020).
51. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847-2849 (2016).
52. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797 (2004).