Structural and functional analysis of the interaction between the human microbiota and CEACAMs

Dissertation

zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

(Dr.rer.nat.)

vorgelegt von

Patrizia Bonsignore

an der

Mathematisch-Naturwissenschaftliche Sektion

Fachbereich Biologie

Konstanz, 2021

Tag der mündlichen Prüfung: 11.05.2021

1. Referent: Prof. Dr. Christof Hauck

2. Referent: Prof. Dr. David Schleheck

3. Referent: Prof. Dr. Thomas Mayer

Table of Contents

Chapter III

CEACAM-binding yeasts are recognized by the innate immune receptor CEACAM3

Chapter IV

Classification of epithelial CEACAMs – their role in human-specific colonization by commensal gut bacteria

Acknowledgements

An dieser Stelle möchte ich mich recht herzlich bei allen bedanken, die mir diese Doktorarbeit ermöglicht haben. Mein besonderer Dank gilt Herrn Prof. Dr. Christof Hauck dafür, dass er mir die Möglichkeit gegeben hat an diesem sehr spannenden Thema zu arbeiten, für seine großartige Betreuung und die Möglichkeit auch eigene Ideen einzubringen und umzusetzen. Vielen Dank, dass ich Teil der CEACAM-Familie sein durfte.

Zudem möchte ich mich auch recht herzlich bei Herrn Prof. Dr. David Schleheck und Prof. Dr. Thomas Mayer für die bereitwillige Übernahme des Zweitgutachtens und des Prüfungsvorsitzes bedanken.

Ein großer Dank geht auch an meine Arbeitsgruppe, von der mir jeder Einzelne sehr ans Herz gewachsen ist und die mir eine unvergesslich schöne Zeit bereitet hat. Ich werde euch alle schrecklich vermissen.

Besonders bedanke ich mich auch bei Susanne Feindler-Boeckh, Claudia Hentschel, Petra Zoll-Kiewitz, Anne Keller und Petra Schnurr, den *guten Seelen* des Labors, die immer alles am Laufen gehalten haben und jeder Zeit ein offenes Ohr für mich hatten. Außerdem danke ich meinem Labornachbarn Timo Baade für die vielen Gespräche, Diskussionen und die Organisation vieler schöner, feucht-fröhlicher Labor-Grillfeste, unserem Lieblings-Post-Doc Johannes Kuiper für seine offene, fröhliche und hilfsbereite Art und Tanja Grimm und Clovis Hugues Seumen Tiogang für den regelmäßigen Einblick in ein anderes, nicht-CEACAM-beinhaltendes, wissenschaftliches Themengebiet. Besonders möchte ich mich auch bei meinen CEACAM-Artgenossen Jonas Adrian und Tamara Schuhmacher bedanken. Zum einen für ihre Unterstützung und die tolle Zusammenarbeit, die schließlich mit dir, Jonas, zu dieser gemeinsamen, fantastischen Publikation geführt hat. Wir waren ein großartiges Team! Zum anderen möchte ich mich bei den Beiden für unser allmorgendliches Frühstück, unsere vielen anregenden Gespräche, Kaffeepausen und den ganzen Spaß bedanken, den wir zusammen hatten. In diesem Zusammenhang komme ich nicht umhin Kollegin Goob zu erwähnen. Liebe Griseldis, auch dir möchte ich von ganzem Herzen für ALLES danken, vor allem aber für deine Freundschaft.

Bedanken möchte ich mich auch bei all meinen Freunden, die mich während dieser langen Zeit begleitet haben. Ein ganz besonderer Dank geht hierbei an Joana Thiel, meine ständige Wegbegleiterin im Studium, die schnell von der *Ökö-Tante aus dem ersten Semester* hin zu meiner *engsten Freundin* avanciert ist. Ich bin so stolz, dass wir das Studium und jetzt auch noch unsere Promotion gemeinsam durchgezogen und vor allem durchgestanden haben.

Von ganzem Herzen möchte ich mich auch noch bei meiner Familie und besonders bei meinen Eltern bedanken, die mir das alles erst ermöglicht haben. Vielen Dank für eure grenzenlose Unterstützung und eure bedingungslose Liebe. Wenn ich auf etwas noch stolzer bin als auf diesen Abschluss, dann darauf eure Tochter zu sein.

Zum Schluss möchte ich noch meinem Mann Raffael danken, der mich immer unterstützt, an mich glaubt und mir auch in schwierigen Zeiten unbeirrt zur Seite steht.

Summary

The human body represents an excellent ecological niche for a versatile community of microbes such as bacteria, archaea, fungi, and protozoa. Besides commensal and mutualistic residing microorganisms, pathogens regularly attempt colonization, for which they have developed diverse strategies to gain foothold in the human host. One strategy involves a subgroup of the immunoglobulin super family: carcinoembryonic antigen-related cell adhesion molecules (CEACAMs). Epithelial members of the surface expressed CEACAM proteins serve as receptors for certain pathogens such as *Neisseria gonorrhoeae*, *Haemophilus influenzae*, or *Helicobacter pylori*. Remarkably, severe diseases caused by those pathogens occur rarely due to a sophisticated defense provided by the same protein family: the neutrophil granulocyte-expressed innate immune receptor CEACAM3. CEACAM3 recognizes CEACAM-binding pathogens and initiates a rapid, opsonin-independent clearance.

In this work we provide an overview of the state of the art with respect to CEACAM3, including its structure, functions and evolutionary background (Chapter I). Subsequently, we focus on the evolution of CEACAM3 within the primate lineage. We analyze genomes of higher primates identifying new CEACAM3 orthologs. Comparison of primate CEACAM3 genes reveal an extremely fast evolving extracellular domain, whereas the intracellular signal transducing part appears to be conserved. Testing the ability of different primate CEACAM3 variants to recognize human restricted pathogens demonstrates decreasing binding affinity with increasing phylogenetic distance. Exchanging single amino acids in gorilla CEACAM3 towards human CEACAM3 reestablish recognition of the pathogen *Haemophilus aegyptius*. Remarkably, several pathogens, such as *Haemophilus influenzae* or *Neisseria gonorrhoeae* exhibit adhesins that target CEACAM1, but circumvent recognition by the highly similar Ig_V-like domain of CEACAM3. We unveil a single amino acid variation between both receptors that is crucial for CEACAM1 binding, but prevents CEACAM3 association in *Neisseria gonorrhoeae*. An additional mutation at another site within the extracellular domain of CEACAM3 allows the reestablishment of *Haemophilus influenzae* recognition by the adapted CEACAM3 variant. Interestingly, a human CEACAM3 polymorphism exhibiting these amino acid alterations is found in around 40% of the African population. The selection for CEACAM3 variants with an extended binding spectrum demonstrates ongoing adaptation and counter-adaptation between pathogen and host (Chapter II).

Recognition by CEACAM3 and interaction with epithelial CEACAMs is not exclusively found in bacteria, but is also observed for the opportunistic pathogenic yeast *Candida albicans*. By breaching the mucosal barrier and causing local and systemic infections (Candidiasis), this organism constitutes a serious threat to the health of the human host. We are interested in whether the innate immune receptor CEACAM3 might play a part in defense mechanisms against *C. albicans.* In our study, binding assays reveal that CEACAM3 is able to recognize a broad range of yeasts, which underlines the exceptional protective spectrum of this immune receptor. We observe that CEACAM interaction is depended on growth conditions of *Candida* such as iron limitation or the presence of serum, as commonly experienced for tissue infiltrating microbes. Interestingly, *Candida* is only recognized by human CEACAM3 and the closely related chimpanzee CEACAM3 but not by other primate orthologs, indicating a more recent evolutionary development. Still, the proteinaceous adhesin of *Candida* involved in CEACAM-binding remains elusive. Presumably this adhesin activates the innate immune receptor CEACAM3 and causes intracellular receptor tyrosine phosphorylation but does not result in CEACAM3-dependent uptake. Together, these results implicate CEACAM3 in the recognition and activation of downstream signalling following yeast-binding to host phagocytes (Chapter III).

Each mammalian species has its own characteristic microbiota that is shaped by two major factors: nutrition and host genetics. While the influence of diet has been studied extensively, we aim to investigate the role of host genetics in regard to microbial compositions, by putting members of the CEACAM family in centre of our study. CEACAMs are exploited by human-restricted pathogens for host colonization and have a high diversification in mammals, therefore representing a suitable host-specific factor that could help to shape a characteristic microbiota. In our study, CEACAM-associating commensal gut bacteria are enriched from human stool samples and identified using 16S rRNA gene pyrosequencing. Although the CEACAM-binding spectrum of these commensals is as versatile as their phylogenetic origin, they all bind to CEA, a CEACAM member exclusively expressed in epithelial cells. Interestingly, bacterial species that are preferentially found in humans show restriction to human CEACAMs. In contrast, *Enterococcus faecalis*, known to colonize not only human but also mouse intestine, possesses a broadened binding spectrum that also comprises other mammalian CEACAM proteins. Infection of cells with and without CEACAM-expression reveals no difference in bacterial interaction on cellular surfaces, suggesting a minor contribution during early

colonization events. However, fluorescently labeled sugar residues derived from the highly glycosylated CEACAM molecules can be recovered in *E. faecalis* proteins after CEACAM interaction. We suggest that CEACAMs may not be involved in early cell surface association processes, but that microbe interaction with host-specific CEACAM proteins and subsequent degradation of linked carbohydrate moieties could support long-term colonization by providing an additional docking site and energy source. In this way, CEACAM molecules could contribute to the formation of a host characteristic microbial community (Chapter IV).

Altogether, this study provides insight into the complex interplay between pathogens, host, and the commensal microbiota using the example of the CEACAM family.

Zusammenfassung

Der menschliche Körper bietet eine hervorragende ökologische Nische für eine Vielzahl verschiedener Mikroben, zu denen u.a. Bakterien, Archaeen, Pilze und Protozoen gehören. Neben kommensalen und mutualistischen Mikroorganismen, versuchen auch Pathogene regelmäßig sich anzusiedeln. Um im menschlichen Wirt Fuß zu fassen, haben sie verschiedenste Strategien entwickelt. Eine Strategie umfasst eine Untergruppe der Immunglobulin-Superfamilie: die Carcinoembryonic Antigen-related Cell Adhesion Molecules (CEACAMs). Epitheliale Mitglieder der oberflächenexprimierten CEACAM-Proteine dienen bestimmten Pathogenen wie *Neisseria gonorrhoea*, *Haemophilus influenzae*, oder *Helicobacter pylori* als Rezeptoren. Bemerkenswerterweise lösen diese Erreger nur selten schwerwiegende Erkrankungen aus. Grund dafür ist ein raffinierter Verteidigungsmechanismus, bereitgestellt von derselben Proteinfamilie: der Immunrezeptor CEACAM3, welcher auf neutrophilen Granulozyten exprimiert wird. CEACAM3 erkennt CEACAM-bindende Pathogene und initiiert auf eine opsonin-unabhägige Weise deren rasche Beseitigung.

Diese Arbeit enthält eine Übersicht des aktuellen Wissensstandes über CEACAM3 im Hinblick auf dessen Struktur, Funktion und evolutionären Hintergrund (Kapitel I). Wir betrachten den evolutionären Verlauf von CEACAM3 innerhalb der Primatenlinie und identifizieren bei der Genomanalyse höherer Primaten ein neues CEACAM3-Ortholog. Der Vergleich von CEACAM3-Genen aus verschiedenen Primaten zeigt eine extrem schnell evolvierende extrazelluläre Domäne, während der intrazelluläre, signalvermittelnde Teil sich als stark konserviert herausstellt. Tests zeigen, dass die Fähigkeit verschiedener Primaten-CEACAM3-Varianten human-spezifische Pathogene zu erkennen, mit zunehmender phylogenetischer Distanz abnimmt. Der Austausch einzelner Aminosäuren in Gorilla-CEACAM3 hin zu humanem CEACAM3 resultiert in der Erkennung des Pathogens *Haemophilus aegyptius*. Bemerkenswerterweise verfügen einige Pathogene wie *Haemophilus influenzae* oder *Neisseria gonorrhoeae* über Adhäsine, die zwar auf CEACAM1 abzielen, aber die Erkennung durch die sehr ähnliche Ig_V-Domäne von CEACAM3 vermeiden. Wir ermitteln eine einzelne abweichende Aminosäure zwischen den beiden Rezeptoren, die entscheidend für die Bindung an CEACAM1 ist, aber die Bindung von CEACAM3 an *Neisseria gonorrhoeae* verhindert. Eine zusätzliche Mutation an anderer Stelle erlaubt zudem die Erkennung von *Haemophilus*

influenzae durch die angepasste CEACAM3-Variante. Interessanterweise findet sich ein humaner CEACAM3-Polymorphismus mit genau diesen Aminosäureveränderungen in etwa 40% der afrikanischen Bevölkerung. Eine Selektion auf CEACAM3-Varianten mit einem erweiterten Bindespektrum verdeutlicht die permanente Anpassung und Gegenanpassung zwischen Pathogenen und ihrem Wirt (Kapitel II).

Die Erkennung durch CEACAM3 und die Interaktion mit epithelialen CEACAMs findet nicht ausschließlich bei Bakterien statt, sondern wird auch bei der Hefe *Candida albicans* beobachtet. Durch das Durchdringen der Schleimhautbarriere und das Verursachen von lokalen und systemischen Infektionen (Candidose), stellt dieser Organismus eine ernsthafte Bedrohung für die Gesundheit des menschlichen Wirtes dar. In dieser Studie wird untersucht, ob der Immunrezeptor CEACAM3 an Abwehrprozessen gegen *C. albicans* beteiligt ist. Bindungsstudien zeigen, dass CEACAM3 in der Lage ist eine Auswahl veschiedenster Hefen zu erkennen, was das äußerst protektive Spektrum dieses Immunrezeptors unterstreicht. Außerdem beobachten wir, dass die Interaktion von *C. albicans* mit CEACAMs von Wachstumsbedingungen wie Eisenmangel oder Serumverfügbarkeit abhängt; mit diesen Bedingungen werden üblicherweise Mikroben konfrontiert, die bis ins Gewebe vordringen. Interessanterweise wird *Candida* nur von menschlichem CEACAM3 und dem nahe verwandten Schimpansen CEACAM3 erkannt, jedoch nicht von anderen Primatenorthologen, was auf eine evolutionär neuere Entwicklung hindeutet. Das protein-basierte *Candida* Adhäsin, welches an CEACAM-Interaktionen beteiligt ist, konnte bis jetzt nicht näher definiert werden. Vermutlich aktiviert eben jenes Adhäsin CEACAM3 und verursacht die beobachtete intrazelluläre Tyrosinphosphorylierung des Rezeptors, welche jedoch nicht in einer CEACAM3-abhängigen Aufnahme resultiert. Unsere Ergebnisse zeigen, dass Hefen von CEACAM3 erkannt werden, was die Aktivierung von nachgeschalteten Signalen zur Folge hat (Kapitel III).

Jede Säugetierart verfügt über eine eigene, charakteristische Mikrobiota, die durch zwei Hauptfaktoren beeinflusst wird: die Ernährung und die Genetik des Wirtes. Während der Einfluss der Ernährung schon vielfältig untersucht wurde, zielen wir darauf ab, die Rolle der Wirtsgenetik im Hinblick auf die mikrobielle Zusammensetzung zu untersuchen. Im Zentrum unserer Untersuchung stehen hierfür Mitglieder der CEACAM-Familie. CEACAM-Proteine werden von Pathogenen, die sich ausschließlich auf den Menschen beschränken, für die Wirtskolonisierung ausgenutzt und weisen außerdem eine große Vielfalt in Säugern auf. Das

macht sie, im Hinblick auf Wirtsspezifität, zu Faktoren, die dabei helfen könnten eine charakteristische Mikrobiota zu entwickeln. In unserer Studie werden kommensale Darmbakterien aus menschlichen Stuhlproben angereichert und mittels 16S rRNA Gen-Pyrosequenzierung identifiziert. Obwohl das CEACAM-Bindespektrum der Kommensalen sich als ebenso vielfältig herausstellt wie ihre phylogenetische Herkunft, binden alle an CEA, ein Mitglied der CEACAM-Familie das ausschließlich durch Epithelzellen exprimiert wird. Bakterien, die vorwiegend im Menschen gefunden werden, zeigen eine Beschränkung auf menschliche CEACAMs. Im Gegensatz dazu zeigt *Enterococcus faecalis*, der dafür bekannt ist nicht nur den menschlichen Darm, sondern auch den von Mäusen zu besiedeln, ein erweitertes Bindespektrum, welches auch andere Säugetier-CEACAM Proteine umfasst. Die Infektion von Zellen mit und ohne CEACAM-Expression zeigt keinen Unterschied in der Interaktion von Bakterien mit der Zelloberfläche, was auf einen eher unwesentlichen Beitrag von CEACAMs während des frühen Kolonisierungsgeschehens hinweist. Jedoch können fluoresenz-markierte Zuckerreste aus den stark glykosylierten CEACAM-Molekülen in *E. faecalis* Proteinen nach dessen Interaktion mit CEACAMs wiedergefunden werden. Wir nehmen an, dass CEACAMs zwar nicht in frühe Assoziierungsprozesse mit der Zelloberfläche involviert sind, aber, dass die Interaktion der Mikroben mit wirtsspezifischen CEACAM-Proteinen und der anschließende Abbau der gebundenen Kohlenhydratreste eine langfristige Kolonisierung begünstigen kann, da dadurch eine zusätzliche Andockstelle und Energiequelle vom Wirt zur Verfügung gestellt wird. Auf diese Weise könnten CEACAM-Moleküle einen Beitrag zu der Entstehung einer wirtscharakteristischen Mikrobengemeinschaft leisten (Kapitel IV).

Zusammengefasst gibt diese Studie Einblick in das komplexe Zusammenspiel zwischen Pathogenen, Wirt und der kommensalen Mikrobiota am Beispiel der CEACAM-Familie.

General Introduction

The human microbiota

The human body is colonized by a complex community of different microorganisms composed of bacteria, fungi, archaea, protozoa, and viruses that reside on or within their host. The entity of those microbes is called *the human microbiota* and can mainly be found on human skin, in the nasopharynx, the oral cavity, the urogenital tract, and the gastrointestinal tract. The highest microbe abundance in count, as well as species diversity, is localized in the human gut. Performing a number of essential functions, the gut microbiota has great impact and significance to health of the human body. Sometimes it is called *the forgotten organ*, as the role of the human microbiota was underestimated for a long time (O'Hara and Shanahan 2006). Microorganisms inhabiting the human body were thought to colonize its host randomly, depending on environment and nutrition. Only during the last decades, it emerged that actually a strict regulation determines which organisms can colonize and to which extent they can settle down. It became clear that those microbes represent an entity whose members are capable of communicating with each other and even more important, who are capable of communicating with the human host itself. Furthermore, the same microbes turned out to supply various beneficial functions for the human body including modulation of metabolism, interference with the innate immune system, and regulation of epithelial development by providing an additional set of genes that broaden the enzymatic spectrum of the host (Savage 1977; Whitman, Coleman et al. 1998; Ley, Peterson et al. 2006; Wang and Li 2015). The collective genomes of those microbes are therefore summarized by the term *the human microbiome*. Nowadays, an extensive research field is aiming to understand functions of this complex community of microorganisms and tries to unveil its contribution to the upkeep of a healthy human body.

Acquisition and maturation

Development and maturation of the human microbiota takes place during the first year of life. Experts are divided into two camps regarding the question at which point acquisition actually starts: those who support the *sterile womb paradigm* based on a sterile fetal environment where the first contact with microorganisms occurs during and directly after birth (Funkhouser and Bordenstein 2013), and those who support the *in utero colonization hypothesis* that assume bacterial communities to reside and therefore to colonize the fetus already within the

uterus (Jimenez, Marin et al. 2008; Aagaard, Ma et al. 2014; Collado, Rautava et al. 2016). In both scenarios, however, the neonate's microbiota is influenced first vertically (from the mother) and then horizontally (from other humans or the environment). The extent of influence nowadays depends on the way of delivery though. During natural delivery the baby passes the birth canal and thereby gets in contact with vaginal and gut flora of the mother, which supports the development of a normal and healthy gut microbiota in neonates. During a caesarean delivery the baby is primarily exposed to the skin flora of the midwife, doctors, parents, and environmental microbes, leading to colonization of untypical microorganisms in the gut and the occurrence of pathogenic bacteria such as *Staphylococcus*. Therefore the risk for the development of diseases such as allergies, asthma, or diabetes mellitus increases (Wall, Ross et al. 2009). A second essential influence factor with regard to the development of a normal gut microbiota is the infant's diet. Breast milk builds an important source of the probiotic bacterium *Lactobacillus*, which supports the generation of a healthy gut flora (Soto, Martin et al. 2014). Formula-fed babies that lack this probiotic source also tend to develop the aforementioned diseases. The human gut is an anoxic environment that shelters mainly strict anaerobic microorganisms (>99%). However, in the beginning of life, aerobes inhabit the human gut as first colonizers. Consuming the prevailing oxygen, they create a reduced atmosphere allowing anaerobic organisms to settle down. After one year, an adult gut flora with its characteristic composition has developed (Stark and Lee 1982).

Function of a healthy human gut microbiota

The human gut microbiota has essential functions for the health of the human host reaching from metabolic and trophic tasks, including digestion of nutrition, degradation of host-indigestible polysaccharides, and synthesis of vitamins to host protection by outcompeting pathogens (Grenham, Clarke et al. 2011; Neish 2014; Trompette, Gollwitzer et al. 2014). The relationship between gut microbiota and the human host is not simply a non-harmful coexistence, but rather a mutualistic one where each party profits from the other (Quigley 2013). Gut microorganisms extract energy from various food components including the fermentation of dietary fibers. In return, they provide valuable metabolic products for the human host. One example would be the production of vitamins such as vitamin K_2, as well as B-vitamins that serve as essential coenzymes for metabolic host processes (Hill 1997). Interestingly, vitamin biosynthetic genes are not specific for a certain clade, but they are broadly distributed across different gut species even originating from distinct phyla (Das,

Babaei et al. 2019). Other benefits produced by the gut microbiota are short chain fatty acids (SCFAs), such as butyrate, propionate or acetate, which are taken up by the host cells (Quigley 2013; Clarke, Stilling et al. 2014). Butyrate represents the main energy source for epithelial cells and favors their proliferation. It improves junctional integrity between them and thereby supports the maintenance of the epithelial cell barrier (Mathewson, Jenq et al. 2016). Furthermore, butyrate and propionate influence immune responses by potentiating the regulatory T (Treg) cell population (Arpaia, Campbell et al. 2013; Furusawa, Obata et al. 2013). Treg cells can suppress inflammation by inhibiting T cell proliferation and cytokine production, thereby maintaining self-tolerance and preventing autoimmunity (Kondelkova, Vokurkova et al. 2010). By competing for the same intestinal niches and substrates, the intestinal microbiota also counteracts colonization and distribution of pathogenic microbes, such as the opportunistic pathogenic yeast *Candida albicans* (Kennedy and Volz 1985). Likewise, the commensal *Bacteroides thetaiotaomicron* competitively excludes *Citrobacter rodentium* by consuming carbohydrates in the intestinal lumen that are required for the pathogen's growth (Kamada, Kim et al. 2012). Probiotics like *Lactobacillus* and *Bifidobacterium* protect the human host from overgrowth of the opportunistic pathogen *Clostridium difficile* by limiting the adherence to epithelial cells and decreasing the cytotoxicity of the pathogen (Valdes-Varela, Alonso-Guervos et al. 2016). Some commensals also produce antimicrobial substances such as bacteriocines (cell wall-active bactericidal polypeptides), to challenge the colonization of certain pathogens like multidrug resistant enterococci, *Clostridium difficile*, or *Staphylococccus aureus* (Millette, Cornut et al. 2008; Park, Shin et al. 2008; Crowther, Baines et al. 2013; Hammami, Fernandez et al. 2013; Garcia-Gutierrez, Mayer et al. 2019). Furthermore, due to its metabolic products, the human gut microbiota can expand its influence beyond the gastrointestinal tract to even more distal organs and systems. It has the ability to act on the central nervous system (CNS) by modulating hormones or producing hormone-mimicking molecules, therefore resembling an endocrine organ (Forsythe, Sudo et al. 2010; Evans, Morris et al. 2013). To fulfill all those functions properly and to hinder the development of certain diseases such as inflammatory bowel diseases (IBD), colon cancer, or simply the overgrowth of pathogens, it is crucial to establish and maintain a well-balanced gut microbiota. At this point our immune system plays an essential role: to avoid constant inflammatory response, it needs to tolerate beneficial microbes while still counteracting harmful ones.

The mutual relationship of the human microbiota and the host immune system

The human immune system must detect a wide range of pathogens including viruses, bacteria, fungi, protists, and multi-cellular parasites, and needs to distinguish them from the host's commensal microorganisms. Training for this task starts immediately after birth. During the birth process, the neonate and therefore also the neonate's immune system experiences its first exposure to commensals; a co-development starts. First factors that shape the baby's microbiota and the response to new microbes are derived from mother's milk and contain immune cells, commensals (e.g. *Lactobacillus*), metabolites, IgA, and cytokines. Oligosaccharides present in breast milk promote the growth of certain beneficial organisms such as *Bifidobacterium* (Marcobal, Barboza et al. 2010; Marcobal and Sonnenburg 2012). The neonate's immune system is trained by early regulation and interaction with commensals during the development, which finally leads to the establishment of a stable homeostasis between host and microbes during the first years of life. To maintain this homeostatic state throughout lifetime, communication between microbiota and host is essential. Therefore, microorganisms exhibit conserved microbial-associated molecular patterns (MAMPs), which are recognized by immunosensory cells. Those cells possess a host pattern recognition receptor (PRR) system that allows the distinction of pathogenic from commensal microbes. The host epithelium constitutes the first sensory line of defence. Active sampling of antigens, for instance derived from commensals or pathogens, is conducted by three main types of immunosensory cells. These cells are involved in the perpetuation of homeostasis, not only by recognizing MAMPs, but also by controlling overgrowth and diffusion of residing microbes to avoid tissue invasion and constant inflammatory response (Figure 1). Enterocytes, the first immunosensory cell type, secrete cytokines and chemokines to direct cells of the innate and adaptive immune system to the site of infection. Additionally, they express an immunoglobulin receptor that binds to IgA that is produced and secreted on the apical side by plasma cells. This binding results in the transportation of IgA across the mucosal epithelium to the apical side (Snoeck, Peters et al. 2006). IgA serves as direct defense mechanism and helps to shape the composition of the microbiota (Shanahan 2005; Wei, Shinkura et al. 2011). M cells, the second immunosensory cell type, are specialized epithelial cells that selectively endocytose luminal antigens and transport them to the basal side of the cell. There, antigen-presenting cells such as macrophages, lymphocytes and dendritic cells take over the antigens and migrate to lymph nodes in order to initiate immune response. DCs, the third cell type, can

directly sample gut content by extending dendrites between enterocytes (Rescigno, Urbano et al. 2001). In case of tissue invasion, DCs internalize the microbe and present it to mesenteric lymph nodes to induce immune response (Macpherson and Uhr 2004). The PRR system of those immunosensory cells, that allows microbe distinction, includes receptors such as Toll-like receptors (TLRs), the nucleotide-binding oligomerization domain/caspase recruitment domain isoforms (NOD/CARD) and C-type lectins (CLECs) (Cario 2005; Iliev, Funari et al. 2012). They hold an important role in immune-cell activation, and appear to be essential for microbiota-host communication. NOD proteins and TLRs are expressed by enterocytes and DCs and get activated by commensals. TLRs are involved in enterocyte proliferation, the reparation of the gut, and therefore play an important role for the intestinal barrier function (Rakoff-Nahoum, Paglino et al. 2004; Fukata, Michelsen et al. 2005). The CLEC Dectin-1 targets the β-1,3-glucan structure, a component of the fungal cell wall. Since Dectin-1 regulates commensal fungi colonization, Dectin-1 deficiency elevates the susceptibility to the inflammatory bowel disease colitis (Iliev, Funari et al. 2012). Many PRR ligands are derived from commensals, but in contrast to pathogen-associated molecular patterns (PAMPs), MAMPs do not trigger inflammatory response. In a healthy colonic mucosa, enterocytes express high levels of the TLR inhibitor Toll-interacting protein (Tollip) and the single immunoglobulin IL-1R-related molecule (SIGIRR) (Otte, Cario et al. 2004). SIGIRR is a Toll/interleukin-1 receptor-containing inhibitory molecule. Similar to Dectin-1, its absence leads to higher susceptibility for colitis. Together with Tollip, SIGIRR reduces inflammatory response to the surrounding environment by inhibiting TLR activity. This process seems to mediate mucosal tolerance towards the commensal microbiota (Garlanda, Riva et al. 2004). In addition, tolerance is induced by regulatory T cells and DCs as they are able to inhibit excessive cytokine response to the microflora by T helper 1 cells (Th1) (Rook and Brunet 2005). The inflammatory response to pathogens normally triggers the activation of the nuclear factor (NF)-κB pathway. In contrast, most commensals do not activate NF-κB but are able to inhibit the pathway by inducing nuclear export of the NF-κB subunit, p65, to the cytoplasm (Kelly, Campbell et al. 2004). Another strategy to avoid immune response is the presentation of host-related molecules on the surface of commensals. *Bacteroides fragilis*, for instance, synthesizes fucosylated surface molecules, thereby mimicking fucosylated glycoproteins and glycolipids of intestinal enterocytes resulting in immune hyporesponsiveness (Coyne, Reinap et al. 2005).

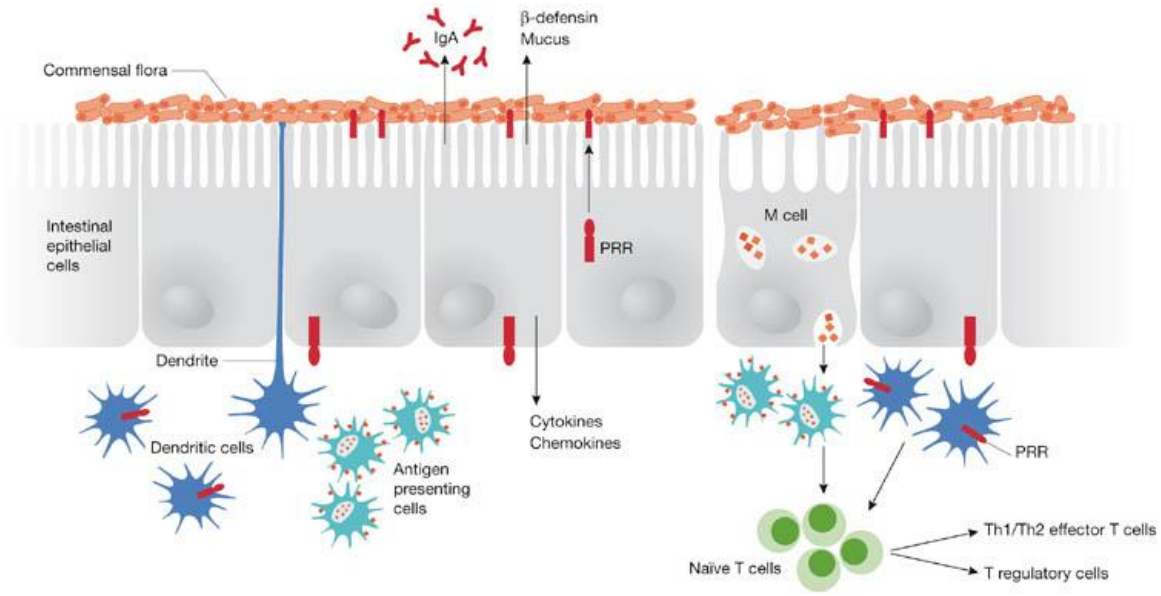


Figure 1: Immunosensory cells control and shape the human gut microbiota. Enterocytes release cytokines and chemokines to recruit immune cells to the site of infection. IgA secretion defends against pathogens and shapes the microbiota composition. M cells endocytose luminal antigens and transport them to the basal side of the cell. There, antigen-presenting cells (macrophages, lymphocytes, dendritic cells (DCs)) take over the antigens and migrate to lymph nodes to initiate immune response. DCs sample gut content by extending dendrites between enterocytes and internalize tissue invading microbes. For bacterial antigen recognition, DCs exhibit pattern recognition receptors (PRRs). Additionally, DCs regulate immune response or tolerance by stimulating either effector or regulatory T cells (O'Hara and Shanahan 2006).

Microbiota composition and diversity in the human gastrointestinal tract

The human gut microbiota is mainly composed of strict anaerobic microorganisms, which clearly outnumber the facultative anaerobes and aerobes by far (Gordon and Dubos 1970; Savage 1970). Although there are more than 50 bacteria phyla in general, only two phyla prevail in the human gut: the Bacteroidetes and the Firmicutes (Schloss and Handelsman 2004). A bacterial minority belongs to other phyla such as Proteobacteria, Actinobacteria, Fusobacteria, and Verrucomicrobia (Eckburg, Bik et al. 2005). Numbers of bacterial species found in the human gut vary dramatically between different studies; it is estimated that up to 1,000 different species colonize this organ (Xu and Gordon 2003). Interestingly, the intestinal

microbiota is a heterogeneous community that varies in number as well as composition depending on the intestine location (Figure 2 A). 10^1 bacteria per gram of content reside in the stomach, increasing to 10^3 to 10^7 bacteria per gram in the duodenum, jejunum, and ileum and finally reaching the maximal bacterial cell number in the colon with 10^{11} to 10^{12} bacteria per gram intestinal content (O'Hara and Shanahan 2006). Along with the bacterial number the bacterial diversity increases, whereas the oxygen amount decreases. In addition to cell number and diversity, also the composition of microorganisms varies (Frank, St Amand et al. 2007). Whereas the stomach shows microaerophilic and aerotolerant genera like *Helicobacter* and *Lactobacillus*, the small intestine accommodates facultative anaerobes like Actinobacteria and the Firmicutes class Bacilli. In contrast, the colon is enriched by strict anaerobic microbes such as the Firmicutes family Lachnospiraceae and members of the Bacteroidetes phylum (Frank, St Amand et al. 2007). Besides the longitudinal microbiota alteration, also the latitudinal bacterial composition varies (Figure 2 B). The intestinal epithelium is lined by a thick mucus layer that separates it from the intestinal lumen. Different areas are inhabited by diverse microbes. *Bacteroides*, *Bifidobacterium*, *Streptococcus*, members of Enterobacteriacea, *Enterococcus*, *Clostridium*, *Lactobacillus*, and *Ruminococcus* are present in the lumen, whereas *Clostridium*, *Lactobacillus*, and *Enterococcus* are predominant in the mucus layer and the epithelial crypts of the small intestine (Swidsinski, Loening-Baucke et al. 2005). A third factor in microbiota composition is a temporal variation (Figure 2 C). Directly at birth, or even within the uterus, colonization by first microbes starts. Depending on the way of delivery and feeding habit, the primary colonizers differ and can affect the development of the intestinal microbiota enduringly. During the first year of life, the number of microorganisms and their diversity increases strongly. By the end of the first year, the microbiota has settled down and resembles that of an adult (Mackie, Sghir et al. 1999). In elderly persons, the microbial diversity decreases again and the phyla ratios shift, resulting in elevated levels of inflammatory markers and enhanced frailty (Bartosch, Fite et al. 2004; van Tongeren, Slaets et al. 2005). Nevertheless, all through life a permanent alteration takes place influenced by external factors such as diet, medication use, or environmental exposures (Turnbaugh, Backhed et al. 2008). Importantly, those alterations only affect the human microbiota on the genera and species level, while the phylum level is stable and only shows variation in the proportions (Eckburg, Bik et al. 2005; Gill, Pop et al. 2006). This suggests that the maintenance of certain phyla is essential, whereas the functional redundancy within a

phylum allows variation in microbiota composition between individuals without impairing proper function.

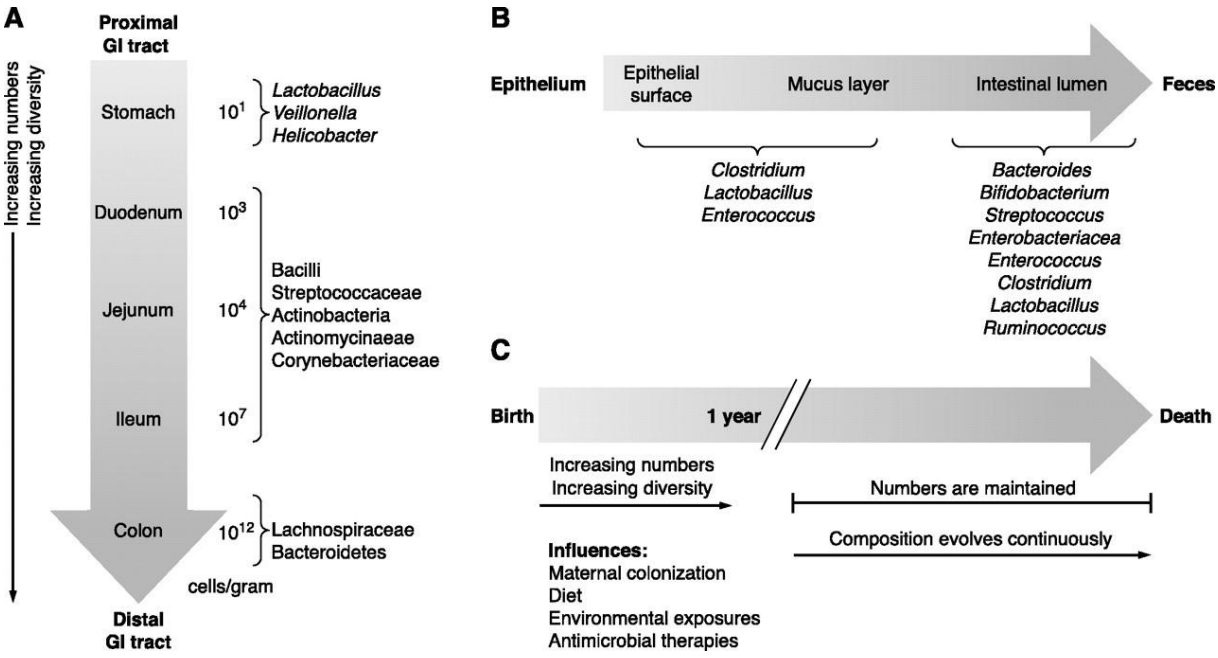


Figure 2: Local and temporal variation of gastrointestinal microbiota composition. A: Longitudinal alterations across the gastrointestinal tract. B: Latitudinal microbial variations in the intestine. C: Temporal course of microbiota establishment and maintenance and the influence of certain factors to microbial composition (Sekirov, Russell et al. 2010).

Certain influence factors modulate microbial composition

The homeostasis of the human gut microbiota establishes early in life, but can later on be affected by various factors that influence composition and diversity. The foundation is laid by the infant's diet (breast or formula milk), but also in later life stages the impact of personal nutrition remains significant. The fecal microbiota of European children eating a typical western diet, high in animal protein, sugar, starch, fat, and low in fiber was compared to that of African children from a rural village with a diet low in sugar, starch, and fat and high in fiber (Figure 3). The study revealed a significant difference in the proportions of the four main phyla. Whereas Actinobacteria and Bacteroidetes dominated the African children's microbiota, Firmicutes and Proteobacteria were prevalent in European children (Figure 3) (De

Filippo, Cavalieri et al. 2010). Within the phylum of Bacteroidetes especially bacteria from the genus *Prevotella* and *Xylanibacter* were enriched. These are known to exhibit genes for cellulose and xylan hydrolysis, which appears to be an adaptation to a polysaccharide-rich diet, allowing a maximized energy intake from fibers. In addition, significantly more short-chain fatty acids were found to be produced by African children's microbiota, known to have a protective function against inflammations which decreases the risk for the development of certain diseases (e.g. inflammatory bowel diseases, colon cancer) (De Filippo, Cavalieri et al. 2010; Liu, Wu et al. 2015). Interestingly, European children showed high abundance of the phylum Firmicutes, which is positively correlated with the typical western disease obesity (Ley, Turnbaugh et al. 2006), as well as the phylum Proteobacteria, that is known to comprise the majority of pathogenic bacteria in humans (Rizzatti, Lopetuso et al. 2017). Not only nutrition, but also medication use can have a severe impact on the microbial community. For instance, antibiotic therapy not only aims at pathogenic bacteria, but also affects commensals, thereby compromising richness and diversity of the human microbiota. Depletion of residing microbes favors intrusion and accumulation of pathogens in unoccupied niches. Resulting perturbation and imbalance of the commensal gut microbiota can cause intestinal problems, such as antibiotic-associated diarrhea (AAD) (Nord, Heimdahl et al. 1986). Besides extrinsic factors, such as diet, age, and medication use, also the intrinsic factor *host genetics* plays an essential role. Investigation of gut microbiota derived from different great ape species, including human, common chimpanzee, bonobo, western gorilla, and eastern gorilla, revealed that the phylogenetic branching order based on the microbial composition is consistent with phylogenetic relationships of the hosts, reflecting a co-evolutionary process (Ochman, Worobey et al. 2010). To better understand co-evolution of the indigenous microbiota and the mammalian host, Ley et al. analyzed and evaluated feces of 60 mammalian species. The results indicated not only diet, but also phylogeny to influence bacterial diversity significantly (Ley, Hamady et al. 2008). The fact that each mammalian species has its own characteristic microbiota points toward the existence of host-specific factors, which allow colonization and composition in a species-specific manner. A well-studied example demonstrating the complex and highly specific interplay between microbe and host is the beneficial symbiosis between the Hawaiian Bobtail Squid (*Euprymna scolopes*) and *Vibrio fischeri*, a bioluminescent bacterium that colonizes the squids light organ. The luminescence of *V. fisheri* at night provides *E. scolopes* with counter-illumination camouflage that prevents the squid from

casting a shadow on the ocean floor and protects its host from predators. At the same time, *V. fisheri* is provided with shelter and a stable source of nutrients delivered by the host (Jones and Nishiguchi 2004). Newly hatched squids are horizontally colonized by *V. fisheri* originating from the surrounding seawater (Nyholm and McFall-Ngai 2004). During the colonization process, the host secrets mucus, that can only be colonized by *V. fisheri* via its type IV pili. Additionally, ciliated host cells within the light organ selectively draw the bacteria in and promote their growth. Other bacterial species are actively rejected. Once the light organ is sufficiently colonized by *V. fisheri*, the bacterium causes the ciliated host cells to die, preventing other microorganisms from colonization (McFall-Ngai and Ruby 1991). This host-association causes luminescence dependent morphological changes in the squids light organ. *V. fisheri* produces and secrets an autoinducer that allows quorum sensing (density-dependent bacterial cell-cell communication), whereby the transcription of the lux operon is activated once a certain concentration of bacteria is reached. This activation results in bioluminescence (Miyashiro and Ruby 2012). During dawn, *E. scolopes* ejects 90% of the symbiotic bacteria in a process called “venting” in a diel rhythm. In this process, the apical surfaces of epithelial cells start blebbing into the crypt spaces and are removed. The epithelium restores, however, within hours after venting, allowing the remaining *V. fisheri* to grow and to recover its population until dusk (Rader and Nyholm 2012).

In contrast to species-specific colonization by commensals in non-mammalian species or the well-studied field of species-specific colonization by pathogens, such factors are widely unknown for commensal microorganisms in humans. Identification of host-specific markers utilized by commensal microbes would represent an important step toward the better understanding of host colonization processes and the formation of characteristic microbiota composition patterns in the human body.

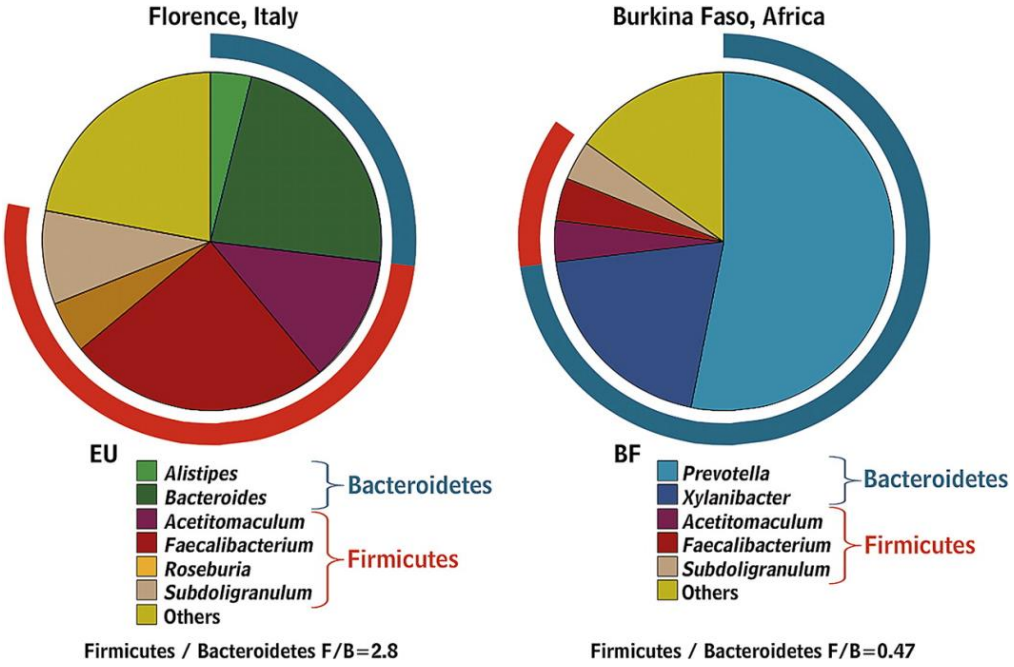


Figure 3: Gut microbiota composition of European children and African children from a rural village with a diet high in fibre (Simren, Barbara et al. 2013).

Host-specific microbe colonization

Microbes usually enter the human host via the digestive tract, the respiratory tract, the conjunctiva, or the urogenital tract. Subsequent adherence to mucosal surfaces is a prerequisite to establish a foothold in the human host. In general, the participation of two factors is required: an adhesin and a receptor. The receptor normally is represented by a eukaryotic cell surface structure such as specific carbohydrates or peptide residues. The microbial adhesin is exposed on the microbe's cell surface and mediates the interaction with the host cell receptor. In host-pathogen interactions a distinction is drawn between generalists, who can infect various species and specialists that exclusively colonize a single host species (Baumler and Fang 2013). Prime examples for human-restricted specialists are *Haemophilus influenzae, Helicobacter pylori, Neisseria gonorrhoeae, Neisseria meningitidis, Mycobacterium leprae, Salmonella* Typhi*, Streptococcus pneumoniae, Streptococcus pyogenes, Vibrio cholerae* and *Treponema pallidum* (Pan, Yang et al. 2014), while some pathogens such as *Legionella pneumophila*, *Campylobacter jejuni*, or *Staphylococcus aureus* can infect various animals (Sakwinska, Giddey et al. 2011; Dearlove, Cody et al. 2016; Boamah, Zhou et al. 2017). Many bacterial adhesins target extracellular matrix (ECM) components such

as collagen, laminin, and fibronectin (Patti, Allen et al. 1994). *Streptococcus pyogenes*, for instance, possess an extensive set of ECM-binding adhesins including collagen-, laminin-, and fibronectin-binding proteins that allow the pathogen to cause respiratory tract infection as well as severe infections like septicemia, necrotizing fasciitis, or streptococcal toxic shock-like syndrome by invading deeper tissues (Yamaguchi, Terao et al. 2013; Brouwer, Barnett et al. 2016). *Treponema pallidum*, the causative agent of syphilis, interacts with fibronectin by utilizing two treponemal proteins: M23B subfamily peptidase and Tp0483 (Peterson, Baseman et al. 1983; Cameron, Brown et al. 2004). *Salmonella* Typhi requires a type-3 secretory system-1 (T3SS-1)-associated lipoprotein called InvH to attach to epithelial cells, but uses its type-IVB pili to bind to cystic fibrosis transmembrane conductance regulator (CFTR) protein for human cell invasion (Altmeyer, McNern et al. 1993; Pier, Grout et al. 1998). Whereas the field of pathogen-human interaction is well studied and understood only little is known about commensal bacteria and the molecular mechanisms behind their host interaction. Commensal microbes are known to interact with human glycoproteins to establish foothold in the human body. One of the few well investigated examples would be the genus *Bacteroides* that could be shown to rely on a unique class of Polysaccharide Utilization Loci (PUL), the so-called commensal colonization factors (*ccf*). Deletion of the *ccf* genes in the highly abundant gram-negative gut bacterium *Bacteroides fragilis* results in a disturbed colonization behaviour (S. Melanie Lee, Mazmanian, 2013). To gain foothold, *B. fragilis* is known to specifically bind to intestinal mucins of the human host (Huang JY1, Lee SM, Mazmanian SK., 2011). Mucins are highly glycosylated proteins that are synthesized by epithelial host tissues. Some mucins are membrane-bound due to transmembrane domains however the majority is secreted and represents the main component of intestinal mucus. Mucins have a protective function by binding to pathogens and building a barrier for bacteria and fungi, thereby counteracting the development of infections. Commensal *Neisseria* species (e.g. *N. lactamica* and *N. subflava*) use a distinct colonization strategy. Expression of the adhesion molecule opacity associated (Opa) protein that is localized in the outer membrane in high numbers and gives Opa-positive colonies an opaque appearance, allows the interaction with members of the carcinoembryonic antigen-related cell adhesion molecule (CEACAM) family (Toleman, Aho et al. 2001). CEACAMs are highly glycosylated proteins that are expressed by epithelial, endothelial and hematopoietic cells. They play a role during cell adhesion, differentiation, proliferation, and survival. In the human gut especially CEA (CEACAM5) is highly abundant. It

is either membrane bound via a glycosylphosphatidylinositol (GPI) anchor, or it occurs in a secreted form. Besides exclusively commensal *Neisseria*, also microbes that reside within the human host and are known to have opportunistic pathogenic behavior (e.g. *Neisseria meningitidis*, *Haemophilus influenzae*, *Helicobacter pylori*) bind to CEACAM proteins in order to colonize in a host-specific manner. Since microbe-CEACAM interaction is mainly investigated with regard to pathogens, further investigations will be required to define the role of CEACAM proteins as potential host factors for colonization by commensals.

The CEACAM family

The carcinoembryonic antigen-related cell adhesion molecule (CEACAM) family is a group of mammalian immunoglobulin-related glycoproteins. It belongs to the carcinoembryonic antigen (CEA) family that is part of the immunoglobulin superfamily. The CEA family comprises two subgroups: CEACAMs, which are encoded by 12 genes (Zebhauser, Kammerer et al. 2005) and pregnancy-specific glycoproteins (PSGs), which are encoded by 11 genes in the human genome (Teglund, Olsen et al. 1994). The members of both subgroups have a similar structure: they consist of an amino-terminal immunoglobulin variable (Ig_V)-like domain followed by varying numbers of immunoglobulin constant (Ig_C)-like domains. While PSGs are soluble proteins that are mainly expressed by the placenta and secreted during pregnancy, the majority of the highly-glycosylated CEACAM molecules are anchored to the membrane by either a helical transmembrane domain or a glycosylphosphatidylinositol (GPI) anchor (Figure 4) (Thompson, Grunert et al. 1991; Chang, Semyonov et al. 2013). They are expressed by epithelial cells, endothelial cells, as well as hematopoietic cells and can be found on various tissues (Virji, Makepeace et al. 1996; Gray-Owen, Dehio et al. 1997). CEA (CEACAM5) is exclusively expressed by epithelial cells, whereas CEACAM3, CEACAM4, and CEACAM8 are only present on the surface of neutrophil granulocytes. In contrast, CEACAM1 and CEACAM6 are expressed in both, epithelial cells and hematopoietic cells (Skubitz and Skubitz 2008). CEACAMs are involved in diverse physiological events. Most participate in the modulation of general cellular processes such as cell adhesion, differentiation, proliferation, and survival (Tchoupa, Schuhmacher et al. 2014). But some have a precise functional role like CEACAM16 that is involved in hearing and is exclusively expressed in the inner ear (Zheng, Miller et al. 2011), or the granulocyte receptor CEACAM3 that initiates phagocytosis of certain bacterial pathogens (Schmitter, Agerer et al. 2004). To perform these functions, CEACAMs need to transmit signals into the cell. CEACAM1 and CEACAM3 possess a cytoplasmic domain, capable

of signal transduction (Gray-Owen and Blumberg 2006; Buntru, Roth et al. 2012). Several other CEACAMs only possess a GPI anchor and lack cytoplasmic domains, but are still able to transduce signals by a yet widely unknown mechanism.

Interestingly, CEACAMs are targeted by certain pathogenic bacteria that utilize the molecules as host receptors. Those microbes exploit the Ig_V-like domain of epithelial CEACAMs (CEACAM1, CEA, and CEACAM6) to attach to the mucosal surface, whereby establishment and maintenance of stable colonization of the human host is facilitated (Swanson, Robbins et al. 1987). Binding by bacterial pathogens results in clustering of CEACAM molecules and subsequent induction of cellular responses including cytoskeleton rearrangement, gene expression, enhanced cell adhesion, and endocytosis (Tchoupa, Schuhmacher et al. 2014). Modulation of host cell signals via the immunoglobulin superfamily receptors improves mucosal surface colonization and enhances the chance for prolonged survival within the host. A prime example is the human-restricted pathogen *Neisseria gonorrhoeae*. *N. gonorrhoeae* expresses different Opa proteins that allow tight connection to CEACAM1, CEA, and CEACAM6. This intimate binding enhances epithelial cell adhesion and allows nitric oxic (NO), produced by gonococci under microaerophilic conditions, to permeate epithelial cell membranes. NO initiates a eukaryotic signaling pathway resulting in the upregulation of components that enhance cell-matrix adhesion. As a consequence, exfoliation and delamination of infected cells is efficiently suppressed, thereby creating stable foothold on the mucosa (Muenzner, Rohde et al. 2005; Muenzner, Bachmann et al. 2010; Muenzner and Hauck 2020). In addition, CEACAM engagement triggers gonococcal internalization into epithelial cells and allows *N. gonorrhoeae* to reach subepithelial space via transcytosis (Wang, Gray-Owen et al. 1998).

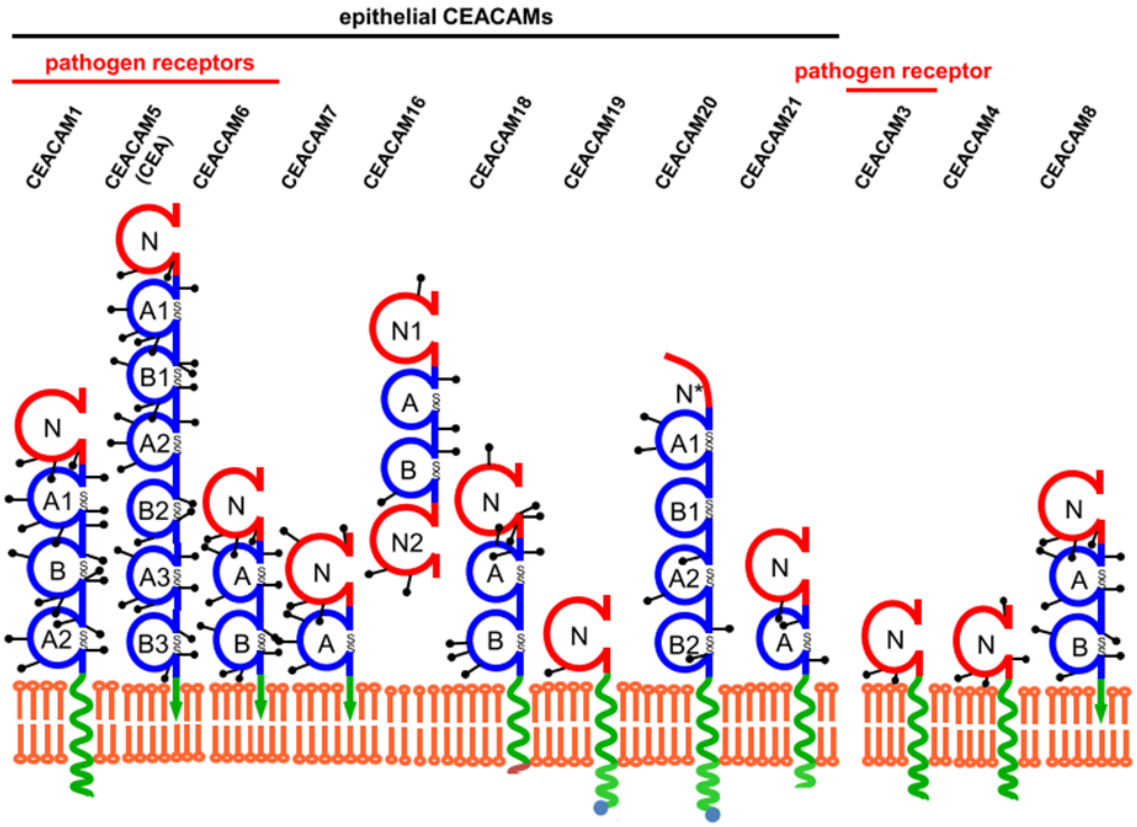


Figure 4: Schema of human carcinoembryonic antigen-related cell adhesion molecule (CEACAM) family members. The immunoglobulin variable (Ig_V)-like domains are indicated by red spheres, the immunoglobulin constant (Ig_C)-like domains by blue spheres that build disulfide bonds (S-S). Transmembrane domains are depicted by green spirals and GPI anchors by green arrows. CEACAM20 exhibits only a partial Ig_V-like domain (N*). The heavy glycosylation is indicated by black sticks. CEACAM1, CEACAM3, CEA, and CEACAM6 serve as pathogen receptor molecules (Tchoupa, Schuhmacher et al. 2014).

CEACAM-binding microorganisms

Due to their broad expression pattern on various host surfaces, epithelial CEACAMs represent an extraordinary advantageous molecular target for colonizing microorganisms. The paradigm for CEACAM-binding microbes is the aforementioned gram-negative diplococcus ***Neisseria gonorrhoeae***. As causative agent of the sexually transmitted disease gonorrhea, it causes 78 million cases per year worldwide (Newman, Rowley et al. 2015). Its high prevalence and rising multi-drug resistance among *N. gonorrhoeae* isolates makes it a serious and imminent threat to human health (Eyre, Sanderson et al. 2018). *N. gonorrhoeae* mainly colonizes the urogenital

tract but can also cause infections in the eye, nasopharynx, and rectum (Noble, Cooper et al. 1979; Lee, Choi et al. 2002; Danby, Cosentino et al. 2016). The hallmark of male urogenital tract infection by gonococci is an inflammatory response, during which massive amounts of neutrophil granulocytes are recruited to the site of infection (Rest and Shafer 1989). Subsequent shedding of colonized epithelial cells helps to remove the neisserial invader. While gonococcal infection in men is mostly accompanied by extensive inflammation, in women gonorrhea more often takes course with only subclinical symptoms. Unnoticed gonococcal infection can lead to pelvic inflammatory disease (PID) that can provoke infertility, ectopic pregnancy, chronic pelvic pain, and cancer (Chan, Seraj et al. 1996; Chang and Parsonnet 2010; Mitchell and Prabhu 2013). *N. gonorrhoeae* can also cause neonatal conjunctivitis by infecting the baby's eyes during passage through the birth canal. Left untreated, the infection can result in blindness of the infant (Sandstrom 1987). In addition, neisserial infection increases the susceptibility to human immunodeficiency virus (HIV), as the massive inflammatory response during gonococcal colonization impairs mucosal integrity (Chen, Boulton et al. 2003).

Neisseria species exhibit 10 to 11 loci in their genome encoding for different Opa variants (Connell, Shaffer et al. 1990; Dempsey, Litaker et al. 1991; Bhat, Gibbs et al. 1992). Opa proteins share the common structure of an outer membrane β-barrel comprised of eight antiparallel β-strands linked by four extended extracellular loops (Malorny, Morelli et al. 1998). One of those four loops is highly conserved between Opa proteins, another loop is semi-variable, and two loops are highly variable due to frequent homologous recombination whereby parts of the variable loops are exchanged between different opa loci (Bilek, Ison et al. 2009). In addition to genomic variation, the number of expressed Opa-proteins within one population can differ strongly due to phase variation. The basis forms a pentameric repeat sequence (CTCTT) within the 5' coding region of *opa* genes (Stern, Brown et al. 1986). Depending on the number of repeats, the opa gene is in frame and therefore expressed. Alternatively, a premature stop codon leads to translation of a non-functional protein. Phase variation allows adaptation to rapidly changing environments and occurs at a frequency of 10^{-3} in *opa* genes, generating a heterogeneous population of gonococci expressing a random combination of multiple Opa proteins, only one single Opa protein, or none at all (Mayer 1982). This flexibility is highly advantageous for immune evasion and therefore increases the virulence of *N. gonorrhoeae* enormously. The importance of Opa proteins in host environment

is highlighted by studies where Opa-negative gonococci grown in culture medium were introduced into male volunteers. Subsequent reisolation from the urethra revealed domination by Opa-positive *Neisseria*, confirming a clear selection for the opaque phenotype *in vivo* (Swanson, Barrera et al. 1988; Jerse, Cohen et al. 1994).

Another species of the genus *Neisseria* that exploits epithelial CEACAMs for colonization is the opportunistic pathogen ***Neisseria meningitidis***. It inhabits up to 10% of the human population as commensal microbe but as soon as it starts to disseminate in the human host, it causes rapid advancing and severe forms of life-threatening septicemia and meningitis (DeVoe 1982; Apicella 2010). In contrast to gonococci, meningococci colonize the mucosal surface of the upper respiratory tract and express only up to four different Opa proteins (Achtman 1995; Virji 1996). Infection of CEACAM-humanized mouse models expressing different CEACAM molecules revealed that expression of human CEACAM1 was essential to establish intranasal colonization. *In vivo,* phase variants that expressed CEACAM1-specific Opa proteins were favored because they allowed mucosal attachment and invasion into the subepithelial space. In line with this finding, Opa-deficient meningococci were unable to colonize CEACAM1-expressing mice (Johswich, McCaw et al. 2013). As mentioned above, the expression of CEACAM-binding Opa proteins is not restricted to pathogenic representatives of the genus *Neisseria* but can also be found in the commensal species *N. lactamica* and *N. subflava* (Toleman, Aho et al. 2001).

Haemophilus influenzae is a gram-negative, human-restricted pathogen that resides within the human nasopharynx alongside with *N. meningitidis.* As a commensal, *H. influenzae* asymptomatically colonizes the mucosal surface of the upper respiratory tract. Under certain circumstances such as immunodeficiency, virus infection, or very young age, it can cause local or even systemic infections such as otitis media, conjunctivitis and sinusitis as well as severe meningitis (Rao, Krasan et al. 1999; Agrawal and Murphy 2011; Norskov-Lauritsen 2014; Van Eldere, Slack et al. 2014). Like *Neisseria*, *H. influenzae* interacts with epithelial CEACAMs to colonize the human host (Bookwalter, Jurcisek et al. 2007). For this purpose, the bacterium expresses a β-barrel shaped adhesin, the outer membrane protein P1 (OMP P1) (Tchoupa, Lichtenegger et al. 2015). OMP P1 is a homolog to the FadL fatty-acid transporter found in *E. coli* and consists of 14 antiparallel β-strands connected by seven extended extracellular loops. Analogous to gonococcal Opa proteins, the extracellular loops dictate the interaction with the Ig_V-like domain of CEACAM proteins. Most *H. influenzae* strains associate with CEACAM1, and

some bind to CEA but so far only *H. influenzae* biotype *aegyptius* is known to interact with the granulocyte receptor CEACAM3 (Tchoupa, Lichtenegger et al. 2015).

Moraxella catarrhalis, the main causative agent of purulent conjunctivitis and sinusitis in adults or otitis media in children, also belongs to the community of respiratory tract inhabiting CEACAM-binders (Hill and Virji 2003). In contrast to the β-barrel adhesin of *Neisseria* and *Haemophilus*, *M. catarrhalis* expresses a structurally distinct adhesion protein: the homotrimeric ubiquitous surface protein A1 (UspA1) (Hill and Virji 2003). UspA1 monomers consist of an extracellular amino-terminal β-sheet-based globular head domain, followed by an elongated stalk region that links the head domain to a membrane immersed β-barrel. This structure allows the formation of a stable docking platform more distant from the bacterial membrane (Hill, Edwards et al. 2005; Koretke, Szczesny et al. 2006). Despite the structural disparity to Opa proteins, UspA1 targets the same amino-terminal region of CEACAMs (Villullas, Hill et al. 2007).

Additionally, pathogenic *Escherichia coli* strains, such as **diffusely adhering *E. coli* (DAEC)** and **adherent-invasive *E. coli* (AIEC)**, exploit epithelial CEACAMs. DAEC contribute to urinary tract infections and exhibit Afa/Dr fimbriae to mediate CEACAM interaction (Guignot, Peiffer et al. 2000; Guignot, Hudault et al. 2009). In contrast, the Crohn's disease-associated *E. coli* strain AIEC expresses FimH, a common type I pilus adhesin, to colonize the intestinal mucosa via CEACAM-binding (Carvalho, Barnich et al. 2009). Sequencing of different AIEC strains isolated from Crohn's disease patients revealed point mutations in the *fimH* gene. These mutations enhanced the ability to adhere to CEACAM-expressing epithelial cells, highlighting the selection for pathoadaptive changes in bacterial pathogens (Dreux, Denizot et al. 2013).

Helicobacter pylori inhabits the upper gastrointestinal tract of more than 50% of the human population (Hooi, Lai et al. 2017). As an opportunistic pathogen, it can cause chronic inflammation, thereby promoting the development of gastric ulcer disease and gastric cancer (Salama, Hartung et al. 2013). *Helicobacter* could be shown to depend on the outer membrane adhesin HopQ that mediates tight association with the gastric epithelium by interacting with CEACAM1. This interaction is essential for translocation of the *H. pylori* toxin CagA into epithelial host cells (Belogolova, Bauer et al. 2013; Javaheri, Kruse et al. 2016). The recently solved crystal structure of HopQ in complex with the N-terminal Ig_V-like domain of CEACAM1 revealed the β-hairpin insertion in HopQ's extracellular 3+4 helix bundle domain to be important for binding to the β-strands C, C', F, and G (CC'FG-face) of CEACAM1. Interestingly,

this CEACAM-interaction side is shared by other adhesins such as Opa proteins or UspA1, although they are structurally unrelated and have evolved independently (Bonsor, Zhao et al. 2018; Moonens, Hamway et al. 2018).

Fusobacterium nucleatum is a gram-negative, anaerobic inhabitant of the human oral cavity and is involved in periodontal diseases. Through a homotrimeric autotransporter adhesin, the CEACAM binding protein of Fusobacterium (CbpF), it mediates binding to CEACAM1 (Brewer, Dymock et al. 2019). The head domain of CbpF consists of a left-handed β-helix and possesses a hydrophobic and highly conserved SSAFG-like sequence allowing the formation of a trimeric β-roll head. A neck region links the head to three strands of α-helices, which coil around each other like known for the stalk region of UspA1, the CEACAM-binding adhesin of *Moraxella catarrhalis* (Brewer, Dymock et al. 2019).

For a long time CEACAM-binding appeared to be an exclusive feature for certain members of gram-negative bacteria. However, the yeast ***Candida albicans*** revealed CEACAM-binding qualities as well and therefore represents the first eukaryote in the community of CEACAM-interacting microorganisms (Klaile, Muller et al. 2017). As an opportunistic pathogen, *Candida albicans* is part of the normal human microflora. However, under predisposing conditions such as antibiotic therapy or in immunocompromised patients it can cause infections reaching from superficial and local up to systemic infections associated with an elevated morbidity and mortality rate (Eggimann and Pittet 2014). To date, it is unknown if CEACAM interaction provides an advantage in regard to host colonization or pathogenicity and by which means *Candida albicans* interacts with human CEACAMs. Some experiments hint to an immunoregulatory function and binding by a proteinaceous *Candida* component, but the specific molecular mechanism has yet to be elucidated (Klaile, Muller et al. 2017).

The convergent evolution of highly diverse CEACAM-binding adhesins in human-restricted pathogens and members of the human microbiota, underlines the benefits those organisms gain by interacting with the receptor family. Exploiting epithelial CEACAMs allows the establishment of host colonization, suppression of epithelial exfoliation, and modulation of immunological processes. However, this advantageous strategy comes at a significant cost, as those microbes become susceptible to the detection by the granulocyte receptor CEACAM3 (discussed extensively in chapter I).

Aim of study

Several human-restricted microbes exploit members of the CEACAM family to colonize mucosal surfaces of the human host. Currently, the majority of the identified CEACAM-binding microorganisms are pathogenic gram-negative bacteria such as *Neisseria gonorrhoeae*, *Neisseria meningitidis* as well as the opportunistic pathogens *Moraxella catarrhalis*, *Haemophilus influenzae*, and *Helicobacter pylori.* But also innocuous commensals such as *Neisseria lactamica* and *Neisseria subflava,* or the opportunistic pathogenic yeast *Candida albicans* were shown to interact with CEACAMs. Microbe overgrowth, infiltration of deeper tissues, or dissemination can cause severe infections. To control CEACAM-binding microorganisms, the human host provides a specialized defense molecule in form of the granulocyte specific innate immune receptor CEACAM3. CEACAM3 recognizes and captures microbes by their CEACAM-targeting adhesins and triggers rapid, opsonin-independent phagocytosis. This study focuses on the molecular details of pathogen recognition by variants of the innate immune receptor CEACAM3 and tries to unravel the role of CEACAMs in general as receptors for commensal bacteria.

I. In the first chapter a review is provided, which summarizes the current state of the art with respect to CEACAM3. Its unique properties including structure and function are highlighted and its emergence during primate evolution is discussed.

II. Subsequently, the evolutionary course of CEACAM3 within the primate lineage was addressed. Genome comparison revealed that CEACAMs, and in particular CEACAM3, are amongst the fastest evolving genes within the human genome. To evaluate the consequences of sequence variations we intended to compare nucleotide sequences of CEACAM3 genes derived from different primate species and analyze them functionally in biochemical and cellular assays. It was of particular interest to understand, how some pathogen adhesins such as certain Opa-proteins of *N. gonorrhoeae* or OMP P1 of *H. influenzae* accomplish to bind CEACAM1, but avoid recognition by CEACAM3. These studies directly led to the question, if particular alleles of CEACAM3 found in distinct human populations determine adhesin-CEACAM3 interactions, which should be investigated by additional biochemical and functional assays.

III. The opportunistic pathogenic yeast *Candida albicans* was identified to interact with epithelial members of the CEACAM family. We asked, if this ability is specific for *C. albicans* or if other yeast species residing within the human host can utilize the receptor family as well. Additionally, we investigated if *C. albicans* recognition by the immune receptor CEACAM3 can initiate immune response. As the yeast ligand for CEACAM3 is currently unknown, we wanted to employ unbiased biochemical and genetic approaches to identify the CEACAM3 target structure on yeast cell surfaces.

IV. The fact that each mammalian species has its own characteristic microbiota points toward the existence of host-specific factors, which determine colonization and composition in a species-specific manner. For commensal microorganism such factors are widely unknown. In contrast, structures and strategies responsible for host-specific colonization by human-restricted pathogens are well studied and provide a starting point for investigation. We hypothesize that epithelial members of the CEACAM family are utilized by different gut bacteria to establish host specific colonization. Therefore, we aimed to identify CEACAM-binding commensals that subsequently should be analyzed for their ability to colonize cell surfaces in a CEACAM-dependent manner.

Chapter I

CEACAM3 – a prim(at)e invention for opsonin-independent phagocytosis

Patrizia Bonsignore[1], Johannes W. P. Kuiper[1], Jonas Adrian[1], Griseldis Goob[1], and Christof R. Hauck[1,2]

[1] Lehrstuhl für Zellbiologie, Universität Konstanz, 78457 Konstanz, Germany

[2] Konstanz Research School - Chemical Biology, Universität Konstanz, 78457 Konstanz, Germany

Abstract

Phagocytosis is one of the key innate defense mechanisms executed by specialized cells in multicellular animals. Recent evidence suggests that a particular phagocytic receptor expressed by human polymorphonuclear granulocytes, the carcinoembryonic antigen-related cell adhesion molecule 3 (CEACAM3), is one of the fastest evolving human proteins. In this focused review we will try to resolve the conundrum, why a conserved process such as phagocytosis is conducted by a rapidly changing receptor. Therefore, we will first summarize the biochemical and structural details of this immunoglobulin-related glycoprotein in the context of the human CEACAM family. The function of CEACAM3 for the efficient, opsonin-independent detection and phagocytosis of highly specialized, host-restricted bacteria will be further elaborated. Taking into account the decisive role of CEACAM3 in the interaction with pathogenic bacteria, we will discuss the evolutionary trajectory of the CEACAM3 gene within the primate lineage and highlight the consequences of CEACAM3 polymorphisms in human populations. From a synopsis of these studies, CEACAM3 emerges as an important component of human innate immunity and a prominent example of a dedicated receptor for professional phagocytosis.

Introduction

The ability to detect and phagocytose microbes is vital to protect multicellular organisms against dangerous infections. In mammals, this important function is accomplished by dedicated immune cells, the so-called professional phagocytes, encompassing macrophages, dendritic cells and polymorphonuclear granulocytes (PMNs). They carry out phagocytosis via two distinct mechanisms: On the one hand, they perform opsonin-independent phagocytosis by utilizing receptors such as mannose receptor, scavenger receptor, Siglecs, DC-SIGN, or Dectin-1, which directly recognize and bind microbial surfaces, which expose characteristic molecular patterns, such as glycan structures with terminal mannose or sialic acid residues, or fungal β-glucans (Underhill and Ozinsky 2002; Crocker, Paulson et al. 2007; Goyal, Castrillon-Betancur et al. 2018). As such types of glycans are found on various microorganisms, including bacteria, fungi, as well as protozoa, and can also occur on endogenous structures, opsonin-independent receptors often have a broad and diverse range of ligands. On the other hand, professional phagocytes are capable of performing opsonin-dependent phagocytosis. Prominent opsonin-dependent receptors are complement- or Fc receptors, which require

prior coating of particles with host-derived complement components or specific antibodies before they are able to initiate phagocytosis (Joshi, Butchar et al. 2006; Nimmerjahn and Ravetch 2008; Degn and Thiel 2013). Therefore, opsonin-dependent receptors can be targeted towards specific microbes, but they cannot support phagocytosis in situations where opsonins are either not present or where they fail to mark the microbial surface, for example in the case of antigenically variable or encapsulated microorganisms.

Recent work has indicated, that at least in primates, a third group of specialized phagocytic receptors operates, which combines pathogen-specific detection with the immediate action of opsonin-independent receptors. The paradigm for this type of phagocytic receptors is the carcinoembryonic antigen-related cell adhesion molecule 3 (CEACAM3). CEACAM3 is a receptor of the immunoglobulin (Ig) superfamily and a member of the CEA subfamily of Ig-domain containing cell adhesion molecules (IgCAMs) (Figure I-1). In humans, CEACAM3 is selectively expressed by PMNs and plays a prominent role in the opsonin-independent detection and elimination of a small set of human-restricted bacteria. In this review, we will place CEACAM3 in the context of a growing list of bacterial pathogens expressing CEACAM-binding adhesins, and discuss the biochemical and functional evidence that this receptor is an effective phagocytosis-initiating protein and granulocyte activator. Further, we will elaborate the evolutionary trajectory of the CEACAM3 gene within the primate lineage and discuss the significance of human CEACAM3 polymorphisms, which appear to accommodate the recognition of variable bacterial surface antigens.

CEACAM family members and their role as microbial targets

Upon the identification of carcinoembryonic antigen (CEA) as a prominent surface protein expressed by human colon carcinomas, it was soon realized that antibodies directed against CEA react with numerous other proteins, especially on granulocytes (Gold and Freedman 1965; Goleski, Rule et al. 1972). According to their apparent molecular weights, these proteins were initially termed non-specific cross-reacting antigen (NCA) -26, -50, -90, -95, and -160 (Kuroki, Matsuo et al. 1990). Screening of a human leukocyte cDNA library with a probe derived from NCA-50 uncovered several transcripts including clone W264 containing a 1,259-base pair insert (Kuroki, Arakawa et al. 1991). The insert encoded a 252 amino acid protein, which was designated Carcinoembryonic Gene family Member 1a (CGM1a) and later grouped, due to its reactivity with monoclonal antibodies, into the CD66 cluster of differentiation.

Besides CGM1a (CD66d), the CD66 antigens comprise biliary glycoprotein (BGP; CD66a), CGM6 (CD66b), NCA-50 (CD66c), and CEA (CD66e), which share 69-92% amino acid sequence identity in their amino-terminal immunoglobulin-variable (Ig_V)-like domains with CGM1a (Benchimol, Fuks et al. 1989; Oikawa, Inuzuka et al. 1989; Kuroki, Arakawa et al. 1991; Oikawa, Inuzuka et al. 1991). Superposition of known crystal structures of CEA, CD66a, and CD66c reveals that these sequence similarities also translate into high structural conservation between these proteins (Bonsor, Zhao et al. 2018; Moonens, Hamway et al. 2018).

The growing awareness of the complexity of the CEA family necessitated a major revision of the nomenclature, which led to CGM1a (CD66d) being renamed CEACAM3 (Beauchemin, Draber et al. 1999) (for current and former nomenclature of CEACAM family members discussed in this review, please consult Figure I-1). While CEACAM3 transcripts and protein have only been detected in human granulocytes and myeloid leukemia cells, the other closely related CD66 antigens are either widely expressed on epithelial and hematopoetic cells (CD66a/BGP/CEACAM1 as well as CD66c/NCA-50/CEACAM6) or exclusively expressed by mucosal epithelial cells (CD66e/CEA/CEACAM5) (Thompson and Zimmermann 1988; Nagel, Grunert et al. 1993; Kuroki, Yamanaka et al. 1995). Furthermore, CEACAM3 is distinct from other CD66 antigens in that its extracellular part does only comprise a single Ig_V-like domain and lacks additional Ig constant (Ig_C)-like domains, which are present in varying numbers (2-6 Ig_C-like domains) in CEACAM1, CEACAM5 and CEACAM6 (Figure I-1) (Thompson, Grunert et al. 1991). This short stature of CEACAM3 might also be the reason why this receptor does not participate in binding interactions with other CEACAM family members, as the Ig_C-like domains can stabilize cis- and trans-interactions between CEACAM extracellular domains and thereby contribute to cell-cell adhesion (Zhou, Fuks et al. 1993; Watt, Teixeira et al. 2001). Indeed, CEACAM1, CEA, CEACAM6 or CEACAM8 engage in CEACAM-CEACAM interactions with each other to support cell-cell binding (Benchimol, Fuks et al. 1989; Zhou, Fuks et al. 1993; Teixeira, Fawcett et al. 1994; Wikstrom, Kjellstrom et al. 1996; Watt, Teixeira et al. 2001).

As CEACAM3 does not participate in these binding interactions, what could then be the function of this particular CEA-related protein on professional phagocytes? Clearly, the physiological role of CEACAM3 can only be reconciled in light of the fact that several pathogenic bacteria and fungi take advantage of epithelial CEACAMs as preferred docking sites on the mucosa (for review see (Tchoupa, Schuhmacher et al. 2014)). Indeed, a growing list of pathogens has been found to express dedicated adhesins to specifically connect to

human CEACAM family members, which are exposed on the apical surface of human epithelial cells such as CEACAM1, CEA, and CEACAM6. These microorganisms comprise *Neisseria gonorrhoeae* (causative agent of the venereal disease gonorrhea), *Neisseria meningitidis* (bacterial meningitis), *Haemophilus influenzae* (pneumonia, bacterial meningitis), *Haemophilus aegyptius* (purulent conjunctivitis), *Helicobacter pylori* (chronic gastritis, stomach cancer), *Moraxella catarrhalis* (otitis media, sinusitis), *Fusobacterium nucleatum* (periodontal disease), pathogenic *Escherichia coli* strains (Adherent-invasive *E.coli*, Diffusely-adherent *E.coli*; involved in Crohn's Disease), and the yeast *Candida albicans* (candidiasis, systemic infections) (Chen and Gotschlich 1996; Virji, Makepeace et al. 1996; Virji, Watt et al. 1996; Hill and Virji 2003; Berger, Billker et al. 2004; Barnich, Carvalho et al. 2007; Tchoupa, Lichtenegger et al. 2015; Koniger, Holsten et al. 2016; Klaile, Muller et al. 2017; Brewer, Dymock et al. 2019). It is important to mention, that almost each of these pathogens employs a structurally distinct adhesive protein to bind human CEACAMs, implying that these bacterial adhesins have evolved independently multiple times in a striking form of convergent evolution (Malorny, Morelli et al. 1998; de Jonge, Bos et al. 2002; Conners, Hill et al. 2008; Korotkova, Yang et al. 2008; Tchoupa, Lichtenegger et al. 2015; Bonsor, Zhao et al. 2018; Moonens, Hamway et al. 2018). Evidently, there must be strong, but not necessarily a uniform selection pressure on these microorganisms to develop CEACAM-binding adhesins. Several non-mutually exclusive explanations have been put forward to explain this exceptional preference of distinct pathogenic microbes to engage human CEACAMs. One finding relates to the fact, that CEACAM1, the target of a large fraction of these adhesins, is also expressed by T cells and that major CEACAM1 isoforms have a negative regulatory role in T cell stimulation and proliferation (reviewed in (Gray-Owen and Blumberg 2006)).

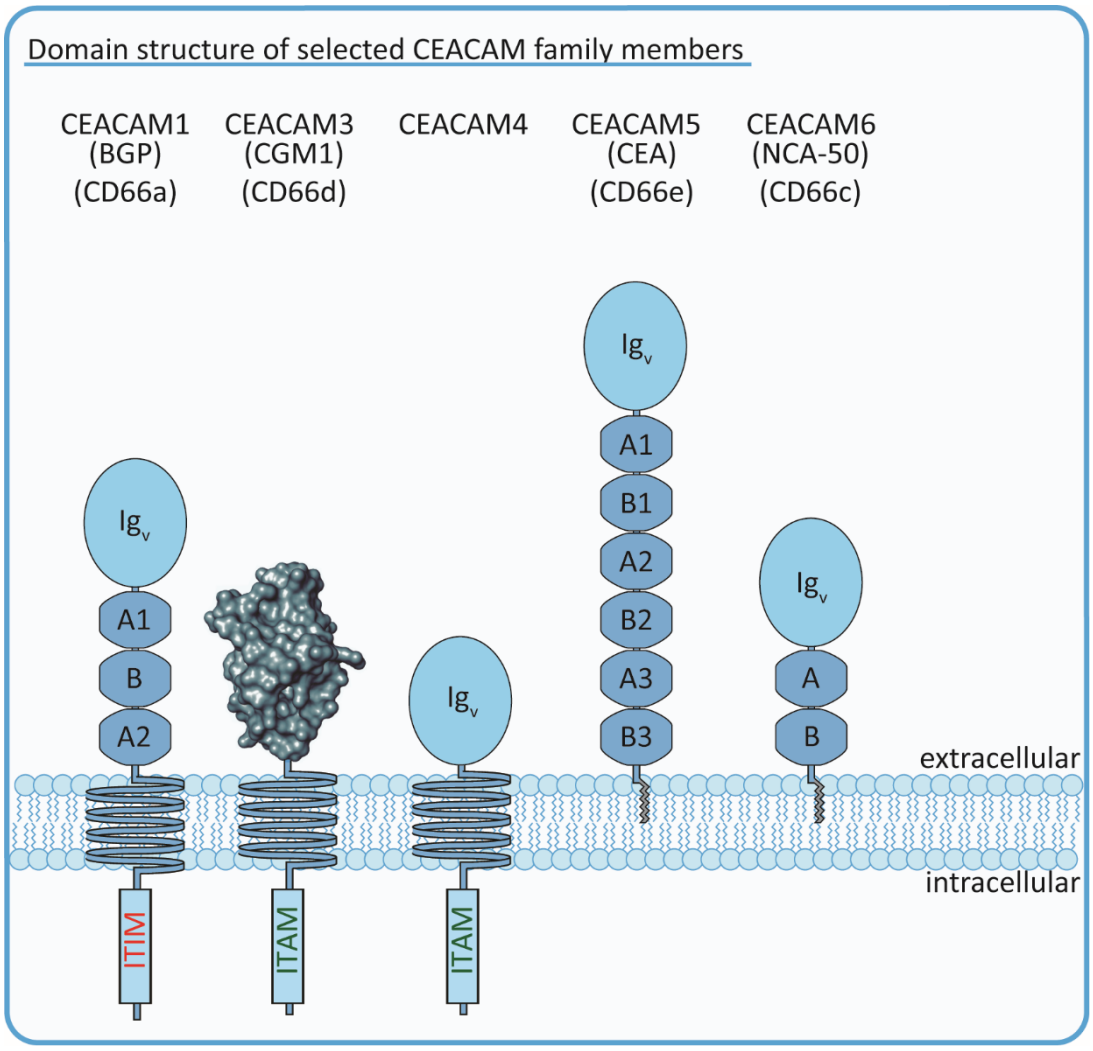


Figure I-1: Schematic drawing of selected members of the human CEACAM family. Schematic outline of several members of the human carcinoembryonic antigen (CEA)-related cell adhesion molecule (CEACAM) receptor family. All CEA-related proteins belong to the immunoglobulin (Ig) superfamily and are characterized by the possession of a homologous amino-terminal Ig variable (Ig_V)-like domain, which is depicted in the case of CEACAM3 as a rendered protein surface according to (Bonsor, Zhao et al. 2018). The blue circles indicate the Ig_V-like domains of CEACAMs other than CEACAM3, while the blue octagons indicate additional Ig constant 2 (Ig_{C2})-like domains occurring in different numbers in particular family members. The transmembrane helices of CEACAM1, CEACAM3, and CEACAM4 connect the extracellular Ig-domains with functional ITIM (CEACAM1), ITAM-like (CEACAM3), or consensus ITAM sequences (CEACAM4). GPI-anchors of CEACAM5/CEA and CEACAM6 are depicted in gray.

A second hypothesis is based on the fact that a unifying theme for all CEACAM-binding microbes is their outstanding ability to colonize, often throughout the lifespan of an individual, the mucosal surface of either the naso-pharynx, the gastro-intestinal, or the urogenital tract. The role of CEACAM-engagement in mucosal colonization has been best worked out in the case of *N. gonorrhoeae* and *N. meningitidis* and demonstrated that both microbes greatly profit from tight association with CEACAMs, which facilitates successful host colonization (Muenzner, Bachmann et al. 2010; Johswich, McCaw et al. 2013; Islam, Anipindi et al. 2018). Aside from their role as a handle by which to anchor to the mucosal epithelia, CEACAM engagement allows the bacteria to suppress the exfoliation and delamination of superficial epithelial cells, thereby creating a stable foothold on the mucosa (Muenzner, Rohde et al. 2005; Muenzner, Bachmann et al. 2010; Muenzner, Kengmo Tchoupa et al. 2016). It becomes obvious that pathogens can immensely profit, potentially in multiple ways, from engaging CEACAMs on epithelial cells and this nicely explains the prevalence and independent evolution of CEACAM-binding adhesins amongst human pathogens. However, why is it then that humans rarely succumb to gonococcal infection or develop severe forms of disease after being colonized by *N. meningitidis* or *Helicobacter pylori*, which are present in a large fraction of the healthy population? Indeed, CEACAM-binding pathogens such as *Fusobacterium nucleatum*, *Moraxella catarrhalis*, *Neisseria gonorrhoeae* or *Helicobacter pylori*, despite being able to efficiently colonize the human mucosa, rarely or only in a minority of the cases lead to fatal disease. It is exactly in the context of this scenario that we can now appreciate the role of CEACAM3, a CEACAM family member that does not participate in cell-cell interactions, but is present on the surface of professional phagocytes. In particular, the capacity of CEACAM3 to trigger rapid phagocytosis of attached particles and to activate bactericidal mechanisms of granulocytes will be discussed in the next sections, as these features provide major clues to understand the specialized function of this protein.

CEACAM3-initiated signal transduction leading to phagocytosis

CEACAM3's notable status within the CEACAM family is not only due to its small extracellular domain and its cell-type specific expression pattern, but is also based on a particular sequence motif within its cytoplasmic domain. Similar to the prototypic opsonin-dependent phagocytic receptors of the Fc receptor family, the carboxy-terminus of CEACAM3 encompasses an immunoreceptor tyrosine-based activation motif (ITAM) (see for review (Weiss and Littman

1994)). To be more precise, the motif found in CEACAM3 does not conform perfectly to the consensus ITAM (*D/Ex$_{(7)}$D/ExxYxxI/Lx$_{(6-8)}$YxxI/L*), but resembles an ITAM-like motif, where the carboxy-terminal leucine/isoleucine residue is substituted by methionine (Chen, Bolland et al. 2001; McCaw, Schneider et al. 2003; Buntru, Roth et al. 2012). The presence of this motif and the expression in professional phagocytes already indicates that CEACAM3 might be involved in phagocytosis of CEACAM-binding bacteria. For several CEACAM-binding pathogens, granulocytes play a major role during symptomatic disease. For example, the purulent exudate containing numerous granulocytes with intracellular, gram-negative diplococci is a diagnostic hallmark of gonorrhea (Figure I-2 A and B). It has long been known that gonococci, which express a CEACAM3-binding adhesin, are recognized and phagocytosed by human granulocytes in an opsonin-independent manner (Figure I-2 C), while isogenic strains lacking a CEACAM-binding adhesin are hardly recognized under these conditions (Chen and Gotschlich 1996; Virji, Makepeace et al. 1996; Gray-Owen, Dehio et al. 1997; Gray-Owen, Lorenzen et al. 1997; Hauck, Meyer et al. 1998). Despite the presence of other CEACAM family members such as CEACAM1 and CEACAM6 on the granulocyte surface and despite the fact that CEACAM3 is expressed at low levels, CEACAM3 is the main driving force behind this rapid and efficient opsonin-independent phagocytosis (Schmitter, Agerer et al. 2004). Evidence for the prominent role of CEACAM3 comes from pharmacological, biochemical, genetic, and microbiological approaches: inhibitors that affect CEACAM1 and CEACAM6-mediated uptake in transfected cell lines (such as cholesterol-depleting agents) do not interfere with the opsonin-independent phagocytosis of CEACAM-binding bacteria by granulocytes (Schmitter, Pils et al. 2007), while inhibitors or blocking antibodies selectively affecting CEACAM3-mediated uptake in transfected cell lines also disrupt this process in granulocytes (Gray-Owen, Dehio et al. 1997). Interference with CEACAM3-specific binding partners or signaling processes by transduction of primary human granulocytes with dominant-negative variants also severely compromises the opsonin-independent uptake of CEACAM-binding bacteria. Selective expression of CEACAM3, but not CEACAM1 or CEACAM6, in murine promyelocytic cells can recapitulate major features of neutrophil activation in response to CEACAM-binding bacteria such as oxidative burst and degranulation (Sarantis and Gray-Owen 2012). Furthermore, microbes expressing particular adhesins, which bind CEACAM1, but not CEACAM3, are hardly phagocytosed by primary human granulocytes in the absence of opsonins (Roth, Mattheis et

al. 2013). Therefore, the immediate and dramatic phagocytic response of human granulocytes exposed to CEACAM-binding bacteria can be mainly attributed to CEACAM3.

A number of studies have addressed the molecular basis of CEACAM3's capability to vigorously trigger opsonin-independent phagocytosis. Most of these investigations, conducted with either transfected human cell lines or primary human granulocytes, have pointed towards a major role of the ITAM-like motif for CEACAM3 functionality in phagocytosis. For example, phosphorylation of the tyrosine residues within this motif (Y-230/Y-241) is critical for CEACAM3-initiated phagocytosis, as mutation of either tyrosine to a phenylalanine significantly decreases internalization and mutation of both residues results in an additive effect (Chen, Bolland et al. 2001; Billker, Popp et al. 2002; McCaw, Schneider et al. 2003; Schmitter, Agerer et al. 2004). Interestingly, a single tyrosine to phenylalanine mutation completely blocked phosphorylation of CEACAM3 (McCaw, Schneider et al. 2003). Whether this points to a cooperative phosphorylation mechanism requiring both tyrosine residues or is due to inadequate sensitivity of the assay is unclear. Besides the ITAM-like motif, additional structural elements within the cytoplasmic domain possibly contribute to phagocytic signaling as the CEACAM3 Y230F/Y241F double mutant exhibits residual phagocytic activity compared to variants, which lack the complete cytosolic domain (Billker, Popp et al. 2002; McCaw, Schneider et al. 2003). In contrast to CEACAM1 and CEACAM6, cholesterol-rich membrane domains (lipid rafts) do not seem to contribute to CEACAM3-mediated phagocytosis, as the CEACAM3-dependent internalization of bacteria is insensitive to severe cholesterol depletion, e.g. by methyl-β-cyclodextrin (Schmitter, Pils et al. 2007; Muenzner, Bachmann et al. 2008; Voges, Bachmann et al. 2012). It has been proposed that a Y to F mutation in the ITAM motif generates a recognition motif for AP-2, which could support an endocytic mode of internalization (McCaw, Schneider et al. 2003). However, regular endocytosis via AP-2 initiated, clathrin-coated vesicles has an upper size limit of 200 nm (Rejman, Oberle et al. 2004), implying alternative endocytic processes upon deletion or mutation of the CEACAM3 ITAM-like sequence. Though it is currently unknown, which specific cellular processes guide the residual, ITAM-independent internalization of bacteria, the ITAM-dependent events upon CEACAM3 stimulation have been extensively analyzed.

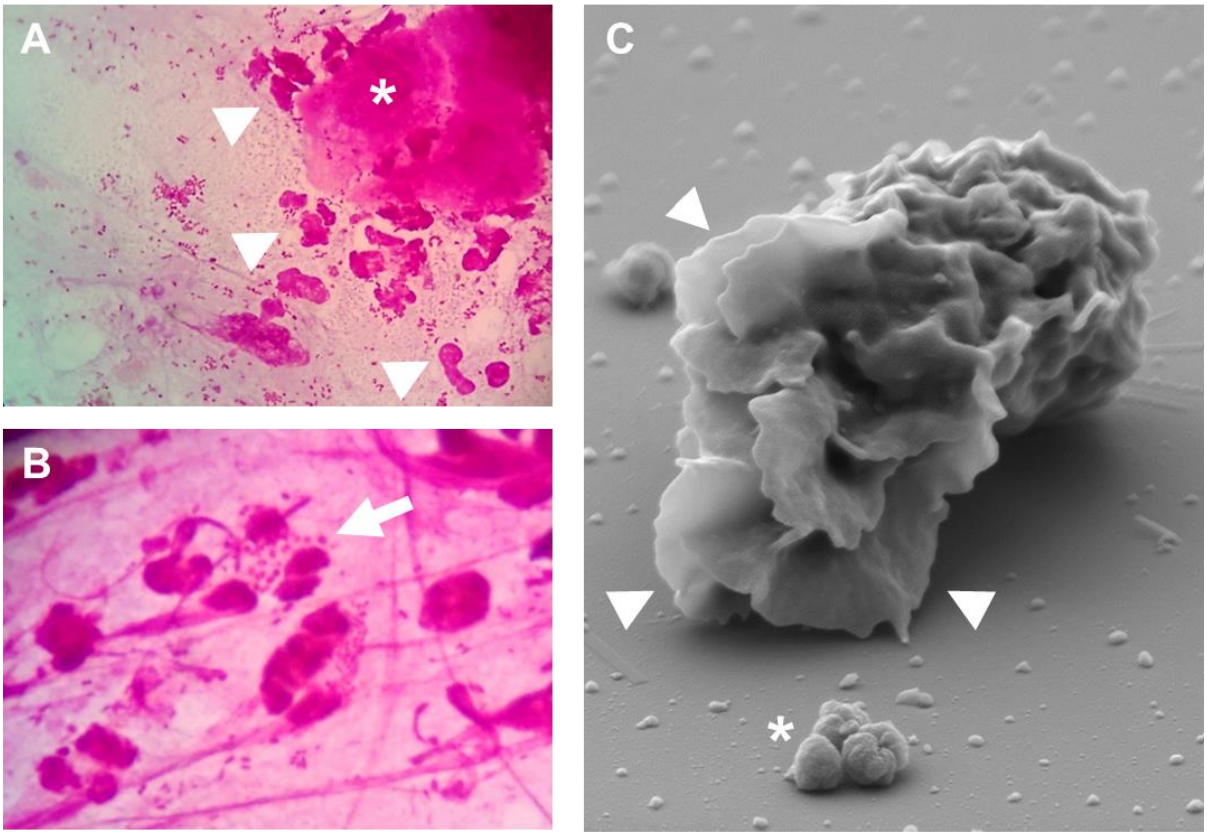


Figure I-2: Granulocytes respond to *Neisseria gonorrhoeae*. (A,B) Gram stain of cervical smears from *N. gonorrhoeae*-infected women. Even at low magnification (A) the presence of granulocytes (arrowheads) and detached epithelial cells (asterisk) can be readily detected. (B) At higher magnification, granulocytes with phagocytosed diplococci (arrow) as well as extracellular bacteria are visible. (C) Scanning electron microscopy highlights the massive lamellipodia (arrowheads) induced on isolated primary human granulocytes upon exposure to CEACAM3-binding *N. gonorrhoeae* (asterisk).

Genetic, biochemical, and pharmacological evidence supports a major role for kinases of the Src family in CEACAM3 phosphorylation (Figure I-3 A). Indeed, the local clustering of CEACAM3 by the multivalent bacteria triggers recruitment and activation of several members of the Src family tyrosine kinases, including Hck and Fgr in granulocytes, while in transfected cell lines Src, Yes and Fyn might take over the respective role (Hauck, Meyer et al. 1998; McCaw, Schneider et al. 2003; Schmitter, Pils et al. 2007; Buntru, Zimmermann et al. 2009). Due to acyl modification, Src family kinases are constitutively associated with the cytoplasmic leaflet of membranes and are therefore in a prime position to initially phosphorylate the ITAM tyrosine residues. Although the tyrosine kinase Syk is also recruited to nascent gonococci-containing

phagosomes in an ITAM-dependent fashion, pharmacological inhibition of Syk did not reduce bacterial internalization (Sarantis and Gray-Owen 2007). Only when polystyrol beads larger than 5 µm (a typical gonococcal diplococcus is about 1-2 µm in size) were coated with anti-CEACAM IgG and used as bacterial surrogate, Syk augmented internalization. The fact that Syk facilitates phagocytosis depending on particle size is not unique to CEACAM3 as the same phenomenon has been observed for FcγR-mediated phagocytosis (Crowley, Costello et al. 1997). Therefore, Syk is dispensable for internalization of gonococci, but it clearly does promote downstream bactericidal activity by enhancing the oxidative burst, degranulation and the NF-κB-mediated inflammatory response (Sarantis and Gray-Owen 2007).

Upon phosphorylation, the ITAM-like motif in CEACAM3 creates a platform for effectors that drive cytoskeletal remodeling required for phagocytic cup formation (Figure I-3 B). The adaptor molecule Nck binds CEACAM3 in a phosphorylation-dependent manner via its SH2 domain and recruits the WAVE2 complex to sites of bacterial attachment (Pils, Kopp et al. 2012). WAVE2 is part of a multiprotein complex that activates Arp2/3-mediated actin nucleation, which drives lamellipodia extension (reviewed in (Rougerie, Miskolci et al. 2013)). Indeed, ablation of Nck or inhibition of WAVE2 impedes lamellipodia formation and bacterial uptake (Pils, Kopp et al. 2012). The WAVE2 complex is a coincidence detector, which is activated by the local and temporal co-occurrence of protein tyrosine phosphorylation, acidic phospholipids and activation of the small GTPase Rac (Lebensohn and Kirschner 2009). Importantly, all of these events are concomitantly initiated by CEACAM3 engagement. Acidic phospholipids such as phosphatidylinositol-(3,4,5)-trisphosphate (PIP_3) and phosphatidylinositol-(3)-phosphate (PI3P) sequentially accumulate at the phagocytic cup during engulfment (Booth, Telio et al. 2003; McCaw, Liao et al. 2004; Voges, Bachmann et al. 2012). WAVE2 activation strictly depends on the recruitment of the small Rho GTPase, Rac, which is crucial for remodeling of the cytoskeleton (Sit and Manser 2011). Indeed, Rac1 is also essential for CEACAM3-mediated phagocytosis and is locally and transiently activated at the phagocytic cup (Hauck, Meyer et al. 1998; Billker, Popp et al. 2002; Schmitter, Agerer et al. 2004). Activation of Rac is catalyzed by specific guanine nucleotide exchange factors (GEFs) that exchange Rac1-bound GDP for GTP and CEACAM3 aggregation triggers Rac activation by the GEF Vav (Schmitter, Pils et al. 2007). Interestingly, Vav itself is activated by Src family kinases (English, Orlicek et al. 1997; Fumagalli, Zhang et al. 2007) and the Vav SH2 domain can interact directly with the phosphorylated Y-230 within the ITAM-like motif of CEACAM3

(Schmitter, Pils et al. 2007). As most other GEFs bind via their pleckstrin homology-domain (PH-domain) to phosphoinositides, their recruitment is intrinsically dependent on PI3K activity. In contrast, Vav's direct interaction with phosphorylated CEACAM3 could render this process PI3K-independent. Indeed, phosphoinositide 3-kinase (PI3K) activity is dispensable for CEACAM3-mediated internalization of gonococci (Booth, Telio et al. 2003; Buntru, Kopp et al. 2011). The direct interaction between phosphorylated CEACAM3 and a Rac GEF as well as the direct linkage to the WAVE complex via Nck could explain why CEACAM3-mediated phagocytosis is particularly efficient and rapid (McCaw, Liao et al. 2004). *In vitro*, ~90% of primary human granulocytes have internalized multiple CEACAM-binding gonococci within 20 minutes of infection, while CEACAM-non-binding gonococci remain almost untouched (McCaw, Schneider et al. 2003; McCaw, Liao et al. 2004). These studies of CEACAM3-initiated signal transduction have also revealed that the ability of this receptor to trigger phagocytosis relies on cellular constituents, such as Src family kinases, Vav, Rac, and Nck-WAVE, which can be found in almost any mammalian cell type. Maybe this universality of CEACAM3 downstream connections explains why transfection of CEACAM3 cDNA in diverse cell types, from cervical epithelial cells to mouse fibroblasts, is sufficient to convert such non-professional phagocytes into cells avidly and efficiently internalizing CEACAM-binding bacteria. The generic nature of these processes also indicates that studies of CEACAM3-initiated phagocytosis are helpful to delineate the core elements necessary for productive internalization of particles.

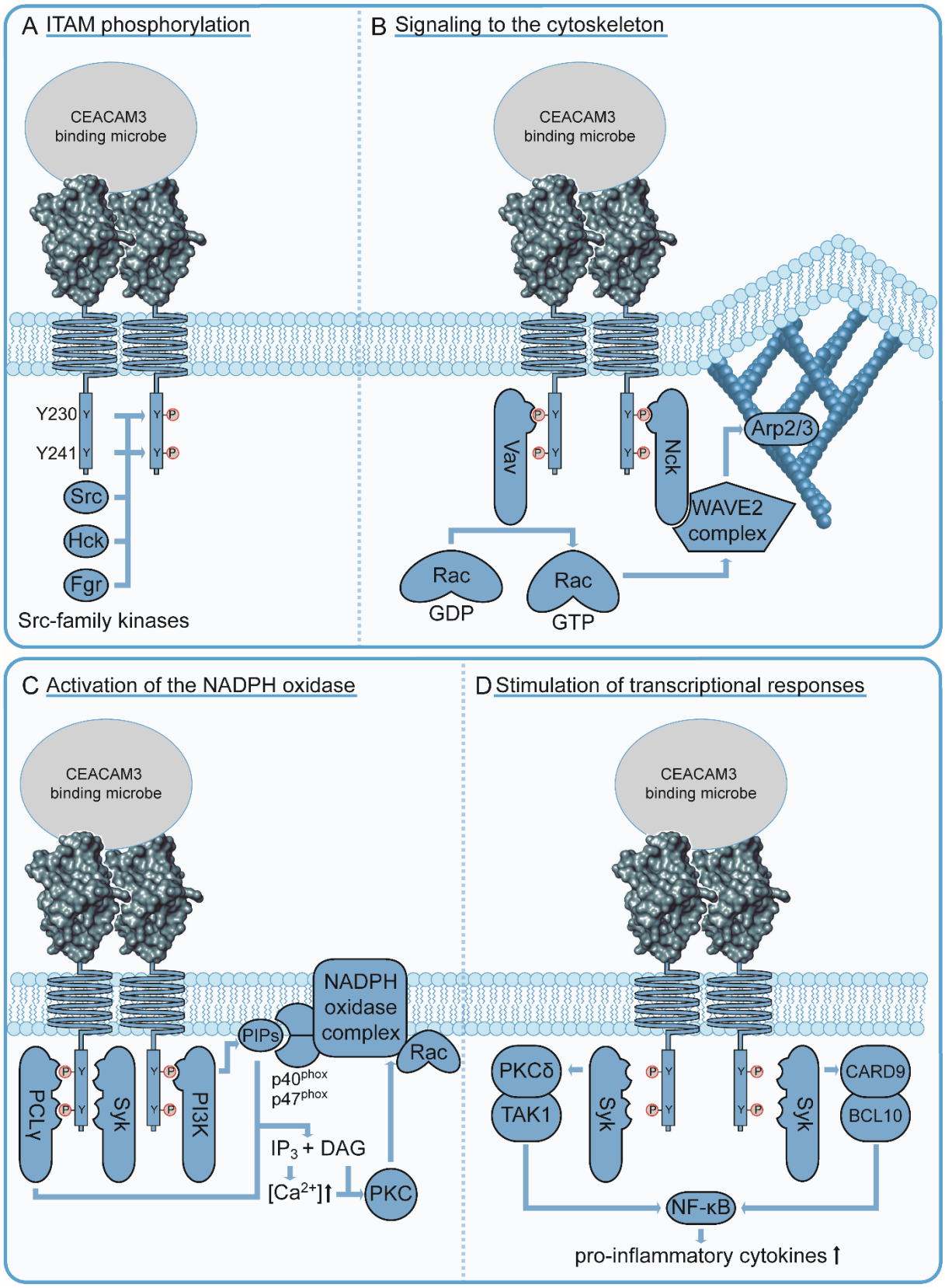


Figure I-3: CEACAM3 signaling connections. (A) Phosphorylation of CEACAM3 upon bacterial engagement. Localized clustering of CEACAM3 receptors by multivalent, CEACAM-binding bacteria trigger recruitment and activation of several Src family tyrosine kinases, which in turn

phosphorylate the ITAM-like sequence in the cytoplasmic domain of the receptor at two tyrosine residues (Y230 and Y241). This phosphorylation event is critical for downstream signaling processes. (B) CEACAM3 signaling to the cytoskeleton. The adaptor protein Nck binds via its SH2 domain to the membrane proximal CEACAM3 phosphotyrosine residue pY230 and recruits the WAVE2 complex to sites of bacterial attachment. Activation of this complex requires a localized increase in GTP-loaded Rac, which is orchestrated by recruitment of the Rac guanine nucleotide exchange factor Vav. Similar to Nck, Vav binds via its SH2 domain to CEACAM3 pY230. Activation of the WAVE2 complex leads to localized activation of Arp2/3 at the site of infection resulting in particle engulfment by actin cytoskeleton-based lamellipodial protrusions. (C) CEACAM3 signaling to the NADPH oxidase. Residue pY230 serves as an interaction platform for the regulatory subunit of phosphatidylinositol 3' kinase (PI3K). PI3Ks are responsible for the generation of phosphoinositides (PIPs), which regulate the oxidative burst by recruitment of NADPH oxidase subunits $p40^{phox}$ and $p47^{phox}$. In addition, PIP hydrolysis by CEACAM3-recruited phospholipase C γ (PLCγ) releases the second messengers inositol-(1,4,5)-trisphosphate (IP_3) and diacyl glycerol (DAG). Together, they activate the Protein kinase C (PKC), which further stimulates the NADPH oxidase complex. (D) CEACAM3 stimulation of neutrophil transcriptional response. CEACAM3 can trigger the NF-κB-mediated transcriptional regulation of IL-8 and other neutrophil chemotactic factors. CEACAM3-initiated NF-κB activation occurs via two distinct routes: the PKCδ/TAK1 pathway as well as the CARD9/BCL10 pathway depends on Syk localization to phosphorylated CEACAM3.

CEACAM3-initiated signal transduction beyond phagocytosis

In professional phagocytes, phagocytosis is tightly coupled to downstream bactericidal processes. One central regulatory switch in this regard is the GTPase Rac, which not only regulates cytoskeletal remodeling, but is also an essential subunit of the NADPH oxidase, the enzyme complex generating the microbiocidal oxidative burst (Figure I-3 C). In contrast to the PI3K-independent Rac activation during engulfment described above, the CEACAM3-induced oxidative burst strictly depends on PI3K activity (Buntru, Kopp et al. 2011). Since there is only a partial temporal overlap between particle engulfment and the oxidative burst, it is possible that Vav-mediated GTP-loading of Rac during engulfment is disconnected from a putative phosphoinositide-dependent GEF, which might activate Rac later during the induction of an

oxidative burst. However, PI3K-generated phosphoinositides are required at additional regulatory steps during NADPH oxidase assembly, such as the recruitment of NADPH oxidase subunits $p40^{phox}$ and $p47^{phox}$ through their PI3P-binding PX-domains (Zhan, Virbasius et al. 2002). Phosphoinositides also serve as substrate for various lipid phosphatases and phospholipases. Interestingly, the SH2 domain of phospholipase C gamma (PLCγ) can bind CEACAM3 *in vitro* and the isolated SH2 domain is enriched around gonococci-containing phagosomes (Streichert, Ebrahimnejad et al. 2001; McCaw, Schneider et al. 2003). PLCγ-mediated hydrolysis of phosphatidylinositol-(4,5)-bisphosphate (PIP_2) produces the second messengers diacyl glycerol (DAG) and inositol-(1,4,5)-trisphosphate (IP_3), which in turn trigger intracellular Ca^{2+} release and PKC activation. Indeed, intracellular Ca^{2+} levels in PMNs rapidly rise upon CEACAM3 engagement and this process does not occur in cells lacking PLCγ (Chen, Bolland et al. 2001). Accordingly, upstream PLCγ activity might be required to allow PKC-mediated phosphorylation of multiple NADPH oxidase subunits, which represents an important regulatory step in activation of the oxidative burst (reviewed in (Belambri, Rolas et al. 2018)).

Though both tyrosine residues within the ITAM-like sequence of CEACAM3 seem to be functionally relevant, the biochemical assays conducted so far have pointed towards the Y230-residue as the central hub for interactions with SH2 domain-containing proteins. This conclusion is based on binding studies with synthetic phospho-peptides and pulldown experiments with recombinant proteins, which demonstrate that the SH2-domains of Src family kinases, PI3K, Nck, or Vav selectively bind of phospho-Y230. Therefore, the CEACAM3 ITAM-like sequence has been likened to a so-called HemITAM sequence found for example in the macrophage receptor Dectin-1, where also a single tyrosine residue conveys the phagocytic function (Rogers, Slack et al. 2005; Underhill, Rossnagle et al. 2005). Interestingly, the only known negative regulator of CEACAM3-mediated signaling, the adaptor protein Grb14, also targets phospho-Y230 (Kopp, Buntru et al. 2012). One can easily envision that Grb14 restricts access for other SH2-domain containing effector proteins and thereby interferes with CEACAM3-mediated phagocytosis. However, the idea that multiple proteins compete for phospho-Y230 of CEACAM3 immediately begs the question how these binding events can be coordinated to allow a productive and orchestrated cellular response. In the future, time-resolved analysis of the various SH2-domain-mediated binding events upon bacterial CEACAM3 engagement might help to answer this question. Nevertheless, the

emerging overall picture of CEACAM3 phosphorylation-initiated events depicts Y-230 within the ITAM-like motif as the minimal structural feature, which directly links CEACAM3 engagement with cytoskeletal remodeling as well as with the initiation of bactericidal responses.

It is interesting to mention that there is a further member of the CEACAM family, CEACAM4, which has a domain architecture similar to CEACAM3, which is selectively expressed in granulocytes, and which harbors a consensus ITAM (*D/Ex$_{(7)}$D/ExxYxxLx$_{(6-8)}$YxxI*) (Figure I-1). CEACAM4 is an orphan receptor, as neither an endogenous nor a microbial ligand for this membrane protein has been detected. However, the cytoplasmic domain of CEACAM4 can trigger particle uptake, with both ITAM-embedded tyrosine residues engaging in SH2-domain interactions (Delgado Tascon, Adrian et al. 2015). Therefore, the human CEACAM family appears to harbor additional phagocytic receptors that might function to eliminate microorganisms in an opsonin-independent manner.

Although the main task of PMNs is to clear pathogens through their bactericidal capabilities, there is mounting evidence that they also can shape the inflammatory response (reviewed in (Nathan 2006; Thomas and Schroder 2013)). Indeed, CEACAM3 activation can trigger an inflammatory program in PMNs (Figure I-3 D). CEACAM3 binding by *Moxarella catarrhalis* activates the CARD9/BCL10 pathway resulting in NF-κB-mediated expression of the potent neutrophil chemotactic factor, interleukin-8 (IL-8) (Heinrich, Heyl et al. 2016). As CEACAM3 expression is restricted to human neutrophils, its contribution to inflammatory responses in an organismal context are difficult to study. Sinstova *et al.* made use of a transgenic mouse model that harbors a 187-kb human bacterial artificial chromosome encoding CEACAM3, CEACAM5, CEACAM6 and CEACAM7 (CEABAC mice) (Chan and Stanners 2004) to study the CEACAM-mediated response to infection with *N. gonorrhoeae* (Sintsova, Sarantis et al. 2014). Global expression profiles were generated from both WT and transgenic neutrophils infected with Opa^{+} gonococci to discern CEACAM-dependent signatures. A pronounced upregulation of pro-inflammatory cytokines was observed in neutrophils from CEABAC mice, which depends on p38 MAPK activity and the PKCdelta/TAK1/NF-κB axis (Sintsova, Sarantis et al. 2014). Though the used transgenic neutrophils and primary human granulocytes express other CEACAMs, which could be engaged by CEACAM-binding bacteria (such as CEACAM6 in the case of transgenic murine neutrophils or CEACAM1 and CEACAM6 in the case of primary human

neutrophils), inhibitor studies point to CEACAM3-ITAM signaling as the main contributor to neutrophil-initiated inflammatory responses. *In vivo*, infection of CEACAM-transgenic, but not wild-type mice with *N. gonorrhoeae* led to increased neutrophil infiltration and increased levels of neutrophil-derived IL-1β and MIP-1α. Accordingly, CEACAM3-signaling on the one hand helps to limit gonococcal survival, but also initiates a vicious cycle, where the bacteria-triggered release of chemotactic cytokines leads to increased neutrophil influx and potentiates the risk of severe damage to the infected tissue (Stevens and Criss 2018). In this light, it will be important to understand how CEACAM3 signaling is kept in check to prevent unrestrained inflammatory signaling. In contrast to the initiation of CEACAM3 signaling, surprisingly little is known about its termination. The negative regulatory role of the adaptor protein Grb14 (Kopp, Buntru et al. 2012) has already been discussed above. Furthermore, phosphorylation of CEACAM3 appears to be counteracted by the cytoplasmic protein tyrosine phosphatase SHP-1, which most likely constrains CEACAM3 effector functions by compromising ITAM functionality (Hauck, Gulbins et al. 1999). Interestingly, neutrophils also express CEACAM1, which contains an immunoreceptor tyrosine-based inhibitory motif (ITIM). However, co-recruitment of CEACAM1 does not seem to have an inhibitory effect on CEACAM3 phagocytic activity (Sarantis and Gray-Owen 2012). Further research is required to address mechanisms that could regulate CEACAM3 activity, including intracellular trafficking of CEACAM3 and cooperation with other phagocytic receptors (e.g. Fc-γ and complement receptors). Considering the potential detrimental effects of excessive CEACAM3 activation, it is highly likely that in human PMNs additional negative modulators of CEACAM3 signaling operate.

CEACAM3 evolution – a red queen scenario at work

Based on the functional studies summarized in the previous sections, it is safe to conclude that CEACAM3 represents an effective detector and eliminator of CEACAM-binding bacteria. In its capacity as a phagocytosis-promoting receptor, CEACAM3 can take care of pathogens expressing CEACAM-binding adhesins designed to exploit the human receptors for mucosal colonization. Thereby, CEACAM3 might help to establish a truce between host and pathogen, maybe one of the reasons why CEACAM-binding pathogens are contained in most instances. Looking at this reality from the viewpoint of a microbe: wouldn’t it then be smart to avoid expressing a CEACAM3-binding adhesin in the first place? Obviously, bacteria unanimously

answered this question with “Yes”, as they appear to optimize their adhesins to allow binding to epithelial CEACAMs such as CEACAM1 and CEA, while evading CEACAM3 recognition. Indeed, most CEACAM-binding bacterial adhesins analyzed in this regard show selective binding to either CEACAM1 or CEA. This is true for the CEACAM-binding Opa protein adhesins of *Neisseria gonorrhoeae* and *N. meningitidis* as well as for their functional homologue, the OMP P1 adhesin of *H. influenzae*, which have been studied most thoroughly in this context. For example, a single gonococcal strain encodes eleven distinct Opa adhesins and the complete compendium of Opa proteins has been functionally tested for binding to CEACAM family members in the case of *N. gonorrhoeae* strains MS11 and VP1. In MS11, 10 out of 11 Opa adhesins bind CEA, while only three of these are also able to be recognized by CEACAM3 (Bos, Grunert et al. 1997; Gray-Owen, Dehio et al. 1997; Gray-Owen, Lorenzen et al. 1997). For strain VP1, 8 out of 10 tested Opa adhesins bound to CEA or CEACAM1, while only a single Opa protein associated with CEACAM3 (Roth, Mattheis et al. 2013). In a complementary approach, Sintosova et al. screened a large collection of primary isolates from gonorrhea patients for their CEACAM-binding capacity. Instead of trying to analyze the complete Opa repertoire of each of these strains, the authors used low passage isolates, which are thought to continue to express *in vitro* the Opa variant previously selected for and expressed *in vivo* (Sintsova, Wong et al. 2015). Also in this study, the overwhelming majority with 74% or 80% of the strains bound CEA or CEACAM1, respectively, and only 27% were found to also recognize CEACAM3 (Sintsova, Wong et al. 2015). While most strains bound CEA and/or CEACAM1, but not CEACAM3, not a single strain could be observed, which showed a reverse binding pattern: recognizing CEACAM3, but not recognizing an epithelial CEA family member (Sintsova, Wong et al. 2015). An even more skewed situation is found in *N. meningitidis*, where 11 of 13 Opa proteins of serogroup A, B, and C strains, which have been tested in binding assays with different CEACAMs associated with either CEA or CEACAM1, but none of those Opa proteins associated with CEACAM3 (Muenzner, Dehio et al. 2000; de Jonge, Hamstra et al. 2003). Similarly, all of the OMP P1 adhesins derived from 13 different strains of *H. influenzae* bound to CEACAM1, but none of those adhesins was recognized by CEACAM3 (Tchoupa, Lichtenegger et al. 2015). Accordingly, bacteria seem to optimize their adhesins to discriminate between these closely related CEACAMs and to avoid recognition by CEACAM3.

If there is such a directed evolution on the side of bacterial CEACAM-binding adhesins, is there a discernable adaptation on the host side? The increasing availability of human and other

primate genomic information has now allowed an unravelling of the evolutionary context of CEACAM3. It has been recognized before that homologues of several human CEACAMs are absent from rodents. More specifically, a CEACAM3 ortholog could not be identified even in Old World primates, such as baboon or African green monkey, by homology searches based on the extracellular Ig_V-like domain (Zhou, Zhang et al. 2001). Large scale genome comparisons between closely related primate species have revealed a high degree of non-synonymous vs. synonymous nucleotide changes in ortholog genes of CEACAM family members (Rhesus Macaque Genome Sequencing and Analysis Consortium, Gibbs et al. 2007; Adrian, Bonsignore et al. 2019). In particular, the bacterial receptors within the CEACAM family (CEACAM1, CEACAM3, CEA, CEACAM6) show an exceptionally strong signature of positive selection suggesting that they belong to the fastest evolving human genes. Especially the *CEACAM3* gene appears to be a recent evolutionary invention. Though ITAM-sequence containing CEACAM-related genes have been described for various mammals (Kammerer and Zimmermann 2010; Kammerer, Mansfeld et al. 2017; Missbach, Aleksic et al. 2018), a *CEACAM3* gene with its specific exon/intron structure and the characteristic ITAM-like sequence seems to occur only after the emergence of Old World monkeys (around 35 million years ago) (Bond, Tejedor et al. 2015). Indeed, a *CEACAM3* ortholog has only been detected at the syntenic locus in the genomes of baboon, macaque, orangutan, gorilla, chimpanzee, and humans (Adrian, Bonsignore et al. 2019). Due to uncertainties in the assembly of some genomes such as tarsier, lemurs and lories, gene synteny is not a valid criterion to rule out the existence of CEACAM3 orthologs in lower primates. However, CEACAM3 orthologs share several discriminative features such as protein length or the presence of the characteristic HemITAM, which does not seem to occur in any lower primate genome supporting the idea that CEACAM3 emerged late in primate evolution. Despite the absence of a *CEACAM3* gene, ITAM-containing CEACAM transcripts are present in lower primates, such as transcript XR_001153184.2 in the mouse lemur. The ITAM encoded by XR_001153184.2 perfectly matches the ITAM consensus (YxxL - 7AS – YxxI) as well as the sequence found in human CEACAM4 (Delgado Tascón et al., 2015). Therefore, this transcript could originate from an ancestral primate CEACAM4 ortholog. It is plausible that CEACAM3 is the result of gene duplication and recombination between an ancestral *CEACAM1* gene (providing the Ig_V-domain encoding exon) and an ancestral lower primate *CEACAM4* gene (providing the transmembrane and the intracellular domain encoding exons) (Pils, Gerrard et al. 2008).

Indeed, the 3' end of the large intron 2 of the *CEACAM3* gene bears similarities to intron 2 of the *CEACAM4* gene, supporting the idea that such a recombination event occurred at the advent of higher primates (Pils, Gerrard et al. 2008). Together, the genomic evidence suggests that CEACAM3, the specialized phagocytic receptor for CEACAM-binding bacteria, emerges relatively late during primate evolution, expanding the human innate immunity arsenal.

An interesting corollary of this continuing bacteria-host arms race can be seen by the occurrence of CEACAM3 polymorphisms in the human population. In fact, while people outside of Africa mainly express the common CEACAM3 allele, particular ethnic groups harbor CEACAM3 variants. For example, in several groups with African ancestry, a third of the population expresses a distinct CEACAM3 allele (minor CEACAM3 allele) (Adrian, Bonsignore et al. 2019). This minor allele carries four non-synonymous single nucleotide changes, which convert the CEACAM3-Ig_V-domain amino acid sequence into a near replica of the CEACAM1 Ig_V-domain. Modeling of the polymorphic CEACAM3 Ig_V-domain according to the known CEACAM3 crystal structure (Bonsor, Zhao et al. 2018) reveals that these residues are surface exposed and three of the modified amino acids contribute to the CEACAM3 binding interface for various bacterial adhesins (Adrian, Bonsignore et al. 2019). It comes as no surprise that this minor CEACAM3 allele, that now mimics CEACAM1, shows a striking gain of function in that it recognizes additional bacterial adhesins, such as the OMP P1 protein of *H. influenzae*, which escape detection by the common CEACAM3 allele (Adrian, Bonsignore et al. 2019).

These recent studies have shed light on an ongoing arms race between several host-restricted CEACAM-binding pathogens and the human immune system. In a type of Red-Queen scenario, the diverse CEACAM-binding pathogens need to optimize their adhesins to escape CEACAM3 detection, while still retaining their ability to bind mucosal CEACAMs. As a consequence, human CEACAM3 rapidly evolves to keep up its function. Viewed from the angle of bacterial adhesin binding to distinct CEACAMs as well as from the perspective of human genomic variation, the phagocytic receptor CEACAM3 appears as a central element in the innate defense against this group of extremely specialized bacteria. However, one has to acknowledge that CEACAM binding preferences of bacterial adhesins as well as analyzes of CEACAM3 evolution only provide circumstantial evidence for the importance of this innate immune receptor. Accordingly, a more direct test of CEACAM3 function in the context of innate immune responses within the intact organism would be desirable. Unfortunately, the

standard genetic approach of employing knock-out animals, such as a “CEACAM3-knock-out” mouse, is not an option in this case. Moreover, CEACAM3-transgenic mice with granulocyte-selective expression of CEACAM3, in the absence of other CEACAM family members, do not exist currently.

In view of the increasing availability of human genomic information and corresponding medical records it does not appear unrealistic to foresee a future, where novel insight into CEACAM3 function might come directly from combined genomic-phenotypic studies in humans (FitzGerald, Botstein et al. 2018). Indeed, the occurrence of an unexpected high proportion of loss-of-function alleles in particular human populations (Saleheen, Natarajan et al. 2017) could help to reveal plausible connections between CEACAM3-deficiency and susceptibility for particular infectious diseases. Despite the inherent limitations of research on a primate-only protein, the study of the phagocytic receptor CEACAM3 will surely continue to unearth fascinating aspects of human biology and of our constant and challenging interplay with the microbial world.

Acknowledgement

The authors are indebted to Michael Laumann (Electron Microscopy Service; University of Konstanz, Germany) for scanning electron microscopy and Nathlee Abbai (University KwaZulu-Natal, Durban, South Africa) for providing histological samples. Work by the authors has been supported by the DFG (Ha 2856/10-1 and Ha2856/11-1) and Baden-Württemberg Stiftung (BWST_WSF-020) to CRH.

Chapter II

Adaptation to host-specific bacterial pathogens drives rapid evolution of a human innate immune receptor

Jonas Adrian[1*], **Patrizia Bonsignore[1*]**, Sebastian Hammer[1], Tancred Frickey[2,3], Christof R. Hauck[1,3,#]

[1]Lehrstuhl für Zellbiologie, Universität Konstanz, 78457 Konstanz, Germany

[2]Forest Industry Informatics, Scion, 3015 Rotorua, New Zealand

[3]Konstanz Research School - Chemical Biology, Universität Konstanz, 78457 Konstanz, Germany

[*]These authors contributed equally

Abstract

The selective pressure by infectious agents is a major driving force in the evolution of humans and other mammals. Members of the carcinoembryonic antigen-related cell adhesion molecule (CEACAM) family serve as receptors for bacterial pathogens of the genera *Haemophilus*, *Helicobacter*, *Neisseria*, and *Moraxella*, which engage CEACAMs via distinct surface adhesins. While microbial attachment to epithelial CEACAMs facilitates host colonization, recognition by CEACAM3, a phagocytic receptor expressed by granulocytes, eliminates CEACAM-binding bacteria. Sequence analysis of primate CEACAM3 orthologs reveals that this innate immune receptor is one of the most rapidly evolving human proteins. In particular, the pathogen-binding extracellular domain of CEACAM3 shows a high degree of non-synonymous vs. synonymous nucleotide exchanges, indicating an exceptionally strong positive selection. Using CEACAM3 domains derived from different primates we find that the amino acid alterations found in CEACAM3 translate into characteristic binding patterns for bacterial adhesins. One such amino acid residue is F62 in human and chimp CEACAM3, which is not present in other primates and which is critical for binding the OMP P1 adhesin of *Haemophilus aegyptius*. Incorporation of the F62-containing motif into gorilla CEACAM3 results in a gain-of-function phenotype with regard to phagocytosis of *H. aegyptius*. Moreover, CEACAM3 polymorphisms found in human subpopulations widen the spectrum of recognized bacterial adhesins, suggesting an ongoing multivariate selection acting on this innate immune receptor. The species-specific detection of diverse bacterial adhesins helps to explain the exceptionally fast evolution of CEACAM3 within the primate lineage and provides an example of Red Queen dynamics in the human genome.

Introduction

Infectious agents are a major driving force during evolution including the evolution of man and other mammals (Fumagalli, Sironi et al. 2011). The increased allele frequency of sickle cell β-hemoglobin found in human populations living in malaria-endemic regions is a prime textbook example of a potentially deleterious gene polymorphism maintained due to selective pressure by a specialized parasite (Allison 1954). Genome sequencing has revealed that several immunity-related human genes are under positive selection such as the viral RNA-editing enzyme APOBEC3H (Harari, Ooms et al. 2009; Nakano, Aso et al. 2017). The concept of a constant molecular ‘arms-race’ between pathogens and their hosts is in-line with the so-

called Red Queen hypothesis (van Valen 1973), but examples for variability in bacterial pathogens and co-occurrence of positive selection in corresponding human defense genes are lacking.

We and others have identified epithelial surface proteins of the human carcinoembryonic antigen-related cell adhesion molecule (CEACAM) family as targets for diverse human-restricted bacteria (Tchoupa, Schuhmacher et al. 2014). Members of this immunoglobulin receptor family are expressed on the apical surface of mucosal cells lining the nasopharynx, the intestine and the genitourinary tract (Thompson, Grunert et al. 1991; Beauchemin, Draber et al. 1999). Several gram-negative pathogens, including *Haemophilus influenzae, H. aegyptius, Helicobacter pylori, Neisseria gonorrhoeae, N. meningitidis*, and *Moraxella catarrhalis*, express specialized adhesins to associate with the amino-terminal immunoglobulin variable (Ig_V)-like domain shared by all CEACAM family members (Chen and Gotschlich 1996; Virji, Makepeace et al. 1996; Hill and Virji 2003; Tchoupa, Lichtenegger et al. 2015; Koniger, Holsten et al. 2016). Attachment to apically expressed CEACAMs appears to facilitate bacterial colonization of the mucosal surface, not only by providing a direct molecular connection to the epithelium, but also by suppressing the exfoliation and detachment of infected superficial cells (Muenzner, Rohde et al. 2005; Muenzner, Bachmann et al. 2010; Muenzner, Kengmo Tchoupa et al. 2016). In contrast to this obvious advantage gained by microbes on the level of the epithelium, CEACAM-mediated recognition by innate immune cells appears to have detrimental consequences for the microorganism. In particular, CEACAM3, a family member expressed on polymorphonuclear granulocytes (PMNs; neutrophils), is responsible for the rapid opsonin-independent phagocytosis and clearance of CEACAM-binding microbes (Schmitter, Agerer et al. 2004; Sarantis and Gray-Owen 2007). While the extracellular Ig_V-like domain of CEACAM3 shares high sequence similarity with CEACAM1, CEA, and CEACAM6, the intracellular domain of CEACAM3 is unique and responsible for the specific function of this protein. Indeed, tyrosine residues within the carboxy-terminal HemITAM-like sequence are phosphorylated upon receptor engagement and recruit several cytoplasmic proteins (Sarantis and Gray-Owen 2007; Schmitter, Pils et al. 2007; Schmitter, Pils et al. 2007; Buntru, Kopp et al. 2011; Pils, Kopp et al. 2012). These trigger an intracellular signaling cascade stimulating bactericidal and immunostimulatory responses of granulocytes, including rapid phagocytosis, activation of the oxidative burst and NF-κB-dependent gene expression (Buntru, Roth et al. 2012; Heinrich, Heyl et al. 2016).

Obviously, detection by granulocyte-expressed CEACAM3 is to the disadvantage of the involved bacteria. It is therefore not surprising that pathogens seem to avoid recognition by CEACAM3, while retaining binding to epithelial CEACAMs. This phenomenon is most obvious in the case of *Neisseria gonorrhoeae*. In this species, a single strain contains up to 11 distinct gene loci encoding CEACAM-binding adhesins, the so-called colony opacity-associated (Opa) proteins (Stern, Brown et al. 1986). For two strains, the CEACAM-binding profile of the complete Opa protein repertoire has been analyzed (Gray-Owen, Lorenzen et al. 1997; Roth, Mattheis et al. 2013). In each case, 10 out of 11 Opa proteins bound to epithelial CEACAMs, while only few adhesins (three or one out of 11) were also recognized by CEACAM3 (Roth, Mattheis et al. 2013). This finding indicates that pathogens evolve adhesins with a preference for binding to epithelial CEACAMs and avoiding the granulocyte phagocytic receptor CEACAM3.

A situation similar to *N. gonorrhoeae* Opa proteins has been observed for *H. influenzae* (Hinf) OMP P1, which functions as the CEACAM-binding adhesin of this microbe. OMP P1 proteins from a variety of Hinf isolates recognize CEACAM1, but at the same time do not bind to CEACAM3 (Tchoupa, Lichtenegger et al. 2015). The only CEACAM-binding OMP P1 variant that is recognized by CEACAM3 has been identified in the related species *H. aegyptius* (Tchoupa, Lichtenegger et al. 2015). Akin to neisserial Opa proteins, OMP P1 proteins isolated from different *Haemophilus* strains show a high degree of sequence variation in their extracellular parts (Tchoupa, Lichtenegger et al. 2015). Again, escape from detection by CEACAM3 could explain this genetic diversity. We reasoned, therefore, that human CEACAM3 could be involved in an evolutionary arms race with human-restricted bacteria and that CEACAM3 phylogeny would provide insight into the co-evolution of this innate immune receptor and its bacterial ligands.

By identifying and comparing CEACAM3 orthologs from different primate species, we present evidence that CEACAM3 is under extraordinary strong positive selection in humans. Amino acid changes in the extracellular domain of CEACAM3 from rhesus monkey to humans translate into the ability of this receptor to detect and phagocytose specific bacteria. In particular, we identify a single amino-acid residue that, when changed from gorilla to the human sequence, supports binding to *Haemophilus aegyptius* OMP P1. Furthermore, the co-occurance of single nucleotide polymorphisms in human CEACAM3, as observed in ~40% of

the African population, determines the binding to distinct neisserial Opa proteins. Together, our results establish CEACAM3 as one of the most rapidly evolving human genes, where frequent amino acid changes allow adaptation to distinct human-restricted pathogens.

Results

The granulocyte receptor CEACAM3 is under exceptionally strong positive selection

DNA and amino acid sequence analysis of 13,000 orthologous proteins in man, chimpanzee and rhesus monkey revealed a set of ~150 genes with elevated ratios of non-synonymous substitution per non-synonymous site and synonymous mutations per synonymous site (dN/dS ratios) indicating positive selection (Gibbs, Rogers et al. 2007). Amongst these fast evolving primate genes were several members of the CEACAM family, in particular CEACAM1, CEACAM5 and CEACAM6, which serve as epithelial docking sites for bacterial pathogens (Tchoupa, Schuhmacher et al. 2014). Compared to the average dN/dS ratio of human vs. chimpanzee genes (0.3) (Wolf, Kunstner et al. 2009), the binding mediating Ig_V-like domains of CEACAM1 and CEACAM5 have dN/dS ratios >1 when comparing the human sequences with the orthologous sequences of chimpanzee, pongo, or macaque indicating positive selection (Figure II-1 A). In comparison, the terminal Ig-fold of VCAM, a member of the Ig-superfamily responsible for binding to $\alpha_4\beta_1$-integrin, is under strong purifying selection with a dN/dS-ratio far below 1 (Figure II-1 A).

In contrast to the epithelium-associated CEACAM family members CEACAM1 and CEA, primate orthologs of CEACAM3 had not been included in prior comparative evolutionary studies (Gibbs, Rogers et al. 2007). As CEACAM3 is the innate immune receptor, which detects and eliminates CEACAM-binding bacterial pathogens, it should be, according to the Red Queen hypothesis, under positive selection.

Therefore, we first searched primate genomes for the presence of CEACAM3 orthologs. Using homology searches, gene synteny, exon/intron structure and the presence of a HemITAM, we were able to identify CEACAM3 orthologs in the genomes of Old World monkeys including *Macaca mulatta*, *Papio anubis*, *Pongo abelii*, *Gorilla gorilla*, and *Pan troglodytes* (Figure II-1 B). With the exception of *Papio anubis*, CEACAM3 orthologous sequences from these species had been reported before (Mikkelsen 2005; Locke, Hillier et al. 2011; Scally, Dutheil et al. 2012; Zimin, Cornish et al. 2014). In all these genomes, orthologs of human CEACAM3 were located

at the syntenic position with the identical transcriptional orientation within the CEACAM gene cluster (Figure II-S1). In contrast, CEACAM3 orthologs were absent from the available genomes of New World monkeys or lemurs (Chang, Semyonov et al. 2013; Sato, Kuroki et al. 2015) suggesting that CEACAM3 first emerged following the divergence of Old and New World monkeys (Figure II-1 B). Furthermore, the syntenic chromosomal locus of the gibbon (*Nomascus leucogenys*) also did not contain a CEACAM3 ortholog, which is presumably lost from this species due to genomic rearrangements occurring on chromosome 19 (Figure II-1 B) (Jauch, Wienberg et al. 1992; Carbone, Harris et al. 2014). Sequence alignment of CEACAM3 orthologs indicated a high degree of conservation in the amino-terminal signal sequence, the transmembrane region, as well as the C-terminal intracellular domain including the ITAM-like motif, which initiates downstream signaling upon receptor engagement (Figure II-1 C). In contrast, considerable amino acid sequence diversity could be seen in the extracellular Ig-variable like domain responsible for ligand binding (Figure II-1 C). Due to the amino acid sequence variation in this region, CEACAM3 orthologs from human and chimp also lack an additional N-glycosylation site present in the other species (Figure II-1 C).

To get insight into potential positive selection acting on CEACAM3, the dN/dS ratio between human, chimpanzee (pan), orang-utan (pongo) and rhesus monkey (macaca) was determined for full length CEACAM1, CEACAM5, CEACAM3 and Dectin-I as well as the terminal binding domain of CEACAM3 and Dectin-I (Figure II-1 D). Strikingly, CEACAM3 displays an exceptionally high dN/dS ratio pointing to an even stronger selective pressure acting on this innate immune receptor, when compared to the previously analyzed epithelial CEACAMs (Figure II-1 D). Interestingly, other phagocyte receptors involved in detecting and eliminating pathogens, such as Dectin-1, which recognizes fungal β-glucans, show a dN/dS ratio which is only slightly elevated from the genome-wide average (Figure II-1 D). CEACAMs as well as Dectin-1 bind their ligands via their terminal extracellular domain. Therefore, we restricted the analysis to this domain. The low dN/dS ratio seen for the Dectin-1 β-glucan binding domain indicates that there is strong purifying selection acting on this phagocytic receptor in higher primates (Figure II-1 D). In striking contrast, the Ig_V-like domain of CEACAM3 shows not only a high sequence divergence between human and chimp (90.6% identity), orang-utan (78.5%), or rhesus monkey (74.7%), but also has an extraordinarily pronounced dN/dS ratio. Comparably high dN/dS ratios can be observed between all analyzed combinations of primate CEACAM3 Ig_V-like domains, while the divergence between Dectin-I CLEC-domains remains constantly low.

(Figure II-1 E) This not only suggests that there is an exceptionally strong selection pressure on all CEACAM family members recognizing pathogenic bacteria, but that in particular CEACAM3 appears as one of the fastest evolving primate genes.

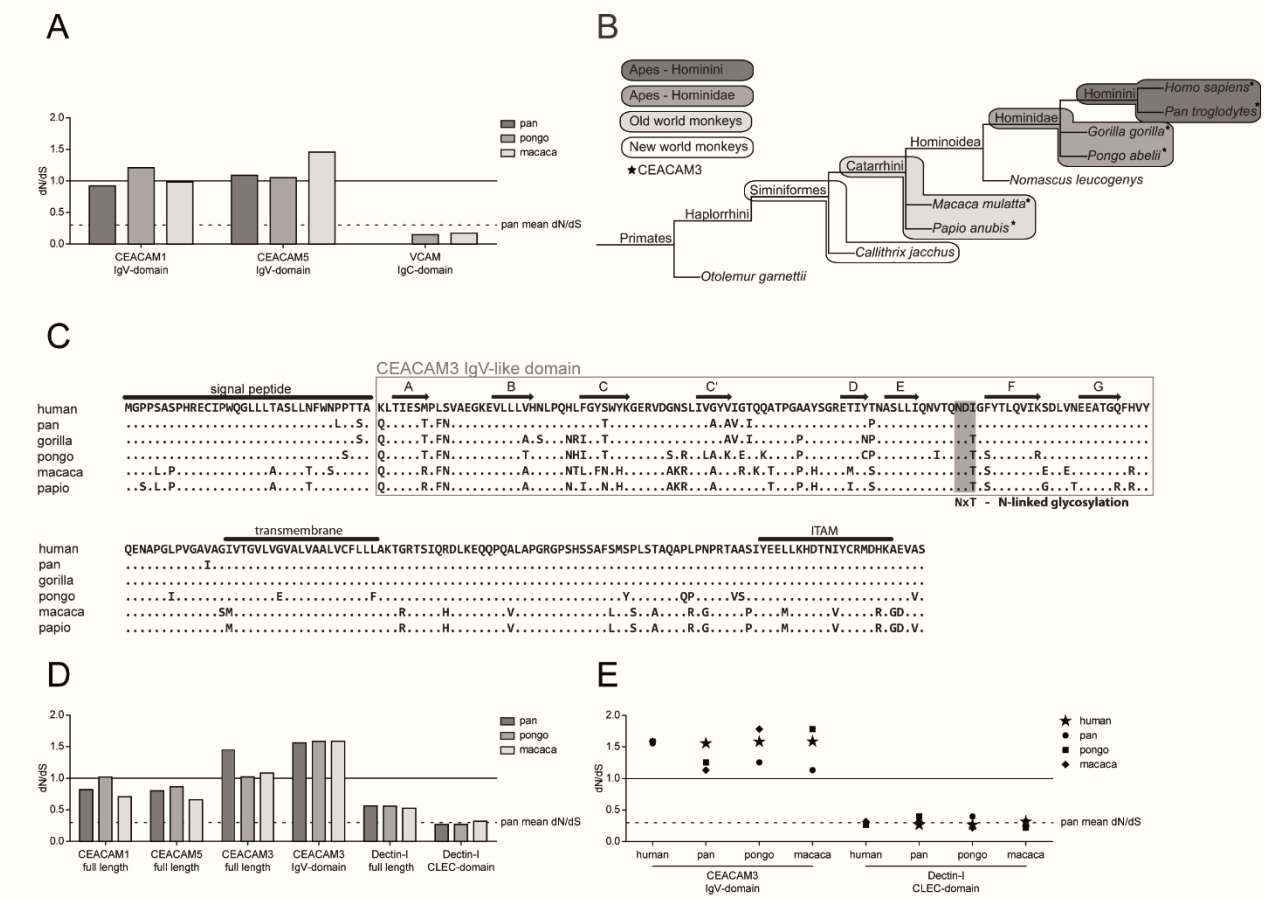


Figure II-1: The granulocyte receptor CEACAM3 is under exceptionally strong positive selection. (A) The ratios of non-synonymous substitution per non-synonymous site and synonymous mutations per synonymous site (dN/dS) in the amino-terminal Ig-domain of human CEACAM1, CEACAM5, and VCAM genes compared to chimp (*Pan troglodytes*), orang-utan (*Pongo abelii*), and macaque (*Macaca mulatta*) using the full length coding region of each gene. The continuous horizontal line indicates a dN/dS ratio of 1, signifying the border between positive selection (>1) and purifying selection (<1) while the dashed line indicates the average dN/dS ratio derived from 17,226 orthologs gene pairs from human and chimp as reported by (Wolf, Kunstner et al. 2009). (B) Human CEACAM3 orthologs identified in primate genomes. Species encoding a CEACAM3 ortholog are indicated by a star. (C) Amino acid sequence alignment of primate CEACAM3 orthologs. Areas highlighted by a horizontal bar

indicate the signal peptide, the β-strands A-G of the Ig_V-like fold, the transmembrane domain, and the cytoplasmic ITAM-like sequence. A predicted additional glycosylation site present in some primates is indicated by a grey box. (D) Plotting of the dN/dS ratios of CEACAM1, CEACAM5, CEACAM3, and Dectin-I full length sequence as well as CEACAM3-Ig_V-like domain and Dectin-I CLEC domain of human to primate orthologs. Lines are as in (A). (E) Comparison of dN/dS ratios for all combinations of human, chimp, orang-utan and macaque sequences of the CEACAM3-Ig_V-domain or the Dectin-I CLEC-domain. Lines are as in (A). See also Figure II-S1 and Table II-S1.

Recognition of human bacterial pathogens by primate CEACAM3 orthologs follows a phylogenetic gradient

As the Ig_V-like domain of CEACAM3 appears to be under strong positive selection, we speculated that the amino acid sequence alterations acquired during Old World monkey evolution should result in functional consequences with regard to bacterial recognition. To test the binding capabilities of primate CEACAMs, we recombinantly expressed the CEACAM3 extracellular Ig_V-like domains of human, chimp, gorilla, rhesus monkey and baboon in the form of soluble GFP-fusion proteins (Figure II-2 A). Though the CEACAM3 orthologs from gorilla, papio, and macaque showed a slightly reduced mobility upon SDS-PAGE (most likely due to the presence of an additional N-glycosylation site), all GFP-fusion proteins were expressed at similar levels (Figure II-2 A). To compare their ability to recognize different CEACAM-binding proteins, we expressed CEACAM3-binding adhesins derived from *Moraxella catarrhalis* (UspA1), *Neisseria gonorrhoeae* (Opa_{52}), and *Haemophilus aegyptius* (OMP P1) in *E. coli*. As laboratory strains of *E. coli* do not express CEACAM-binding adhesins, the heterologous expression allowed a direct comparison of the association of the CEACAM3 orthologs with different adhesins in the same bacterial background. To detect binding, the soluble CEACAM3-GFP-fusions were incubated with intact bacteria, washed, and bacteria-bound protein was detected by immunoblotting with anti GFP-antibodies (see scheme in Figure II-2 B).

Remarkably, the primate CEACAM3 orthologs showed distinct binding patterns to the bacterial adhesins. While human and chimpanzee CEACAM3 was able to interact with all three tested bacterial adhesins, the binding efficiency of gorilla CEACAM3 to neisserial Opa_{52} was clearly reduced, while strong affinity to *Moraxella* UspA1 was retained. Interestingly, gorilla

CEACAM3 completely lacked binding to *H. aegyptius*-derived OMP P1. In the case of catarrhini, rhesus monkey CEACAM3 only interacted with UspA1, while no binding of baboon CEACAM3 to any of the bacterial adhesins was detected (Figure II-2 C). To confirm these data, CEACAM3-binding to the bacterial surface was evaluated by a flow cytometry-based assay. Again, human CEACAM3 showed exceptionally strong binding to all three bacterial adhesins and chimpanzee CEACAM3 recapitulated this binding pattern, albeit with reduced binding to *H. aegyptius* OMP P1 (Figure II-2 D). Confirming the immunoblotting analysis, gorilla CEACAM3 completely lacked binding to OMP P1, while efficiently associating with Opa_{52} and UspA1 (Figure II-2 D). Furthermore, rhesus monkey CEACAM3 bound to UspA1, whereas baboon CEACAM3 did not associate with any of the bacterial adhesins (Figure II-2 D). The derived binding profiles for the different primate CEACAM3s are summarized in Figure II-2 E. Our findings demonstrate that the amino acid sequence alterations observed between primate CEACAM3 orthologs result in differential binding to adhesins derived from human-restricted bacterial pathogens. In this regard, the CEACAM3-mediated recognition of the bacterial adhesins correlates well with the evolutionary distance between these primates. Moreover, the exquisite species-selectivity of some of these interactions could allow the identification of critical molecular determinants required for receptor-ligand binding.

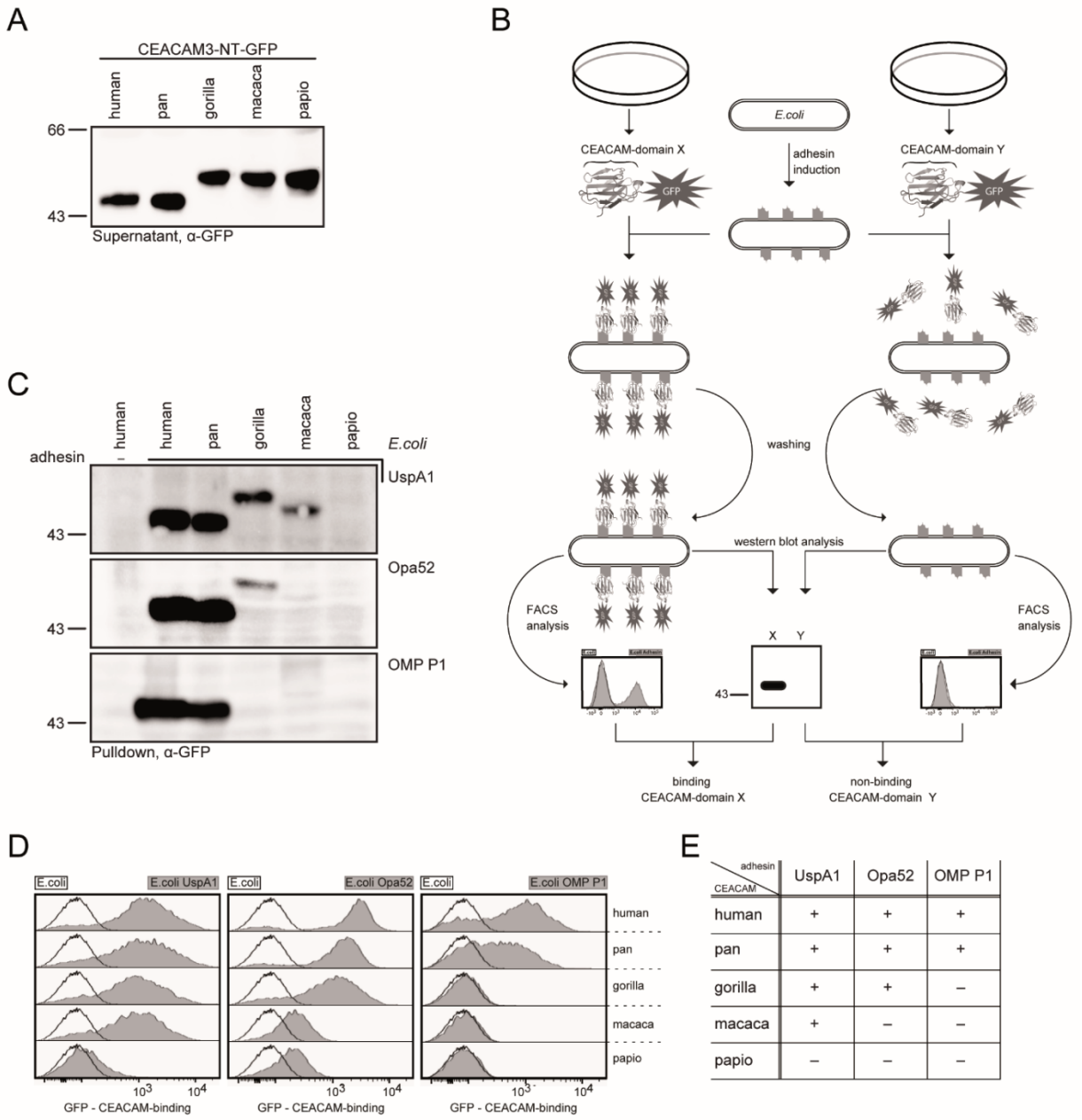


Figure II-2: Recognition of human bacterial pathogens by primate CEACAM3 orthologs follows a phylogenetic gradient. (A) Amino-terminal domains of human, chimpanzee, gorilla, baboon and macaque CEACAM3 were expressed as soluble, GFP-tagged proteins in 293T cells. Equivalent protein amounts in the supernatants were verified by immunoblotting analysis with anti-GFP antibodies. (B) Schematic work-flow of a bacterial pulldown assay. Soluble GFP-tagged CEACAM-Ig_V-like domains are secreted by transfected cells into cell culture supernatant. *E. coli* expressing defined pathogen adhesins are incubated with CEACAM-containing cell culture supernatants ("X" hypothetical binding; "Y" hypothetical non-binding) to allow CEACAM interaction. Subsequently, unbound CEACAM-Ig_V-GFP is removed by repeated washing steps. The surface bound CEACAMs are analyzed via immunoblotting or

flow cytometry. (C, D) Supernatants containing the indicated CEACAM3 variants were incubated with *E. coli* expressing Opa_{52} of *N. gonorrhoeae*, UspA1 derived from *M. catarrhalis*, OMP P1 derived from *H. aegyptius*, or with control *E. coli* expressing no additional adhesin. CEACAM3 binding was determined via bacterial-pulldown and subsequent immunoblotting analysis (C) or analysis of GFP fluorescence via flow cytometry (D). (E) Summary of adhesin-binding pattern of the analyzed primate CEACAM3-Ig_V-like domains. Binding is symbolized by (+), while lack of binding is symbolized by (–).

Minor sequence alterations in CEACAM3 allow recognition of *H. aegyptius* OMP P1 adhesin

As *H. aegyptius* OMP P1 exhibited a striking selectivity for human and chimp CEACAM3, we aligned the human, chimpanzee and gorilla CEACAM3 Ig_V-like domains to reveal sequence alterations exclusively found in gorilla. In previous studies, the binding interface of several bacterial adhesins has been mapped to the non-glycosylated CC'FG-face of the amino-terminal Ig_V-like domain (Virji, Evans et al. 2000; Bonsor, Zhao et al. 2018). Indeed, a three amino acid long stretch (HLF, amino acid residues 27-29 of the mature protein) located at the amino-terminal end of the C-strand differs between human and chimpanzee versus gorilla (Figure II-3 A). 3D-modeling of the human CEACAM3 Ig_V-like domain revealed a hydrophobic cleft flanking the CC'FG-face of the Ig-fold formed by the HLF triplet (Figure II-3 B, C). To assess the impact of this sequence alteration for OMP P1 binding we changed the respective amino acid triplet in the gorilla CEACAM3 Ig_V-like domain to the corresponding human sequence (humanized gorilla; hu-gorilla) and expressed hu-gorilla as a GFP-fusion protein (Figure II-3 D). Binding assays with *E.coli* expressing Opa_{52} or OMP P1 revealed that introduction of the HLF-motif into gorilla CEACAM3 not only permitted binding to OMP P1, but also clearly increased the affinity to *N. gonorrhoeae* Opa_{52} (Figure II-3 E). To confirm these data, CEACAM3-binding to the bacterial surface was measured by an independent flow cytometry-based assay. In accordance with the previous results, human and gorilla CEACAM3 were able to interact with *N. gonorrhoeae* Opa_{52}, while only hu-gorilla, but not gorilla CEACAM3 was able to interact with *H. aegyptius* OMP P1 with an efficiency comparable to the human ortholog (Figure II-3 F). To further narrow down the critical residue(s) within the HLF-motif, we reverted single residues in the HLF triplett of hu-gorilla CEACAM3 back to the gorilla sequence and created hu-gorilla

CEACAM3 NLF, CEACAM3 HRF, and CEACAM3 HLI (Figure II-3 G). Analysis of their binding to OMP P1 revealed that only the hu-gorilla CEACAM3 HLI, lacking the phenylalanine residue F29, was unable to interact with the OMP P1 adhesin, while retaining its affinity for Opa_{52} (Figure II-3 H-I). These results demonstrate that within a three amino acid motif found in human and chimpanzee CEACAM3 Ig_V-like domain a single amino acid residue, F29, which generates a hydrophobic notch on the CC'FG-face, is responsible for the remarkable specificity for selected pathogen adhesins. It is highly suggestive that this minor sequence change represents an evolutionary adaptation that allows CEACAM3-mediated recognition and internalization of *H. aegyptius*.

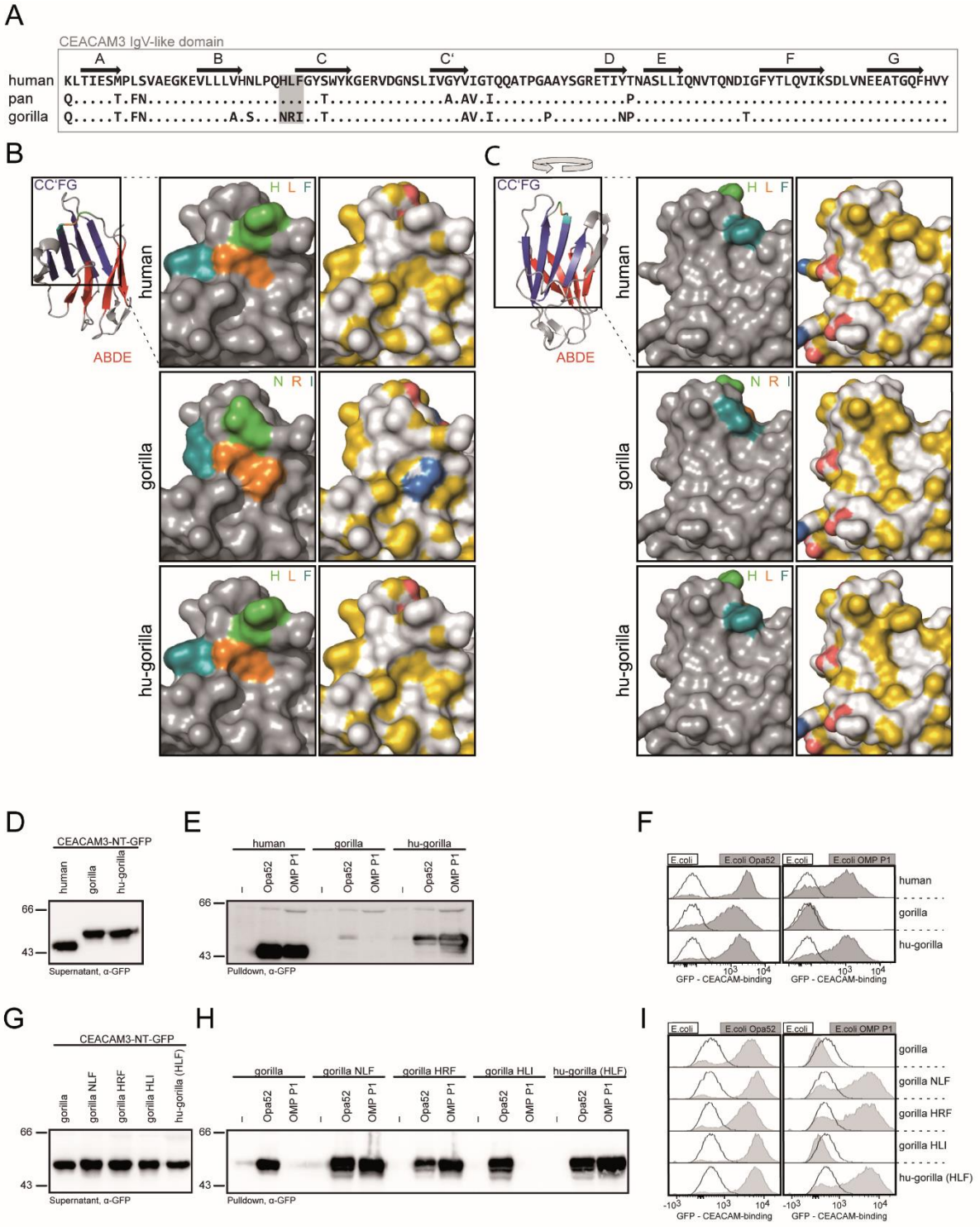


Figure II-3: Minor sequence alterations in CEACAM3 allow recognition of *H.aegyptius* OMP P1 adhesin. (A) Amino acid sequence alignment of the Ig_V-like domain of human, chimpanzee and gorilla CEACAM3 orthologs. Areas highlighted by a horizontal bar indicate the β-strands A-G of the Ig_V-like fold. A grey box marks an amino acid triplet that is identical in human and chimpanzee, but distinct in gorilla CEACAM3. (B) 3D-model of human and gorilla CEACAM3 Ig_V-like domain based on CEACAM3 crystal structure (6AW1). Domain orientation and detail

area are indicated (left). The respective HLF and NRI amino acid triplet is colorized for human, gorilla, and humanized gorilla CEACAM3 Ig_V-like domain (hu-gorilla), in which the NRI-triplet was replaced by HLF (center). Surface charge (red (-), blue (+)) and hydrophobicity (yellow) mapping is marked separately (right). (C) 3D-model as in (B) of the other side on the Ig-fold, with the CC'FG-face to the front. (D) GFP-tagged N-terminal domains of human, gorilla and hu-gorilla CEACAM3 were generated as secreted proteins in 293T cells. Equivalent amounts of soluble proteins in the employed supernatants were verified by immunoblotting analysis. (E, F) Supernatants containing the indicated CEACAM3 variants were incubated with *E. coli* expressing *N. gonorrhoeae* Opa_{52}, *H. aegyptius* OMP P1 or no additional adhesin. CEACAM3 binding was determined via bacterial-pulldown and immunoblotting analysis (E) or flow cytometry (F). (G) GFP-tagged N-terminal domains of gorilla CEACAM3 (NRI) and hu-gorilla CEACAM3 HLF, NLF, HRF or HLI were generated and equivalent amounts of soluble protein in the supernatants were verified by immunoblotting. (H, I) Supernatants containing the indicated CEACAM3 variants were incubated with *E. coli* expressing *N. gonorrhoeae* Opa_{52}, *H. aegyptius* OMP P1, or no additional adhesin. CEACAM3 binding was determined via bacterial-pulldown assays and immunoblotting analysis (H) or flow cytometry (I).

The CC'FG-cleft architecture determines species-specific CEACAM3 binding and pathogen internalization

To test the functional significance of HLF-motif-mediated binding to OMP P1 of *H. aegyptius* in a cellular context, we expressed GFP-tagged human, gorilla or hu-gorilla CEACAM3 or GFP alone in 293T cells. The cells were then infected for 1 h with *Neisseria gonorrhoeae* expressing Opa_{52} (Figure II-4 A) or *Haemophilus aegyptius* (Figure II-4 B). Remarkably, differential staining of total and extracellular bacteria unveiled numerous intracellular gonococci in all cells expressing CEACAM3, irrespective of the species origin of the receptor (Figure II-4 A). These results were in line with the observed receptor binding pattern and indicated that gorilla CEACAM3 is functional in mediating internalization of Opa adhesin-expressing bacteria. When the same set of transfected cells was infected with OMP P1 bearing *H. aegyptius*, gorilla CEACAM3 did not promote uptake of *H. aegyptius*, while cells expressing the human or the humanized gorilla CEACAM3 protein readily internalized this pathogen (Figure II-4 B). Equivalent receptor expression levels were verified by Western blotting (Figure II-4 C). To evaluate the extent of functional restoration by this three amino acid exchange, we infected

cells expressing the different CEACAM3 variants and infected them for 1 h with *E. coli* expressing *H. aegyptius* OMP P1 or control *E. coli*. Internalized bacteria were quantified by gentamicin protection assays (Figure II-4 D). In line with our previous results, *E. coli* expressing OMP P1 could not be internalized by gorilla CEACAM3, while expression of hu-gorilla CEACAM3 promoted bacterial internalization as efficient as the human CEACAM3 ortholog (Figure II-4 D). These results in 293T cells could be confirmed in neutrophil-like HL60 cells that were stably transduced to express either human, gorilla, or hu-gorilla CEACAM3 fused to mKate2. CEACAM3-mKate expression in the derived clonal cell lines was verified by immunoblotting with a mKate specific antibody as well as via flow cytometric analysis (Figure II-4 E). As expected, an enhanced internalization of *N. gonorrhoeae* could be detected in HL60 cells expressing human, gorilla, or hu-gorilla CEACAM3 demonstrating the functionality of all receptor constructs. In contrast, enhanced internalization of *H. aegyptius* could only be observed in human or hu-gorilla CEACAM3 expressing HL60 cells, but not in cells expressing the gorilla variant (Figure II-4 F). No internalization could be observed for *H. influenzae* by any of the CEACAM3-expressing HL60 lines demonstrating that CEACAM1 does not contribute to the observed uptake (Figure II-4 F). Together, these data demonstrate that the gorilla as well as the human variant mediates pathogen internalization. However, due to the absence of the hydrophobic notch provided by F29, the gorilla ortholog is unable to internalize *H. aegyptius*. Humanization of amino acids in the gorilla variant permits efficient *H. aegyptius* binding and restores pathogen internalization. These findings further support the notion that CEACAM3 sequence alterations are evolutionary adaptations that enable granulocytes to respond to particular pathogens that would otherwise escape this innate immune surveillance.

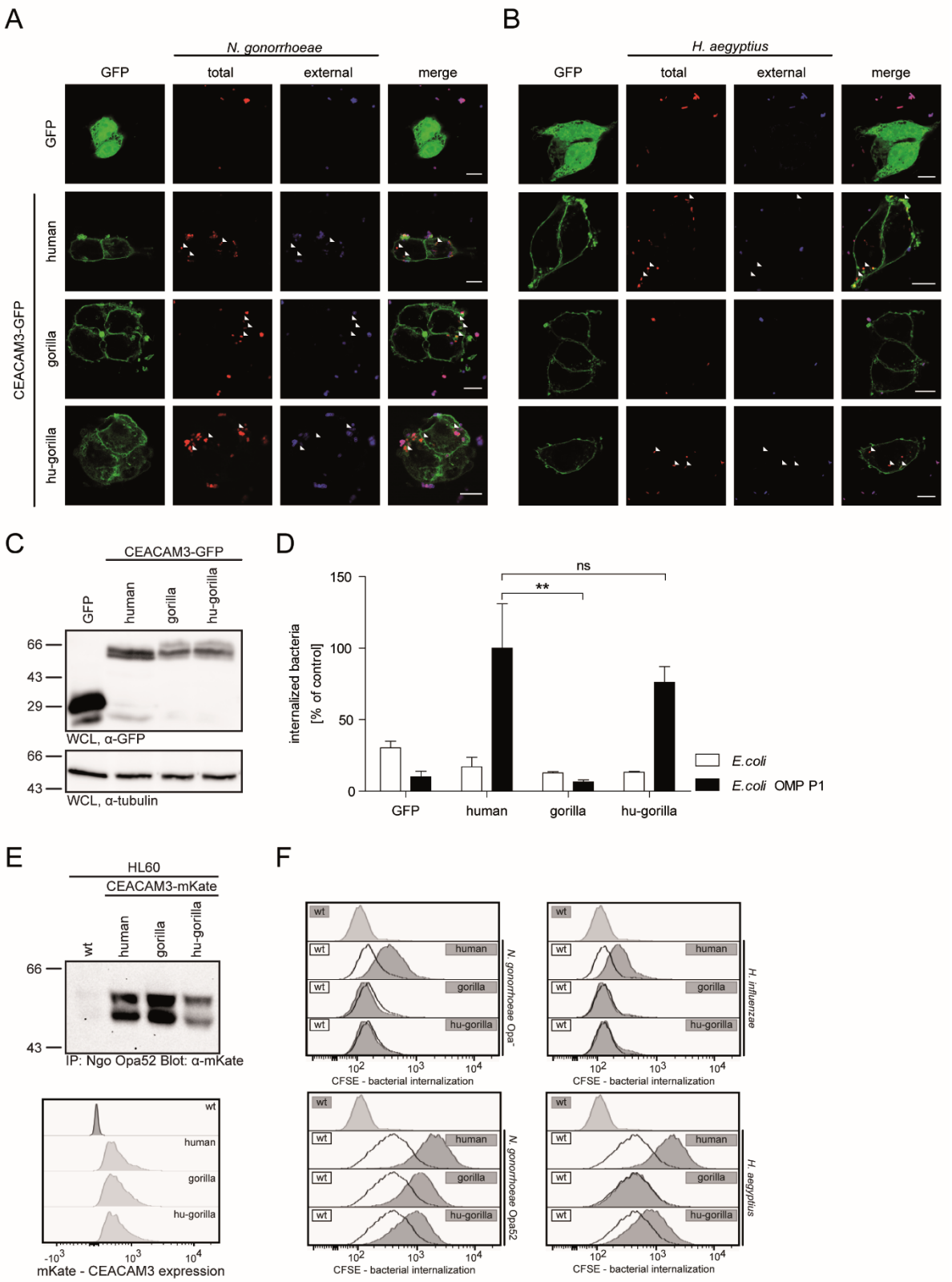


Figure II-4: The CC'FG-cleft architecture determines species-specific CEACAM3 binding and pathogen internalization. (A, B) 293T cells were transfected with GFP or the indicated CEACAM3-GFP variants (green) and infected for 1 h at MOI 40 with biotin- and rhodamine-labeled bacteria (red) either Opa_{52} expressing *N. gonorrhoeae* (A) or *H. influenzae* (B). Fixed

samples were stained with Cy5-streptavidin (blue) to label extracellular bacteria. White arrowheads point to internalized bacteria. Scale bars 10 µm. (C) Equivalent expression of CEACAM3 variants was verified by Western blotting of whole cell lysates (WCL) with a GFP-antibody (upper panel), similar amounts of loaded lysates were controlled with a tubulin-antibody (lower panel). (D) Transfected cells as in (A) were infected for 1 h at MOI of 100 with *E. coli* or *E. coli* expressing OMP P1. The amount of internalized bacteria was subsequently evaluated by gentamicin protection assay. Bars represent means ± SEM from three independent experiments done in triplicate. n.s., not significant; **p <0.01. (E) Parental HL60 cells (wt) or HL60 cells stably expressing the indicated CEACAM3-mKate2 fusion proteins were lysed and expression of human, gorilla, and hu-gorilla CEACAM3-mKate was analyzed after bacterial pulldown by immunoblotting with anti-mKate antibodies (upper). Flow cytometric analysis of mKate2 fluorescence in HL60 cells (lower). (F) HL60 cells were infected for 15 min with fluorescein-labeled *N. gonorrhoeae* Opa_{52} or fluorescein-labeled, non-opaque *N. gonorrhoeae* (Opa^-), *H. influenzae*, or *H. aegyptius*. Fluorescein signals derived from extracellular bacteria were quenched by trypan blue. Signals from phagocytosed bacteria were detected by flow cytometry.

Pathogen adhesins discriminate between CEACAM3 and CEACAM1 based on specific sequence motifs

If human CEACAM3 rapidly evolves to enable the elimination of various host-restricted bacterial pathogens via phagocytosis, then the adhesins on the bacterial side should also evolve to evade CEACAM3-dependent detection, while retaining binding to epithelial CEACAMs, such as CEACAM1, for efficient host colonization. Accordingly, we screened Opa adhesins and OMP P1 adhesins of different *N. gonorrhoeae* and *H. influenzae* isolates for their ability to bind CEACAM1 versus CEACAM3. Analysis of individual Opa or OMP P1 proteins showed that some adhesins (Opa_{52}, Opa_{70}, Opa_{72}, OMP $P1_{Hae}$) associated with both CEACAM1 and CEACAM3, while a larger fraction of the adhesins (Opa_{65}, Opa_{73}, Opa_{74}, Opa_{75}, OMP $P1_{Hinf}$) exclusively bound to CEACAM1 (CEACAM1-optimized adhesins; Figure II-5 A). No adhesins were detected, which associated with CEACAM3, but not CEACAM1. These binding assays were corroborated by FACS-based analysis, which confirmed that most tested CEACAM-binding adhesins are optimized for CEACAM1 association, while evading detection by

CEACAM3 (Figure II-5 B). Amino acid sequence comparison between human CEACAM3 and CEACAM1 indicated a limited set of amino acid changes that could form the basis for selective CEACAM1 binding (Figure II-5 C). We focused on six individual amino acid motifs, based on their predicted positions in the Ig-fold, and mutated these residues in CEACAM3 either individually or in combination (Figure II-5 C). While all CEACAM3 variants were able to recognize neisserial Opa_{52} and *H. aegyptius* OMP P1, the CEACAM1-optimized OMP P1 of *H. influenzae* exclusively bound to CEACAM1 and to the CEACAM3 variant, which combined all six mutated locations (Figure II-5 D and E and Figure II-S2 A). Binding of OMP $P1_{Hinf}$ was not observed for variants lacking the V49A and T69P mutation (Figure II-5 D and E and Figure II-S2 B). For CEACAM1-optimized neisserial Opa-proteins the observed binding pattern was more complex. Interaction with Opa_{65} could be re-established by simultaneous mutation of the RQ-site together with the Q-, PaN-, or T-site. Mutation of the CEACAM3-RQ site alone was not sufficient to permit association with Opa_{65}, illustrating the synergistic nature of these binding interactions. While the CEACAM3 Q-RQ-PaN-T variant was either incapable of mediating interaction with CEACAM1-optimized adhesins (Opa_{73}, Opa_{74}, Opa_{75}) or mediated only weak interaction (Opa_{65}), the additional mutation of V49A and T69P strongly promoted the binding to the bacterial adhesins (Figure II-6 E). These findings illustrate the fine tuning of CEACAM3 binding to particular bacterial adhesins by single amino acid changes. From our results we would predict that polymorphisms at the CEACAM3 locus might promote binding to microbes, which otherwise evade detection by the common CEACAM3 allele. This would further point to pathogenic bacteria as the driving force behind the accelerated evolution of CEACAM3.

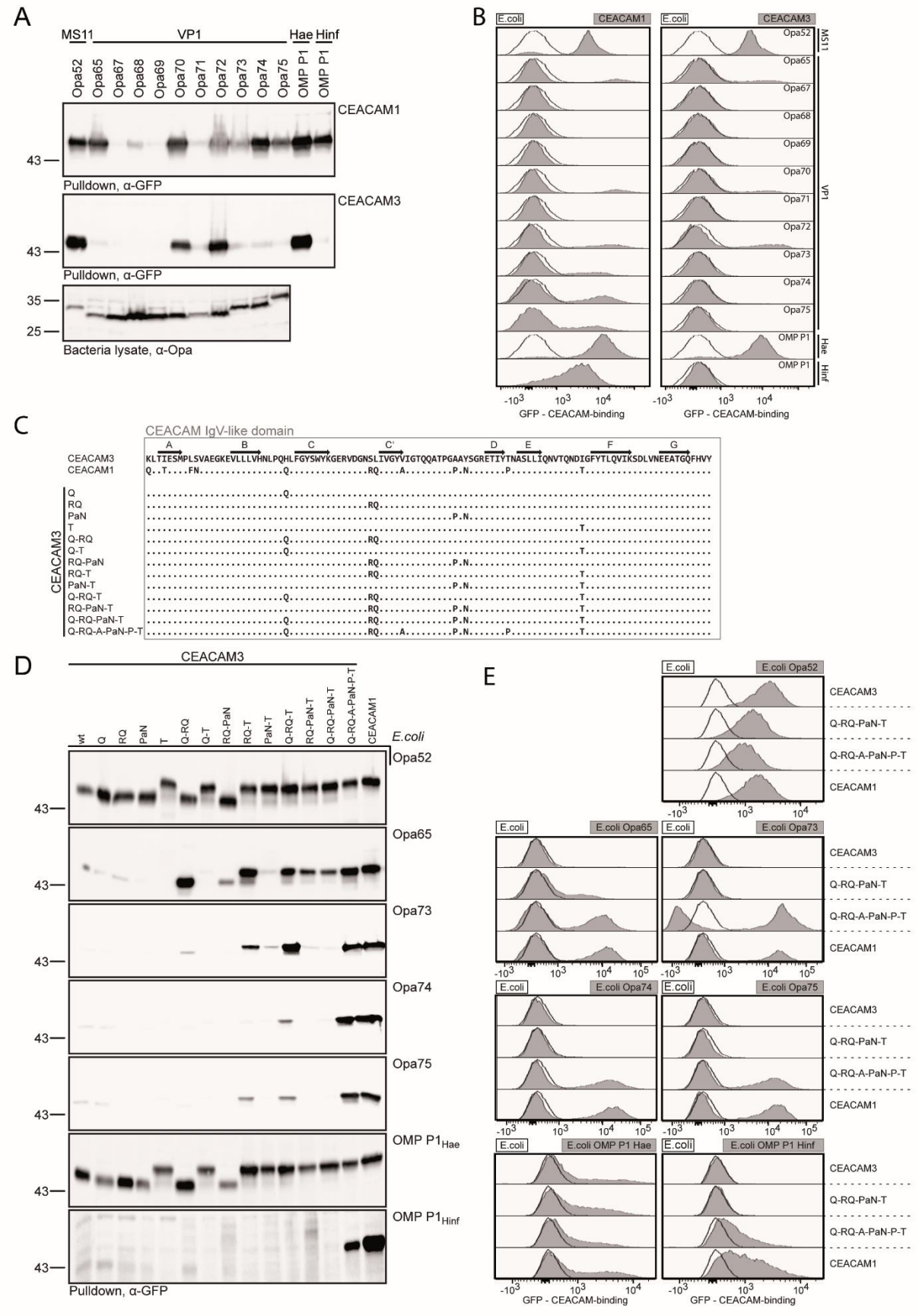


Figure II-5: Pathogen adhesins discriminate between CEACAM3 and CEACAM1 based on specific sequence motifs. (A, B) Soluble GFP-tagged CEACAM3 or GFP-CEACAM1 were

incubated with *E. coli* expressing the individual Opa-variants derived from *N. gonorrhoeae* strains MS11 or VP1 or with OMP P1 derived from *H. influenzae* (Hinf) or *H. aegyptius* (Hae). CEACAM binding was determined via bacterial-pulldown and subsequent immunoblotting analysis with anti-GFP antibodies (A) or analysis of GFP fluorescence via flow cytometry (B). (C) Amino acid sequence alignment of the Ig_V-like domains of CEACAM3 and CEACAM1. Areas highlighted by a horizontal bar indicate the β-strands A-G of the Ig_V-like fold. In a step-wise fashion, the indicated 1-8 amino acid changes are introduced into CEACAM3 to emulate CEACAM1. (D, E) Soluble GFP-tagged CEACAM3 variants described in (C) were incubated with *E. coli* expressing the indicated adhesins. CEACAM binding was determined via bacterial-pulldown and subsequent immunoblotting (D) or by flow cytometry (E). See also Figure II-S2.

Human CEACAM3 polymorphisms broaden the spectrum of recognized adhesins

Whole genome sequencing has uncovered numerous single nucleotide polymorphisms (SNPs) in the human population, which can result in missense mutations. We speculated that affected CEACAM3 alleles could have a functional role in broadening the spectrum of CEACAM3-regonized adhesins. Indeed, four SNPs are present in the CEACAM3 Ig_V-like domain coding region with each of those sequence variants occuring in about 10% of the human population (9.9% (S43R), 9.9% (L44Q), 10.4% (V49A) and 10.8% (T69P); 2543 individual genotypes; Ensembl Database 2018) (Figure II-6 A). Remarkably, these SNPs not only convert the amino acids at these positions to the corresponding CEACAM1 sequence, but they occur at exactly those sites, which appeared to be highly relevant for adhesin-CEACAM selectivity (see Figure II-5 D and E). All of these SNPs are strongly enriched in Africa and are present there in 35.1% (S43R), 35.1% (L44Q), 37.4% (V49A) and 38.1% (T69P) of individuals (Figure II-6 A and Figure II-S3). Closer inspection of the distribution of these SNPs revealed that they, almost exclusively, co-occur in the same individuals as they are presumably present in the same allele (Figure II-6 B). Moreover, the loci show extended haplotype homozygosity as evaluated with the hapbin program, where an integrated haplotype score (|iHS|) of >2 is indicative of recent selection (Voight, Kudaravalli et al. 2006; Maclean, Chue Hong et al. 2015). The fact that an |iHS| >2 for these four SNPs is found only in several African populations, including the Esan in Nigeria, the Luhya in Kenya, and Yoruba in Nigeria, as well as people with African Ancestry

from Southwest America or the Carribean (total of 463 genomes analyzed), indicates a puzzling regional selection on CEACAM3.

As R43, Q44, and A49 also contributed to enhanced recognition of CEACAM1-optimized Opa-proteins by CEACAM3 (see Figure II-5 D and E), we directly assessed the functional relevance of these naturally occurring CEACAM3 variants. GFP-fusion proteins of CEACAM3-Ig_V variants reflecting either the common allele or the SNP-based variants were compared to CEACAM1-Ig_V for their binding to different pathogen adhesins. As before, CEACAM1 and CEACAM3 as well as all SNP-based variants of CEACAM3 efficiently interacted with Opa_{52} and OMP $P1_{Hae}$ (Figure II-6 C). Strikingly, already the V49A substitution in CEACAM3 was sufficient to permit interaction with all tested CEACAM1-optimized Opa-proteins (Figure II-6 C). Furthermore, the combination of A49 with R43/Q44 allowed CEACAM3 to also bind to OMP P1 of *H. influenzae*, a pathogen not recognized by the common allele of CEACAM3 (Figure II-6 D). The differences in binding of CEACAM3 variants to Opa and OMP P1 adhesins were even more obvious in the FACS-based evaluation, where the combination of the R43, Q44, and A49 polymorphisms transformed the non-binding common allele of CEACAM3 to a strong binder for OMP $P1_{Hinf}$ (Figure-II 6 E and Figure-II S3 B and S3 C). 3D-modeling of the various CEACAM3 Ig_V-like domains showed that the S43R, L44Q, and V49A residues are positioned at the CC'FG-face, while the T69P substitution is placed on the opposite face of the Ig_V-like domain explaining the lack of a major contribution from the T69P polymorphism (Figure II-6 F). Together, these results demonstrate that CEACAM3 alleles in the human population reflect the molecular adaptations and counter adaptations of functionally connected host and pathogen proteins, and provide an example for a Red Queen scenario embedded in the human genome.

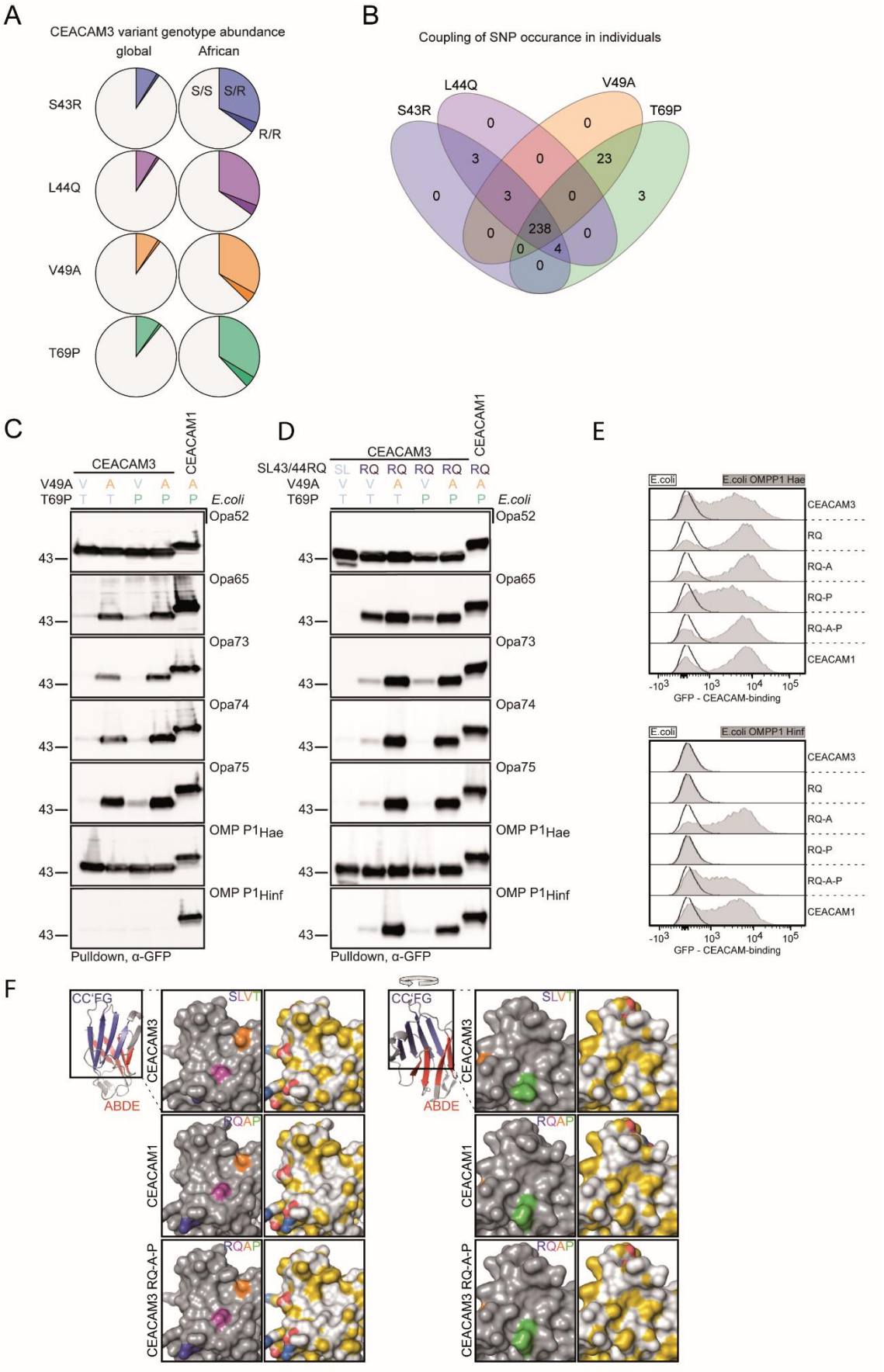

A
CEACAM3 variant genotype abundance
global
African
S43R
L44Q
V49A
T69P
S/S
S/R
R/R
B
Coupling of SNP occurance in individuals
L44Q
V49A
S43R
T69P
0
0
3
0
23
0
3
0
3
0
238
0
0
4
0
C
CEACAM3
CEACAM1
V49A V A V A A
T69P T T P P P
E.coli
Opa52
Opa65
Opa73
Opa74
Opa75
OMP P1Hae
OMP P1Hinf
43
Pulldown, α-GFP
D
CEACAM3
CEACAM1
SL43/44RQ SL RQ RQ RQ RQ RQ
V49A V V A V A A
T69P T T T P P P
E.coli
Opa52
Opa65
Opa73
Opa74
Opa75
OMP P1Hae
OMP P1Hinf
43
Pulldown, α-GFP
E
E.coli
E.coli OMPP1 Hae
CEACAM3
RQ
RQ-A
RQ-P
RQ-A-P
CEACAM1
-10^3
10^3
10^4
10^5
GFP - CEACAM-binding
E.coli
E.coli OMPP1 Hinf
CEACAM3
RQ
RQ-A
RQ-P
RQ-A-P
CEACAM1
-10^3
10^3
10^4
10^5
GFP - CEACAM-binding
F
CC'FG
ABDE
SLVT
RQAP
CEACAM3
CEACAM1
CEACAM3 RQ-A-P

Figure II-6: Human CEACAM3 polymorphisms broaden the spectrum of recognized adhesins. (A) CEACAM3 variant genotype abundance in 2543 individual genotypes from the Ensembl Database 2018. The SNPs are found in 9.9% (S43R), 9.9% (L44Q), 10.4% (V49A) and 10.8% (T69P) of all individuals (left) and are more abundant in African populations at 35.1% (S43R), 35.1% (L44Q), 37.4% (V49A) and 38.1% (T69P) (right). Homozygotes (darker shade) and heterozygotes (lighter shade) for the minor variant are colorized. (B) Venn-diagram depicting the coupled occurrence of the four most common missense SNPs in the human CEACAM3-protein. Total number of individuals is noted within the respective area. (C, D, E) Soluble CEACAM3-GFP variants as indicated were incubated with *E. coli* expressing different Opa or OMP P1 adhesins. CEACAM3 binding to bacteria was determined via bacterial pulldown and immunoblotting (C, D) or by flow cytometry (E). See also Figure II-S2. (F) 3D-model of human CEACAM3 Ig_V-like domain based on CEACAM3 crystal structure (6AW1). For each panel, Ig-domain orientation and zoom is indicated (upper left), depicting the CC'FG-face (blue) and the ABDE-face (red). Relevant regions on CEACAM3 (WT; upper), CEACAM1 (middle) and CEACAM3 SL43/44RQ V49A T69P variant (lower) are colored: Amino acid 43 (blue), 44 (violet), 49 (orange) and 69 (green). Surface charge (red (-), blue (+)) and hydrophobicity (yellow) mapping is marked separately (right). See also Figure II-S3.

Discussion

Innate immune receptors act as sensors and a first line of defense against pathogens. Several pattern-recognition-receptor families such as TLRs, NOD-like receptors, and RIG-like receptors are directed against broadly conserved microbial structures and are mainly under purifying selection in primates (Kimbrell and Beutler 2001; Barreiro, Ben-Ali et al. 2009; Vasseur, Patin et al. 2011; Mukherjee, Ganguli et al. 2014). In contrast, the innate immune receptor CEACAM3, which recognizes adhesins expressed by a group of human-restricted bacterial pathogens, emerges as one of the fastest evolving human genes. Interestingly, CEACAM3 appears as a rather recently evolved CEACAM family member, as *bona fide* CEACAM3 orthologs could not be detected in the genomes of lemurs and New World monkeys. All orthologous *CEACAM3* genes identified occupy the syntenic locus, share the same exon-intron structure, encode a similar sized protein and harbor the YxxL --- YxxM HemITAM signature, which is critical for CEACAM3 function and discriminates CEACAM3 from other ITAM-encoding CEACAM family members (Schmitter, Pils et al. 2007; Schmitter, Pils et al. 2007; Buntru, Kopp et al. 2011; Buntru, Roth et al. 2012; Kopp, Buntru et al. 2012; Pils, Kopp et al. 2012; Bond, Tejedor et al. 2015; Delgado Tascon, Adrian et al. 2015). Therefore, our findings suggest that CEACAM3 appeared as a novel family member around 35 million years ago when Old World monkeys branched off the other simiiniformes (Bond, Tejedor et al. 2015).

The appearance of novel genes has occurred multiple times in the CEACAM family, as this group of immunoglobulin-related proteins has a distinct assortment of membrane receptors in each mammalian lineage (Kammerer and Zimmermann 2010; Kopp, Buntru et al. 2012; Bond, Tejedor et al. 2015; Delgado Tascon, Adrian et al. 2015). It has been noted earlier that, similar to the situation in primates, expansion of CEACAMs in marsupials, bats and carnivores resulted in the occurrence of paired receptors, which have similar extracellular domains, but encode either immunoreceptor tyrosine-based inhibitory motifs (ITIMs) or activatory motifs (ITAMs) in their cytoplasmic part (Kopp, Buntru et al. 2012; Bond, Tejedor et al. 2015; Delgado Tascon, Adrian et al. 2015). This phenomenon has been explained by the role CEACAMs could play in immunoregulation and by pathogen driven selection, but direct evidence for the latter has been lacking. Clearly, epithelial CEACAMs serve as a mucosal docking site for viruses and bacteria, providing a plausible explanation for their rapid diversification in mammals. For example, epithelial CEACAMs are exploited by several potentially deadly microbes such as

meningococci and *H. influenzae*, which can cause bacterial meningitis (Wenger, Hightower et al. 1990). Moreover, gonococci, one of the most common sexually transmitted bacterial pathogens and the causative agent of gonorrhea, have led to neonatal blindness and infertility in a pre-antibiotic world (Laga, Meheus et al. 1989; Cates, Rolfs et al. 1990). At the moment, we can only speculate about the selective forces behind the appearance of CEACAM3 within the primate lineage. However, fatal bacterial infections (meningitis) as well as reduced reproduction resulting from venereal diseases (gonorrhea) are strong selective pressures, which could favor the development of a germ line encoded innate immune receptor.

Pathogenic *Neisseria* (the gonococcus *N. gonorrhoeae* and the meningococcus *N. meningitidis*) as well as *Haemophilus influenzae* are specialized pathogens strictly adapted to their human host. Indeed, while they can infect apes and have occasionally been isolated from captive chimpanzees or gorillas (Mugisha, Kondgen et al. 2014), they are considered to be naturally restricted to humans. It is in line with the idea that CEACAM3 targets host-restricted bacteria, that human, but not monkey CEACAM3 can bind at least a portion of these human-adapted pathogens. To investigate, if CEACAM3 orthologs in primates other than man play a similar role in defense against specialized bacterial pathogens, one would need to study pathogenic bacteria isolated from wild chimpanzees, gorillas or rhesus monkeys. However, there is little knowledge about host-restricted bacteria in apes and higher monkeys, and bacterial isolates from these endangered species are not available in common strain repositories. Moreover, focused searches for several important sexually transmitted microbes in wild chimps and gorillas have failed to detect bacterial pathogens known from the human genital tract such as *Chlamydia* spp. or *Treponema pallidum* (Rushmore, Allison et al. 2017). Therefore, only detailed microbiome analysis and isolation of bacteria from wild apes will provide the appropriate starting point to test, if primate CEACAM3 orthologs are better in recognizing strains derived from the autochthonous microbial flora of the respective apes.

As CEACAM-binding pathogens exploit epithelial members of the CEACAM family, such as CEACAM1 and CEA, to facilitate colonization of host mucosal surfaces (Muenzner, Bachmann et al. 2010), binding to these epithelial CEACAMs should be favored. On the other hand, avoidance of recognition by CEACAM3 allows evasion of opsonin-independent phagocytosis mediated by this granulocyte receptor. In line with this idea, our detailed analysis of numerous pathogen adhesins detects a clear bias of CEACAM1 versus CEACAM3-binding in Opa and OMP

P1 adhesins and confirms earlier studies (Roth, Mattheis et al. 2013; Sintsova, Wong et al. 2015). A strong indication that CEACAM3-based detection is to the disadvantage of the pathogen is the fact that selective CEACAM3-binding adhesins, which lack binding to the epithelial CEACAMs such as CEACAM1 or CEA, do not seem to exist. To identify the molecular determinants that allow pathogens to discriminate between CEACAM1 and CEACAM3, we converted the amino acid sequence of the CEACAM3 Ig_V-like domain in a stepwise fashion to reflect the CEACAM1 primary structure. This extensive mutagenesis approach revealed a minimal set of three amino acid changes on the CC'FG-face of the Ig_V-like domain (S43R, L44Q and V49A), which dramatically broaden the binding spectrum of CEACAM3 to not only recognize CEACAM1-optimized neisserial Opa proteins, but also to detect the OMP P1 adhesin of *H. influenzae* (Figure II-7). Strikingly, all three amino acid alterations are also carried by polymorphic human CEACAM3 alleles, which show signatures of positive selection in African populations. These findings are in line with the idea that CEACAM3 polymorphisms could provide a selective advantage, by controlling a wider range of CEACAM-binding pathogens, and that this positive selection could form the basis for the rapid evolutionary divergence of CEACAM3 seen in different primate species.

Recent crystallographic studies have revealed the binding mode of another CEACAM-specific adhesin, the HopQ protein of *Helicobacter pylori*. This adhesin also mediates a direct protein-protein interaction with the CC'FG face of the CEACAM Ig_V-like domain, where it covers a large surface area (Bonsor, Zhao et al. 2018; Moonens, Hamway et al. 2018). In contrast to the CEACAM1-optimized neisserial Opa proteins and the OMP $P1_{Hinf}$, HopQ exhibits increased affinity for V49 (as found in the common allele of CEACAM3) compared to A49, as it is found in CEACAM1 (Moonens, Hamway et al. 2018). Therefore, the same amino acid change from V49 to A49, which is instrumental to allow CEACAM3-mediated recognition of several Opa-proteins of *N. gonorrhoeae* and of *H. influenzae* OMP P1, might come at the cost of reducing affinity for *Helicobacter* HopQ. In the long run, the multivariate selection pressure exerted on CEACAM3 by different pathogens and their CEACAM-binding adhesins could lead to balancing selection, which maintains distinct alleles in the population. Based on prevalence and severity of the corresponding infectious diseases, the allele frequencies in a

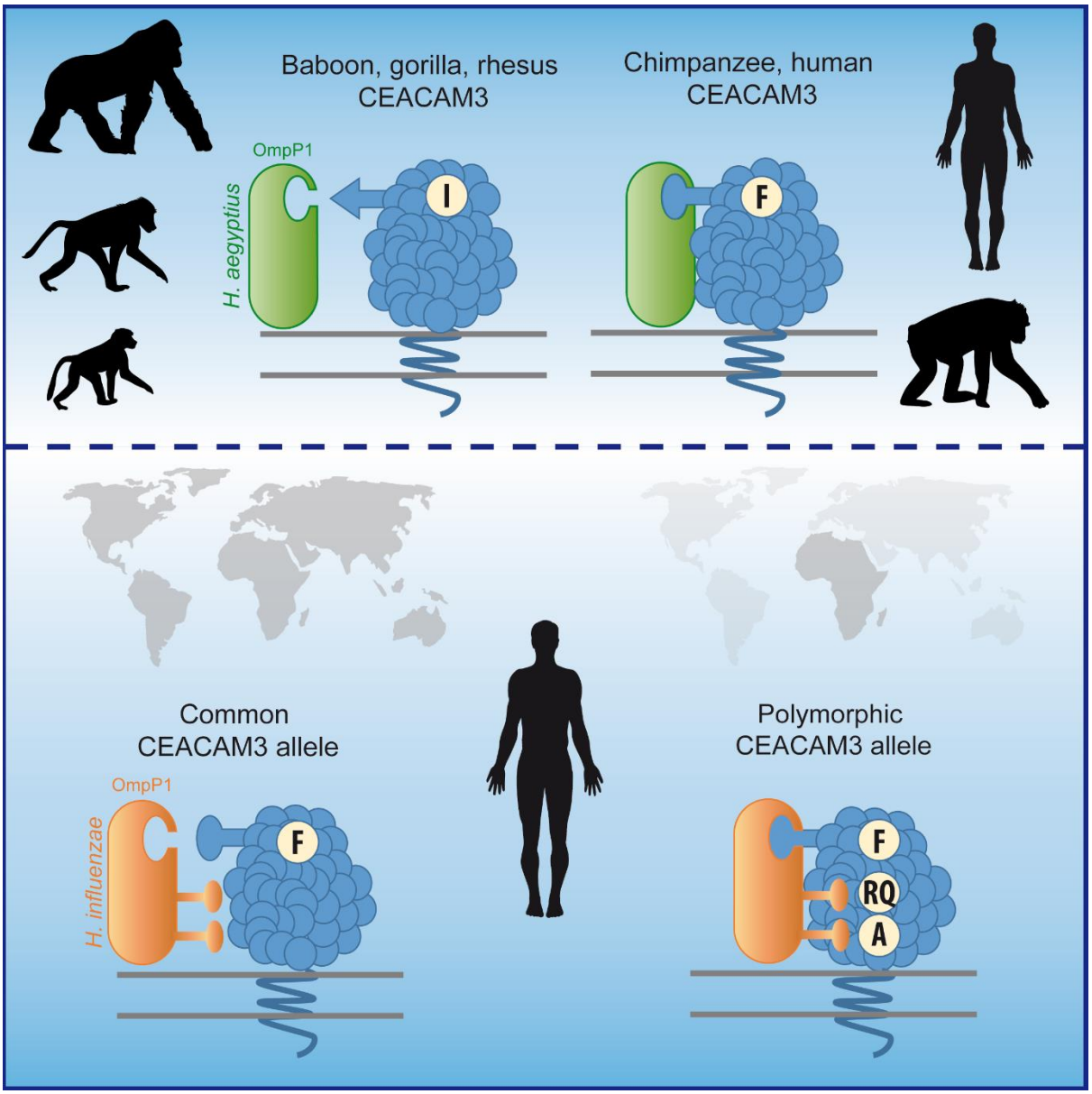


Figure II-7: Recognition of *Haemophilus aegyptius* and *Haemophilus influenzae* by CEACAM3 variants. (Upper) The Ig_V-like domain of CEACAM3 variants identified in baboon, gorilla and rhesus monkey are unable to recognize the OMP P1 adhesin of *Haemophilus aegyptius*, while chimpanzee and human CEACAM3 binds to OMP P1 expressing *Haemophilus aegyptius*. This difference in binding affinity is based on a phenylalanine residue at position 28 in the human and chimpanzee variant. (Lower) While the major human CEACAM3 allele lacks binding affinity to the OMP P1 adhesin of *Haemophilus influenzae* an allele variant, frequently found in African-ancestry populations, is able to mediate this interaction based on three amino acid exchanges.

given population could differ, with heterozygotes able to efficiently recognize a broader range of bacterial pathogens via granulocyte-expressed CEACAM3.

Accordingly, combining the common CEACAM3 allele with a second CEACAM3 allele, which at three positions mimics the CEACAM1 sequence and which occurs in about a third of the African population, could eliminate a blind spot in innate immune protection against CEACAM-binding microbes and could afford an advantage for people heterozygous at the CEACAM3 locus.

Together, our study uncovers a hitherto underappreciated immune receptor, CEACAM3, as one of the most rapidly evolving human genes. Amino acid alterations in this receptor are the basis for recognizing a structurally diverse set of adhesins from gram-negative, human-restricted bacteria. Importantly, CEACAM3 polymorphisms occurring in the human population expand the repertoire of recognized pathogen adhesins suggesting an ongoing arms race in line with the Red Queen hypothesis.

Methods

Experimental model and subject details

Cell lines and cell culture

Human embryonic kidney 293T cells (293T cells) were originally obtained from the German collection of microorganisms and cell cultures, DSMZ, Braunschweig, Germany (cell line number ACC-635). The cells were maintained in Dulbecco's modified Eagle's medium (DMEM) containing 10% calf serum and used between passages 15 and 35. The cell line was not further authenticated. HL60 cells were originally obtained from the American Type Culture Collection, ATCC, Manassas, VA (cell line number CCL-240). The cells were maintained in RPMI 1640 supplemented with 10% fetal calf serum and used between passages 10 and 30. The cell line was not further authenticated.

Bacterial strains and growth conditions

Non-piliated *N. gonorrhoeae* MS11-B2.1 strains expressing Opa_{52} (strain N309) or no Opa proteins (strain N302) were described previously (Schmitter, Agerer et al. 2004). Bacteria were grown at 37°C, 5% CO_2 on GC-Agar (Gibco BRL, Praisley, UK) supplemented with GC-vitamin mix (1:100) and appropriate antibiotics. GC-vitamin mix is prepared by solubilizing 20 g glucose, 2 g glutamine, 5.2 g L-cysteine, 20 mg cocarboxylase, 50 mg nicotinamide adenine dinucleotide (NAD), 4 mg $Fe(NO_3)_3$, 30 mg arginine, 0.6 mg vitamin B_1, 2 mg vitamin B_{12}, 2.6 mg p-aminobenzoic acid, 220 mg cystine, 200 mg adenine, 100 mg uracil, 6 mg guanine in 200 ml H_2O, pH adjusted to pH 7 with HCl. *Haemophilus influenzae* RD KW20 obtained from J. Reidl (Karl-Franzens University, Graz, Austria) and *Haemophilus aegyptius* ATCC11116 (DSMZ-German Collection of Microorganisms and Cell Cultures) were grown on brain-heart-infusion agar supplemented with Levinthal base (Tchoupa, Lichtenegger et al. 2015). Levinthal base was produced by adding 25 ml defibrillated horse blood to 250 ml autoclaved brain-heart-infusion (BHI) medium at ~60°C for 5 min and cooling to RT for 1 h. Levinthal base was cleared by centrifugation for 45 min at 4°C, 16,000 rpm and supplemented with sterile filtered 100 µg/ml NAD. *E. coli* strains expressing neisserial Opa proteins or *Haemophilus* OMP P1 have been used previously (Roth, Mattheis et al. 2013; Tchoupa, Lichtenegger et al. 2015; Rushmore, Allison et al. 2017). The UspA1 coding sequence was amplified from chromosomal DNA of *M. catarrhalis* (strain ATCC25238) using primers 5'

GCGCGCGCCCATGGATGAACAAAATTTATAAAGTGAAGAAAAATGCCGCAGGTC 3' and 5' CGCCCTCGAGCTATTTCCAGCGGTAACTGC 3'. The cDNA was inserted via NcoI/XhoI (restriction sites in primers underlined) into pET-28a (Novagen, Madison, Wisconsin). The pET28a vector encoding UspA1 was transformed in *E. coli* Rosetta(DE3) (Novagen, Madison, Wisconsin). Expression of Opa proteins, OMP $P1_{Hae}$, OMP $P1_{Hinf}$ or UspA1 in *E. coli* Rosetta(DE3) was induced by 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) for 3 h. All *E. coli* strains were cultured at 37°C in Lysogeny Broth (LB) supplemented with appropriate antibiotics.

Method details

Recombinant DNA

The pDNR-Dual LIC vector was created by modification of the pDNR-Dual vector (Clontech) by replacing the multiple cloning site with a corresponding LIC-site (primers 5' TCGACTCCTCCCCCGGGTTAGTGGGGCC 3' and 5' CCACTAACCCGGGGGAGGAG 3') via ApaI/SalI. The pLL3.7 LIC mKate2 vector was created by modification of pLL3.7 (Addgene #11795) by replacing the multiple cloning site with a corresponding LIC-site (primers 5' CTAGCGACTCTCCCCCGGGTTAGTGGGGCA 3' and 5' CCGGTGCCCCACTAACCCGGGGGAGAGTCG 3') via NheI/AgeI. The CEACAM3 Ig_V-like domain of *Pan troglodytes*, *Gorilla gorilla*, *Macaca mulatta* and *Papio anubis* were custom synthesized (Eurofins Scientific) based on published sequences (see Table S1). The coding sequences were amplified using species-specific primers (see Table S2). The CEACAM coding sequence was inserted via Ligation-Independent-Cloning (LIC) into pDNR-Dual LIC. The cDNAs were subsequently transferred to pLPS3'EGFP by Cre-mediated recombination to yield primate-CEACAM3-NT-EGFP. Alterations in human or gorilla CEACAM3s were introduced by Splicing by Overlapping Extension (SOEing) PCR using universal LIC-based CEACAM3 Ig_V-forward and reverse primer together with a mutation specific primer set (see Table S2). After SOEing PCR transfer into pLPS3'EGFP from pDNR-Dual LIC was performed as described above. Gorilla and hu-gorilla full length receptors were produced by SOEing PCR combining gorilla/hu-gorilla Ig_V-like domains and human CEACAM3 C-terminal domain that is identical to the gorilla WT receptor on the amino acid level (see Table S2). After SOEing PCR, cDNA was transferred via LIC into either pDNR-Dual LIC and subsequently into pLPS3'EGFP or into pLL3.7 mKate2-LIC to create primate-CEACAM3-mKate2.

Transfection of 293T cells

Transfection of 293T cells with pLPS3'EGFP primate CEACAM3 was performed by calcium phosphate co-precipitation using 5 µg of plasmid DNA for each 10 cm culture dish at ~20% confluence. Transfection solution was produced by adding 5 µg of plasmid DNA to 500 µl H_2O before adding 500 µl 2x HBS buffer (274 mM NaCl, 42 mM HEPES, 1.4 mM Na_2HPO_4, pH 7.05) and 50 µl 2.5 M $CaCl_2$.

Lentivirus production and generation of stable cell lines

293T cells were transfected with 7 µg pMD2.G, 10 µg psPAX2 and 13 µg pLL3.7 (harboring the corresponding mKate2-tagged primate CEACAM3 construct) by calcium phosphate co-precipitation. After 72 h the virus-containing supernatant was cleared by centrifugation and sterile filtered with a 0.45 µm syringe filter. Spinfection was performed on batches of $1x10^6$ HL60 cells in 1 ml culture medium by addition of 1 ml virus containing supernatant with 8 µg/ml hexadimethrine bromide and subsequent centrifugation at 800 xg for 1 h at RT. Positive clonal cell lines were obtained after flow cytometric single cell sorting of cells showing expression of the red fluorescent protein mKate2.

Binding assay with soluble CEACAMs

Soluble recombinant GFP-tagged CEACAM-Ig_V proteins were produced in HEK293T cells. 24 h after transfection, culture medium was replaced by OptiMEM (Gibco BRL). Culture supernatants were collected 4 days after transfection and purified from cell debris by centrifugation (2500 xg, 4°C, 10 min). Supernatants were adjusted to equal levels of soluble CEACAMs and monomeric CEACAM-GFP fusion proteins were clustered with polyclonal anti-GFP antibodies (tag-tools GmbH, Konstanz, Germany) overnight at 4°C. Then $1x10^8$ bacteria were incubated with 1 ml supernatant containing the indicated GFP-fusion proteins for 1 h at 37°C under gentle rotation. Next, bacteria were washed twice with 1x phosphate-buffered saline (PBS) and samples were processed for either immunoblotting analysis with GFP antibodies or flow cytometry to detect bacteria-associated fluorescence.

Gentamicin protection assay

One day after transfection, 293T cells were seeded in individual wells of 24-well plates ($5x10^5$ cells/well) coated with gelatine. Next day, cells were infected with *E. coli* or *E. coli* expressing Omp $P1_{Hae}$ at a multiplicity of infection (MOI) of 100. 2 h after infection, extracellular bacteria were killed by 1 h incubation with 100 µg/ml gentamicin in cell culture medium. Internalized bacteria were released by cell lysis with 0.5% (w/v) saponin in PBS for 15 min and viable bacteria were enumerated after plating in suitable dilutions. For each sample three technical replicates were enumerated. Statistical evaluation of three independent experiments was performed via two-tailed t-test.

Analysis of bacterial invasion by flow cytometry

Prior to infection, bacteria were suspended in 1 ml PBS, 2 µg carboxyfluorescein-succinimidylester (CFSE) was added and bacteria were labeled for 20 min, at 37°C under constant shaking. The residual staining solution was removed by centrifugation (5000 rpm, 5 min) and three times washing with PBS.

$1x10^6$ HL60 cells were suspended in 1 ml phagocytosis buffer (PBS, 0.9 mM $CaCl_2$, 0.5 mM $MgCl_2$, 5 mM glucose, 1% heat inactivated fetal calf serum) and infected with fluorescein-labelled bacteria for 15 min at 37°C under gentle rotation. Infection was stopped by washing once with ice cold PBS. By addition of trypan blue to a final concentration of 0.2 mg/ml intracellular bacteria were selectively measured by quenching extracellular fluorescein-signal (Pils, Schmitter et al. 2006). Raw data were processed with FlowJo.

Immunofluorescence staining for confocal microscopy

One day after transfection, 293T cells ($4x10^4$ cells/well) were seeded on gelatin-coated glass coverslips in 24-well plates and cultured for one day. Prior to infection, bacteria were labelled with biotin- and rhodamine-succinimidyl ester (Sulfo-NHS-SS-Biotin / TAMRA-SE) as described above. Cells were infected for 1 h at 37°C at a MOI of 30. After infection, the cells were fixed for 20 min in 3% paraformaldehyde in PBS at RT and subsequently washed three times with PBS. Extracellular bacteria were stained with streptavidin-Cy5. As a result, extracellular bacteria were marked by both rhodamine and Cy5, while intracellular bacteria were only stained with Cy5.

Immunoblotting and antibodies

Westernblot was performed as described earlier (Hauck, Hunter et al. 2001) using a monoclonal antibody (mAB) against GFP (clone JL-8, Clonetech, Palo Alto, CA), mAB against tubulin (clone E7, DSHB, University of Iowa), and polyclonal antibodies against mKate2 produced in the local animal facility at the University of Konstanz. All peroxidase and fluorescence-labeled secondary antibodies were purchased from Jackson ImmunoResearch (West Grove, PA).

Identification of primate CEACAM3 orthologs

Bioinformatic identification of primate CEACAM3 orthologs was performed by alignment of the human CEACAM3 protein sequence with genomes listed in Table S3 using NCBI BLAST. Matching regions were sorted based on their genomic positions and subsequently aligned to the human CEACAM3 gene sequence to allow mapping of intron-exon boundaries. By taking into account overall domain structure, domain order, exon length and intron length, this procedure permits reliable differentiation between human CEACAM-family members and is therefore also able to prevent misidentification of other primate CEACAMs. Genomes containing all human CEACAM3 derived exons in the correct order and spacing as well as encoding a protein with a cytoplasmic HemITAM motif (Buntru, Roth et al. 2012) were assumed to encode a CEACAM3 ortholog. Futhermore, for all genomes, where a CEACAM3 orthologs gene could be identified, we investigated gene synteny of the CEACAM locus and could detect the orthologs CEACAM3 gene placed in the same position and identical transcriptional orientation relative to the CEACAM1 and CEACAM4 orthologs (Figure II-S1).

Analysis of human CEACAM3 polymorphisms

High abundance CEACAM3 polymorphisms were located using the 1000 genomes project browser Variation table. We focused on missense variants encoded by SNPs with a global minor allele Frequency >0.05 over all 1000 Genome Phase I populations. Analysis with the 1000 Genome browser population genetics module permitted quantification of polymorphism occurrence as well as homozygote/heterozygote diversity by continent and continent-derived populations.

CEACAM enrichment from cell lysates

To enrich Opa_{52} binding CEACAMs from HL60 cell lysates, *Neisseria gonorrhoeae* expressing Opa_{52} were harvested from an agar plate and resuspended in a fixation solution (4% PFA in PBS) for 20 minutes. After three washing steps with PBS, ~$2.5x10^9$ fixed bacteria were added to 1 ml HL60 cell lysate (~$5x10^7$ cells) and rotated at 4°C for 1 h. *Neisseria* were collected by centrifugation at 5000 rpm for 5 min at 4°C and washed twice with PBS.

Structure modeling of CEACAM3 Ig_V-like domains

CEACAM3 protein sequences were modeled on the crystal structure of the N-terminal domain of human CEACAM3 (6AW1) (Bonsor, Zhao et al. 2018) utilizing "SWISS MODEL" automated protein structure homology-modeling (Peitsch 1996). Hydrophobicity and charge mapping was performed using a previously published YRB-Script (Hagemans, van Belzen et al. 2015).

Quantification and statistical analysis

The internalization of adhesin expressing bacteria by HEK cells expressing CEACAM constructs was quantified by counting of colony forming units in three technical replicates. Average-data from three independent experiments, each performed in triplicate, were compared using the two-tailed t-test in GraphPad Prism 5.0.1. A significance threshold was set at $p < 0.05$. See figures and figure legends for the statistical details.

Key resources table

Reagent or Resource	Source	Identifier
Antibodies		
Mouse monocolonal anti-GFP JL-8	BD Biosciences	AB_2314359
HRP-conjugated Goat anti-mouse	Jackson ImmunoResearch	AB_2307392
Mouse monoclonal Anti-Tubulin E7	DSHB, University Iowa	AB_528499
Mouse monoclonal Anti-pan-CEACAM	Aldevron	Cat# GM-05-05
Rabbit polyclonal Anti-mKate2	University of Constance	N/A
Mouse monoclonal Anti-PTyr PY72	Upstate Biotechnology	PY72AB_448291
Mouse monoclonal Anti-Opa 4B12/C11	(Achtman, Neibert et al. 1988)	4B12/C11
Rabbit polyclonal Anti-GFP-tag	tag-tools	N/A
HRP-conjugated Goat Anti-rabbit	Jackson ImmunoResearch	AB_2313567
Streptavidin, Alexa Fluor® 647 conjugate antibody	Thermo Fisher Scientific	AB_2336066
Bacterial and Virus Strains		
Non-piliated *N.gonorrhoeae* MS11-B2.1 Opa_{52}	(Makino, van Putten et al. 1991)	N309
Non-piliated *N.gonorrhoeae* MS11-B2.1 Opa^-	(Makino, van Putten et al. 1991)	N302
Haemophilus influenzae	A.Wright (Tufts University, Boston, USA)	RD KW20
Haemophilus aegyptius	DSMZ-German Collection of Microorganisms and Cell Cultures	ATCC11116
Escherichia coli pTrc99A *Opa_{52}*	(Kupsch, Knepper et al. 1993)	N/A
Escherichia coli Rosetta(DE3) *pet28a OMP P1-Hinf*	(Tchoupa, Lichtenegger et al. 2015)	N/A
Escherichia coli Rosetta(DE3) *pet28a OMP P1-Hae*	(Tchoupa, Lichtenegger et al. 2015)	N/A
Escherichia coli Rosetta(DE3) *pet28a UspA1*	This Paper	N/A
Escherichia coli Rosetta(DE3) *pet28a Opa_{65}*	(Roth, Mattheis et al. 2013)	N/A
Escherichia coli Rosetta(DE3) *pet28a Opa_{67}*	(Roth, Mattheis et al. 2013)	N/A

Escherichia coli Rosetta(DE3) *pet28a Opa*$_{68}$	(Roth, Mattheis et al. 2013)	N/A
Escherichia coli Rosetta(DE3) *pet28a Opa*$_{69}$	(Roth, Mattheis et al. 2013)	N/A
Escherichia coli Rosetta(DE3) *pet28a Opa*$_{70}$	(Roth, Mattheis et al. 2013)	N/A
Escherichia coli Rosetta(DE3) *pet28a Opa*$_{71}$	(Roth, Mattheis et al. 2013)	N/A
Escherichia coli Rosetta(DE3) *pet28a Opa*$_{72}$	(Roth, Mattheis et al. 2013)	N/A
Escherichia coli Rosetta(DE3) *pet28a Opa*$_{73}$	(Roth, Mattheis et al. 2013)	N/A
Escherichia coli Rosetta(DE3) *pet28a Opa*$_{74}$	(Roth, Mattheis et al. 2013)	N/A
Escherichia coli Rosetta(DE3) *pet28a Opa*$_{75}$	(Roth, Mattheis et al. 2013)	N/A
Chemicals, Peptides, and Recombinant Proteins		
EZ-Link™ Sulfo-NHS-SS-Biotin	Thermo Fisher Scientific	Cat# 21331
Trypan blue	AppliChem	Cat# A0668
IPTG	Roth	Cat# 206-703-0
gentamicinsulfate	AppliChem	Cat# A1492
saponin	AppliChem	Cat# A2542
Carboxyfluorescein SE (CFSE)	Molecular Probes	Cat# C1157
TAMRA-SE	Molecular Probes	Cat# C1171
GC-Agar	Becton Dickinson	Cat# 228950
BHI-Medium	Becton Dickinson	Cat# 237500
Agar	AppliChem	Cat# A2113
NcoI-FD	Thermo scientific	Cat# FD0574
XhoI-FD	Thermo scientific	Cat# FD0694
NheI-FD	Thermo scientific	Cat# FD0973
AgeI-FD	Thermo scientific	Cat# FD1464
hexadimethrine bromide	Sigma-Aldrich	Cat# 107689
OptiMEM	Gibco	Cat# 11058-021
DMEM	Merck	Cat# FG0435
RPMI	Merck	Cat# FG1215
Calf serum	Biochrom	Cat# S 0125
Fetal calf serum	Biochrom	Cat# S 0115

Deposited Data		
Data for primate genome analysis see Table S1		
SNP: T69P	1000 Genomes Project	rs61737014 http://phase3browser.1000genomes.org/Homo_sapiens/Variation/Explore?source=dbSNP;v=rs61737014;vdb=variation
SNP: V49A	1000 Genomes Project	rs61737019 http://phase3browser.1000genomes.org/Homo_sapiens/Variation/Explore?source=dbSNP;v=rs61737019;vdb=variation
SNP: S43R	1000 Genomes Project	rs61738270 http://phase3browser.1000genomes.org/Homo_sapiens/Variation/Explore?source=dbSNP;v=rs61738270;vdb=variation
SNP: L44Q	1000 Genomes Project	rs61738269 http://phase3browser.1000genomes.org/Homo_sapiens/Variation/Explore?source=dbSNP;v=rs61738269;vdb=variation
Hapbin-dataset	Hapbin-dataset	*_chr19.bg.gz https://datashare.is.ed.ac.uk/handle/10283/714

Experimental Models: Cell Lines		
HEK293T	DSMZ	RRID:CVCL_0063
HL60	DSMZ	RRID:CVCL_0002
HL60 CEACAM3-mKate2	This study	
HL60 CEACAM3-mKate2 Cerulean	This study	
HL60 CEACAM3-mKate2 Cerulean sgCerulean	This study	
Oligonucleotides		
Primers for PCR, see Table S2	This paper	
Recombinant DNA		
pDNR-dual	Clonetech	N/A
pLPS3'EGFP	Clonetech	N/A
pLL3.7	Addgene	#11795
Software and Algorithms		
Swiss model	(Waterhouse, Bertoni et al. 2018)	N/A
SNAP	(Korber, Muldoon et al. 2000)	v2.1.1
Image Lab	BioRad	5.2.1 build 11
YRB-script	(Hagemans, van Belzen et al. 2015)	DHS065
FlowJo	FlowJo LLC	v10.0.7

Supplementary information

Supplementary Figures

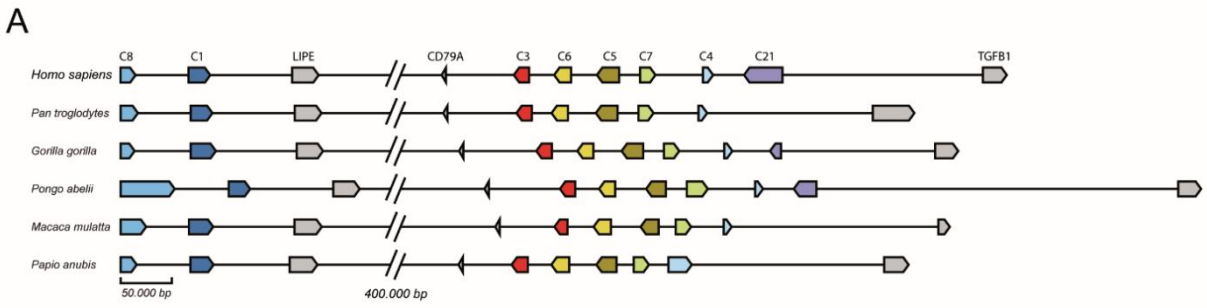


Figure II-S1: Syntenic relationship of the CEACAM gene locus in higher primates. Related to Figure II-1

(A) Comparative analysis of gene synteny in the CEACAM locus in all species for which a CEACAM3 ortholog was identified. Depicted are the CEACAM-encoding loci (colorized) CEACAM8 (C8), CEACAM1 (C1), CEACAM3 (C3), CEACAM6 (C6), CEACAM5 (C5), CEACAM7 (C7), CEACAM4 (C4), CEACAM21 (C21) and the Non-CEACAM marker loci (grey) encoding lipaseE (LIPE), CD79A and transforming growthfactor b1 (TGFB1). Transcriptional gene orientation is indicated by arrowheads. All identified CEACAM3 (C3, red) orthologs are located between the syntenic CEACAM6 (C6) and CD79A encoding loci and the transcriptional gene orientation of all shown CEACAM family member is conserved within the indicated species. The observed synteny supports the proper identification of CEACAM3 orthologs in higher primates.

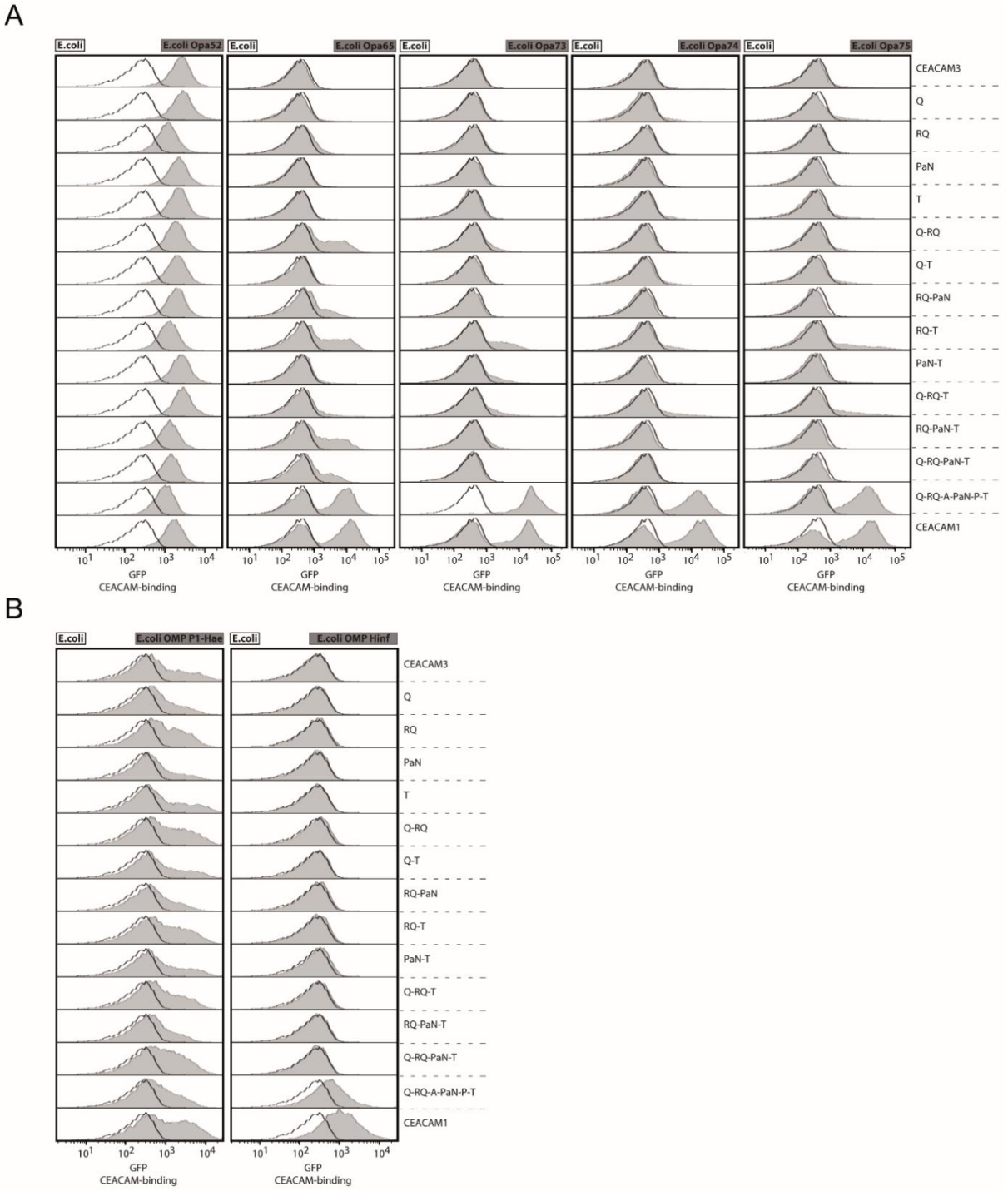


Figure II-S2: Pathogen adhesins discriminate between CEACAM3 and CEACAM1 based on specific sequence motifs. Related to Figure II-5

(A, B) Soluble GFP-tagged CEACAM3 variants (described in Figure II-5 C) were incubated with *E. coli* expressing the indicated adhesins originating from *N. gonorrhoeae* (A), *H. influenzae* or *H. aegyptius* (B) respectively. CEACAM binding was determined via bacterial-pulldown and subsequent flow cytometry.

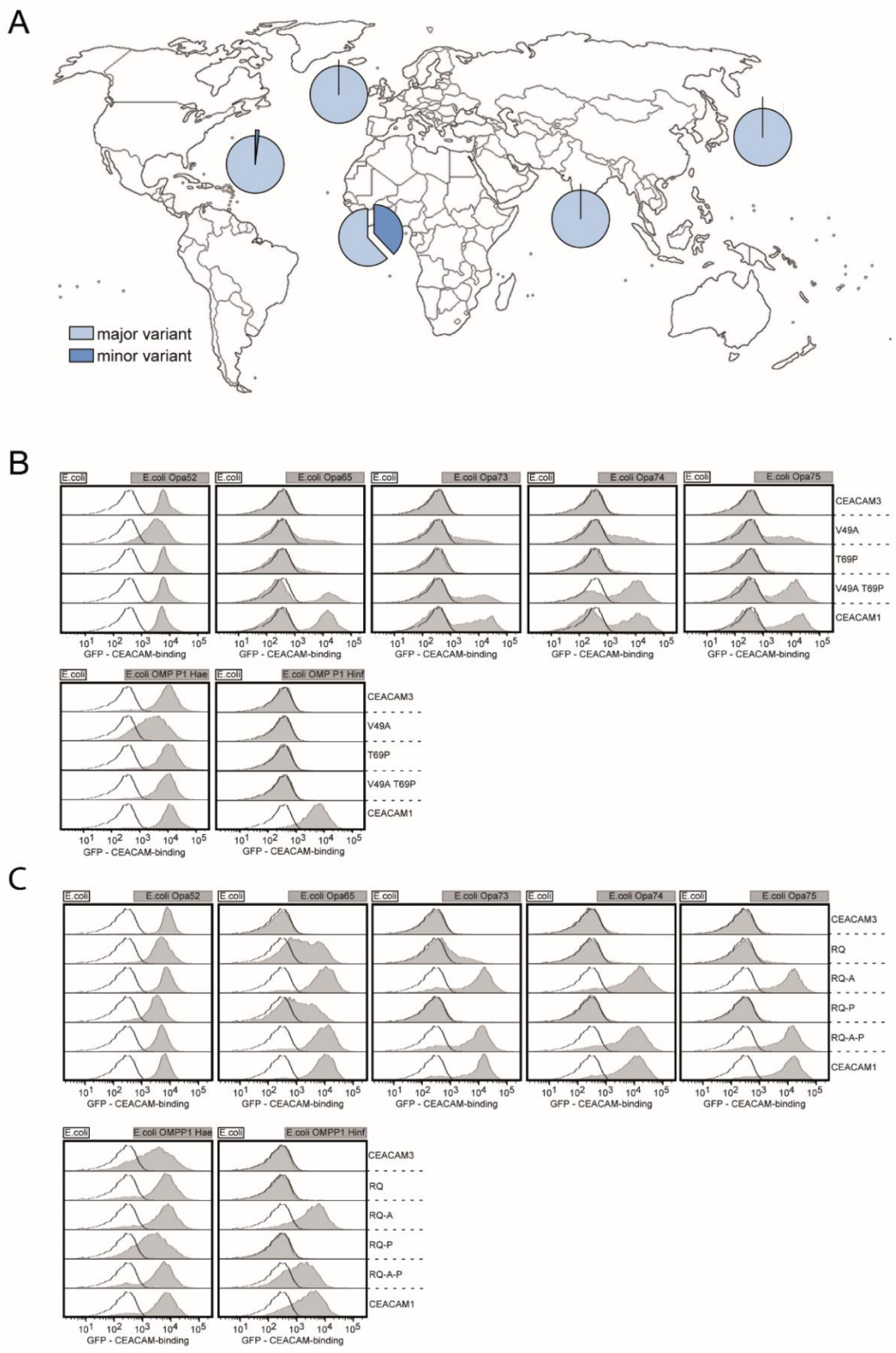


Figure II-S3: Human CEACAM3 polymorphisms broaden the spectrum of recognized adhesins. Related to Figure II-6

(A) World-scale map. Pie charts indicate the abundance of individuals possessing at least one allele (darker) or no allele (lighter) of the minor variant. On a global scale, the CEACAM3 V49A or T69P SNP-containing variants are found nearly exclusively in African populations or derived

populations (21% allele abundance over all African subpopulations) and are absent from sequences of European and Asian populations. (B, C) Soluble GFP-tagged CEACAM3 variants (described in Figure II-6 E and F) were incubated with *E. coli* expressing the indicated adhesins originating from *N. gonorrhoeae*, *H. influenzae* or *H. aegyptius* respectively. Adressing the effects of the residues V49A and T69P alone (B) or in combination with the SL43/44RQ SNP (C) CEACAM binding was determined via bacterial-pulldown and subsequent flow cytometry.

Chapter III

CEACAM-binding yeasts are recognized by the innate immune receptor CEACAM3

Patrizia Bonsignore[1], Jonas Adrian[1], Christof R. Hauck[1,2]

[1] Lehrstuhl für Zellbiologie, Universität Konstanz, 78457 Konstanz, Germany

[2] Konstanz Research School - Chemical Biology, Universität Konstanz, 78457 Konstanz, Germany

Abstract

The human mucosa is colonized by a diverse microbial community comprising not only prokaryotes and viruses but also eukaryotes such as protozoa, parasites and fungi. Among those fungi the pathobiontic yeast *Candida albicans* plays a remarkable role. *C. albicans* constitutes a serious threat to the health of immunocompromised patients as it can breach the mucosal barrier and cause local as well as systemic infections, known as Candidiasis. The immune system of a healthy host renders those tissue and blood intruding yeasts harmless via efficient phagocyte mediated clearance. This is accomplished by specialized phagocytic receptors on the cell surface such as Dectin-1 which recognizes fungal β-glucans and triggers the uptake and elimination of phagocyte-engaged fungi. A recent study implicated another innate immune receptor to bind *C. albicans*: the carcinoembryonic antigen-related cell adhesion molecule 3 (CEACAM3). CEACAM3 is exclusively expressed on granulocytes and belongs to the CEACAM family which also comprises epithelial members. This family has been extensively studied in its role as receptors for pathogenic bacteria such as *Neisseria* or *Haemophilus*. Microbial attachment to epithelial CEACAMs facilitates colonization of the host epithelium, while recognition by CEACAM3, results in phagocytosis of tissue infiltrating CEACAM-binding microbes. In this study we explore the effect of yeast-recognition by CEACAM3 and its contribution to host defence during potential Candidiasis. Remarkably, CEACAM interaction is depended on growth conditions of *Candida* such as iron limitation or the presence of serum, as are commonly experienced by tissue infiltrating microbes. Binding assays revealed that CEACAM3 is able to recognize a broad range of fungi, which might underline the exceptional protective spectrum of this immune receptor. Interestingly, *Candida* recognition can only be accomplished by human CEACAM3 and the closely related chimpanzee CEACAM3 but is absent in other primate orthologs hinting to a more recent evolutionary occurrence. Still, the proteinaceous component of *Candida* involved in CEACAM-binding remains elusive. Presumably via this adhesin *C. albicans* activates the innate immune receptor CEACAM3 and causes intracellular receptor tyrosine phosphorylation but does not result in CEACAM3-dependent uptake. Together, our results implicate CEACAM3 in the recognition and thereby activation of downstream signalling following yeast-binding to host phagocytes.

Introduction

Fungi such as *Saccharomyces*, *Aspergillus*, or *Candida* colonize the human body as commensals including skin and mucosal surfaces. Especially the gastrointestinal tract is inhabited by a variety of fungal species. However, this group of commensals also contains opportunistic pathogens (Richard and Sokol 2019). Certain environmental, health or therapeutic circumstances such as antibiotic therapy, or indirect, immune system compromising factors of the host can favour the induction of pathogenic fungal characteristics. A prime example is *Candida albicans* which readily switches from its commensal yeast cell phenotype to an aggressive filamentous form allowing *C. albicans* to breach the intestinal epithelial barrier (Kim and Sudbery 2011). Invasion of the surrounding tissue causes fatal local infections. By extension into the bloodstream, *C. albicans* can rapidly spread and infect multiple organs causing life-threatening systemic infections. To counteract this potentially fatal dissemination, specialised phagocytic immune cells engage tissue and blood intruding microbes, rendering them harmless via phagocytosis (Qin, Zhang et al. 2016). The prerequisite for this defence is high affinity recognition of the invader. Therefore phagocytes possess specialized receptors such as the membrane bound C-type lectin receptors (CLECs). These pattern recognition receptors target conserved microbial structures. The innate immune receptor Dectin-1 of the CLEC family targets invading fungi by recognition of β-glucan structures in their fungal cell wall. Dectin-1 can initiate elimination of microbes by inducing phagocytosis and intracellular killing (Schorey and Lawrence 2008).

A recent study has indicated that an additional, more specialized receptor aiming for a certain adhesin instead of a universal surface pattern could also contribute to the defence against invading fungi: the carcinoembryonic antigen-related cell adhesion molecule 3 (CEACAM3) (Klaile, Muller et al. 2017). CEACAM3 is a human glycoprotein, exclusively expressed on polymorphonuclear granulocytes (PMNs, neutrophils) and belongs to the carcinoembryonic (CEA) family, which is part of the immunoglobulin superfamily. The CEA family subdivides in two groups, the pregnancy specific glycoproteins (PSGs) and the CEA cell adhesion molecules (CEACAMs). In addition to numerous physiological functions CEACAM1, CEACAM3, CEA (CEACAM5), and CEACAM6 are also implicated in a pathophysiological context. These cell surface molecules serve as docking site for several human restricted pathogens such as *Neisseria gonorrhoea*, *Moraxella catarrhalis, Helicobacter pylori*, or *Haemophilus influenzae*.

These pathogens express unique adhesins to interact with their molecular target: the amino-terminal immunoglobulin variable (Ig_V)-like domain of CEACAM molecules (Chen and Gotschlich 1996; Hill and Virji 2003; Tchoupa, Lichtenegger et al. 2015; Koniger, Holsten et al. 2016). CEACAM1, CEA, and CEACAM6 are exposed on the apical surface of epithelial cells and utilized by CEACAM-binding microbes to enhance mucosal colonization. In addition, CEACAM engagement also permits pathogens to suppress innate immune functions, such as the mucosal exfoliation response (Muenzner, Rohde et al. 2005; Muenzner, Kengmo Tchoupa et al. 2016). Remarkably, CEACAM-binding pathogens such as *Neisseria meningitidis* (bacterial meningitis) or *Helicobacter pylori* (chronic gastritis, stomach cancer) inhabit large parts of the healthy population, rarely leading to severe diseases. The cause is found in the granulocyte-restricted innate immune receptor CEACAM3. The extracellular Ig_V-like domain of CEACAM3 exhibits high sequence congruence with the pathogen target CEACAMs CEACAM1, CEA, and CEACAM6 and therefore functions as bait for CEACAM-exploiting microorganisms. CEACAM3 engages tissue penetrating microbes via their CEACAM interacting adhesins and triggers rapid opsonin-independent phagocytosis and stimulates microbicidal pathways of granulocytes (Schmitter, Agerer et al. 2004; Sarantis and Gray-Owen 2007). CEACAM3 engagement leads to receptor-clustering and subsequent phosphorylation of its intracellular immunoreceptor tyrosine-based activation motif (ITAM)-like sequence. In contrast to a canonical ITAM sequence ($YxxL/Ix_{(6-12)}YxxL/I$) found in immune signalling receptors, such as T-cell receptors or Fcγ-receptors, CEACAM3 possess an HemITAM ($YxxLx_{(7)}YxxM$) sequence such as the phagocytic innate immune receptor Dectin-1 (Buntru, Roth et al. 2012). While CEACAM3 expression is restricted to neutrophils, Dectin-1 is expressed by various myeloid cells and serves as pattern-recognition receptor for β-glucans derived from fungal cell walls. Remarkably, both receptors share striking similarities. Like the immune-signalling downstream of CEACAM3, detection of β-glucan by Dectin-1 leads to phagocytosis of engaged particles, the production of reactive oxygen species (ROS), and NFκB-mediated release of proinflammatory cytokines. A similar immune response can be observed for CEACAM3-activation. For both receptors phosphorylation of the membrane proximal tyrosine residue within their HemITAM sequence is essential and can initiate an all-embracing immune response. Phosphorylation of the membrane distal tyrosine residue, however, is dispensable for signal transduction (Rogers, Slack et al. 2005; Fuller, Williams et al. 2007).

Candida albicans specifically binds to human CEACAMs but not to other mammalian CEACAM family members (e.g. mouse, rat). It directly interacts with the amino-terminal Ig_V-like domain of human CEACAM1, CEACAM3, CEA, and CEACAM6; the absence of Ig_C-like domains has no effect, in contrast, depletion of the Ig_V-like domain abrogates binding. The CEACAM-binding adhesin of *Candida albicans* is currently unknown (Klaile, Muller et al. 2017).

In this study we investigate how CEACAM3 might contribute to the elimination of tissue-infiltrating CEACAM-engaging pathogenic yeast and by which molecular factor *C. albicans* is able to engage with epithelial and neutrophil-restricted CEACAM family members. We reveal several environmental factors that strongly alter the affinity of *C. albicans* to CEACAM-receptors, and explore the relevance of its different organismal states to this process. Further, we show that human CEACAM3 gets activated in response to *C. albicans* recognition and that this recognition is highly specific.

Results

The innate immune receptor CEACAM3 recognizes different yeast species

The hematopoietic receptor Dectin-1 plays an important role in antifungal innate immunity, as it mediates immune cell interaction with β-glucan structures of the fungal cell wall and thereby triggers phagocytosis and related responses. Interestingly, previous studies observed an interaction of the pathobiontic yeast *Candida albicans* and *Candida glabrata* with extracellular domains of CEACAM family members including the granulocyte-specific immune receptor CEACAM3. CEACAM3 is known to trigger opsonin-independent phagocytosis of bound particles and induce immune response after clustering. In this study we aspire to determine the potential role of CEACAM3 in the recognition of CEACAM binding yeast cells and its capability to trigger associated immune responses. Therefore, CEACAM-binding analysis was conducted using soluble Ig_V-like amino-terminal (NT) domains of human CEACAMs. Remarkably, not only *Candida albicans* and its close relative *Candida dubliniensis*, but also the commensal yeast species *Saccharomyces cerevisiae* and the laboratory model yeast species *Pichia pastoris* were able to interact with epithelial CEA as well as the innate immune receptor CEACAM3. In contrast, none of the tested yeast species was able to bind the presumed innate immune receptor CEACAM4 (Figure III-1 A). The CEACAM binding pattern consistency between the yeast strains suggests a common conserved surface structure which is responsible for the interaction. Environmental changes such as high temperature, pH conditions, iron starvation or the presence of serum are known to modulate *Candida* gene expression patterns and can therefore be utilized to uncover interaction determining factors. Those environmental factors are also known to turn the commensal fungus into its virulent form by triggering yeast-to-hyphae transition (Si, Hernday et al. 2013) therefore all conditions were tested with the filament formation-deficient *C. albicans* strain Can34. Clearly, preculture in alkaline medium or under iron starvation decreases, or even completely abrogates CEACAM binding whereas higher temperatures do not alter the binding behaviour (Figure III-1 B). *C. albicans* grown in fetal calf serum (FCS)-enriched medium showed a concentration dependent increase in CEACAM association hinting to an enhanced presence of the CEACAM-binding protein on the fungal surface (Figure III-1 C).

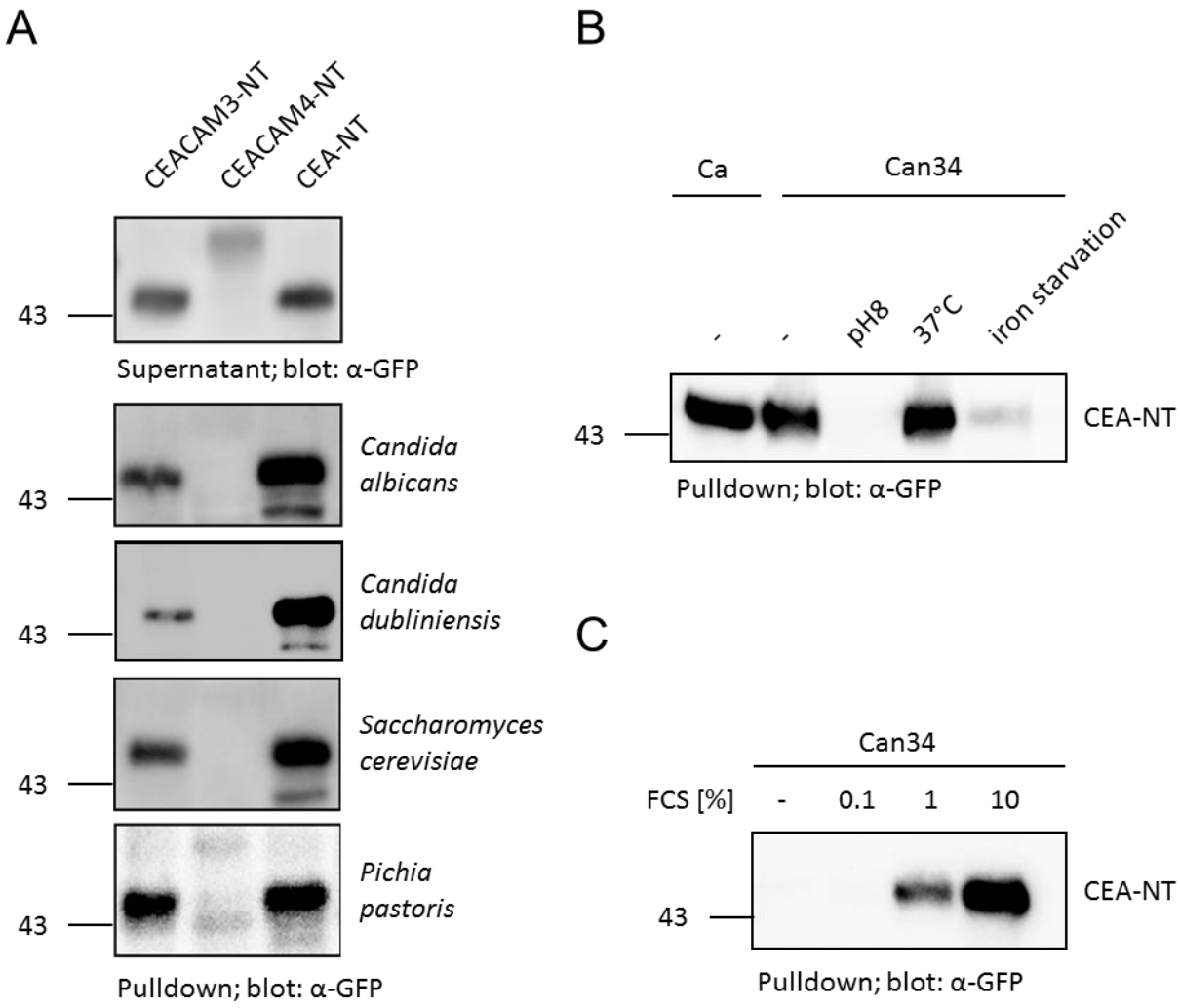


Figure III-1: Recognition of human carcinoembryonic antigen-related cell adhesion molecules (CEACAMs) by various yeast species in a growth condition dependent manner. (A) GFP-tagged amino-terminal (NT) domains of human CEACAM3, CEACAM4, and CEA were generated as secreted proteins in HEK293T cells. Equivalent amounts of the different soluble N-terminal domains (supernatant – topmost panel) were incubated with *C. albicans*, *C. dubliniensis*, *S. cerevisiae*, or *P. pastoris*. Bound CEACAM-Ig_V domains were recovered in a yeast pulldown and evaluated via immunoblotting analysis with anti-GFP antibodies. (B, C) Receptor binding assays as described in (A) were performed with accordingly precultured (3 h) *C. albicans* (Ca: wild-type strain; Can34: filament formation-deficient strain). The impact of the respective growth conditions on CEACAM-association was evaluated via immunoblotting analysis.

***C. albicans* morphological and mating type variation has no impact on CEACAM-binding**

C. albicans is a polymorph organism which readily switches from its commensal yeast cell phenotype to a virulent hyphal phenotype and also can switch between different mating phases: a white phase and an opaque phase (Slutsky, Staebell et al. 1987). The white-phase phenotype describes a spherical yeast cell type that is able to induce hyphae formation and appears as white hemispherical colonies when grown on solid growth medium. The opaque-phase phenotype represents an elongated mating-competent form of *C. albicans* and exhibits grey flat colonies (Miller and Johnson 2002; Sasse, Hasenberg et al. 2013). To test if CEACAM-interaction is associated with a certain phenotype of *C. albicans*, binding studies were performed utilizing yeast cells and hyphae (Figure III-2 A) as well as opaque and white phase cells (Figure III-2 B). *C. albicans* was grown overnight under varying culture conditions comprising different temperatures, presence or absence of serum and iron. Hyphae were induced in conditioned cell culture medium. Binding-studies with yeast cells or hyphae indicated no significant difference in CEACAM interaction. However, in contrast to hyphae, commensal yeast cells provided an additional band of a lower molecular mass. The commensal form of *C. albicans* seems to process associated CEACAM molecules (Figure III-2 A). CEACAM Ig_V-like domains are highly glycosylated, therefore the lower band observed could result from sugar degradation or partial cleavage of the N-terminus by fungal proteases. However, elucidation of processing disparities is not object of this study. Interestingly, iron starvation seems to severely compromise CEACAM-binding of hyphae but not yeast type *C. albican*s (Figure III-2 A).

The switch between white and opaque phase occurs spontaneously in *C. albicans*. However, certain factors favour one phase over another. For instance, temperature drives switching between the two different phases. Elevated (37°C) and low (4°C) temperatures promote the white morphology whereas room temperatures (20-25°C) stabilize the opaque phase (Rikkerink, Magee et al. 1988). Yeast cells stored at 4°C for one week generated a predominantly white phase culture. Conversely, overnight culture at RT results in a predominantly opaque cell culture. Fresh culture medium was inoculated with white or opaque yeast cells and grown for 3 h at 37°C or RT, respectively. Before yeast cells were applied for binding-assays, their phenotype was microscopically confirmed (Figure III-2 B). Both phenotypes showed a comparable CEACAM binding spectrum. White and opaque cells could both bind to CEACAM1, CEACAM3 and CEA. The opaque cells showed a slightly elevated

interaction with CEACAM3 compared to the white phenotype. The CEACAM-binding adhesin of *Candida albicans* seems to be regulated in a phenotype independent manner.

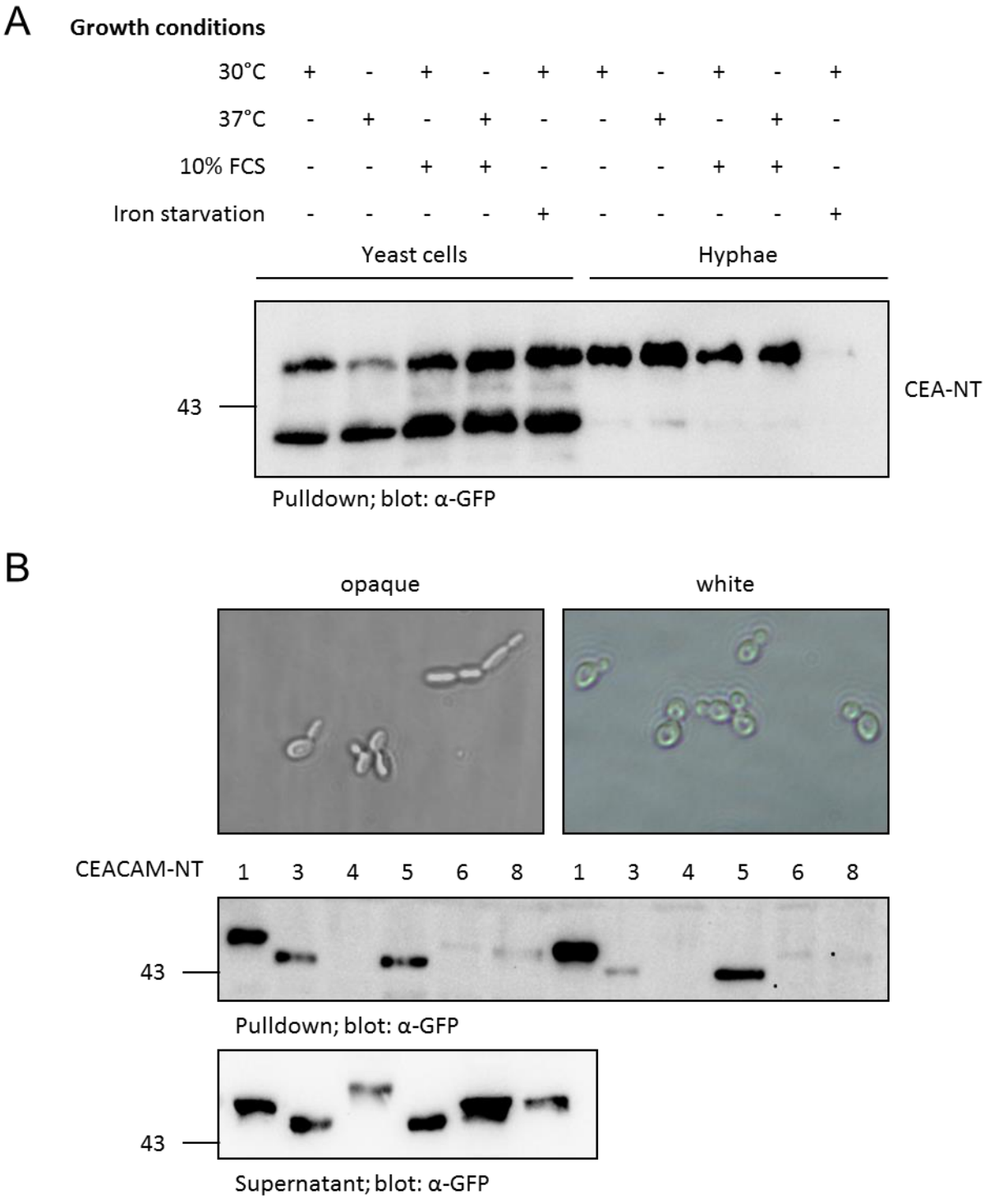


Figure III-2: Phenotype variation of *C. albicans* had no observable effect on CEACAM-binding. (A) *C. albicans* yeast cells were grown overnight under the indicated conditions. To induce yeast-to-hyphae transition, equivalent amounts of yeast cells were incubated with conditioned cell culture medium. Subsequently, binding studies were performed by adding

soluble GFP-tagged human CEA-NT. (B) Binding studies with different CEACAM constructs were performed by utilizing *C. albicans* yeast cells in white or opaque phase. Bound CEACAM-Ig_V-like domains were recovered in a yeast pulldown and evaluated via immunoblotting analysis with anti-GFP antibodies.

Enrichment and identification of potential CEACAM-associating *C. albicans* proteins

A previous study showed that CEACAM-interaction with *C. albicans* was reduced after pretreatment with proteinase K, hinting to a critical proteinaceous component involved in CEACAM-binding (Klaile, Muller et al. 2017). To allow proteomic based identification of potential CEACAM ligands of *C albican*s, CEACAM-coated beads were used to enrich interacting proteins. For this purpose, the yeast cell wall was digested by the β-glucan targeting enzyme zymolyase, releasing cell wall associated proteins to be probed with CEA-GFP-coated beads (Figure III-3 A). To distinguish unspecific interactions, GFP-coated beads were utilized as control. Remarkably, CEA-GFP-beads enriched four protein bands compared to the controls (Figure III-3 B). Those sections were extracted and the protein identities were determined by mass spectrometry (Suppl. Table III-1). Identified *C. albicans* proteins were all predicted to be localized in the yeast cytosol instead of the cell surface, limiting their probability as a native target. The absence of a promising candidate could either be explained by a potential absence of the adhesin due to its fluctuating up and downregulation, or the protein of interest was not part of the analyzed fractions but was present in a less prominent, non-excised band. In a further approach the slightly stronger binding partner CEACAM1 was utilized to enrich *C. albicans* proteins. The presence of the CEACAM-binding adhesin was confirmed in parallel by binding studies (Figure III-3 C). Mass spectrometry revealed an extensive set of proteins bound to CEACAM1 and the non-*C. albicans* binder CEACAM8 (Suppl. Table III-2). To determine enriched proteins, the score ratios of CEACAM1 and CEACAM8 were compared. The higher the protein score the more peptides of a certain protein have been identified. Four proteins that are localized in the *Candida* cell wall were chosen out of the list for further investigation: The two highest scored candidates that only appeared in the CEACAM1 sample (*Fructose-bisphosphate aldolase (FBA1)*, *Phosphoglycerate mutase (GPM1)*) and the two highest ratios of appearance in CEACAM1 and CEACAM8-samples (*Enolase1 (Eno1)*, *Pyruvate decarboxylase (PDc)*) (Table III-3).

Table III-3: CEACAM1-enriched cell wall proteins of *C. albicans*. Identification of potential CEACAM-binding proteins by mass spectrometry. Protein scores of the *C. albicans*-interacting molecule CEACAM1 and the non-binder CEACAM8 were compared and sorted by their CEACAM1-to-CEACAM8 ratios. Four proteins that revealed the highest scores are listed. * indicates CEACAM1 unique hits.

	Score		**Ratio**
Description	**CEACAM1**	**CEACAM8**	**CC1/CC8**
Fructose-bisphosphate aldolase	1135.41	zero	*
Phosphoglycerate mutase	559.99	zero	*
Enolase 1	2975.45	370.42	8.03
Pyruvate decarboxylase	2320.41	391.19	5.93

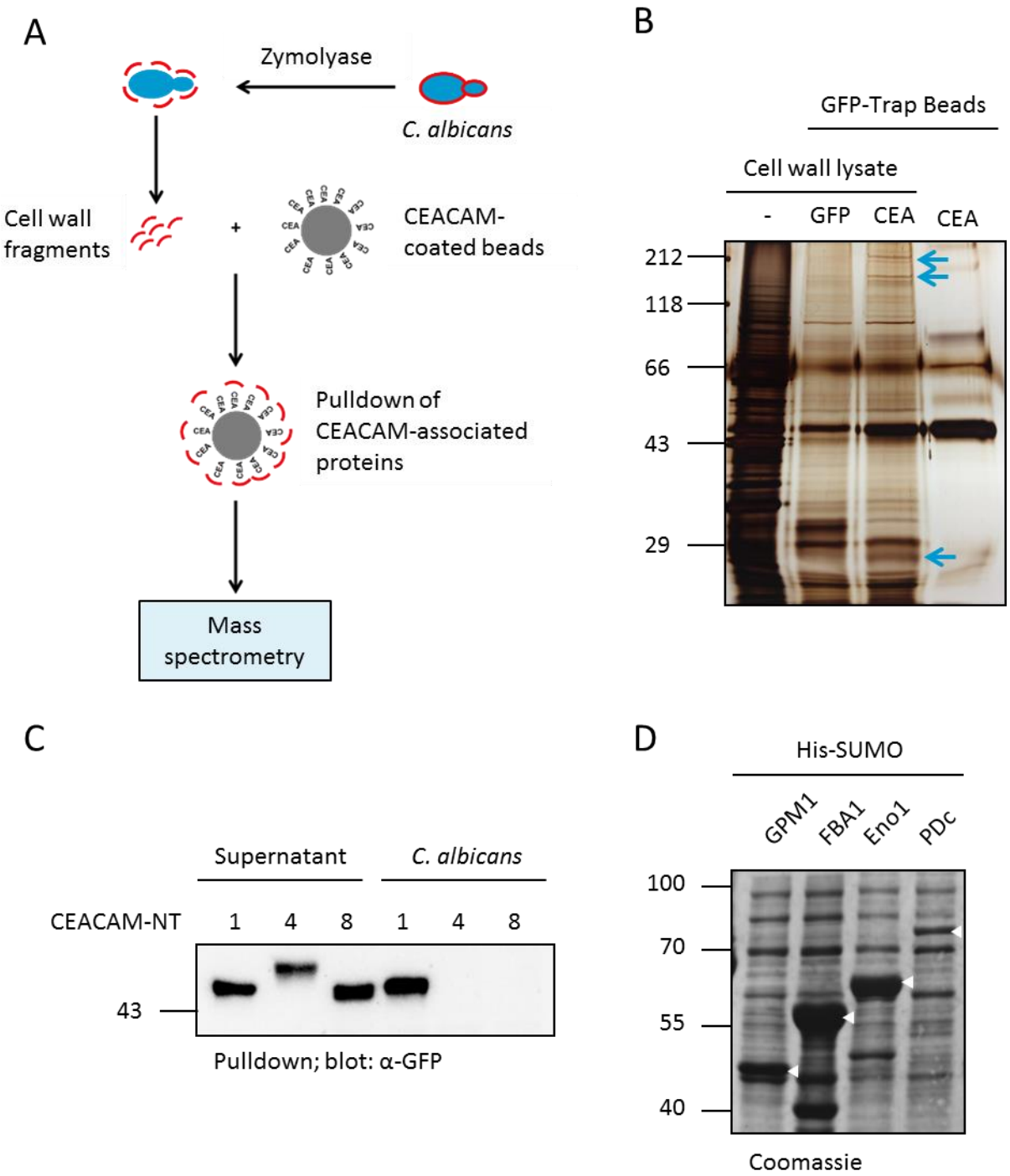

A
Zymolyase
C. albicans
Cell wall fragments
+
CEACAM-coated beads
Pulldown of CEACAM-associated proteins
Mass spectrometry
B
GFP-Trap Beads
Cell wall lysate
-
GFP
CEA
CEA
212
118
66
43
29
C
Supernatant
C. albicans
CEACAM-NT
1
4
8
1
4
8
43
Pulldown; blot: α-GFP
D
His-SUMO
GPM1
FBA1
Eno1
PDc
100
70
55
40
Coomassie
E
Supernatant
GPM1
FBA1
Eno1
PDc
CEACAM-NT
1
8
1
8
1
8
1
8
1
8
43
Ni-beads: α-His; blot: α-GFP

Figure III-3: Identification of CEACAM-interacting proteins. (A) Schematic workflow of the adhesin identification. *C. albicans* is treated with zymolyase to release cell wall components. Cell wall lysate is incubated with CEA-NT-coated beads and CEACAM-associated proteins are identified via mass spectrometry. (B) Silver staining of CEACAM-enriched *C. albicans* proteins. Cell wall lysate is added to GFP-coated beads (lane 2) or CEA-NT-GFP-coated beads (lane 3). After overnight incubation at 4°C, the beads were loaded on an SDS-gel. Whole cell wall lysate (lane 1) and CEA-NT-GFP-coated beads without cell wall lysate (lane 4) were loaded as control. To make protein bands visible, the gel was incubated in silver nitrate solution. (C) Before yeast cell wall proteins were isolated, the presence of the CEACAM-interacting *C. albicans* adhesin was confirmed by binding studies. Therefore, soluble GFP-tagged CEACAM1-NT and the two non-binders, CEACAM4-NT and CEACAM8-NT, were added to *C. albicans* and incubated for 1 h. CEACAM-association was evaluated via immunoblotting analysis with anti-GFP antibodies. (D) The genes of the four most promising identified adhesin candidates of *C. albicans* (Phosphoglycerate mutase (GPM1), Fructose-bisphosphate aldolase (FBA1), Enolase1 (Eno1), and Pyruvate decarboxylase (PDc)) were cloned into an IPTG-inducible His-SUMO expression vector, respectively and transformed into *E. coli*. Gene expression of His-SUMO-tagged GPM1, FBA1, Eno1, or PDc was induced by the addition of IPTG and equal amounts of *E. coli* were analyzed for their protein expression level via Western blotting and subsequent Coomassie staining. (E) Binding studies were conducted to test the CEACAM-binding ability of GPM1, FBA1, Eno1, or PDc. Therefore His-SUMO-tagged GPM1, FBA1, Eno1, or PDc were purified from *E. coli* lysates and coupled to Nickel beads. Soluble GFP-tagged CEACAM1-NT and CEACAM8-NT were added to the respective adhesin candidate coated beads and CEACAM-association was evaluated via immunoblotting analysis with anti-GFP antibodies.

Investigation of potential CEACAM-binding proteins of *Candida albicans*

The four *Candida* cell wall proteins GPM1, FBA1, Eno1, and PDc were tested for their ability to interact with CEACAMs. Genomic DNA of *C. albicans* served as template to amplify and clone the respective gene into the bacterial expression vector pET24α-His-SUMO. Proteins were expressed as His-SUMO (His_6-small ubiquitin-related modifier)-fusion proteins. After protein expression was verified (Figure III-3 D), cell lysates were incubated with nickel beads to purify His-SUMO fusion proteins from the remaining components. Nickel beads coated with the

potential adhesins were incubated with the amino terminal part of CEACAM1 or CEACAM8 to test binding. However, none of the four cloned candidates were able to interact with CEACAM1 (Figure III-3 E). To exclude that the His-SUMO tag disrupts adhesin-ligand interaction, GPM1, FBA1, Eno1, and PDc were purified and His-SUMO tags were proteolytically removed. The respective protein was incubated with CEACAM1-NT-GFP-coated beads and dimethyl pimelimidate (DMP), a crosslinking reagent which covalently couples directly interacting proteins to each other. In case of CEACAM-binding, the CEACAM-adhesin complex would run higher on the SDS gel due to an enhanced molecular weight. Subsequent immunoblotting analysis showed monomeric CEACAM1-NT-GFP and its homodimeric form but no additional band. The tested adhesin candidates of *C. albicans* were not able to directly associate with CEACAM1 (Suppl. Figure III-1).

Infection with yeast cells result in CEACAM3 phosphorylation *in vitro*

The human immune system defends against certain CEACAM-binding pathogens with the neutrophil granulocyte expressed molecule CEACAM3. Recognition of CEACAM-binding microbes by the innate immune receptor results in clustering and thereby activation of the receptor. The CEACAM3 cytoplasmic domain comprises an immunoreceptor tyrosine-based activation motif (ITAM)-like sequence that becomes phosphorylated by Src-family kinases on two tyrosine residues. Upon phosphorylation, CEACAM3 tyrosine residues create a platform for effector proteins that drive cytoskeletal remodelling required for phagocytosis (Figure III-4 A) (Pils, Kopp et al. 2012). This rapid, opsonin-independent internalization is best studied for the human pathogen *Neisseria gonorrhoeae* (Schmitter, Agerer et al. 2004). Upon bacteria engagement CEACAM3 is rapidly activated and facilitates the internalization of the associated bacteria. To determine if CEACAM3-recognition of yeast cells stimulates the same pathway CEACAM3-expressing HEK293T cells were infected with *C. albicans.* Remarkably, CEACAM3 displayed significantly enhanced tyrosine-phosphorylation after 15 and 30 minutes of infection. Afterwards phosphorylation levels gradually decline (Figure III-4 B). As consequence, CEACAM3 activation could lead to CEACAM3 based signal transduction and result in immune response in form of phagocytosis, oxidative burst, or cytokine release similar as those observed for its interaction with pathogenic bacteria.

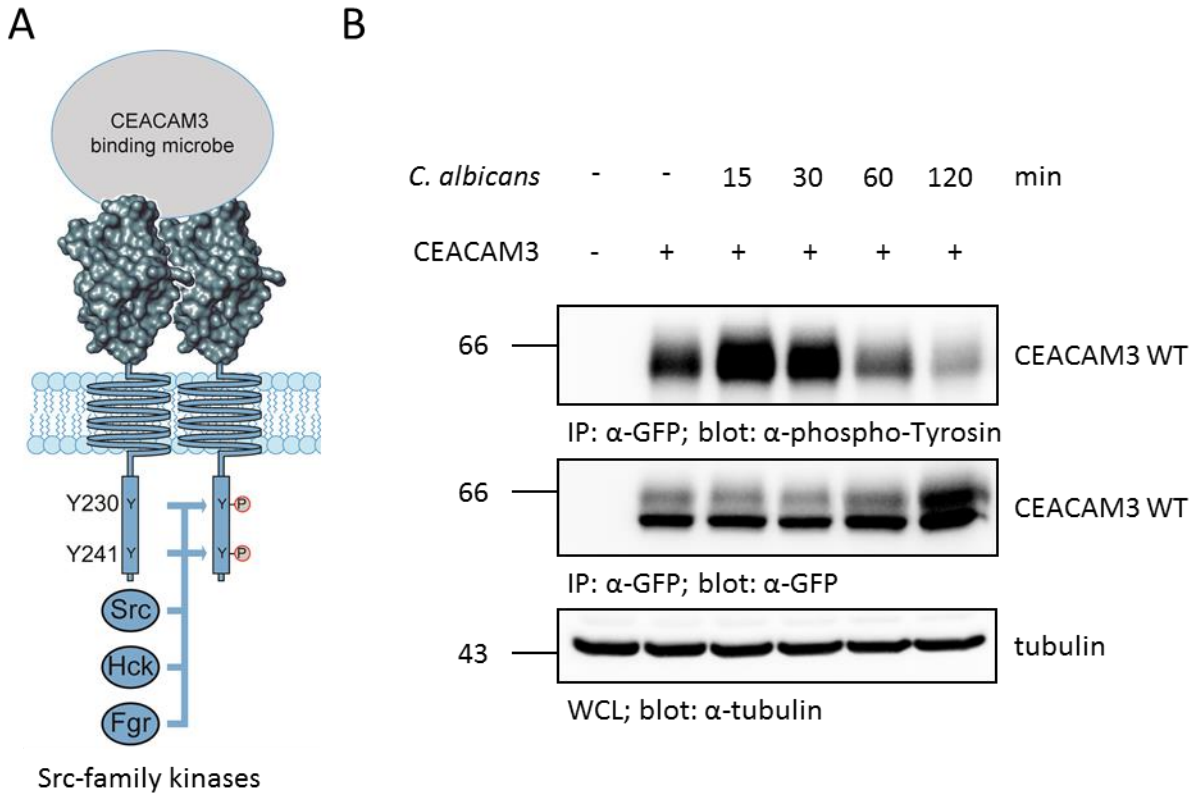
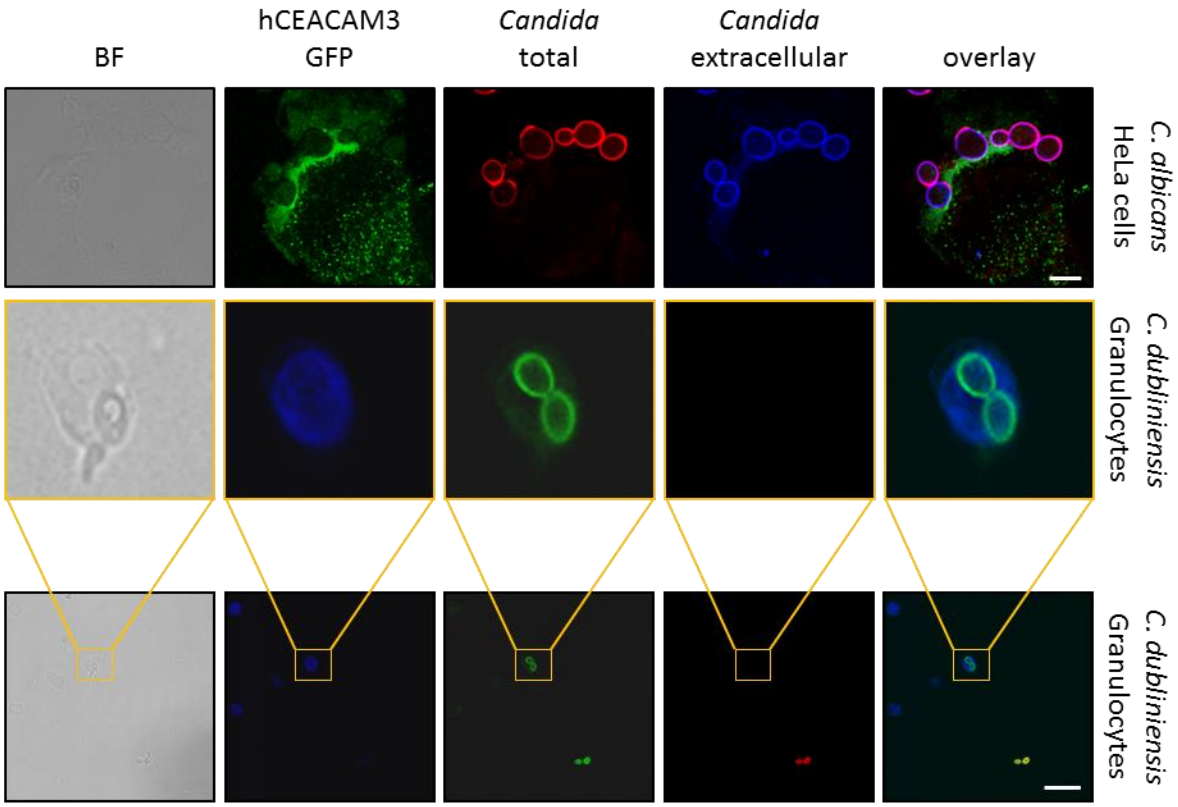


Figure III-4: *C. albicans* infection of CEACAM3-expressing 293T cells results in tyrosine phosphorylation of the CEACAM3 cytoplasmic domain. (A) Phosphorylation of CEACAM3 upon microbe engagement. Clustering of CEACAM3 molecules by CEACAM-binding microbes trigger recruitment and activation of certain Src family tyrosine kinases, which in turn phosphorylate the ITAM-like sequence in the cytoplasmic domain of the receptor. This phosphorylation event is critical for downstream signalling processes. (B) GFP-tagged CEACAM3-expressing HEK293T cells were infected with *C. albicans* and subsequently lysed. CEACAM3-GFP was captured and enriched by anti-GFP-antibody coated beads. The tyrosine phosphorylation level was determined by utilizing anti-phospho-tyrosine antibodies. Equivalent levels of HEK cell lysate for different infection time points were verified by probing whole cell lysates with anti-tubulin antibody. The recovered CEACAM3-GFP amount of each sample was analyzed.

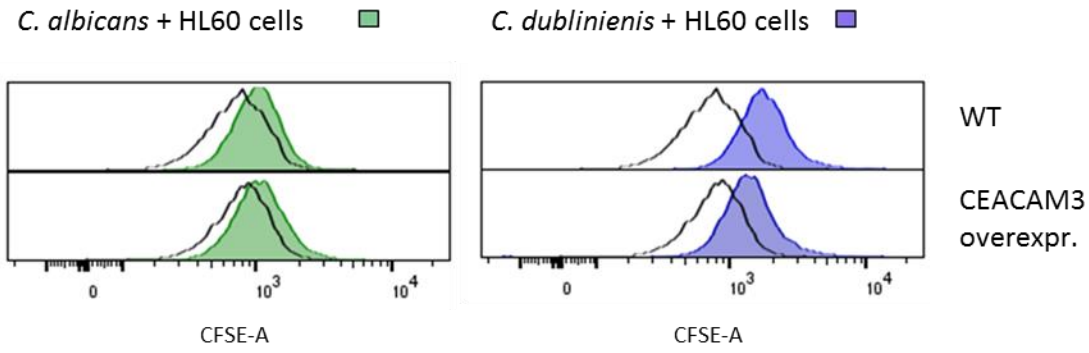
CEACAM3 does not promote yeast cell adhesion and phagocytosis

The recognition of microbes by the granulocyte receptor CEACAM3 leads to clustering at the side of engagement, initiation of downstream signalling and finally internalization of the bound particle (Schmitter, Agerer et al. 2004). Infection of CEACAM3-expressing HeLa cells with *C. albicans* showed clustering of the innate immune receptor at surface associated yeast cells but no internalization could be observed (Figure III-5 A). Moreover, granulocytes, which display greater phagocytic proficiency than Hela cells, were isolated from fresh blood samples and infected with *Candida dubliniensis*. *C. dubliniensis* shares many phenotypic characteristics with *C. albicans* and is phylogenetically a close relative (Gilfillan, Sullivan et al. 1998). In comparison to *C. albicans,* neutrophils internalize *C. dubliniensis* to greater extent and therefore ease CEACAM3 uptake studies with *Candida* (Svobodova, Staib et al. 2012). The freshly isolated primary granulocytes phagocytosed *Candida dubliniensis* within 30 min after infection (Figure III-5 A). Although granulocytes innately express CEACAM3, the observed phagocytosis event cannot solely be pinpointed to this receptor. Besides CEACAM3, granulocytes also express the antifungal innate immune receptor Dectin-1 which promotes fungi recognition and internalization. To discriminate between the uptake contribution of CEACAM3 and Dectin-1, neutrophil-like HL60 cells that express CEACAM3 on a low level were compared to HL60 cells that overexpress human CEACAM3 in a *Candida* infection experiment. HL60 cells revealed a higher internalization rate for *C. dubliniensis* compared to *C. albicans*. However, an enhanced CEACAM3 abundance in HL60 cells did not increase phagocytosis of either yeast but even showed a slightly decreased internalization (Figure III-5 B). Interestingly, infection of CEACAM3-expressing human embryonic kidney (HEK) cells with *Candida* species showed neither uptake into the cell nor binding to the cellular surface. CEACAM3-expression alone seems not to be sufficient to enhance host cell association of *C. albicans* (Figure III-5 C), however, it might promote the process of cell adhesion and thereby serve as a cofactor for further interaction based processes.

A



B



C

CEACAM3-expressing cell type		Surface binding	Candida uptake
HEK 293T		-	-
HeLa		+	-
Granulocytes		+	+
HL60	WT	+	+
	CEACAM3 overexpr.	+	+

Figure III-5: CEACAM3 overexpression does not enhance phagocytosis of *C. albicans*. (A) CEACAM3-GFP-expressing HeLa cells or primary granulocytes were infected for 1 h with biotinylated and TAMRA- (red) or CFSE- (green) stained *Candida* species respectively. Subsequently, extracellular yeasts were stained with streptavidin-Cy5 (blue) or Rhodamine (TRITC)-streptavidin (red). Nuclei of granulocytes were stained with DAPI (blue). Cells were analyzed by confocal microscopy. White arrowheads point to internalized yeast cells. Scale bar_{HeLa} = 5 µm, scale $bar_{Granulocyte}$ = 20 µM. (B) Wild-type HL60 cells (WT) or CEACAM3-overexpressing HL60 cells (CEACAM3) were infected for 30 min with CFSE-labelled *C. albicans* or *C. dubliniensis*. Fluorescein signals derived from extracellular yeasts were quenched by addition of trypan blue. Signals from phagocytosed *Candida* were detected by flow cytometry. (C) Overview of different CEACAM3-expressing human cell lines and their ability to interact and phagocytose *C. albicans*.

Recognition of *Candida albicans* by primate CEACAM3 orthologs exhibits a phylogenetic characteristic

Most receptors of the innate immune system involved in the recognition of foreign molecular patterns such as TLRs and NOD-like receptors are highly conserved (Kimbrell and Beutler 2001; Barreiro, Ben-Ali et al. 2009; Vasseur, Patin et al. 2011; Mukherjee, Ganguli et al. 2014). However, the innate immune receptor CEACAM3, directed against an array of microbial adhesins, has evolved at an exceptional rate. Interestingly, this receptor exclusively occurs in Old World monkeys and seems to be absent in the genomes of lemurs and New World monkeys (Figure III-6 A) (Sato, Kuroki et al. 2015; Adrian, Bonsignore et al. 2019). Sequence alignment of CEACAM3 orthologs revealed a high conservation in the signal sequence as well as the transmembrane region and the intracellular domain of the receptor, which initiates signal transduction after microbe binding. However, the ligand interaction domain of CEACAM3 orthologs exhibits a considerable variability in its amino acid sequence which also affects the recognition of different pathogens (Figure III-6 B). To address whether those amino acid alterations also influence the association with *C. albicans* a binding study with various primate CEACAM3 constructs was conducted. The CEACAM3 extracellular Ig_V-like domains of human, chimpanzee, gorilla, baboon, and rhesus monkey were recombinantly expressed in HEK cells as soluble GFP-fusion proteins (Figure III-6 C).

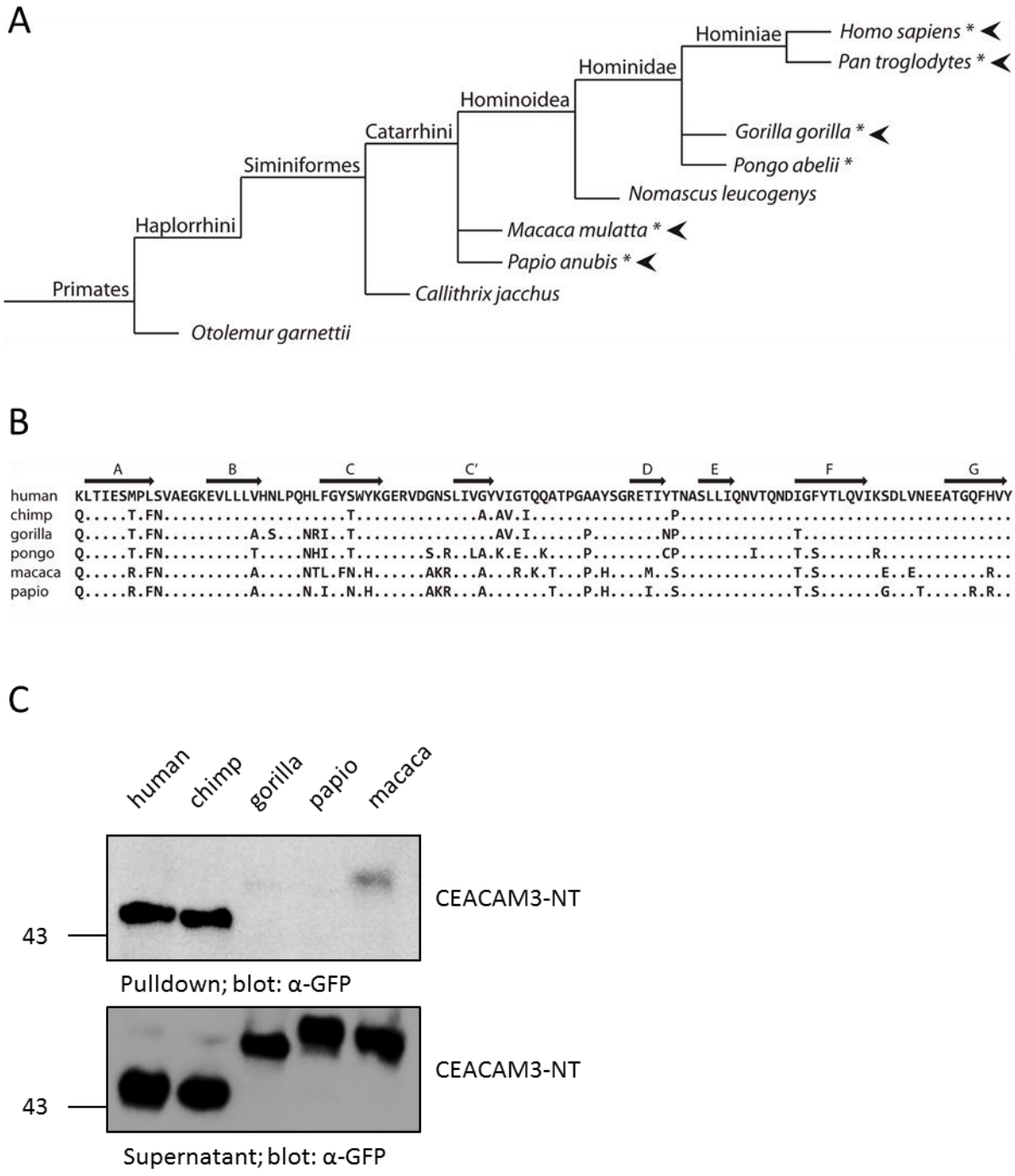


Figure III-6: *C. albicans* shows a striking species specificity for human and chimpanzee CEACAM3. (A) Primate phylogenetic tree. CEACAM3 encoding species are highlighted by an asterisk. CEACAM3 orthologs tested in *C. albicans* binding studies are indicated by an arrow. (B) Alignment of primate CEACAM3 amino acids sequences, β-strands (A-G). (C) Primate CEACAM3-NT-GFP constructs were incubated with *C. albicans*. CEACAM3 binding was determined via yeast pulldown and analyzed by immunoblotting.

Remarkably, besides human CEACAM3 only the relatively closely related chimpanzee CEACAM3 recognized *C. albicans*. Rhesus monkey indicated only a slight binding while gorilla and baboon showed no interaction at all (Figure III-6 C). The amino acid sequence alterations between primate CEACAM3 constructs result in binding to the orthologs in correlation with evolutionary distance.

Discussion

The commensal yeast *Candida albicans* can cause infections in the human host ranging from local to life-threatening systemic infections. To counteract or prevent this incidence, the immune system exhibits a variety of anti-fungal defence functions like phagocytes that capture, internalize and kill *Candida*. The yeast cell recognition is accomplished with the aid of innate immune receptors such as the membrane bound receptor Dectin-1 that targets β-glucan structures of fungal cell walls. Dectin-1 is known to play a major role in non-opsonic fungal uptake and intracellular killing of associated fungi. Additionally it triggers the release of cytokines to recruit further immune cells to the site of infection and the production of reactive oxygen species to dispose internalized fungi (Heinsbroek, Taylor et al. 2008). However, infection of dectin-1-deficient immune cells revealed that depletion of Dectin-1 is not sufficient to abrogate fungi internalization completely hinting to the contribution of other players (Saijo, Fujikado et al. 2007). The functionally related neutrophil granulocyte receptor CEACAM3 might be such a player. It is able to recognize *C. albicans* and is known to initiate a signal cascade similar to Dectin-1-mediated responses. In this study we investigated the role of CEACAM3 in yeast recognition, *Candida* elimination and the involved fungal adhesion structure. During testing of additional yeast species a common CEACAM-binding pattern was found. The close *C. albicans*-relative *Candida dubliniensis*, as well as the commensal yeast *Saccharomyces cerevisiae* and the laboratory species *Pichia pastoris,* all associated with the epithelial CEACAM CEA and were recognized by the granulocyte receptor CEACAM3. In contrast to the various CEACAM-adhesins previously identified in bacteria, a more conserved structure seems to establish CEACAM-interaction in yeast. This interaction appears to be mediated directly by the CEACAM protein part rather than its carbohydrate moieties, as the granulocyte receptors CEACAM4 and CEACAM8, which possess similar glycosylation patterns, do not associate with yeast cell surfaces. This notion is supported by previous studies on well-known CEACAM-binding pathogens which all interact, carbohydrate independently, with the

CEACAM Ig_V-like domain (Bos, Kuroki et al. 1998). Strikingly, the adhesin expression in *C. albicans* seems to be tightly regulated. Certain environmental conditions such as higher temperatures and presence of serum increased yeast-associated CEACAM amounts. This might be due to either enhanced binding affinity or, more likely, upregulation of the adhesion structure. In contrast, iron starvation, a condition that counteracts the virulence of *C. albicans* (Ramanan and Wang 2000), led to decreased interaction. Those conditions are all known to trigger yeast-to-hyphae transition which is accompanied by considerable gene expression alteration (Nantel, Dignard et al. 2002). Especially certain *C. albicans* adhesins including Als3 and Hwp1, that are known to mediate host cell attachment and contribute to *C. albicans* virulence, are upregulated during this phase. We used a *C. albicans* knockout strain that lacked the two main regulators of hyphae initiation, *cph1* and *efg1* (Lo, Kohler et al. 1997). Interestingly, the hyphae-deficient *C. albicans* strain still responded to varying environmental conditions with an altered CEACAM-binding pattern, suggesting a hyphae independent adhesin to be involved. This assumption could further be confirmed by binding studies that revealed yeast cells as well as hyphae to associate with CEACAMs similarly. We observed that the growth rate of *C. albicans* was accelerated during the presence of serum and during higher temperatures, both CEACAM-interaction favouring conditions. In contrast, iron starvation, which negatively influenced CEACAM-binding, slowed the proliferation rate down. The growth rate has a tremendous effect on the expression of different genes. More precisely, half of all yeast genes are affected. Changes in gene expression that are found during an accelerated growth rate are similar to those associated with different types of stress (>80% overlap) such as environmental stress (e.g. high temperature, high/low pH) or nutrient starvation (Regenberg, Grotkjaer et al. 2006; Castrillo, Zeef et al. 2007; Brauer, Huttenhower et al. 2008; Fazio, Jewett et al. 2008). Assuming the growth rate rather than a particular morphology to be decisive for the presence of the CEACAM-adhesion molecule, it is not surprising that different mating phases (white, opaque) and different virulence forms (commensal yeast cell, pathogenic filamentous cell) of *C. albicans* had no impact and showed the same CEACAM-binding pattern. While we could elucidate several interaction determining factors, the molecular identity of the CEACAM-associating adhesin of *C albicans* remains elusive. Utilizing a bead-based isolation method, we were able to enrich and identify certain CEACAM-associating *C. albicans* cell wall proteins such as Fructose-bisphosphate aldolase, Phosphoglycerate mutase, Enolase1, and Pyruvate decarboxylase. In subsequent binding

studies, the expressed and purified cell wall proteins showed no direct CEACAM-interaction. Fructose-bisphosphate aldolase, Phosphoglycerate mutase, and Enolase1 are involved in the glycolysis pathway. Many other enzymes of this pathway were picked up in the screen as well. One explanation for their presence could be the association with sugar moieties of CEACAM1. However, the complete absence of Fructose-bisphosphate aldolase and Phosphoglycerate mutase in the likewise glycosylated CEACAM8 sample argues against it. Remarkably, all investigated adhesin candidates are known to be moonlighting proteins, meaning that they perform more than one function. Besides the glycolysis pathway, Fructose-bisphosphate aldolase and Phosphoglycerate mutase were both observed to be involved in human host interaction (Crowe, Sievwright et al. 2003). Enolase1 was shown to mediate *Candida* binding and thereby enhances invasion of human endothelial cells (Jong, Chen et al. 2003). Interestingly, these three candidates were also identified in a proteomic analysis of human plasminogen-binding cell wall proteins (CWP) of *C. albicans* (Crowe, Sievwright et al. 2003). Plasminogen is the precursor protein of the serine protease plasmin that degrades many blood plasma proteins including fibrin clots. It is a serum component and could theoretically be present during CEACAM-binding studies as CEACAMs are produced in serum containing cell culture medium. Plasminogen, or another serum component, could indirectly mediate *Candida*-CEACAM association by functioning as adapter molecule. This theory would explain the lack of binding when CEACAM-coated beads were incubated with the purified *Candia* proteins after His-SUMO tag removal in the absence of serum. Remarkably, a similar pattern of plasminogen-binding proteins was observed for CWP samples from *C. dubliniensis*, *C. glabrata* and the commensal yeast *S. cerevisiae* (Crowe, Sievwright et al. 2003) which is in line with our finding of CEACAM-associating yeast species.

When it comes to an absent binding, the aspect of correct folding has to be taken into account. *Candida* adhesin candidates were expressed in *E. coli*. Bacteria cannot perform all post-translational modifications that are required for proper folding and activity. All four tested proteins naturally occur as homodimers (FBA1, Eno1) or homotetramers (GPM1, PDc). When we added crosslinking molecules to maintain quaternary structures and analyzed the samples in gel electrophoresis, in all cases a size shift could be observed hinting to native conformations. However, to exclude protein misfolding completely, analysis of protein functionality (e.g. enzyme assay) would be required. An alternative reason why the effort to coprecipitate a CEACAM-binding fungal adhesin failed could arise from altered binding

conditions. CEACAMs are produced in OptiMEM medium, the culture medium of CEACAM-secreting HEK cells. This medium is later collected and used for binding studies as it contains the released CEACAM constructs. Cell wall protein isolation of *C. albicans*, however, is conducted in yeast cell growth medium. For *Candida* protein enrichment, CEACAM-coated beads were added to this medium as it contained the cell wall proteins. Altered pH values or varying salt concentrations could influence and impair proper CEACAM-interaction and therefore the enrichment of the actual adhesin.

The second part of the study focused on CEACAM3 and its contribution to *C. albicans* elimination via phagocytosis. To date, only Gram-negative bacteria are known to activate CEACAM3 and thereby trigger a rapid, opsonin-independent internalization. Here we could show that the innate immune receptor not only recognizes *Candida* but also is activated via phosphorylation at cytoplasmic tyrosine residues in response to yeast cell infection. The rapid CEACAM3-mediated pathogen uptake only takes around 15 minutes. This coincides with the timing of the observed tyrosine phosphorylation maximum. However, neither CEACAM3-expressing HEK cells nor HeLa cells were able to internalize yeast cells. Both cell types are not specialized for phagocytosis therefore the uptake of particles with increased size might not be possible. A former study showed the limitation of HeLa cells to internalize particles bigger than 1-2 µm via CEACAM3 in absence of the cytoplasmic tyrosine kinase Syk (Sarantis and Gray-Owen 2007). The uptake of bacteria like *N. gonorrhoeae* with a size of 1 µm can be accomplished. In contrast, yeast cells such as *Candida* with a size of 5-10 µm might be beyond the phagocytic capabilities of this cell typ. In contrast, infection of phagocytic proficient primary granulocytes or the granulocyte precursor cell line HL60 led to internalization of *Candida*. In those cells the proteins CEACAM3, Syk as well as Dectin-1 are naturally expressed. HL60 cells show only low amounts of CEACAM3, therefore a CEACAM3 overexpressing HL60 cell line was generated. Higher abundance of the innate immune receptor did not result in enhanced levels of phagocytosed yeast cells. This observation suggests that CEACAM3 expression is not crucial in fungal phagocytosis events. Our conclusion is supported by a study in which dectin-1 deficient macrophages were challenged with *C. albicans.* Instead of complete phagocytosis abrogation, *Candida* internalization only showed a reduction by around 50%. While this suggests the participation of additional receptors, it excludes CEACAM3 as it is only expressed in neutrophil granulocytes but not in macrophages (Gales, Conduche et al. 2010). Furthermore, this theory is strengthened by the observation that

dectin-1 knockout mice are still able to cope with *C. albicans* infections in same extent as wild-type mice (Saijo, Fujikado et al. 2007). Again highlighting the contribution of additional receptors involved in yeast cell recognition, however excluding CEACAM3 as it first emerged in primates and therefore is absent in mice. Nevertheless, CEACAM3 gets activated via intracellular tyrosine phosphorylation providing a signalling platform which could contribute to fungal immune response via the activation of non-phagocytic effector pathways such as the production of reactive oxygen species or the release of cytokines. Both would support and contribute to the elimination of the pathogenic yeast.

A further interesting finding was the observation that only human CEACAM3 and the close relative chimpanzee CEACAM3 recognize *C. albicans* with high affinity. Despite the high sequence similarities, phylogenetically more distant CEACAM3 variants such as gorilla, baboon and rhesus monkey exhibited no or only weak interactions. This finding is also known for pathogenic bacteria such as *Neisseria gonorrhoeae*, *Moraxella catarrhalis*, and *Haemophilus aegyptius* and suggests co-evolution of pathogens and the human host (Adrian, Bonsignore et al. 2019). If co-evolution is the mechanism behind this species-specific binding pattern, there has to be an advantage either for the innate immune receptor to recognize the fungus or for *Candida* to associate with the immune receptor CEACAM3. The way this interaction affects yeasts is currently ambiguous. It could result in microbe killing by cytokine-recruited immune cells or oxidative burst induction, or hampering of fungal virulence. However, further investigations are required to provide a clear answer to this question.

Material and Methods

Transfection of HEK293T cells

The human embryonic kidney cell line 293T (HEK293T cells) (DSMZ, Braunschweig, Germany) was grown in Dulbecco's modified Eagle's medium (DMEM) (Merck) supplemented with 10% calf serum (CS) (Biochrom) at 37°C, 5% CO_2. The cells were transfected with pLPS3'EGFP CEACAM plasmid via calcium-phosphate co-precipitation. Therefore 5 µg plasmid DNA was added to 500 µl of H_2O followed by 500 µl of 2x HBS buffer (274 mM NaCl, 42 mM HEPES, 1.4 mM Na_2HPO_4, pH 7.05) and 50 µl 2.5 M $CaCl_2$.

Fungal strains and growth conditions

Candida albicans SC5314 and GFP-expressing *Candida albicans* SC5314 were kindly provided by Prof. Dr Joachim Morschhäuser (Zentrum für Infektionsforschung, Universität Würzburg, Würzburg, Germany). All fungal strains were grown in 10% yeast extract, 20% peptone, and 20% glucose (YPG) medium over night at 30°C. For testing the effect of different growth conditions, the YPG medium was supplemented with 1 M NaOH for an alkaline pH (pH 8), with 400 µM deferoxamin (Sigma-Aldrich) for iron starvation (Fe-Chelator), or with 0.1-10% fetal calf serum (FCS) (Biochrom). Alternatively fungi were grown at 37°C.

Binding studies of yeast species to soluble CEACAMs

Soluble CEACAM-ectodomains fused to GFP were produced in HEK293T cells. 500 µl of the CEACAM-GFP containing supernatant was clustered with rabbit anti-GFP serum (1:1000) (local animal facility at the University of Konstanz, Germany) overnight at 4°C. The overnight grown yeast cultures were harvested by centrifugation (3000 xg, RT, 5 min), the medium was discarded and the cell pellet suspended in 1 ml 1x phosphate-buffered saline (PBS) buffer. The optical density at a wavelength of 600 nm (OD_{600}) was determined and OD_{600} 1 of the respective fungal species was incubated with the clustered CEACAM-GFP constructs for 1 h at RT under gentle rotation. Afterwards, yeast cells were pelleted (18,000 xg, 4 °C, 1 min) and washed twice with PBS buffer. Finally, yeasts were either suspended in 4x SDS sample buffer for immunoblotting, or added to FACS buffer (1x PBS, 5% heat-inactivated FCS) for flow cytometry.

Immunoblotting

Samples from binding studies were analyzed by SDS-PAGE and Western blot. After separating proteins by their molecular weight via SDS-PAGE, they were transferred to polyvinylidene difluoride (PVDF) membranes (Millipore, Bedford, MA). The membranes were blocked (2% bovine serum albumin in Tris-buffered saline containing 0.05% Tween 20) for 1 h at RT and incubated in monoclonal antibody (mAB) against GFP (clone JL-8, Clontech, Palo Alto, CA) overnight at 4°C. Next day, the membrane was incubated with polyclonal HRP-conjugated goat anti-mouse antibody (Jackson ImmunoResearch) diluted 1:10.000 for 1 h at RT, and visualized by chemiluminescence (ChemiDoc Touch, BioRad).

CEACAM3 phosphorylation assay

CEACAM3-GFP-expressing HEK293T cells ($3x10^6$) were seeded on gelatine coated 6 cm plates and starved overnight (DMEM, 0.5% CS) to reduce phosphorylation levels. Next day, heat inactivated medium (DMEM, 0.5% CS) containing 10 µM pervanadate was added and cells were infected with yeasts at a multiplicity of infection (MOI) of 30 at 37°C. At the indicated times, cells were washed with 1 ml ice-cold PBS and lysed in modified radioimmune precipitation assay (RIPA) buffer (25 mM Hepes (pH 7.4), 0.1% SDS, 0.5% sodium deoxycholate, 1% Triton X-100, 150 mM NaCl, 20 mM MgCl2, 10% glycerol, 10 mM sodium pyrophosphate, 100 mM NaF, 1 mM Na3VO4, 0.1 mM sodium orthovanadate, 1 mM para-Nitrophenylphosphate, 10 µg/ml of each benzamidine, aprotinin, leupeptin, pefabloc and pepstatin). GFP-specific monobodies coupled to sepharose beads were added to the cleared cell lysates and incubated overnight at 4°C to enrich CEACAM3-GFP. 1x SDS sample buffer was added to the beads. The samples were boiled for 15 min at 98°C and analyzed by SDS-PAGE and subsequent Western blotting as described before.

Analysis of fungal invasion by flow cytometry

Yeast cells were harvested and suspended in 1 ml PBS buffer. 2 µg carboxyfluorescein-succinimidylester (CFSE) (Molecular Probes) was added and yeast cells were labelled for 45 min at 37°C under constant shaking and protected from light. After the staining process, yeast cells were washed three times with 1 ml PBS buffer to remove the remaining staining solution (18,000 xg, 4°C, 1 min). Alternatively, GFP-expressing yeast cells were utilized.

$1x10^6$ CEACAM-expressing human cells (HL60 (DSMZ), granulocytes (freshly isolated)) were suspended in 1 ml phagocytosis buffer (1x PBS, 0.9 mM $CaCl_2$, 0.5 mM $MgCl_2$, 5 mM glucose, 1% heat inactivated FCS) and infected with $1x10^7$ 5-Carboxy-tetramethylrhodamine N-succinimidyl ester (TAMRA-SE)-stained yeast cells (MOI of 10) for 30 min at 37°C. The cells were washed once with ice cold PBS buffer and afterwards kept on ice. The total amount of associated yeast cells was measured by flow cytometry. To selectively measure internalized yeast cells, 0.2 mg/ml trypan blue was added to quench extracellular CFSE-signal (Pils, Schmitter et al. 2006). Raw data were evaluated with FlowJo (FlowJo LLC).

Primary granulocyte isolation

Primary human granulocytes were isolated from citrated blood of single donors by gradient centrifugation on Ficoll layer. The granulocyte layer was collected and suspended in 0.96% NaCl containing 1% polyvinyl alcohol for 45 min to remove erythrocytes by sedimentation. Remaining erythrocytes were lysed by hypotonic shock in ddH_2O for 1 min (Brandt, Van Damme et al. 1991). Lysis was stopped by adding an equal volume of 2x PBS. Granulocytes were transferred into phagocytosis buffer (1x PBS, 0.9 mM $CaCl_2$, 0.5 mM $MgCl_2$, 5 mM glucose, 1% heat-inactivated fetal calf serum).

Immunofluorescence staining for confocal microscopy

CEACAM3-expressing HeLa cells were seeded on poly-L-lysine coated coverslips in 24-well plates ($8x10^4$/well). Next day, yeasts were labelled with 2 µg 5-Carboxy-tetramethylrhodamine N-succinimidyl ester (TAMRA-SE) (Sigma-Aldrich) and 0.1 mg/ml EZ-Link™ Sulfo-NHS-LC-LC-Biotin (Thermo Fisher Scientific) in PBS for 45 min at 37°C, light protected and under constant shaking. Yeasts were washed three times with 1 ml PBS to remove the staining solution. Cells were infected with an MOI of 6 for 1 h at 37°C. After infection, unbound yeasts were washed off with 350 µl PBS^{++} (1x PBS, 0.9 mM $CaCl_2$, 0.5 mM $MgCl_2$). The cells were fixed in 4% paraformaldehyde for 20-30 min at RT. 300 µl of blocking solution (PBS^{++}, 10% FCS) were added for 10 min at RT and subsequently extracellular yeasts were stained with streptavidin-Cy5 (Molecular Probes) for 45 min at RT. Cells were stained with 0.2 µg/ml 4',6-diamidino-2-phenylindole (DAPI) for 5 min at RT. Afterwards the staining solution was removed, the coverslips were transferred on glass slides and preserved by adding Dako Fluorescence Mounting Medium (Dako, Glostrup, Denmark) in between.

Primary granulocytes were seeded on poly-L-lysine coated coverslips in 24-well plates ($4x10^4$/well). Yeasts were labelled with 2 µg CFSE and 0.1 mg/ml Sulfo-NHS-SS-Biotin in PBS for 45 min at 37°C, light protected and under constant shaking. Cells were infected with yeast (MOI 3) in phagocytosis buffer for 1 h at 37°C. After infection, unbound yeasts were washed off with 350 µl PBS^{++}. The cells were fixed in 4% paraformaldehyde for 20-30 min at RT. 300 µl of blocking solution were added for 10 min at RT and subsequently extracellular yeasts were stained with Rhodamine (TRITC)-streptavidin (Jackson) for 45 min at RT. Afterwards the coverslips were transferred on glass slides and preserved by adding Dako Fluorescence Mounting Medium in between.

Cell wall protein isolation

60 ml YPG medium was inoculated with *C. albicans* and grown overnight. Next day yeast cells were harvest (4487 xg, RT, 10 min), resuspended in 3.5 ml prewarmed (30°C) Resuspension Buffer (50 mM Tris pH 7.5, 30 mM DTT) and incubated for 20 min at 30°C on a shaker (60 rpm). Afterwards cells were pelleted (4487 xg, 5 min, RT) and resuspended in 2 ml YPD containing 1 M Sorbitol (RT). 1.5 ml 2 M Sorbitol (RT) and 100 µl ice cold Zymolyase (3 mg Zymolyase (amsbio), 50 mM Tris pH 7.4, 2x protease inhibitors, 2 mM PMSF) were added and incubated for 3-4 h at 30°C on a rotor (60 rpm). Yeast spheroblasts were pelleted (4487 xg, 5 min, 4°C) and the supernatant, containing released cell wall proteins, was collected. Zymolyase was inactivated for 5 min at 60°C.

Cell wall protein enrichment by CEACAM-coated beads

GFP-Trap beads were incubated with equivalent amounts of soluble amino-terminal GFP-tagged domains of CEACAM1, CEACAM4 or CEACAM8 overnight at 4°C. After the removal of unbound CEACAMs by several washing steps with PBS buffer, cell wall protein lysate was added to the beads and incubated overnight at 4°C under constant rotation. Next day, beads were washed thoroughly with PBS buffer to remove unbound or only weak associated proteins. Subsequently, 1x SDS sample buffer was added to each bead sample and boiled at 98°C to release bead-associated proteins. Proteins were separated by SDS-PAGE.

Silver staining of proteins in polyacrylamide gels

After the electrophoretic separation of proteins on polyacrylamide gels, they were visualized via silver staining. Therefore the gel was incubated in solution 1 (40% ethanol, 10% acetic acid)

and afterwards in solution 2 (30% ethanol) for 1h respectively to denature and fix the proteins and to remove SDS from the gel. Subsequently, 0.2 g/l $Na_2S_2O_3$ x 5 H_2O was added for 1 min and removed by washing the gel with ultrapure water. During an incubation with silver nitrate (2 g/l $AgNO_3$) solution for 20 min, silver ions associate with negatively charged protein side chains. Unbound silver ions were removed by a washing step with ultrapure water. To visualize protein bands, the gel was incubated in developing solution containing 6% Na_2CO_3, 5 mg $Na_2S_2O_3$ x 5 H_2O and 0.05% formaldehyde. The alkaline formaldehyde reduces silver ions to elemental silver wherefore proteins are stained brown to black. To stop the staining reaction, a solution containing 50 mM EDTA and 1% acetic acid was added.

Generation of different *C. albicans* morphologies

To generate white phase yeast cells, a *C. albicans* culture grown at 30°C in YPG medium was stored at 4°C for one week. To obtain opaque phase yeast cells, *C. albicans* was grown overnight at RT in YPG medium. The phase morphology was confirmed microscopically. Yeast-to-hyphae transition was induced by growing white yeast cells in OptiMEM (Gibco BRL) medium at 37°C for 2-3 h.

Cloning and expression of potential adhesin candidates

Genomic DNA was used as template to amplify GPM1, FBA1, ENO1 and PDC. The genes were cloned into the BamHI and XhoI site of pET24α-His-SUMO (Novagen, Madison, Wisconsin). The pET24α-His-SUMO vectors containing the respective gene were transformed in *E.coli* BL21DE3 Rosetta (Novagen, Madison, Wisconsin) and protein expression was induced by adding 1 mM IPTG (Roth) for 3-4 h. All expressed *Candida* proteins (GPM1, FBA1, ENO1 and PDC) were fused to His-SUMO (His_6-small ubiquitin-related modifier). The plasmid-containing *E. coli* strains were grown in LB-Kanamycin (50 µg/ml) medium at 37°C.

Testing CEACAM-binding of potential *Candida* adhesins

E. coli strains expressing the respective *Candida* proteins were suspended in 10 mM HEPES buffer supplemented with protease inhibitors and lysed by sonication. Ni Sepharose High Performance beads (nickel beads) (Amersham Biosiences) were added to cell lysates to purify His-SUMO tagged proteins. Unbound cell lysate components were removed using washing buffer (20 mM phosphate buffer pH 7.4, 0.5 M NaCl, 20 mM imidazole). Nickel beads coated with the potential adhesins were incubated with the amino terminal part of CEACAM1-GFP or

CEACAM8-GFP to test interaction. After thorough washing steps, nickel beads were boiled with 1x SDS buffer to release attached proteins and the samples were analyzed by immunoblotting utilizing anti-GFP antibodies.

Adhesin proteins were purified on a column. 4 µg of the His-SUMO cleaving enzyme Ulp1 was added per 1 mg protein. After an overnight incubation at 4°C, nickel beads were added to remove His-SUMO fragments. To detect associated adhesins, either biotinylated His-SUMO-free adhesins were incubated with CEACAM-GFP-coated beads, or soluble CEACAM-GFP constructs were incubated with a *Candida* adhesin for 1 h at RT and the non-cleavable crosslinker DMP (dimethyl pimelimidate x2 HCl) (Thermo Scientific) was added afterwards. Therefore 0.1 mg DMP was solved in 500 µl PBS buffer and 53 µl of 0.2 M triethanolamine, pH 8.0 was incubated with CEACAM-adhesin complexes for covalent linkage. The crosslinking reaction was stopped by 50 µl 1 M Tris buffer, pH 8.1 and the complexes were purified via GFP-Trap beads.

Supplementary information

Supplementary Figures

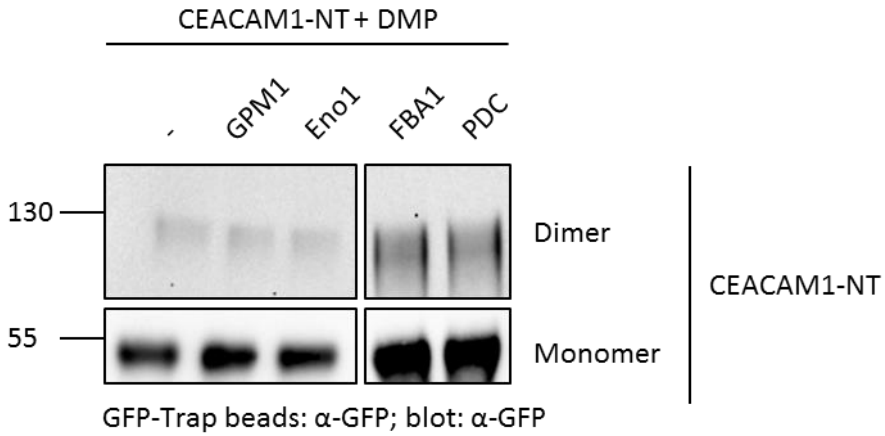


Figure III-S1: CEACAM1-NT-enriched adhesin candidates do not directly interact with CEACAM1-NT. Adhesin candidates identified by mass spectrometry were cloned into an IPTG-inducible His-SUMO expression vector, respectively and transformed into *E. coli*. After IPTG-induction, expressed His-SUMO fusion proteins were purified and the His-SUMO tag was removed to exclude CEACAM-binding interference. CEACAM1-NT-GFP coated beads were incubated with the respective adhesin candidate and directly interacting proteins were covalently coupled by the addition of a crosslinking reagent dimethyl pimelimidate (DMP). CEACAM-linkage was evaluated via immunoblotting analysis with anti-GFP antibodies.

Supplementary Tables

Supplementary Table III-1: Mass spectrometry results of CEACAM-interacting adhesin candidates derived from *C. albicans*. Single bands were cut out and analyzed.

	Description	**MW [kDa]**
Band 1+2 (> 212 kDa)	type I glyceraldehyde-3-phosphate dehydrogenase [Escherichia coli]	35.8
	probable translation elongation factor EF-1 alpha [Candida albicans SC5314]	49.9
	Putative acetyl-coenzyme-A carboxylases [Candida albicans SC5314]	253.3

	Ubiquitin precursor (polyubiquitin) [Candida albicans SC5314]	25.8
	HSP90 [Candida albicans SC5314]	80.8
Band 3 (118 kDa < protein < 212 kDa)	hypothetical protein CaO19.12474 [Candida albicans SC5314]	40.2
	ubiquitin-40S ribosomal protein S27b, partial [Candida albicans P37005]	7.9
	type I glyceraldehyde-3-phosphate dehydrogenase [Escherichia coli]	35.8
	potential mitochondrial iron transporter Mrs3 [Candida albicans SC5314]	35.3
	hypothetical protein CaO19.4868 [Candida albicans SC5314]	13.5
Band 4 (~ 27 kDa)	likely cytosolic ribosomal protein L8 [Candida albicans SC5314]	28.5
	60S ribosomal protein L8-B [Candida albicans P94015]	28.2
	cytosolic ribosomal protein S1 (rp10) [Candida albicans SC5314]	29.0
	BMH1 (14-3-3 protein)[Candida albicans SC5314]	29.5
	SSB1 (HSP70 family heat shock protein) [Candida albicans SC5314]	66.4
	RecName: Full=Guanine nucleotide-binding protein subunit beta-like protein; AltName: Full=Cytoplasmic antigenic protein 1	34.5
	potential mitochondrial inner membrane ATP/ADP translocator [Candida albicans SC5314]	32.7

Supplementary Table III-2: Mass spectrometry results of CEACAM-interacting adhesin candidates derived from *C. albicans*. All CEACAM-bead associated proteins were analyzed. Proteins are sorted by the CEACAM1 protein scores in a descending order. The investigated proteins are highlighted in blue.

Accession	Description	Score CC1	Score CC8	MW [kDa]
P0CY35	Elongation factor 1-alpha 1 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=TEF1 PE=3 SV=1 - [EF1A1_CANAL]	19431.8	19177.7	50.0
P40910	40S ribosomal protein S1 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=RPS1 PE=3 SV=3 - [RS3A_CANAL]	4563.48	4857.24	29.0

Q5A0M4	Elongation factor 2 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=EFT2 PE=1 SV=2 - [EF2_CANAL]	4525.85	5595.21	93.3
Q9UVJ4	60S ribosomal protein L10a OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=RPL10A PE=3 SV=2 - [RL10A_CANAL]	3503.45	3916.90	24.4
Q96W54	40S ribosomal protein S22-A OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=RPS22A PE=3 SV=3 - [RS22A_CANAL]	3214.88	3152.55	14.8
P30575	Enolase 1 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=ENO1 PE=2 SV=1 - [ENO1_CANAL]	2975.45	370.42	47.2
P83774	Guanine nucleotide-binding protein subunit beta-like protein OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=ASC1 PE=1 SV=2 - [GBLP_CANAL]	2961.81	3228.26	34.5
P25997	Elongation factor 3 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=CEF3 PE=1 SV=3 - [EF3_CANAL]	2564.64	2744.13	116.9
Q59KI0	UTP--glucose-1-phosphate uridylyltransferase OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=UGP1 PE=1 SV=2 - [UGP1_CANAL]	2468.18	2583.65	55.5
P83779	Pyruvate decarboxylase OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=PDC11 PE=1 SV=2 - [PDC1_CANAL]	2320.41	391.19	62.4
O93827	Mannose-1-phosphate guanyltransferase OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=MPG1 PE=1 SV=1 - [MPG1_CANAL]	1497.03	1320.93	40.0
Q5A4E2	ATP-dependent RNA helicase DED1 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=DED1 PE=3 SV=1 - [DED1_CANAL]	1301.50	1297.52	72.8
Q9URB4	Fructose-bisphosphate aldolase OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=FBA1 PE=1 SV=2 - [ALF_CANAL]	1135.41		39.2
P46273	Phosphoglycerate kinase OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=PGK1 PE=3 SV=1 - [PGK_CANAL]	1132.01	301.97	45.2
P87206	ATP-dependent RNA helicase eIF4A OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=TIF1 PE=3 SV=1 - [IF4A_CANAL]	1004.28	707.37	44.6
Q59S06	Nucleolar protein 58 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=NOP58 PE=3 SV=2 - [NOP58_CANAL]	955.82	970.52	57.1
P84149	mRNA export factor MEX67 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=MEX67 PE=1 SV=2 - [MEX67_CANAL]	904.46	972.21	68.2

Q59LU0	ATP-dependent RNA helicase DBP2 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=DBP2 PE=3 SV=2 - [DBP2_CANAL]	878.69	1622.87	61.2
O42817	40S ribosomal protein S0 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=RPS0 PE=2 SV=2 - [RSSA_CANAL]	852.67	972.52	28.7
Q9HGT6	Serine--tRNA ligase, cytoplasmic OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=SES1 PE=1 SV=1 - [SYSC_CANAL]	742.51	709.93	53.0
Q5AI15	Polyadenylate-binding protein, cytoplasmic and nuclear OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=PAB1 PE=3 SV=1 - [PABP_CANAL]	695.31	688.65	70.4
P41797	Heat shock protein SSA1 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=SSA1 PE=1 SV=2 - [HSP71_CANAL]	688.76	568.22	70.3
Q59MV9	Flavohemoprotein OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=YHB1 PE=2 SV=1 - [FHP_CANAL]	669.05	489.34	45.8
Q5AQ76	Protein transport protein SEC24 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=SEC24 PE=3 SV=2 - [SEC24_CANAL]	660.80	901.19	102.0
P46614	Pyruvate kinase OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=CDC19 PE=1 SV=3 - [KPYK_CANAL]	635.98	304.30	55.4
O43101	Centromere/microtubule-binding protein CBF5 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=CBF5 PE=3 SV=1 - [CBF5_CANAL]	632.20	1022.90	54.3
P0CU34	Peroxiredoxin TSA1-B OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=TSA1B PE=2 SV=1 - [TSA1B_CANAL]	615.53	457.20	21.8
O13426	Serine hydroxymethyltransferase, cytosolic OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=SHM2 PE=1 SV=4 - [GLYC_CANAL]	569.50	467.74	52.0
Q5AGV4	Eukaryotic translation initiation factor 3 subunit B OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=PRT1 PE=3 SV=1 - [EIF3B_CANAL]	567.77	515.00	84.2
P82612	Phosphoglycerate mutase OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=GPM1 PE=1 SV=3 - [PMGY_CANAL]	559.99		27.4
P53704	Glutamine--fructose-6-phosphate aminotransferase [isomerizing] OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=GFA1 PE=1 SV=3 - [GFA1_CANAL]	543.51	502.15	79.2

Q5AGZ9	RuvB-like helicase 2 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=RVB2 PE=3 SV=1 - [RUVB2_CANAL]	539.99	597.42	54.5
Q5AML1	Eukaryotic translation initiation factor 3 subunit C OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=NIP1 PE=3 SV=2 - [EIF3C_CANAL]	502.68	1021.97	99.8
O94038	Alcohol dehydrogenase 2 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=ADH2 PE=3 SV=1 - [ADH2_CANAL]	497.87	197.77	36.8
P46587	Heat shock protein SSA2 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=SSA2 PE=1 SV=4 - [HSP72_CANAL]	479.09	444.49	70.0
Q5AAW3	ATP-dependent RNA helicase DHH1 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=DHH1 PE=3 SV=1 - [DHH1_CANAL]	423.65	447.64	62.1
Q5AAU3	Protein transport protein SEC31 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=PGA63 PE=3 SV=2 - [SEC31_CANAL]	417.63	352.28	136.2
P83776	Hexokinase-2 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=HXK2 PE=1 SV=2 - [HXKB_CANAL]	410.62		53.4
P48989	Histone H2B.1 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=HTB1 PE=3 SV=3 - [H2B1_CANAL]	398.02	276.64	14.1
Q8NJN3	Acetyl-coenzyme A synthetase 2 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=ACS2 PE=3 SV=3 - [ACS2_CANAL]	388.68	409.55	73.8
Q5ANP2	Nascent polypeptide-associated complex subunit alpha OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=EGD2 PE=3 SV=1 - [NACA_CANAL]	386.77	370.01	19.5
P46598	Heat shock protein 90 homolog OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=HSP90 PE=3 SV=1 - [HSP90_CANAL]	352.92	226.70	80.8
Q9P940	Triosephosphate isomerase OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=TPI1 PE=1 SV=3 - [TPIS_CANAL]	351.60	166.82	26.6
O59931	60S ribosomal protein L13 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=RPL13 PE=3 SV=1 - [RL13_CANAL]	346.96	404.64	23.0
Q59SU5	Histone H2A.1 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=HTA1 PE=3 SV=3 - [H2A1_CANAL]	292.42	333.50	14.0
Q5ACM9	Eukaryotic translation initiation factor 3 subunit J OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=HCR1 PE=3 SV=1 - [EIF3J_CANAL]	273.24	262.03	32.0

Q59LF9	Methionine aminopeptidase 2 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=MAP2 PE=3 SV=2 - [MAP2_CANAL]	271.39	276.65	50.3
Q5A0W7	RuvB-like helicase 1 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=RVB1 PE=3 SV=1 - [RUVB1_CANAL]	266.24	333.10	50.0
Q59Q46	Inosine-5'-monophosphate dehydrogenase OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=IMH3 PE=1 SV=2 - [IMDH_CANAL]	254.18	179.69	56.2
O42766	14-3-3 protein homolog OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=BMH1 PE=3 SV=2 - [1433_CANAL]	253.36	152.34	29.5
Q59VP2	Histone H2A.2 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=HTA2 PE=3 SV=3 - [H2A2_CANAL]	252.62	399.54	13.8
Q59PT0	V-type proton ATPase subunit B OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=VMA2 PE=3 SV=1 - [VATB_CANAL]	232.93	281.46	57.2
P83780	Glucose-6-phosphate isomerase OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=PGI1 PE=1 SV=2 - [G6PI_CANAL]	232.14		61.1
Q59VR3	FK506-binding protein 3 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=FPR3 PE=3 SV=1 - [FKBP3_CANAL]	227.82	291.52	47.6
Q96VB9	Heat shock protein homolog SSE1 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=MSI3 PE=1 SV=2 - [HSP7F_CANAL]	206.75	111.18	78.5
Q5A455	Protein transport protein SEC23 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=SEC23 PE=3 SV=2 - [SEC23_CANAL]	196.37	488.12	85.6
Q59MQ0	Myosin-5 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=MYO5 PE=1 SV=2 - [MYO5_CANAL]	137.52	194.96	146.9
Q5A416	tRNA (adenine(58)-N(1))-methyltransferase catalytic subunit TRM61 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=TRM61 PE=3 SV=1 - [TRM61_CANAL]	124.04	106.35	38.0
Q5AM60	Chitinase 4 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=CHT4 PE=3 SV=1 - [CHI4_CANAL]	86.15	59.04	44.3
Q59S78	Small COPII coat GTPase SAR1 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=SAR1 PE=3 SV=2 - [SAR1_CANAL]	85.60	213.36	21.5

Q59PL9	Eukaryotic translation initiation factor 3 subunit A OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=RPG1A PE=3 SV=2 - [EIF3A_CANAL]	80.45	224.01	106.0
Q5AJD0	ATP-dependent RNA helicase DBP5 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=DBP5 PE=3 SV=2 - [DBP5_CANAL]	78.63	164.64	60.2
Q59PR9	Transcriptional regulator HMO1 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=HMO1 PE=2 SV=1 - [HMO1_CANAL]	65.14	107.03	24.8
P83783	Adenosylhomocysteinase OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=SAH1 PE=1 SV=2 - [SAHH_CANAL]	59.76		49.0
Q5APF2	GMP synthase [glutamine-hydrolyzing] OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=GUA1 PE=3 SV=1 - [GUAA_CANAL]	59.45		58.8
P39827	Cell division control protein 10 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=CDC10 PE=2 SV=2 - [CDC10_CANAL]	55.65		40.7
Q5AK59	ATP-dependent RNA helicase HAS1 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=HAS1 PE=3 SV=1 - [HAS1_CANAL]	54.97	56.71	63.0
Q5AF03	Glyoxalase 3 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=GLX3 PE=1 SV=1 - [HSP31_CANAL]	53.95		25.8
Q92210	Phosphoribosylaminoimidazole carboxylase OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=ADE2 PE=3 SV=2 - [PUR6_CANAL]	40.92		62.4
Q92410	Alpha,alpha-trehalose-phosphate synthase [UDP-forming] OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=TPS1 PE=1 SV=1 - [TPS1_CANAL]	38.27	40.80	54.4
Q5A8X9	IMP-specific 5'-nucleotidase 1 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=ISN1 PE=3 SV=1 - [ISN1_CANAL]	37.60		52.4
Q9Y872	Sulfate adenylyltransferase OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=MET3 PE=3 SV=2 - [MET3_CANAL]	28.64		58.8
P43063	Cyclin-dependent kinase 1 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=CDC28 PE=1 SV=1 - [CDK1_CANAL]	15.96		36.6
P22274	ADP-ribosylation factor OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=ARF1 PE=3 SV=5 - [ARF_CANAL]		192.28	20.2

Q59TU0	Nascent polypeptide-associated complex subunit beta OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=EGD1 PE=3 SV=1 - [NACB_CANAL]		55.12	17.0
Q5AHH4	Small heat shock protein 21 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=HSP21 PE=1 SV=1 - [HSP21_CANAL]		33.90	21.5
Q5AEF2	Protein transport protein SEC13 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=SEC13 PE=1 SV=2 - [SEC13_CANAL]		41.60	33.0
Q59X38	Pescadillo homolog OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=NOP7 PE=1 SV=2 - [PESC_CANAL]		26.64	67.8

Chapter IV

Classification of epithelial CEACAMs – their role in human-specific colonization by commensal gut bacteria

Patrizia Bonsignore[1], Alexandra Roth[1], Tancred Frickey[2,3], Christof R. Hauck[1,3]

[1] Lehrstuhl für Zellbiologie, Universität Konstanz, 78457 Konstanz, Germany

[2] Forest Industry Informatics, Scion, 3015 Rotorua, New Zealand

[3] Konstanz Research School - Chemical Biology, Universität Konstanz, 78457 Konstanz, Germany

Abstract

The human gut harbours an enormous amount and diversity of microorganisms that live in a symbiotic relationship with their host. To perform essential tasks, such as nutrition digest, the human microbiota requires a characteristic composition. The basis for a human-specific microbe community is formed by two major factors: nutrition and host genetics. In search of host-specific parameters, we investigate the carcinoembryonic antigen-related cell adhesion molecule (CEACAM) family. This human protein family is known to provide receptors for host colonization by human-restricted pathogens and has a high diversification in mammals, therefore representing a factor that might help to shape a human-characteristic microbiota. In this study, CEACAM-associating commensal gut bacteria were enriched out of human stool samples employing magnetic beads coated with recombinant human CEACAM1 and CEA. The enriched bacteria were identified using 16S rRNA gene pyrosequencing. Two major phyla were detected: Bacteroidetes and Firmicutes. Most identified species within the Bacteroidetes belonged to the genus *Prevotella*, while the Firmicutes exhibited a very heterogeneous set of genera and classes. In a biochemical approach, the CEACAM-binding spectrum of identified candidates was analyzed using different recombinant CEACAM constructs. Bacterial species that are almost exclusively found in the human body showed restriction to human CEACAMs. In contrast, *Enterococcus faecalis*, known to colonize not only human but also mouse intestine, possessed a broadened binding spectrum that comprised also other mammalian CEACAM proteins. Infection of CEACAM-expressing cells with commensal bacteria and analysis via confocal microscopy revealed no difference in bacterial interaction on the cellular surfaces. However, our data clearly show a specific binding of commensal microorganisms to soluble CEACAM family members. Remarkably, those microorganisms turned out to modify CEACAMs during interaction resulting in lower CEACAM molecule masses. Fluorescently labeled sugar residues incorporated by CEACAM molecules were retrieved from *E. faecalis* proteins after CEACAM interaction. We suggest that CEACAMs may not be involved in early cell surface association processes, but that interaction with host-specific CEACAM molecules and subsequent degradation of linked carbohydrate moieties could support long-term colonization by providing an additional docking site and energy source.

Introduction

The human body is colonized by a complex community of microorganisms composed of bacteria, fungi, archaea, protozoa, and viruses that reside on or within their host. The highest number and diversity of the so-called *human microbiota* is found in the human gut, where it fulfills various physiological functions such as nutrition digest, immune system development or defence against pathogens. The additional set of genes provided by commensal microorganisms broadens the enzymatic spectrum of the digestive system, allowing host-indigestible polysaccharide degradation, the production of short chain fatty acids, and the synthesis of vitamins. Thereby, energy sources are acquired that otherwise would be inaccessible for human metabolism. The human microbiota stimulates and calibrates the human immune system and builds a protective barrier on mucosal surfaces impeding tissue invasion by pathogens and furthermore hindering their colonization by competing for the same ecological niches (Neish 2014; Trompette, Gollwitzer et al. 2014) Discrimination of commensal from pathogenic microbes is an important ability, acquired by the human immune system during development, that allows maintenance of the symbiotic relationship between microbiota and host and defense against harmful pathogens (Belkaid and Hand 2014).

The human gut microbiota is mainly composed of strict anaerobic organisms, which outnumber aerobes and facultative anaerobes by far (Gordon and Dubos 1970; Savage 1970). Although, there are more than 50 bacterial phyla, only two main phyla prevail in the human gut: the Bacteroidetes and the Firmicutes (Schloss and Handelsman 2004). Few bacteria affiliate to other phyla such as Proteobacteria, Actinobacteria, Fusobacteria, and Verrucomicrobia (Eckburg, Bik et al. 2005). Homeostasis of the human gut microbiota establishes early in life but can be affected by various factors during the whole life time of the host. While geography, sex, age, social structures and medication use have been associated with gut microbiota diversity and number, the identity of microbiota members is defined by personal nutrition and host evolutionary history (Ley, Peterson et al. 2006; Groussin, Mazel et al. 2017). Although diet rapidly and reproducibly alters the microbiome (Faith, McNulty et al. 2011; David, Maurice et al. 2014), individuals possess unique microbiota compositions that are determined by host genetics rather than nutrition (Goodrich, Davenport et al. 2017; Sasson, Kruger Ben-Shabat et al. 2017). Many studies have shown a co-evolutionary process between host and microbiota (Ley, Hamady et al. 2008; Ochman, Worobey et al. 2010;

Goodrich, Davenport et al. 2017; Sasson, Kruger Ben-Shabat et al. 2017). The fact that each mammalian species has its own characteristic microbial community, points toward the existence of host-specific factors, which allow colonization and composition in a species-specific manner. In contrast to the well-studied field of species-specific colonization by pathogens, such factors are widely unknown for commensal microorganisms. Human-restricted bacteria such as *Haemophilus influenzae, Helicobacter pylori, Neisseria gonorrhoeae,* or *Neisseria meningitidis* utilize epithelial surface-expressed members of the carcinoembryonic antigen-related cell adhesion molecule (CEACAM) family for this purpose. The CEACAM family is a fast evolving group of mammalian immunoglobulin-related glycoproteins that has a high diversification in mammals and therefore represents a suitable species-specific factor. They consist of an amino-terminal immunoglobulin variable (Ig_V)-like domain followed by varying numbers of immunoglobulin constant (Ig_C)-like domains. They are expressed by epithelial cells, endothelial cells, as well as hematopoietic cells, and can be found on various tissues (Virji, Makepeace et al. 1996; Gray-Owen, Dehio et al. 1997). The Ig_V-like domain of epithelial CEACAM1, CEA and CEACAM6 is bound by microbes and enables host colonization. Interestingly, CEACAM-associating bacteria express structurally distinct adhesive proteins to associate with human CEACAM molecules (e.g. Opa-proteins in *Neisseria*, OMP P1 in *Haemophilus*, UspA1 in *Moraxella*, HopQ in *Helicobacter*, CbpF in *Fusobacterium*), suggesting convergent adhesin evolution (Chen, Grunert et al. 1997; Hill and Virji 2003; Tchoupa, Lichtenegger et al. 2015; Javaheri, Kruse et al. 2016; Brewer, Dymock et al. 2019). CEACAMs are involved in diverse physiological events. Most participate in the modulation of general cellular processes such as cell adhesion, differentiation, proliferation, and survival (Tchoupa, Schuhmacher et al. 2014). To perform these functions, CEACAMs interact with other receptors to transmit signals into cells. This signaling property is utilized by CEACAM-binding bacteria. CEACAM-clustering around cell surface-associated microbes initiates a similar cellular response as it is observed during physiological CEACAM stimulation processes, including cytoskeleton rearrangement, gene expression, enhanced cell adhesion, and endocytosis (Tchoupa, Schuhmacher et al. 2014). Modulation of host cell signals via the immunoglobulin receptors improves mucosal surface colonization and enables prolonged survival within the host. Besides pathogens, some commensal species of the genus *Neisseria*, such as *N. lactamica* and *N. subflava,* express different opacity-associated (Opa) proteins that allow tight connection to human CEACAM1, CEA, and CEACAM6 which enhances epithelial cell

adhesion (Toleman, Aho et al. 2001; Muenzner, Rohde et al. 2005; Muenzner, Bachmann et al. 2010).

In this study we focus on the potential of epithelial CEACAM family members as species-specific colonization factors. First, we aimed to enrich novel CEACAM-binding bacteria derived from the human gut microbiota. Therefore, magnetic beads coated with human CEACAM1 and CEA were incubated with a pool of human stool samples to enrich CEACAM binding bacteria. Subsequently, the identity of bead-associated bacteria was determined by next generation sequencing of the 16S rRNA genes. Especially members of the classes Clostridia, Bacilli and Bacteroidia were identified. In CEACAM-binding studies, a versatile set of the identified commensal bacteria were analyzed for their CEACAM-binding capacity by incubating them with different CEACAM constructs. They revealed distinct binding specificities for CEACAM1, CEACAM3, and CEACAM6. However, all species interacted with CEA, a CEACAM-member exclusively expressed by epithelial cells. Infection of human CEA-expressing cells with commensal microbes exhibited no elevated numbers of bacteria on host cell surfaces, suggesting a minor role for CEACAMs during the early host colonization process itself. Strikingly, CEACAM molecules showed a reduced molecular weight after association with commensals, indicating receptor modifications. We could show that carbohydrate moieties of the highly glycosylated CEACAM proteins were degraded during bacteria interaction. Fluorescence labelling of CEACAM-integrated sugar residues allowed us to retrace these sugar molecules after degradation and revealed that the gut commensal *E. faecalis* integrated CEACAM-derived carbohydrates partially into own proteins. Degradation of those carbohydrate moieties could serve as food source, opening a niche for CEACAM-binding commensals. Our data clearly show the binding of CEACAM family members by commensal microorganisms, and suggest that those receptors may not be involved in early cell surface association processes, but instead in long-term colonization by providing an additional docking site and energy source.

Results

Enrichment of CEACAM-associating bacteria from human stool samples

In this study we focus on members of the fast evolving CEACAM glycoprotein family and their potential to serve as commensal microbe interaction partners. First, we aimed to enrich novel CEACAM-binding bacteria derived from the human gut microbiota. For this purpose, we established an enrichment technique based on magnetic beads coated with the bait protein. Amino-terminal domains of human CEACAM1 and CEA fused to GFP were bound to anti-GFP-coupled magnetic beads and incubated with a pool of human stool samples. Subsequently, the beads were loaded onto a column placed in a magnetic field and were washed thoroughly to remove unbound bacteria (Figure IV-1). The identity of bead-associated bacteria was determined by bacterial DNA isolation and amplification of the variable regions 1 and 2 (V1-V2) of the 16S rRNA genes via PCR, utilizing the universal primers 27F and 338R. The 16S rRNA amplicons were pyrosequenced with a Roche 454 FLX Titanium sequencer resulting in about 170.000 high quality sequences with an average length of 360 bp (Dissertation Roth, 2013).

Bioinformatics analysis of 454 pyrosequencing results revealed human CEACAM-binding bacteria to originate from diverse classes

To identify human CEACAM1 and CEA-enriched bacteria we compared the bacterial 16S rRNA 454 pyrosequencing results to 16S rRNA sequences present in the SILVA database. All SILVA small subunit (SSU)-rRNA sequences were clustered using CD-hit at 95% identity. Representatives of each cluster and human CEACAM1 and CEA-enriched sequences were combined and analyzed using CLuster ANalysis of Sequences (CLANS) (Frickey and Lupas 2004). CLANS uses a graph layout algorithm to visualize pairwise sequence similarities (all-against-all BLAST sequence comparison) and depicts them as dots in a three-dimensional space (Figure IV-2 A). Based on the reciprocal BLAST hits, each sequence is assigned an 'attraction' value to every other sequence. In a three-dimensional space, graph sequences similar to each other are located in close proximity whereas less similar sequences appear more distantly from each other. Utilizing the three-dimensional CLANS map model, we were able to place the human CEACAM1 and CEA-enriched 16S rRNA sequences into a taxonomic

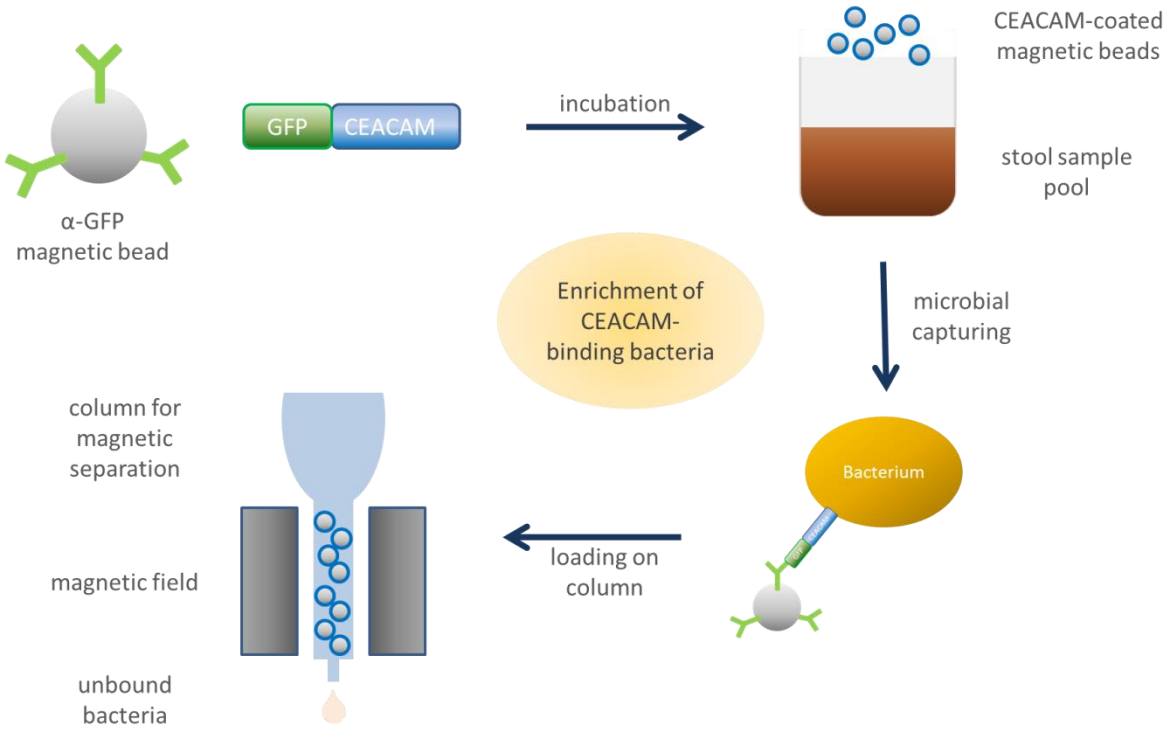


Figure IV-1: Schematic work-flow of CEACAM-binding bacteria enrichment from human stool samples using magnetic beads. GFP-tagged Ig_V-like domains of human CEACAMs were linked to anti-GFP antibody-coupled magnetic beads and incubated with a pool of fresh stool samples to capture CEACAM-recognizing bacteria. The magnetic beads were loaded onto a column placed in a magnetic field. Unbound bacteria were washed away and bead-associated bacteria were analyzed.

context (Figure IV-2 A). Accordingly, our sequences clustered into a diverse set of classes such as *Bacteroidia* or *Clostridia*, that mainly belonged to the two gut main phyla Bacteroidetes and Firmicutes. Some clusters were located within the phylum Proteobacteria which comprises many opportunistic pathogens including all currently known CEACAM-binding bacteria. In further investigations, members of the phylum Proteobacteria were excluded and we focused on the gut main phyla Bacteroidetes and Firmicutes. Remarkably, most identified Bacteroidetes species belonged to the genus *Prevotella*. In contrast, species of the Firmicutes phylum exhibited a heterogeneous group of genera and classes. In subsequent investigations we focus on the genus *Prevotella* as representative for the phylum Bacteroidetes and a diverse set of genera such as *Clostridium, Enterococcus* and *Lactobacillus* as representatives for the phylum Firmicutes.

A versatile set of commensal microbes interact with human CEACAM molecules

The human gut is a reduced environment sheltering mainly strict anaerobe microbes. The genus *Prevotella* also belongs to this obligate anaerobic community, therefore species had to be cultured under oxygen-free conditions. Seven *Prevotella* species as well as the identified closely related species *Bacteroides fragilis* were tested for their ability to interact with different human members of the CEACAM family. Therefore, soluble Ig_V-like domains of human CEACAMs fused to GFP were utilized for binding studies. Bacteria species were incubated with the soluble CEACAM constructs and washed extensively to remove unbound receptors. Associated molecules were detected via anti-GFP antibodies in immunoblotting. The tested bacteria species had distinct binding specificities for CEACAM1, CEACAM3, and CEACAM6. However, all were able to interact with CEA, a protein exclusively expressed by epithelial cells (Figure IV-2 B) (Table IV-1). In contrast to the CEACAM-enriched Bacteroidetes species, the phylum Firmicutes revealed not a single dominant genus but a highly diverse set of different genera belonging to different classes. We investigated representatives of the two most prominent Firmicutes classes Bacilli (*Lactobacillus*, *Enterococcus*) and Clostridia (*Clostridium*), as well as the classes Erysipelotrichia (*Catenibacterium*) and Negativicutes (*Megasphaera*). Similar to Bacteroidetes species, all tested Firmicutes species had distinct binding specificities for CEACAM1, CEACAM3, and CEACAM6, but were all able to interact with CEA (Figure IV-2 C) (Table IV-1). The phylum Bacteroidetes mainly includes Gram-negative bacteria, while the phylum Firmicutes mainly comprises Gram-positive members. To date, only Gram-negative bacteria were known to bind CEACAM molecules. Here, we identify for the first time Gram-positive CEACAM-interacting species.

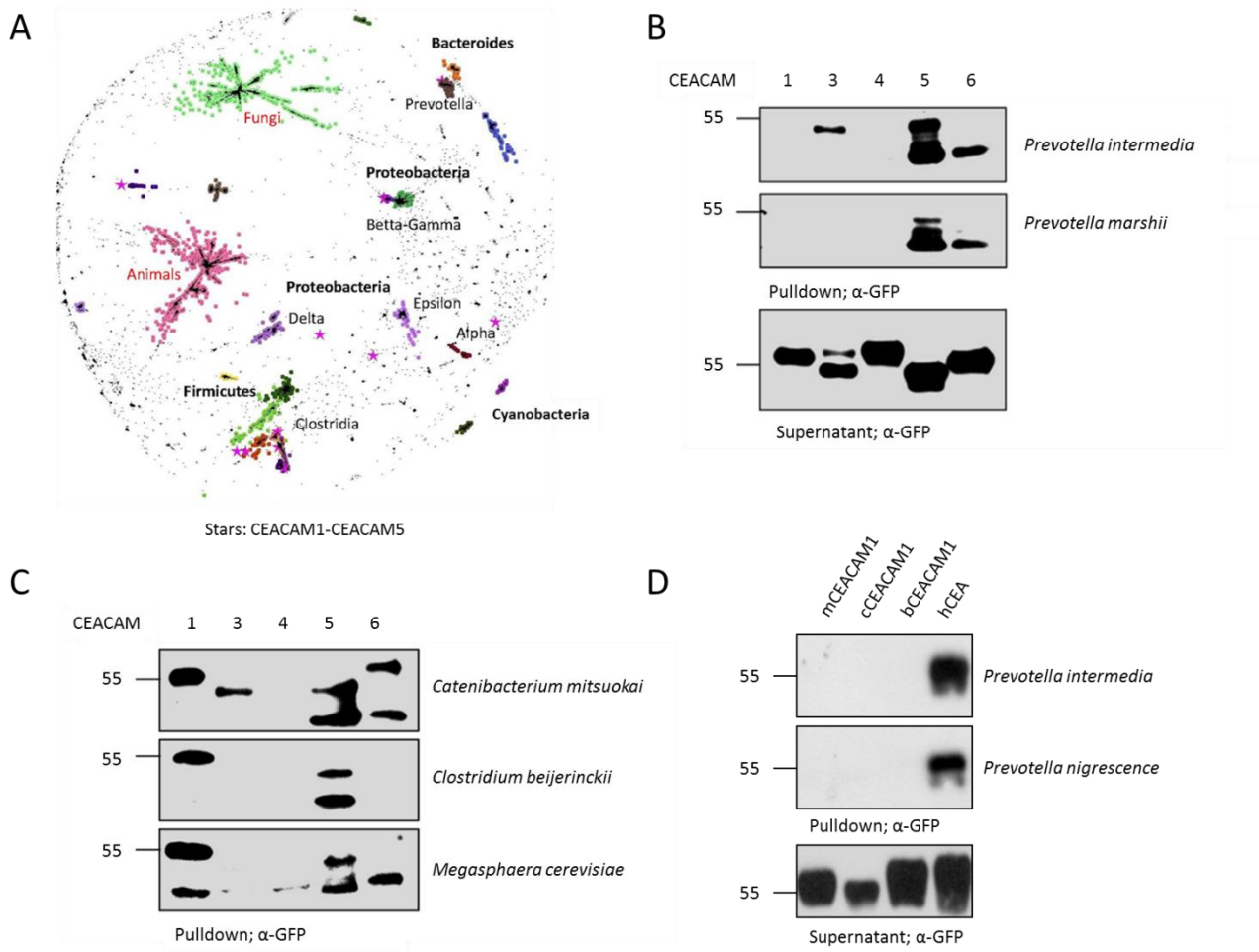


Figure IV-2: Commensal microbes identified from human stool samples specifically interact with human CEACAM molecules. (A) 3D-CLANS map of the SILVA database. Utilizing the three-dimensional CLANS map model, human CEACAM1 and CEA-enriched 16S rRNA sequences were analyzed and compared to the SILVA database and placed into a taxonomic context. Identified clusters are highlighted by purple stars. (B, C) GFP-tagged Ig_V-like domains of human CEACAM1, CEACAM3, CEACAM4, CEA and CEACAM6 were generated as secreted proteins in HEK293T cells. Equivalent amounts of the different soluble Ig_V-like domains (supernatant – lowest panel) were incubated with commensal representatives of Bacteroidetes (*P. intermedia*, *P. marshii*) and Firmicutes (*C. mitsuokai*, *C. beijerinckii*, *M. cerevisiae*). (D) GFP-tagged Ig_V-like domains of murine, canine, bovine CEACAM1 and human CEA were incubated with the commensal bacteria *Prevotella intermedia* and *Prevotella nigrescence* (Dissertation A. Roth, 2013). CEACAM-binding was determined upon bacterial pulldown and anti-GFP antibodies in Western blot.

Prevotella targets CEACAM proteins in a species-dependent manner

Although the same bacterial phyla largely predominate in mammalian guts, the relative abundance of different clades varies between different mammals. Thus, each mammalian species has its own characteristic microbiota. At lower taxonomic levels (family, genus), microbes are more likely to associate with a particular mammalian lineage (Nishida and Ochman 2018). Some members of the microbiota are distributed equally in different hosts. For instance, *Enterococcus faecalis* can be found as a prominent inhabitant in a variety of mammalian gut microbiotas (e.g. human, dog, mouse) (Hammerum 2012). However, other bacteria such as *Prevotella intermedia* (human) or *Prevotella ruminicola* (cattle, sheep) are predominant in particular host species (Batt and Tortorello 2015). We tried to answer the question if commensal bacteria that are regularly found in the human microbiota, but are mainly absent in other mammalian species, bind to CEACAM proteins in a human-specific manner. Therefore, binding studies were conducted by incubating the human commensals *Prevotella intermedia* and *Prevotella nigrescence* with CEACAM constructs derived from human and other mammalian species. *Prevotella* species bind to human CEA, however, homologues of the CEACAM5 gene (CEA protein) are absent in non-primate mammals. Therefore the most widely expressed member of the CEACAM family, CEACAM1, was deployed for non-human species for interaction studies (Kammerer and Zimmermann 2010). Only the soluble Ig_V-like domain of human CEA but not CEACAM1 domains of mouse, dog and cattle were recognized by *P. intermedia* or *P. nigrescence* hinting to human CEACAM-specific interactions (Figure IV-2 D).

Table IV-1: Overview of commensal microbes and their CEACAM-binding spectrum. Gram-negative (Bacteroidetes, Proteobacteria) and Gram-positive (Firmicutes) bacteria were tested for their interaction capacity with epithelial (CEACAM1, 5, 6) and granulocyte CEACAMs (CEACAM1, 3, 4, 6, 8).

Phylum	Genus	Species	CEACAM					
			1	3	4	5	6	8
Bacteroidetes	*Bacteroides*	*fragilis*	-	NC	-	+	NC	NC
	Prevotella	*intermedia*	-	+	-	+	+	+
		nigrescens	-	+	-	+	-	+
		corporis	-	+	-	+	-	+
		marshii	-	-	-	+	+	NC
		copri	-	NC	-	+	NC	NC
		timonensis	+	NC	-	+	NC	NC
		nanceiensis	NC	NC	-	+	NC	NC
Firmicutes	*Catenibacterium*	*mitsuokai*	+	+	-	+	+	NC
	Clostridium	*beijerinckii*	+	-	-	+	-	NC
	Enterococcus	*faecalis*	-	-	-	+	-	NC
		hirae	-	NC	NC	+	NC	NC
	Megasphaera	*cerevisiae*	+	-	-	+	+	NC
	Lactobacillus	*brevis*	NC	NC	NC	+	-	NC
		murinus	NC	NC	NC	+	-	NC
Proteobacteria	*Escherichia*	*coli*	-	-	-	-	-	-
	Haemophilus	*influenzae*	+	-	-	+	-	-
	Moraxella	*catarrhalis*	+	+	-	+	+	-
	Neisseria	*meningitidis*	+	+	-	+	+	-

(+): binding (-): no binding NC: not yet characterized

Human CEA alone is not sufficient to mediate cell surface colonization by commensal bacteria

The ability of commensal gut bacteria to interact with human CEACAMs appears to be a phylogenetically widely distributed feature. To understand the physiological role of this phenomenon we investigated the capacity of bacteria-CEACAM interaction to mediate cell surface colonization. The epithelial receptor, human CEA, constitutes the family member with the highest frequency in the human gut, thus HEK cells were transfected with a wild-type human CEA construct. Subsequently, human CEA-expressing HEK cells were infected with *Enterococcus faecalis*, a Gram-positive resident of the human gut that exclusively binds to human CEA but not to other human derived CEACAMs (Figure IV-3 A). For further detection, *E. faecalis* was stained using carboxytetramethylrhodamine succinimidyl ester (TAMRA), a dye that covalently links to free amine groups on the bacterial surface. To exclude any impact on CEACAM-interaction by the modified lysine residues on the *Enterococcus* surface after staining, previous to HEK cell infection, binding studies were conducted utilizing TAMRA stained *E. faecalis* and soluble human CEAs. Unstained *E. faecalis*, human CEACAM6 and the non-binder *Escherichia coli* served as controls (Figure IV-3 B). Although human CEA-binding was not hampered by TAMRA-staining, infection of human CEA-expressing HEK cells showed no increased cell surface association of *E. faecalis* compared to wild-type HEK cells. Interestingly, human CEA molecules seem to cluster around the surface attached bacteria (Figure IV-3 C).

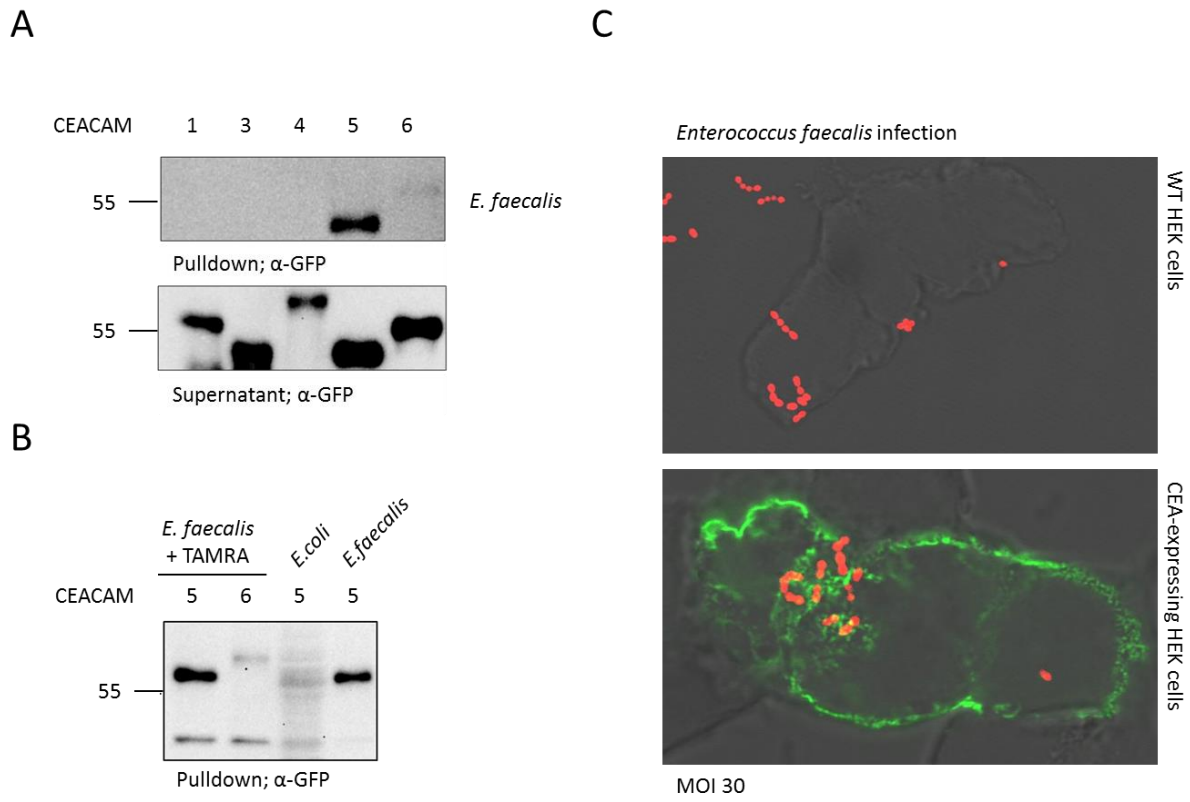


Figure IV-3: *Enterococcus faecalis* colonizes CEACAM-expressing cells and wild-type cells equally. (A) GFP-tagged Ig_V-like domains of human CEACAM1, CEACAM3, CEACAM4, CEA and CEACAM6 were generated as secreted proteins in HEK293T cells. Equivalent amounts of the different soluble Ig_V-like domains (supernatant – lower panel) were incubated with *Enterococcus faecalis*. (B) *E. faecalis* was stained with TAMRA and incubated with GFP-tagged Ig_V-like domains of human CEA or CEACAM6. *E. coli* (negative control) and unstained *E. faecalis* (positive control) were incubated with GFP-tagged CEA. CEACAM-binding was determined upon bacterial pulldown and anti-GFP antibodies in Western blot. (C) Wild-type HEK293T cells or human CEA-expressing HEK293T cells (green) were infected with TAMRA-stained *E. faecalis* (red) for 1 h (MOI 30). Cells were analyzed by confocal microscopy.

Commensal bacteria process CEACAM molecules upon binding

The conducted CEACAM-binding studies often exhibited a striking byproduct: a second protein signal with a lower molecular weight (roughly 15 kDa smaller) that was completely absent in CEACAM loading controls (Figure IV-4 A). This additional signal was visualized by anti-GFP antibodies and most likely represents a modified CEACAM-GFP construct, lacking a part of the CEACAM glycoprotein, degraded by commensal bacteria during the binding process. Microorganisms express a variety of enzymes such as proteases or glycosidases e.g. for their own protection or for the acquisition of energy from nutrition. The highly glycosylated CEACAM proteins either loose a part of the protein backbone or their carbohydrate chains are affected in the observed degradation process. In further investigations, additionally to the anti-GFP antibody, a CEACAM-specific antibody, that solely detects human CEACAMs, was used to visualize CEACAM molecules during immunoblotting. *E. faecalis* was incubated with human CEA-NT-GFP, mouse CEACAM1-NT-GFP and GFP alone (negative control). Immunoblotting with anti-GFP antibodies showed that *E. faecalis* interacted with both receptors human CEA and mouse CEACAM1. Furthermore, a signal of the degraded CEACAM construct could be observed for both, human and mouse CEACAM. However, the anti-CEACAM antibody only recognized the intact receptor of human CEA at around 55 kDa. Expectedly, the mouse CEACAM1 protein was not detected by the anti-human CEACAM antibody but also a signal for the degraded CEA form with a molecular weight of around 40 kDa was absent (Figure IV-4B). The processed CEACAM molecules seem to have lost the part that serves as epitope for the human CEACAM-specific antibody hinting to protein degradation. To confirm this hypothesis, we next tried to block or increase protein degradation by inhibiting or promoting protease activity, respectively. For this purpose, the bacteria-CEACAM incubation time was decreased from 1 h to 10 min and conducted at different temperatures. Lower temperature should slow protease activities down, thus the amount of the degradation product should be smaller compared to higher temperatures. In contrast, higher temperatures should accelerate protease activity and therefore enhance the amount of the degradation product. In addition, we blocked protease activity completely by adding the metalloprotease inhibitor EDTA, the aspartic protease inhibitor Pepstatin A (Figure IV-4 C), the serine protease inhibitors Pefabloc, Aprotinin, and PMSF as well as the cysteine, serine and threonine protease inhibitor Leupeptin (data not shown). *E. faecalis* degraded the

CEA-NT-GFP construct under all conditions similarly. Neither temperature nor protease inhibition showed any effect on CEACAM degradation levels (Figure IV-4 C).

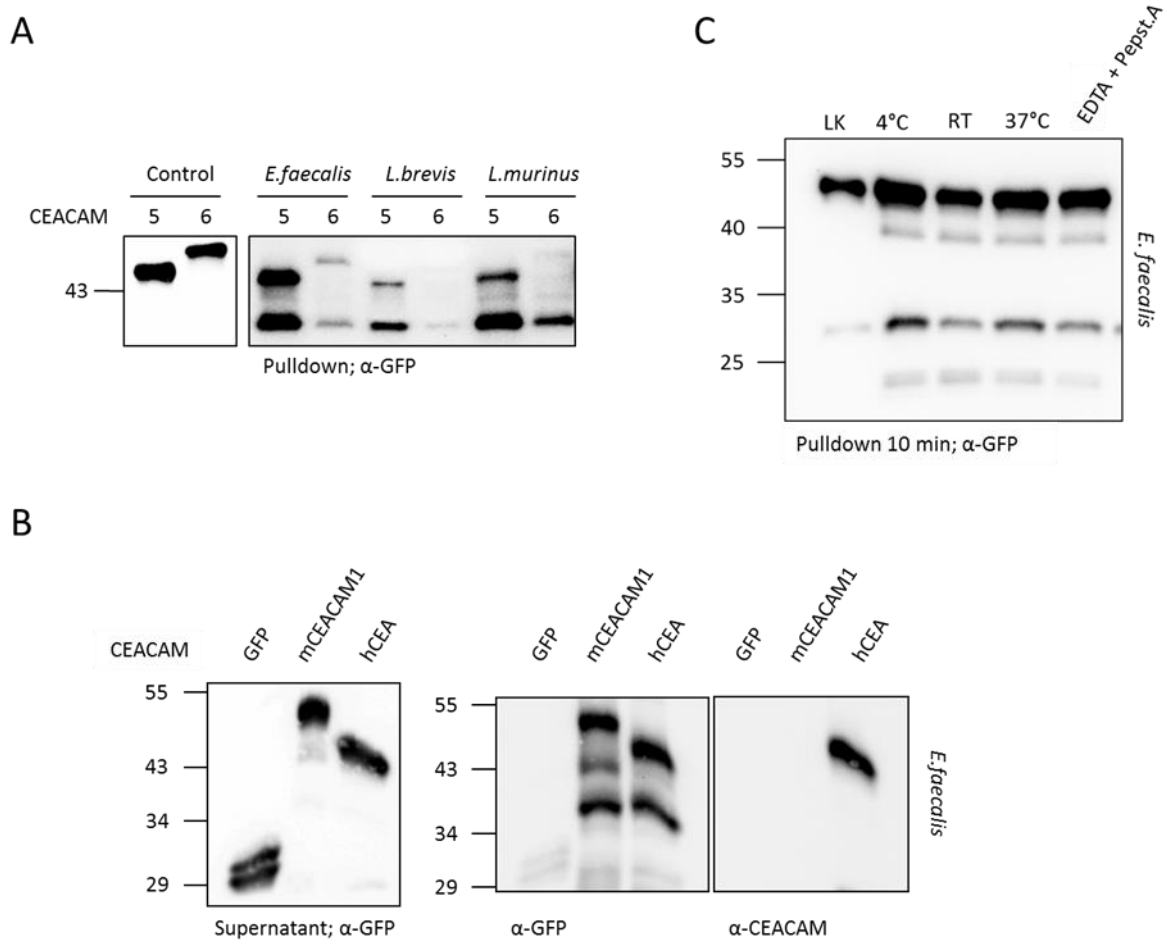


Figure IV-4: CEACAM degradation by commensal bacteria cannot be blocked by protease inhibitors. Equivalent amounts of GFP-tagged Ig_V-like CEACAM domains were used for binding studies with commensal bacteria. (A) *Enterococcus faecalis*, *Lactobacillus brevis* and *Lactobacillus murinus* were incubated with human CEA and CEACAM6. CEACAM-binding was determined upon bacterial pulldown and anti-GFP antibodies in Western blot (B) *E. faecalis* was incubated with GFP (negative control), mouse CEACAM1 and human CEA. CEACAM-binding was determined upon bacterial pulldown and subsequent Western blot either by using anti-GFP antibodies or anti-human CEACAM antibody. (C) *E. faecalis* was incubated with human CEA for 10 min under varying temperatures (4°C, RT, 37°C) in the presence of the protease inhibitors EDTA and Pepstatin A. CEACAM-binding was determined upon bacterial pulldown and anti-GFP antibodies in Western blot.

Commensal bacteria bind the CEACAM protein backbone and degrade carbohydrate structures of the glycoprotein

Carbohydrate residues of host glycoproteins are regularly degraded by commensal bacteria to acquire energy for the bacterial metabolism (Ouwerkerk, de Vos et al. 2013; Crouch, Liberato et al. 2020). CEACAM constructs show a clear reduction in molecular weight after bacteria interaction that seems to be independent of protease activities (Figure IV-4 C). Therefore, the possibility of sugar degradation in the highly glycosylated CEACAM proteins was investigated by conducting binding studies with the commensal gut bacterium *Enterococcus faecalis*. *E. faecalis* was grown in freshly prepared culture medium, rich in nutrition, and harvested in the exponential growth phase (after 3 h). At this particular time there should be no need to degrade CEACAM-carbohydrates for energy acquisition. A second *E. faecalis* sample was taken in stationary phase (after 24 h), when nutrients are mostly depleted. Both samples, the well fed and the starving *E. faecalis* were incubated with human CEA-NT-GFP and analyzed via immunoblotting. *E. faecalis* originating from the exponential growth phase bound to human CEA without processing it. *E. faecalis* harvested from stationary phase, however, showed two bands: the unprocessed CEACAM molecule (upper band) and the degraded form (lower band) (Figure IV-5 A), supporting the theory of CEACAM degradation for the purpose of energy acquisition. Interestingly, the exponentially grown *E. faecalis* was still binding to human CEA although it had no requirement for an additional food source suggesting a dual function for CEACAM molecules.

The CEACAM-NT-GFP constructs utilized in all binding studies are composed of the GFP-protein (27 kDa) fused to the N-terminal Ig_V-like domain of CEACAM proteins (12 kDa) and varying amounts (2-4) of N-linked carbohydrate chains (roughly 6-7 kDa each), depending on the CEACAM member. To strengthen the hypothesis of deglycosylation and exclude protein degradation, human CEACAM3 and the highly similar gorilla CEACAM3 construct were employed for binding studies. Human CEACAM3 exhibits two N-glycosylation sides, whereas gorilla CEACAM3 possesses a third N-glycosylation side in close proximity to the others (Figure IV-5 B). If parts of the CEACAM protein are cleaved, the molecular weight of the resulting constructs should differ due to the additional glycosylation side in gorilla CEACAM3 which roughly accounts for additional 6 kDa of the molecular weight. For the binding approach *Prevotella intermedia* was chosen as the genus *Prevotella* is specialized to carbohydrate

digestion and exhibits a large enzyme set for this purpose (Wu, Chen et al. 2011). Incubation of *Prevotella intermedia* with human CEACAM3 and gorilla CEACAM3 for 10 min, 30 min and 60 min showed a decreasing amount of the unprocessed CEACAM molecule over time but a nearly constant level of the processed variant (Figure IV-5 B). Human CEACAM3 and gorilla CEACAM3 both showed an additional signal for CEACAM-GFP with a molecular weight of roughly 40 kDa, excluding protein degradation but strengthening carbohydrate removal. Interestingly, human CEACAM3 and especially gorilla CEACAM3 showed additional bands with intermediate molecular weights between the upper, unprocessed CEACAM3 and the lower, processed one (Figure IV-5 B).

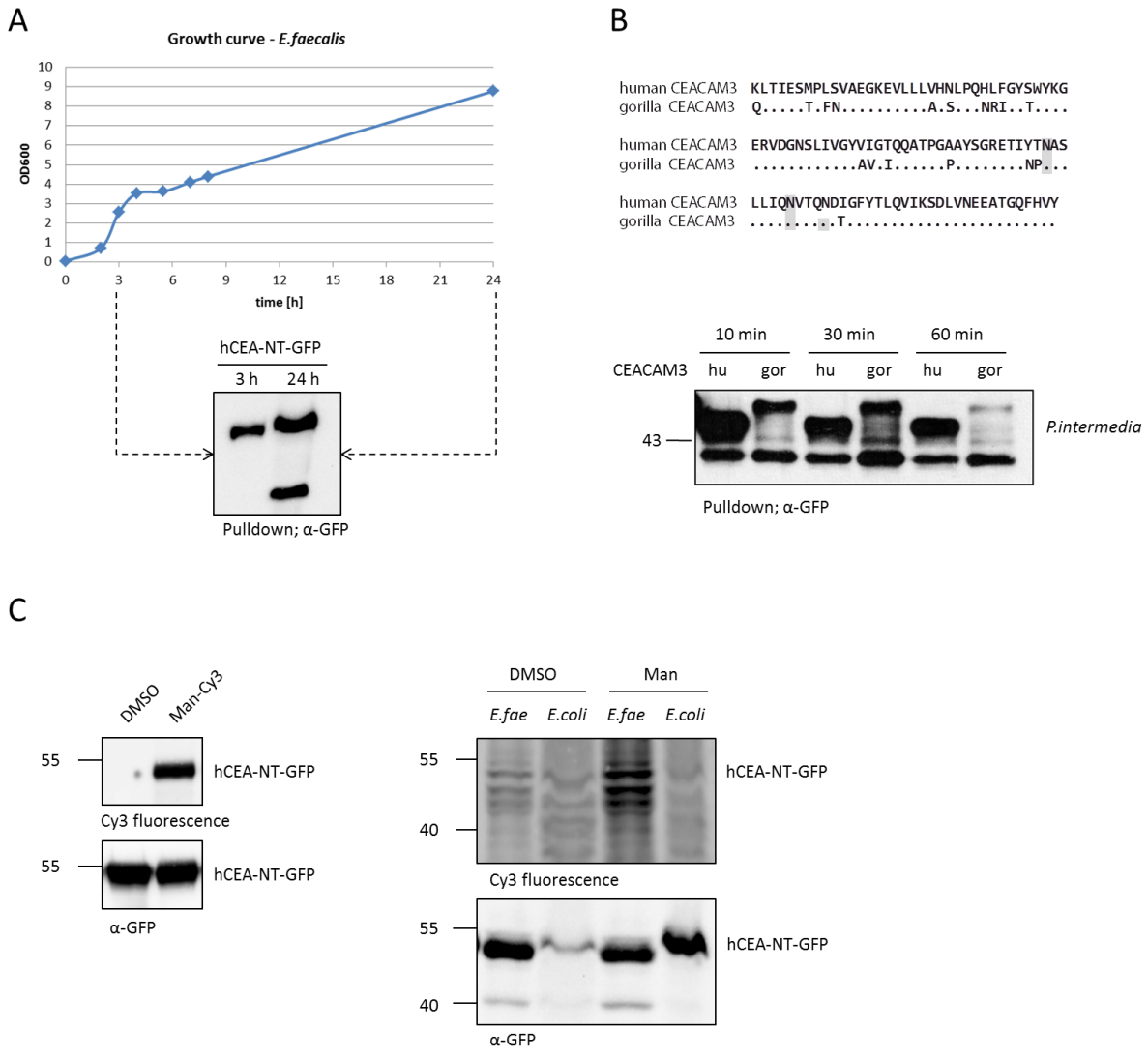


Figure IV-5: Commensal bacteria degrade carbohydrate structures of CEACAMs and utilize them in metabolism. (A) *E. faecalis* was grown for 24 h in fresh culture medium. Bacteria

samples were taken after 3 h (exponential growth phase) and 24 h (stationary growth phase). Equivalent amounts of CEA-NT-GFP domains were used for binding studies with *E. faecalis*. (B) Amino acid sequence alignment of the Ig_V-like domain of human CEACAM3 and gorilla CEACAM3. Grey boxes mark N-glycosylation sites. Equivalent amounts of human CEACAM3-NT-GFP domains or gorilla CEACAM3-NT-GFP domains were used for binding studies with *Prevotella intermedia* for different incubation times (10 min, 30 min, 60 min), respectively. (C) Mannose-C_4H_6 or DMSO was added to CEA-NT-GFP expressing HEK293T cells. Incorporated mannose-C_4H_6 residues were fluorescently labelled by Cy3-tetrazine and analyzed either by immunoblotting using anti-GFP antibodies (α-GFP) or by the fluorescence imaging tool Typhon FLA 9500 (Cy3 fluorescence). Equivalent amounts of CEA-mannose-C_4H_6 or CEA-DMSO (control) were used for binding studies with *E. coli* (CEACAM non-binder) or *E. faecalis*. CEACAM-binding was determined upon bacterial pulldown and anti-GFP antibodies in Western blot.

Enterococcus faecalis metabolizes carbohydrates derived from human CEA

To investigate the fate of CEACAM-derived carbohydrates after bacteria interaction, we applied a method that allowed fluorescent labelling of sugar residues and subsequent detection by a fluorescence imaging tool. For this purpose, HEK293T cells were transfected with the human CEA-NT-GFP containing plasmid. The day after transfection, cells were fed with the monosaccharide mannose that was linked to a methylcyclopropene (mannose-C_4H_6). Since mannose-C_4H_6 was solved in DMSO, we added the same amount of DMSO without mannose-C_4H_6 to CEA-NT-GFP-expressing control cells. After 48 hours the cells had inserted mannose-C_4H_6 residues into human CEA-NT-GFP glycoproteins and secreted the constructs into cell culture medium. The medium was harvested and CEACAM molecules were incubated with Cy3-tetrazine to label the incorporated mannose-C_4H_6 residues. Tetrazine reacts with methylcyclopropene in the inverse electron demand Diels–Alder reaction and thereby covalently links the fluorescent dye Cy3 to mannose residues. CEACAM molecules were blotted to a membrane and were either visualized via immunoblotting (α-GFP) or via a fluorescence imaging tool (Cy3 fluorescence) (Figure IV-5 C). CEACAM constructs synthesized in the presence of mannose-C_4H_6 depicted similar expression levels as constructs synthesized with DMSO-only. Expectedly, only CEACAMs with incorporated mannose-Cy3 showed

fluorescence signals (Figure IV-5 C). In a next approach, regular CEA-NT-GFP and mannose-CEA-NT-GFP both were incubated with the non-binding bacterium *Escherichia coli* and the CEA-binder *Enterococcus faecalis*. After removal of unbound CEACAM molecules, the samples were boiled and all proteins (bacteria and associated CEACAMs) were blotted onto a membrane for further analysis. Probing with anti-GFP antibodies showed *E. faecalis* to bind to both, regular CEA and mannose-CEA. Expectedly, the CEACAM non-binder *E. coli* did not interact with regular CEA. However, *E. coli* was able to interact with mannose-CEA. Remarkably, imaging Cy3-fluorescence of the samples revealed some protein bands of *E. faecalis* to be fluorescently labelled after the interaction with CEACAM-mannose-Cy3 (Figure IV-5 C). Therefore, *E. faecalis* seems to deglycosylate human CEA and utilize mannose-Cy3 molecules in metabolism. Mannose-Cy3 appears to be incorporated into *E. faecalis* proteins during bacterial glycoprotein synthesis.

Discussion

The fact that each mammalian species has its own characteristic microorganism community points toward the existence of host-specific factors, which allow host colonization and variety in the microbe composition in a species-specific manner (Nishida and Ochman 2018). In search of host-specific colonization parameters, this study investigated the carcinoembryonic antigen-related cell adhesion molecule (CEACAM) family. CEACAMs are in constant change as they evolve rapidly. The ancestral *CEACAM1* gene, for instance, underwent multiple duplications, which resulted in widely species-specific distribution of CEACAM1-related members of the CEA gene family in mammals (Kammerer, Popp et al. 2007). Furthermore, CEACAMs are already known to serve as colonization factor in the human host for Gram-negative pathogens such as *Neisseria gonorrhoea* or *Haemophilus influenzae,* which infect humans exclusively (Virji, Makepeace et al. 1996; Norskov-Lauritsen 2014; Landig, Hazel et al. 2019). We investigated the role of human CEACAMs in regard to interaction and colonization by commensal bacteria. Enrichment of human CEACAM-binding bacteria out of human stool samples revealed a diverse set of classes within the phylum Firmicutes such as *Clostridia* and *Bacilli.* Interestingly, the phylum Bacteroidetes was mainly represented by the genus *Prevotella,* a member of the class *Bacteroidia. Prevotella spp*. are Gram-negative bacteria and part of the human oral, vaginal, and gut microbiota. They are genetically equipped for glycan degradation and associated with vegetarian or Mediterranean diets rich in fruits and

vegetables (De Filippo, Cavalieri et al. 2010). This study is the first to identify a broad range of commensal bacteria, including Gram-positive microorganisms such as *Bacilli* and *Clostridia*, in the context of CEACAM interaction. The tested representatives of the identified genera and classes had distinct binding specificities for CEACAM1, CEACAM3, and CEACAM6. Interestingly, all were able to interact with CEA, a protein exclusively expressed by epithelial cells. CEA has a high abundance in the human gut where it is either linked to epithelial cell membranes or occurs in a soluble form in the adjacent mucus layer. The mucus layer constitutes a physical barrier that minimize infiltration of the underlying epithelium by microorganisms and thereby helps to avoid constant inflammatory responses (Johansson, Gustafsson et al. 2010). Genera such as *Clostridium*, *Lactobacillus* and *Enterococcus* colonize both, the epithelial surface and the mucus layer which is why human CEA represents an excellent host factor for interaction between host and its microbiota (Sekirov, Russell et al. 2010). Commensal microbes are known to preferentially bind to host glycoproteins, because they provide two main advantages: First, the attachment to a host receptor enables colonization. Second, carbohydrate residues of the bound glycoprotein serve as energy source (Ouwerkerk, de Vos et al. 2013; Crouch, Liberato et al. 2020). In line with this, our studies show that commensal microbes not simply bind to but also modulate CEACAM molecules. A clear molecular weight reduction of the receptors (roughly 15 kDa) indicated degradation that followed the initial microorganism binding step. Interaction studies with *Enterococcus faecalis* and fluorescently labeled sugar residues linked to CEACAM molecules, hinted to cleavage of the carbohydrate side chains and subsequent utilization in the bacterial metabolism, as *Enterococcus* proteins appeared to be fluorescently labeled afterwards. In agreement with this finding, nutrient excess in an early growth phase of *E. faecalis* led to CEA-binding, without its alteration. Nutrient deficiency in a later growth phase, however, resulted in degradation of CEA (Figure IV-5 A) (Li, Li et al. 2020). We speculate that CEA fulfills a similar function as reported for other host glycoproteins such as mucins (Crouch, Liberato et al. 2020): The highly glycosylated CEACAM proteins may provide docking sides for commensal microorganisms, while supplying nutrient sources in form of the covalently bound carbohydrate chains. *In vitro*, infection of CEA-expressing cells with commensal bacteria revealed no difference in bacterial interaction on cellular surfaces, hinting to a minor contribution during early colonization events. We suggest that CEACAMs may not be involved in early cell surface association processes, but that microbe interaction with host-specific CEACAM molecules and subsequent degradation of linked carbohydrate

moieties supports long-term colonization by providing an additional niche. Furthermore, occupation of CEACAM molecules by commensals reduces the availability of binding sites for pathogenic microbes. For instance, *Neisseria lactamica*, a commensal inhabitant of the upper respiratory tract, is known to protect against its pathogenic close relative *Neisseria meningitidis*, the causative agent of meningitis (Bennett, Griffiths et al. 2005; Evans, Pratt et al. 2011). Both, *N. lactamica* and *N. meningitidis* express Opa proteins; adhesive structures that target and tightly bind CEACAM molecules (Toleman, Aho et al. 2001). Likewise, other commensal, CEACAM-associating bacteria can mediate a similar effect providing the human host with the benefit of the resulting colonization resistance. According to that, simultaneous administration of a CEA-exploiting pathogen and CEA-associating commensals to human CEA-transgenic mice should result in less successful pathogen colonization event than infection of human CEA-transgenic mice with the same pathogen alone.

To illuminate the question of host-specificity, the two human-residing commensal bacteria *Prevotella intermedia* and *Prevotella nigrescence* were tested for their ability to bind to other mammalian CEACAM proteins such as mouse, dog and cattle CEACAM1. In compliance with our hypothesis, both *Prevotella* species recognized human CEACAM only, and showed no binding to other mammalian constructs. This observation hints to a species-specific CEACAM interaction and also strengthens the hypothesis that CEACAMs contribute to host specific colonization. In contrast, *Enterococcus faecalis* that colonizes not only human, but also other mammalian gastrointestinal tracts bound to human CEA and mouse CEACAM1 similarly. The specific CEACAM-affinities of different commensals such as *Prevotella* (human CEACAM proteins) or *Enterococcus* (human and mouse CEACAMs), imply direct interaction with CEACAM protein backbones and excludes an exclusive interaction through carbohydrate chains. Thus, interaction studies including commensal bacteria and sugar-free CEACAM molecules either generated via glycosidase treatment or by production in bacteria should result in similar binding patterns. The degradation process of sugar moieties only seems to be an advantageous byproduct during the binding process.

Both parties, the host and the associated microorganisms, contribute to the formation of host-specific microbiota compositions. On the one side, microorganisms are provided with a certain set of genes that enable them to mediate cell surface interactions, to metabolize available nutrition, and to respond sufficiently to various stresses (Powell, Leonard et al. 2016). On the

other side, the host influences the microbiota composition directly by daily nutrition, but also indirectly via genetic factors. Our data demonstrate that a versatile group of commensal bacteria interacts with epithelial members of the CEACAM family, exhibiting a certain host-specificity. Identification of structures that contribute to the formation of a host-specific microbial community has the potential to build a basis for the development of novel therapeutic strategies in regard to diseases promoted by dysbiosis (e.g. inflammatory bowel diseases), or for the cure of infections caused by pathogens (e.g. *N. meningitidis*, *H. pylori*).

Material and Methods

Enrichment of CEACAM-binding bacteria via magnetic bead-based isolation

Fresh stool samples were diluted in PBS and filtrated through an Acrodisc PSF syringe filter (PALL Corporation) with a pore size of 10 µm. The filtrated stool samples were incubated with CEACAM-coupled beads for 30 min at RT with head-over-head rotation. A MS 25 column (Miltenyi Biotec) was placed in the µMACS separator and equilibrated with 200 µl lysis buffer. Afterwards, the stool sample-CEACAM-bead solution was applied onto the column. The column was rinsed three times with 200 µl PBS-T and two times with 200 µl PBS. The column was removed from the separator and the CEACAM-bound bacteria were eluted with 50 µl H_2O. For isolation of bacterial DNA, the bacteria were boiled and cell debris was removed by centrifugation.

Pyrosequencing of barcoded 16S rRNA gene amplicons

The hypervariable regions V1 and V2 of the 16S rRNA genes were amplified from each sample with the universal primers 27F and 338R. The forward primer (5'-**CCTATCCCCTGTGTGCCTTGGCAGTCTCAG**TCAGAGTTTGATCCTGGCTCAG-3') contained the 454 Life Sciences GS FLX Titanum primer sequences B (bold letters), a TC linker and the broadly conserved bacterial primer 27F (underlined). The reverse primer (5'-**CCATCTCATCCCTGCGTGTCTCCGACTCAGNNNNNNNNNN**CATGCTGCCTCCCGTAGGAGT-3) contained the 454 Life Sciences GS FLX Titanum primer sequences A (bold letters) a unique 10-base barcode a CA linker sequence and the broad-range bacterial primer 338R (underlined). PCR reactions were carried out in 50 µl reactions containing 1 µl Pfu, 5 µl of 10x buffer, dNTP mix, 10 µM forward and reverse primer and 3 µl template DNA. The PCR proceeded at 94°C for 2 min, followed by 30 cycles of 94°C for 20 s, 55°C for 30 s and 72°C for 60 s. Amplification products were gel purified (Qiagen) and quantified using the PicoGreen Assay Kit (Invitrogen). The purified amplicons were pooled in equal ratios and sequenced by 454 FLX Titanum pyrosequencer (Genomic Center, University of Konstanz, Germany), using 454 Life science primer A.

Bioinformatical analysis

Identification of CEACAM1/CEA-enriched bacteria was performed first by identifying 16S rRNA fragments (hypervariable region V1 and V2) that showed significant differences in abundance over the control vs. experimental conditions (stool sample input and CEACAM8-enriched bacteria vs. CEACAM1/CEA-enriched bacteria). All sequences were clustered using CD-hit (95% identity filtering), with the number of sequences combined in each cluster providing the respective abundance of each fragment-type (cluster representative). By taking into account the respective abundances of fragments over the experimental conditions and replicate experiments, a standard T-test (using the Statistics::TTest module in Perl, testing for unequal variance) was used to assign each change in abundance a P-value. Fragments with P-values below 0.05 were regarded as significantly different in abundance between the control and experimental conditions (sequences of interest).

Using the 16S rRNA sequences present in the SILVA database (V104) (Quast, Pruesse et al. 2013) we then attempted to identify from which group of bacterial species these fragments originated from. First, all SILVA small subunit (SSU)-rRNA sequences were clustered using CD-hit at 80% identity. The cluster representatives (4480 sequences) were combined with our fragment sequences of interest and the combined set was then analyzed using CLuster ANalysis of Sequences (CLANS). We checked whether any of our sequences of interest were not in the bacterial group. After this step we only used bacterial sequences from the SILVA dataset. Taking the full set of bacterial sequences from the SILVA database we excluded sequences annotated as 'unidentified', 'environmental sample', 'unknown', 'metagenome', 'uncultured' (113900 sequences). This set was filtered (using CD-hit at 95% identity resulting in a set of 8105 sequences) and the sequences of interest were added to them. The combined set of sequences was clustered using CLANS. CLANS performs an all-against-all BLAST comparison of sequences and represents them as dots in 3D-space. Each sequence is assigned an 'attraction' value to every other sequence based on their reciprocal BLAST hits; in this case, the Score of the hit was divided by the length of the hit to provide a Score-per-column value. The Score-per-column value was used as the similarity metric and determined the respective attraction of sequences to each other. By equilibrating the graph in 2D or 3D space, sequences similar to each other move into close proximity while sequences of lesser similarity are located more distantly from each other. This leads to a clustered representation of sequence-space,

with large clouds of dots representing large groups of sequences with greater than average similarity to each other. Using this 3D/2D CLANS map, we were able to identify all major bacterial families/groupings from the SILVA database and were able to place our fragments of interest into a taxonomic context.

Binding studies of commensal bacteria species to soluble CEACAMs

Soluble CEACAM-ectodomains fused to GFP were produced in HEK293T cells. 500 µl of the CEACAM-GFP containing supernatant was clustered with rabbit anti-GFP serum (1:1000) (University of Konstanz, Germany) overnight at 4°C. The overnight grown bacteria cultures were harvested by centrifugation (3000 xg, RT, 5 min), the medium was discarded and the cell pellet suspended in 1 ml 1x phosphate-buffered saline (PBS) buffer. The optical density was determined at a wavelength of 600 nm (OD_{600}) and OD_{600} 0.5 of the respective bacteria species was incubated with the clustered CEACAM-GFP constructs for 1 h at RT under gentle rotation. Afterwards, bacterial cells were pelleted (3000 xg, RT °C, 1 min) and washed twice with PBS buffer. Finally, bacteria were suspended in 1x SDS sample buffer for immunoblotting.

Immunoblotting

Samples from binding studies were analyzed by SDS-PAGE and Western blot. After separating proteins by their molecular weight via SDS-PAGE, they were transferred to polyvinylidene difluoride (PVDF) membranes (Millipore, Bedford, MA). The membranes were blocked (2% bovine serum albumin in Tris-buffered saline containing 0.05% Tween 20) for 1 h at RT and incubated in monoclonal antibody (mAB) against GFP (clone JL-8, Clontech, Palo Alto, CA) or against human CEACAM1, 3, 4, 5, 6 (clone D14HD11, Genovac, Freiburg, Germany) overnight at 4°C. Next day, the membrane was incubated with polyclonal HRP-conjugated goat anti-mouse antibody (Jackson ImmunoResearch) diluted 1:10.000 for 1 h at RT, and visualized by chemiluminescence (ChemiDoc Touch, BioRad).

Transfection of HEK293T cells

The human embryonic kidney cell line 293T (HEK293T cells) (DSMZ, Braunschweig, Germany) was grown in Dulbecco's modified Eagle's medium (DMEM) (Merck) supplemented with 10% calf serum (CS) (Biochrom) at 37°C, 5% CO_2. The cells were transfected with pLPS3'EGFP CEACAM plasmid via calcium-phosphate co-precipitation. Therefore 5 µg plasmid DNA was

added to 500 µl of H_2O followed by 500 µl of 2x HBS buffer (274 mM NaCl, 42 mM HEPES, 1.4 mM Na_2HPO_4, pH 7.05) and 50 µl 2.5 M $CaCl_2$.

Immunofluorescence staining for confocal microscopy

Human CEA-expressing HEK293T cells were seeded on poly-L-lysine coated coverslips in 24-well plates ($8x10^4$/well). Next day, *E. faecalis* was labelled with 2 µg 5-Carboxy-tetramethylrhodamine N-succinimidyl ester (TAMRA-SE) (Sigma-Aldrich) in PBS for 30 min at 37°C, light protected and under constant shaking. Bacteria were washed three times with 1 ml PBS to remove the unbound staining solution. Cells were infected with an MOI of 30 for 1 h at 37°C. After infection, unbound *E. faecalis* were washed off with 350 µl PBS^{++} (1x PBS, 0.9 mM $CaCl_2$, 0.5 mM $MgCl_2$). The cells were fixed in 4% paraformaldehyde for 20-30 min at RT. 300 µl of blocking solution (PBS^{++}, 10% FCS) were added for 10 min at RT. Human CEA was stained for 1 h at RT using 25 µg/ml of a polyclonal anti-CEACAM antibody produced in the local animal facility at the University of Konstanz, Germany. Afterwards, the antibody was removed by washing cells three times with 350 µl PBS^{++}. The cells were incubated with 300 µl blocking solution for 5 min at RT, the secondary antibody Alexa Fluor 488-conjugated AffiniPure Goat Anti-Mouse IgG (Jackson ImmunoResearch, West Grove, PA) was added and incubated for 45 min at RT in the dark. Cells were washed again three times with PBS^{++} to remove unbound antibody and the coverslips were transferred on glass slides and preserved by adding Dako Fluorescence Mounting Medium (Dako, Glostrup, Denmark) in between.

Fluorescent labelling of CEACAM carbohydrate residues

HEK293T cells were transfected with pLPS3'EGFP CEA-NT plasmids. Next day, 10 µl of 100 mM mannose-methylcyclopropene (mannose-C_4H_6) (synthesized and provided by AG Wittmann, University of Konstanz, Germany) diluted in DMSO was directly added into the culture medium of the transfected HEK cells (final concentration: 100 µM). 10 µl DMSO without labelled mannose-C_4H_6 was added to a second plate of transfected HEK cells for negative control. After an incubation time of 48 h, the CEA-NT-GFP containing cell culture medium was collected. For binding studies, 250 µl of the CEA-NT-GFP containing supernatant (with and without mannose-C_4H_6) was clustered with rabbit anti-GFP serum (1:1000) (University of Konstanz, Germany) overnight at 4°C. Next day, OD_{600} 2 of *Enterococcus faecalis*, grown over night, was added to the supernatants and incubated for 45 min at RT. Bacteria were pelleted, the supernatant was removed and the pellet was washed with 1 ml 10 mM Tris buffer, pH 7.4 containing 150 mM

NaCl in MilliQ. After centrifugation and removal of the washing buffer, the bacteria pellet was resuspended in 60 µl washing buffer and 10 µl of 100 µM Cy3-tetrazine (AG Wittmann, University of Konstanz, Germany) was added. The samples were incubated for 1 h, at 25°C, on a shaker (400 rpm), in the dark to allow the **inverse electron demand Diels–Alder reaction** (**DA_{INV}**) to take place. Afterwards the bacteria were pelleted and solved in 1x SDS sample buffer before they were boiled and immunoblotting was conducted. Once, the samples were transferred onto a PVDF membrane, the Cy3-fluorescence was analyzed via Typhoon FLA 9500 (GE Healthcare).

Silver staining of proteins in polyacrylamide gels

After the electrophoretic separation of proteins on polyacrylamide gels, they were visualized via silver staining. Therefore the gel was incubated in solution 1 (40% ethanol, 10% acetic acid) and afterwards in solution 2 (30% ethanol) for 1h respectively to denature and fix the proteins and to remove SDS from the gel. Subsequently, 0.2 g/l $Na_2S_2O_3$ x 5 H_2O was added for 1 min and removed by washing the gel with ultrapure water. During an incubation with silver nitrate (2 g/l $AgNO_3$) solution for 20 min, silver ions associate with negatively charged protein side chains. Unbound silver ions were removed by a washing step with ultrapure water. To visualize protein bands, the gel was incubated in developing solution containing 6% Na_2CO_3, 5 mg $Na_2S_2O_3$ x 5 H_2O and 0.05% formaldehyde. The alkaline formaldehyde reduces silver ions to elemental silver wherefore proteins are stained brown to black. To stop the staining reaction, a solution containing 50 mM EDTA and 1% acetic acid was added.

Bacterial species and growth conditions

Bacterial species were ordered at DSMZ (Table IV-2), isolated from human stool sample (*Bacteroides fragilis*) or isolated from a swab of an eye infection (*Enterococcus faecalis*). Strict anaerobic bacteria were grown in PYG medium (DSMZ protocol 104.) or chopped meat medium (BD BBL™). Facultative anaerobes were cultivated in brain-heart-infusion (BHI) medium (BD™) (Table IV-2). To create anaerobic conditions PYG medium was filled into culturing glass bottles sealed with a butyl rubber plug and a screw cap. The medium was autoclaved and the oxygen-containing atmosphere was replaced by nitrogen. Once the medium was cooled down the heat labile components, haemin and vitamin K1, were added via a syringe. Strict anaerobes were applied to the respective medium as well via a syringe.

Facultative anaerobic bacteria were grown in BHI medium under constant shaking. All species were cultivated at 37°C.

Table IV-2: Bacteria strains and growth media. Strict anaerobes are highlighted by an asterisk.

Species	**Medium**	**Source**
*Bacteroides fragilis **	PYG	isolated from stool sample
*Catenibacterium mitsuokai **	PYG	DSMZ
*Clostridium beijerinckii **	Chopped meat	DSMZ
Enterococcus faecalis	BHI	isolated from eye infection
Enterococcus hirae	BHI	DSMZ
Lactobacillus brevis	BHI	Prof. Dr. Thomas Böttcher, University of Konstanz, Germany
Lactobacillus murinus	BHI	isolated from macaque stool sample
*Megasphaera cerevisiae **	PYG	DSMZ
*Prevotella copri **	PYG	DSMZ
*Prevotella corporis **	PYG	DSMZ
*Prevotella intermedia **	PYG	DSMZ
*Prevotella marshii **	PYG	DSMZ
*Prevotella nanceiensis **	PYG	DSMZ
*Prevotella nigrescens **	PYG	DSMZ
*Prevotella timonensis **	Chopped meat	DSMZ

General Discussion

The human body is constantly exposed to a multitude of different microbes originating from the environment and our daily nutrition. However, only a fraction of these microorganisms survive and multiply in the human host. The prerequisite for persistent colonization is the expression of genes that allow attachment to host structures and a sufficient energy acquisition. Only microbes that are adapted optimally to niches within the human body can persist for prolonged time periods and withstand the competition with other microorganisms. To gain a foothold and overcome obstacles such as colonization resistance, epithelial barriers, or the host immune system, pathogens are provided with powerful tools in form of virulence factors that qualify them for adhesion, invasion, and immune evasion (Vaca, Thibau et al. 2020). These factors also enable them to compete against the prevailing microbiota and colonize the human host. One colonization strategy involves epithelial members of the CEACAM family that have been identified as targets for diverse human-restricted pathogens such as *Neisseria gonorrhoeae, Heamophilus influenzae,* or *Helicobacter pylori* (Tchoupa, Schuhmacher et al. 2014; Javaheri, Kruse et al. 2016; Klaile, Muller et al. 2017). Those receptors increase the pathogenicity of microbes as they allow the establishment of a more intimate interaction with cell surface structures of the human host and resistance to host clearance mechanisms. CEACAM-interaction seems to provide the pathogen with a multitude of advantages, which is underlined by the fact that a set of highly diverse adhesins, all aiming for the same target structure (the CC'FG-face of the CEACAM Ig_V-like domain), evolved independently several times (e.g. Opa-proteins in *Neisseria*, OMP P1 in *Haemophilus*, UspA1 in *Moraxella*, HopQ in *Helicobacter*, CbpF in *Fusobacterium*) (Chen and Gotschlich 1996; Virji, Makepeace et al. 1996; Virji, Watt et al. 1996; Hill and Virji 2003; Berger, Billker et al. 2004; Barnich, Carvalho et al. 2007; Tchoupa, Lichtenegger et al. 2015; Koniger, Holsten et al. 2016; Klaile, Muller et al. 2017; Brewer, Dymock et al. 2019). One explanation for the preference of various pathogenic microbes to engage human CEACAMs, relates to the fact that CEACAM1, the target of a large fraction of these adhesins, is also expressed by T-cells and that CEACAM1 isoforms have a negative regulatory role in T-cell stimulation and proliferation (Gray-Owen and Blumberg 2006). Furthermore, interaction of pathogens such as *M. catarrhalis* or *N. meningitidis* with CEACAM1 results in reduced TLR2-initiated inflammatory responses of human epithelial cells (Slevogt, Zabel et al. 2008). Tyrosine phosphorylation of the intracellular

immunoreceptor tyrosine-based inhibitory motif of CEACAM1 mediates these inhibitory effects and represents an excellent immune-evasion strategy. The role of CEACAM-binding in mucosal colonization is best studied for *N. gonorrhoeae* and *N. meningitidis,* and demonstrates that both microbes greatly profit from association with CEACAM proteins (Muenzner, Bachmann et al. 2010; Johswich, McCaw et al. 2013; Islam, Anipindi et al. 2018). Both organisms express different Opa proteins that allow tight connection to CEACAM1, CEA, and CEACAM6. This intimate binding enhances epithelial cell adhesion and allows nitric oxide, produced by gonococci, to permeate epithelial cell membranes and trigger the upregulation of components that enhance cell-matrix adhesion. As a consequence, exfoliation and delamination of infected cells is efficiently suppressed, thereby creating stable foothold on the mucosa (Muenzner, Rohde et al. 2005; Muenzner, Bachmann et al. 2010; Muenzner and Hauck 2020). Furthermore, CEACAM engagement triggers gonococcal internalization into epithelial cells and allows *N. gonorrhoeae* to reach subepithelial space via transcytosis (Wang, Gray-Owen et al. 1998). Pathogens can immensely profit, potentially in multiple ways, from engaging CEACAMs on epithelial cells which nicely explains the prevalence and independent evolution of CEACAM-binding adhesins amongst human pathogens.

In this work we show for the first time that a broad range of commensal bacteria, belonging to the two intestine main phyla, Bacteroidetes and Firmicutes, interact with molecules of the CEACAM family as was previously known for pathogenic species only (Figure IV-2). The examined commensal microorganisms (e.g. *B. fragilis, Prevotella* spec., *Enterococcus* spec., *Lactobacillus* spec., etc.) had distinct binding specificities for CEACAM1, CEACAM3, CEA, and CEACAM6 (Figure IV-2). Remarkably, all species interacted with CEA (Table IV-1), a CEACAM-member strongly expressed by intestinal epithelial cells. Therefore CEA is available for commensals such as *B. fragilis,* which not only resides in the intestinal lumen and mucus but also intimately associates with epithelial cells and colonizes e.g. crypts of the large intestine (Lee, Donaldson et al. 2013; Fung, Artis et al. 2014). Besides the membrane-associated version, CEA exists in a secreted form and is present in the adjacent mucus layer where many commensals reside. This mucus layer constitutes a physical barrier that minimizes infiltration of the underlying epithelium by microorganisms to avoid constant inflammatory responses (Johansson, Gustafsson et al. 2010). Glycoproteins such as mucins or potentially CEA restrain a high number of both the normal microbiota (e.g. *B. fragilis* (Huang, Lee et al. 2011)) and enteric pathogens (e.g. *H. influenzae, M. catarrhalis* (Reddy, Murphy et al. 1997)) in the

mucus. Commensal microbes are known to preferentially bind to host glycoproteins, because they provide two main advantages. First, the attachment to a host receptor enables colonization. Second, carbohydrate residues of the bound glycoprotein serve as energy source (Crouch, Liberato et al. 2020). The commensal gut bacterium, *Bacteroides fragilis* for instance, binds specifically to mucins, highly glycosylated proteins and the major constituent in intestinal mucus and degrades them. Several studies have shown that mucins provide a nutrient source for growth of *B. fragilis* (Roberton and Stanley 1982; Tsai, Sunderland et al. 1992) and can serve as sole carbon and nitrogen source (Macfarlane and Gibson 1991). Expression of SusC/SusD (starch utilization system) homologs in *Bacteroides* species allows them to bind mucin carbohydrates and import them across the outer membrane for further digestion (Zhu, Kwiatkowski et al. 2015). Mucus binding serves as a physical mechanism to maintain host colonization and provides *B. fragilis* with nutrients for bacterial cell growth. Different studies suggest that specific interaction with the hosts mucus (mucins and other components) is relevant for colonization of *B. fragilis in vivo* (Huang, Lee et al. 2011). In line with this, our studies show that commensal microbes not simply bind to, but also modulate CEACAM molecules. A clear molecular weight reduction of the receptors (roughly 15 kDa) indicated degradation that followed the initial microorganism binding step (Figure IV-2 B+C; Figure IV-4, Figure IV-5 B). Interaction studies with *Enterococcus faecalis* and fluorescently labeled sugar residues linked to CEACAM molecules, hinted to cleavage of the carbohydrate side chains and subsequent utilization in the bacterial metabolism, as *Enterococcus* proteins appeared to be fluorescently labeled afterwards (Figure IV-5 C). In agreement with this finding, nutrient excess in an early growth phase of *E. faecalis* led to CEA-binding, without its alteration. Nutrient deficiency in a later growth phase, however, resulted in degradation of CEA (Figure IV-5 A). In this regard, we speculate that similar to other host glycoproteins, the highly glycosylated CEACAM proteins can constitute an additional niche by providing docking sides for commensal microorganisms, while supplying nutrient sources in form of the covalently bound carbohydrate chains (Crouch, Liberato et al. 2020). *In vitro*, infection of CEA-expressing cells with commensal bacteria revealed no difference in bacterial interaction on cellular surfaces, hinting to a minor contribution during early colonization events (Figure IV-3 C). A similar situation for CEACAMs is found during *Neisseria gonorrhoeae* infection *in vivo*: To infect the male urethra, gonococci first utilize a pilus for the initial adherence to mucosal epithelia. In a second step, interaction with CEACAM molecules via neisserial Opa proteins

mediates a tight association with the cell surface (Wang, Gray-Owen et al. 1998). We suggest that CEACAMs may not be involved in this initial binding step to associate with cell surfaces, but that microbe interaction with host-specific CEACAM molecules allow a more intimate binding and that subsequent degradation of linked carbohydrate moieties support long-term colonization by providing an additional nutrient supply.

Interestingly, the majority of CEACAM-targeting pathogens recognize the epithelial receptor CEA as well (Tchoupa, Schuhmacher et al. 2014; Koniger, Holsten et al. 2016; Klaile, Muller et al. 2017; Brewer, Dymock et al. 2019). During *Neisseria gonorrhoeae* infection, CEACAM1-interaction enhances gonococcal association and tissue penetration. Opa-CEA-binding on vaginal epithelia, however, was shown to facilitate long-term colonization of the lower genital tract (Islam, Anipindi et al. 2018). CEA molecules and their occupation by commensal bacteria may contribute to colonization resistance and impede the permanent settlement of CEACAM-targeting pathogens by competition for a finite amount of binding sites and nutrition. Probiotics like *Lactobacillus* and *Bifidobacterium*, for instance, protect the human host from overgrowth of the opportunistic pathogen *Clostridium difficile* by limiting the adherence to epithelial cells and decreasing its cytotoxicity (Valdes-Varela, Alonso-Guervos et al. 2016). The human commensal *Neisseria lactamica* competes for the same niche and binding sites as the opportunistic pathogen, *Nesseria meningitidis*; the causative agent of meningitidis. Both inhabit the respiratory tract, share antigenic structures such as Opa proteins on their cell surfaces and target CEACAM proteins for colonization purposes (Toleman, Aho et al. 2001). Therefore the presence of *N. lactamica* is able to prevent against *N. meningitids* infections (Virji 2009).

Pathogens that succeed to colonize the host cell surface, or opportunistic pathogens that spread and overgrow, can invade host cells and the surrounding tissues. At this point, the human immune system intervenes to counteract and control local and systemic infections. One such control mechanism involves another member of the CEACAM family: the innate immune receptor CEACAM3. This receptor is expressed on the cell surface of human neutrophil granulocytes, which are recruited to the site of infection. CEACAM3 recognizes a broad range of CEACAM-associating bacterial adhesins and mediates rapid opsonin-independent uptake and killing of bound microorganism (Bonsignore, Kuiper et al. 2019).

CEACAM interaction and the immune response that is initiated by CEACAM3 are extensively studied for gram-negative pathogens, but are yet unexplored for other CEACAM-binding microorganisms. We focused on the opportunistic, pathogenic yeast *Candida albicans* that constitutes a serious threat to the health of immunocompromised patients as it can breach the mucosal barrier and cause local as well as systemic infections, known as Candidiasis. *C. albicans* targets epithelial CEACAM1, CEA, and CEACAM6. For CEACAM1 and CEACAM6, it was shown that association results in an altered epithelial immune response including the secretion of IL-8 (Klaile, Muller et al. 2017). We hypothesized that CEACAM3 also contributes to immune response and defense against human challenging fungi such as *C. albicans*. Upon infection of *C. albicans*, we observed clustering of the innate immune receptor at the host cell attachment side of yeast cells (Figure III-5 A) and tyrosine phosphorylation in response to fungi recognition (Figure III-4 B). However, activated CEACAM3 alone was not sufficient to trigger internalization of bound yeast cells (Figure III-5). The innate immune receptor Dectin-I targets β-glucan structures in fungal cell walls and initiates the same downstream signaling by its HemITAM-motif as CEACAM3. We assumed that activated CEACAM3 could collaborate with Dectin-I and lead to an increased yeast cell uptake by intensifying signaling. Expression of Dectin-I and CEACAM3 resulted in yeast cell phagocytosis, but overexpression of CEACAM3 had no enhancing effect on uptake of fungi (Figure III-5 B). Identification of the CEACAM-targeting adhesin of *Candida* may help to better understand the underlying binding mechanisms, and would allow investigating the physiological role of yeast-CEACAM interaction by generating *Candida* knock-out strains that lack the CEACAM-binding adhesin. If CEACAM3 is involved in yeast cell defense processes, *Candida* strains without the CEACAM-targeting adhesin should have advantages in survival compared to wild-type strains. Similar to the immune response described for CEACAM1 and CEACAM6, the signal pathway initiated by CEACAM3 normally results in the release of cytokines to recruit more immune cells to the site of infection, which might help to engulf and kill yeast cells (Antachopoulos and Roilides 2005). Further investigations need to clarify whether *Candida* binding and activation of CEACAM3 is followed by cytokine release, and which kind of cytokines are involved. To support our predication, *in vivo* studies would be pivotal to show that *Candida*-infection is better handled by an organism that offers CEACAM3-expressing PMNs compared to an organism without CEACAM3-expresssing neutrophils. Unfortunately, this model organism is currently not available. In nature, CEACAM3 can only be found in primates and CEACAM3 knock-out animals

do not exist yet. Attempts to generate transgenic mice that express CEACAM3 solely in their hematopoietic cells as it occurs naturally in primates have failed so far.

Besides *C. albicans*, we found other fungal members of the human microbiota (*Saccharomyces cerevisiae, Candia dubliniensis*) and a laboratory model organism (*Pichia pastoris*) to associate with epithelial CEACAMs (Figure III-1 A). Remarkably, all tested yeast species were recognized by CEACAM3 and exhibited the same CEACAM binding specificity, which hints towards a conserved adhesive structure on the yeast cell surface. Alternatively, during evolution CEACAM-targeting molecules developed independently in different yeast species and represent a versatile set of CEACAM-binding adhesins similar to the diverse structures found in bacteria. Once more, identification of CEACAM-binding adhesins expressed by different yeast species would help to answer the question of the underlying molecular mechanism behind this interaction. The receptor CEACAM3 is not well conserved and constitutes one of the fastest evolving primate genes (Figure II-1 D). In line with this, we found only human and chimpanzee CEACAM3 to target *C. albicans*, but not the more distantly relatives: gorilla, baboon, or rhesus monkey (Figure III-6), suggesting a rather novel development of yeast cell recognition by the immune receptor in evolution. One could speculate that *Candida albicans* has opened up the niche of CEACAM proteins in humans and chimpanzee for interaction purposes more recently. Possession of an adhesin that targets human or chimpanzee CEACAMs simultaneously makes *Candida* prone to the recognition by the immune receptor CEACAM3. In fact, CEACAM3 is a recent evolutionary invention and occurs only after the emergence of Old World monkeys (Chang, Semyonov et al. 2013; Sato, Kuroki et al. 2015). The driving force behind the evolution of CEACAM3 as immune receptor seems to be the selective pressure by infectious and potentially deadly microbes such as meningococci and *H. influenzae*, which can both cause bacterial meningitis (Wenger, Hightower et al. 1990). Moreover, gonococci, one of the most common sexually transmitted bacterial pathogen and the causative agent of gonorrhea, has led to neonatal blindness and infertility in a pre-antibiotic world (Laga, Meheus et al. 1989; Cates, Rolfs et al. 1990). Fatal bacterial infections (meningitis) as well as reduced reproduction resulting from venereal diseases (gonorrhea) are strong selective pressures, which could have favored the development of a germ line encoded innate immune receptor like CEACAM3 (Adrian, Bonsignore et al. 2019).

CEACAM1, which is targeted by a large fraction of pathogens, exhibits a splice variant that contains immunoreceptor tyrosine-based inhibitory motifs (ITIM) in the cytoplasmic domain. Clustering and oligomerization of CEACAM1 by pathogen binding triggers signaling into the cell and results in an immune modulation such as decreased production of cytokines and other inflammatory mediators (Nagaishi, Chen et al. 2008). Additionally, CEACAM1 engagement initiates a characteristic uptake pathway that promotes transcytosis of microorganisms through intact epithelial layers, which allows immune evasion and systemic dissemination. In contrast, the receptor CEACAM3 contains an immunoreceptor tyrosine-based activation motif (ITAM) that is activated via pathogen recognition. CEACAM3 clustering results in enhanced immune response in form of elevated cytokine release and opsonin-independent phagocytosis followed by intracellular killing of bacteria (Schmitter, Pils et al. 2007). In this regard, it is not surprising that CEACAM-binding pathogens seem to target epithelial members to facilitate host colonization (Muenzner, Bachmann et al. 2010), while avoiding recognition by CEACAM3 to evade opsonin-independent phagocytosis. In our studies, we analyzed numerous pathogen adhesins and found a clear preference for CEACAM1 compared to CEACAM3 in Opa and OMP P1 adhesins (Figure II-5) (Roth, Mattheis et al. 2013; Sintsova, Wong et al. 2015; Adrian, Bonsignore et al. 2019). Adhesins selectively interacting with CEACAM3, but lacking the ability to binding to epithelial CEACAMs do not seem to exist. This fact underlines the disadvantage of pathogens, which are detected by the immune receptor. In our studies, stepwise conversion of the CEACAM3 amino acid sequence toward CEACAM1, revealed that three amino acid changes (S43R, L44Q and V49A) broaden the binding spectrum of CEACAM3 enormously, and enable not only the recognition of CEACAM1-optimized neisserial Opa proteins, but also detection of the *H. influenzae* adhesin OMP P1 (Figure II-6 C+D). Interestingly, those alterations naturally occur in a polymorphic human CEACAM3 allele, carried by ~10% of all individuals and even ~40% of the African population (Figure II-6 A) (Adrian, Bonsignore et al. 2019). CEACAM3 polymorphisms might provide a selective advantage by detecting and defending a broader spectrum of CEACAM-binding pathogens, promoting the rapid evolutionary divergence of the innate immune receptor in primates.

Interestingly, while all pathogens recognize CEACAM1, only a few commensal bacteria were able to interact with this member of the CEACAM family. The selective pressure for bacterial adhesins to target the immune inhibitory receptor CEACAM1 seems to be less pronounced in commensal bacteria than in pathogens. Presumably due to the less invasive behavior of

commensal microorganism, they are not permanently challenged by the human immune system and are not selected for their ability to modulate or evade it to survive within the human host.

To keep up with a constantly changing host environment, both, pathogens and commensals, need to adapt rapidly. Food habits provoke short term alterations in the microbiota composition. For instance, nutrition high in fiber promotes an enhanced abundance of the genus *Prevotella* and *Xylanibacter* in the human gut microbiota. Both genera are known to exhibit genes for cellulose and xylan hydrolysis and provide the host with a maximized energy intake from fibers (De Filippo, Cavalieri et al. 2010). In some cases, nutrition also causes long term alterations by affecting the microbiome via horizontal gene transfer. The marine bacterium *Zobellia galactanivorans* inhabits marine red algae *Porphyra* spp. (nori), the most important nutritional seaweed, traditionally used to prepare sushi. *Z. galactanivorans* transferred carbohydrate-active enzymes (porphyranases) to the commensal gut bacterium *Bacteroides plebeius* in Japanese individuals allowing energy acquisition from red algae derived carbohydrates (Hehemann, Correc et al. 2010). In individuals alteration in nutrition can lead to a rapid, but often transient change in species composition of the gut microbiome (Singh, Chang et al. 2017). Random mutation or recombination events in the host genome, however, can establish more permanent and heritable changes in microbiota composition (Benson, Kelly et al. 2010). Presentation of novel host structures to the cell environment provides new challenges, but also new possibilities for the microbial community. CEACAMs represent such constantly changing host structures. The ancestral *CEACAM1* gene, for instance, underwent multiple duplications, which resulted in species-specific expansion of CEACAM1-related members of the CEA gene family in mammals reaching from less than 5 genes in lagomorphs, around 30 in Old World monkeys up to more than 100 in bats (Kammerer, Popp et al. 2007; Kammerer and Zimmermann 2010). The fact that each mammalian species has its own characteristic microorganism community points toward the existence of host-specific factors, which allow microbial colonization and variation in microbe composition in a species-specific manner. Many studies have shown a co-evolutionary process between host and its microbiota (Ley, Hamady et al. 2008; Ochman, Worobey et al. 2010; Lerner, Matthias et al. 2017; Jeong, Arif et al. 2019). For instance, investigation of gut microbiota derived from different great ape species, including human, common chimpanzee, bonobo, western gorilla, and eastern gorilla, revealed that the phylogenetic branching order

based on the microbial composition is consistent with phylogenetic relationships of the hosts, reflecting co-evolution (Ochman, Worobey et al. 2010). We hypothesized that mammalian CEACAM proteins due to their species-specific appearance could contribute to the characteristic microbiota composition in mammals. In this work, binding studies conducted with human commensal *Prevotella* species and CEACAM proteins, originating from different mammalian species, revealed binding preferences for human-derived constructs (Figure IV-2 D). Interestingly, *Enterococcus faecalis,* a resident of human and mouse intestines, associated with human as well as with murine CEACAMs (Figure IV-4 B) (Dubin and Pamer 2014; Barnes, Dale et al. 2017). The specific affinities of commensals, such as *Prevotella* for human CEACAM proteins or *Enterococcus* for human and mouse CEACAMs but not for CEACAMs derived from other mammalian species, imply direct interaction with CEACAM protein backbones. The degradation of sugar moieties only seems to be an advantageous byproduct during this binding process. In this way, CEACAM molecules may contribute to the formation of a host characteristic microbial community. To gain a better understanding and validate this hypothesis, the microbiota and microbiome of wild-type and human CEA-expressing transgenic mice could be analyzed and compared. Differences in the composition would give first hints about which species gain advantages by the presence of this particular receptor.

In conclusion, our data extent the spectrum of CEACAM-binding microorganisms by pathogenic and commensal fungi, as well as commensal bacteria of the two gut main phyla, Bacteroidetes and Firmicutes. We show that commensal microbes interact with CEACAMs and process the covalently linked carbohydrate moieties. We hypothesize that CEACAM glycoproteins represent an additional niche for commensals and that the human host may profit from this interaction due to the effect of colonization resistance. Certain pathogenic microbes manage to colonize the mucosal surface by binding to epithelial CEACAMs, followed by invasion of host cells and surrounding tissues. The unique innate immune receptor CEACAM3, responsible for the specific elimination of CEACAM-targeting pathogens, handles this inconvenience by impressively rapid, opsonin-independent phagocytosis. In response, bacterial adhesins such as OMP $P1_{Hinf}$, which targets epithelial CEACAMs but simultaneously avoids detection by CEACAM3, appear to be favored. CEACAM3 holds a distinct set of polymorphisms, thereby broadening the spectrum of adhesins that can be recognized by the immune receptor. Alteration of single amino acids to adapt to changing microbe structures seems to be fostered, as it provides the advantage of detecting also pathogens that evolved

to evade recognition by CEACAM3. However, those variations might entail the loss of other target structures. Presumably, heterozygotes have an advantage, because their two different CEACAM3 variants broaden the spectrum of pathogens that are eliminated after binding these receptors.

Identification of host specific structures like CEACAMs have the potential to build a basis for the development of novel therapeutic strategies in regard to diseases promoted by dysbiosis (e.g. inflammatory bowel diseases), or for the cure of infections caused by pathogens (e.g. *N. meningitidis*, *H. pylori*). Colonization resistance is a vital phenomenon generated by the extremely complex interactions of microorganisms in the human body with each other and with the human host. Understanding colonization resistance and the underlying mechanisms is essential to prevent and treat a wide range of human diseases, from infection to autoimmunity to metabolic disorders (Pickard and Nunez 2019). One approach to enhance colonization resistance is the administration of prebiotics. Prebiotics are dietary nutrients that target a subset of the microbiota and aim to enhance their beneficial functions and growth. Our studies hinted to the participation of commensal bacteria in colonization resistance against CEACAM-targeting pathogens. Therefore, identification of prebiotics that specifically support growth of those commensals may be a powerful tool.

The emerging field of microbiome reprogramming aims to find new approaches that allow creating a healthy microbiome to cure diseases derived from imbalanced microbial compositions or pathogens. For instance, infection of mice with an exogenous *Bacteroides* strain harboring a rare gene cluster for porphyrin utilization and administration of the marine polysaccharide, porphyrin, generated an exclusive metabolic niche in mice and allowed successful colonization (Shepherd, DeLoache et al. 2018). On this basis, it is thinkable to generate commensal bacteria strains that possess CEACAM-targeting adhesin genes derived from human pathogens (e.g. OMP P1 of *H. influenzae*) to boost their competition ability for the same binding sites. Simultaneous administration of the genetically modified bacteria strains and prebiotics that selectively supports the growth of those strains may help to cure or prevent infections caused by CEACAM-targeting pathogens. The expression of genes in commensal microbes that support colonization of niches preferentially utilized by pathogens can extenuate pathogenicity of the invader by limiting their chances to gain proper foothold and to spread. Indication for the functionality of this approach is observed in nature: the

commensal *Neisseria lactamica* that prevents against *N. meningitids* infections in the human host (Toleman, Aho et al. 2001; Virji 2009). One possible explanation might be the expression of similar antigenic structures such as CEACAM-targeting Opa proteins by both, *N. meningitidis* and *N. lactamica,* which target the same binding sites.

Using the example of CEACAMs, these data nicely illustrate the complex interplay between pathogens, host, and the host-associated microbiota and provides an additional therapeutic point of action.

Declaration of author's contribution

Chapter I

"CEACAM3 – a prim(at)e invention for opsonin-independent phagocytosis of bacteria"

***Patrizia Bonsignore**, Johannes W. Kuiper, Jonas Adrian, Griseldis Goob, Christof R. Hauck*

P.B., J.K., J.A., G.G., and C.R.H. wrote the review

Chapter II

"Adaptation to host-specific bacterial pathogens drives rapid evolution of a human innate immune receptor"

*Jonas Adrian, **Patrizia Bonsignore**, Sebastian Hammer, Tancred Frickey, Christof R. Hauck*

J.A. and C.R.H. conceived the study and designed the experiments; J.A. and P.B. performed the experiments; S.H. and T.F. established and performed the genome analysis; J.A., P.B., and C.R.H. evaluated the data; J.A. and C.R.H. wrote the paper.

Chapter III

"CEACAM-binding yeasts are recognized by the innate immune receptor CEACAM3"

***Patrizia Bonsignore**, Jonas Adrian, Christof R. Hauck*

P.B. and C.R.H. conceived the study and designed the experiments; P.B. and J.A. performed the experiments; P.B. and J.A. evaluated the data; P.B. wrote the manuscript

Chapter IV

"Classification of epithelial CEACAMs – their role in human-specific colonization by commensal gut bacteria"

***Patrizia Bonsignore,** Alexandra Roth, Tancred Frickey, Christof R. Hauck*

P.B., A.R. and C.R.H. conceived the study and designed the experiments. A.R. identified CEACAM-binding commensal bacteria (Figure IV-2 D); T.F. performed bioinformatical analysis (Figure IV-2 A), P.B. performed binding and colonization experiments; P.B. wrote the manuscript

List of Publications

Patrizia Bonsignore, Johannes W. Kuiper, Jonas Adrian, Griseldis Goob, Christof R. Hauck

"CEACAM3 – a prim(at)e invention for opsonin-independent phagocytosis of bacteria"

Submitted to Journal: Frontiers in Immunology

Specialty Section: Molecular Innate Immunity

Jonas Adrian*, **Patrizia Bonsignore***, Sebastian Hammer, Tancred Frickey, Christof R. Hauck

* These authors contributed equally

"Adaptation to Host-Specific Bacterial Pathogens Drives Rapid Evolution of a Human Innate Immune Receptor. "

Curr Biol. 2019 Feb. doi: 10.1016/j.cub.2019.01.058.

Abbreviations

AAD	antibiotic-associated diarrhea
AIEC	adherent-invasive *E. coli*
BHI medium	Brain Heart Infusion medium
Afa	afimbral adhesin
Ca^{2+}	calcium
CARD	caspase recruitment domain
CbpF	CEACAM binding protein of Fusobacterium
CD	cluster of differentiation
CEA	carcinoembryonic antigen
CEABAC	CEA bacterial artificial chromosome
CEACAM	carcinoembryonic antigen-related cell adhesion molecule
CFTR	cystic fibrosis transmembrane conductance regulator
CGM1a	Carcinoembryonic Gene family Member 1a
CFSE	carboxyfluorescein succinimidyl ester
CLANS	CLuster ANalysis of Sequences
CLEC	C-type lectin receptors
CNS	central nervous system
COPD	chronic obstructive pulmonary disease
CS	calf serum
CW lysate	cell wall lysate
CWP	cell wall proteins
DAEC	diffusely adhering *Escherichia coli*
DAG	diacyl glycerol
DAPI	4’,6-diamidino-2-phenylindole
DC	dendritic cell
Δ (delta)	gene deletion
DMEM	Dulbecco’s modified Eagle’s medium
DMP	dimethyl pimelimidate x2 HCl

dN/dS	ratio of non-synonymous substitution per non-synonymous site and synonymous mutations per synonymous site
DNA	deoxyribonucleic acid
DSMZ	German collection of microorganisms and cell cultures
ECM	extracellular matrix
Eno1	Enolase 1
FACS	fluorescence-activated cell sorting
FBA1	Fructose-bisphosphate aldolase
GDP	guanosine diphosphate
GEF	guanine nucleotide exchange factor
GFP	green fluorescent protein
GPI	glycosylphosphatidylinositol
GPM1	Phosphoglycerate mutase
GTP	guanosine triphosphate
HEK cells	Human embryonic kidney cells
HemITAM	hemi-immunoreceptor tyrosine-based activation motif
Hinf	*Haemophilus influenzae*
HIV	human immunodeficiency virus
HRP	horse radish peroxidase
HSPG	heparan-sulfate proteoglycan
hu-gorilla	humanized gorilla
Ig	immunglobulin
Ig_C	immunoglobulin constant
Ig_V	immunoglobulin variable
IgCAM	Ig-domain containing cell adhesion molecule
IL	interleukin
IP3	inositol-(1,4,5)-trisphosphate
IPTG	isopropyl β-D-1-thiogalactopyranoside
ITAM	immunoreceptor tyrosine-based activation motif

ITIM	immunoreceptor tyrosine-based inhibitory motif
kDa	kilo Dalton
LB	Lysogeny Broth
MAMP	microbe-associated molecular pattern
M cell	Microfold cells
mAB	monoclonal antibody
MOI	multiplicity of infection
NAD	nicotinamide adenine dinucleotide
NADPH	nicotinamide adenine dinucleotide phosphate
NCA	non-specific cross-reacting antigen
NF-κB	nuclear factor 'kappa-light-chain-enhancer' of activated B-cells
Ngo	*Neisseria gonorrhoeae*
Ni	nickel
NOD	nucleotide-binding oligomerization domain
OM	outer membrane
OMP P1	outer membrane protein P1
Opa	opacity-associated
pAB	polyclonal antibody
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PDc	pyruvate decarboxylase
PFA	paraformaldehyde
PH	pleckstrin homology
PI3K	phosphatidylinositol 3'-kinase
PI3P	phosphatidylinositol-(3)-phosphate
PID	pelvic inflammatory disease
PIP_2	phosphatidylinositol-(4,5)-bisphosphate
PIP_3	phosphatidylinositol-(3,4,5)-trisphosphate
PKC	protein kinase C
PLCγ	phospholipase C gamma
PMN	polymorphonuclear granulocytes

PRR	pattern-recognition receptor
PSG	pregnancy-specific glycoprotein
PVDF	polyvinylidene difluoride
RIPA	radioimmune precipitation assay
RT	room temperature
ROS	reactive oxygen species
SCFA	short chain fatty acids
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SIGIRR	single immunoglobulin IL-1R-related molecule
SH2	Src-homology 2
SOEing	Splicing by Overlapping Extension
SNP	single nucleotide polymorphism
SSU	small subunit
SUMO	small ubiquitin-related modifier
SusC	starch utilization system C
Syk	spleen tyrosine kinase
T cell	thymus cell
T3SS-1	type-3 secretory system-1
TAMRA-SE	5-(and-6)-Carboxytetramethylrhodamine, Succinimidyl Ester
Th1	T helper 1 cells
TLR	toll-like receptor
Tollip	TLR inhibitor Toll-interacting protein
Treg	regulatory T cell
UspA1	ubiquitous surface protein A1
WASP	wiskott-aldrich syndrome protein
WAVE	WASP-family verprolin homologous protein
YPG	yeast, peptone, glucose

References

Aagaard, K., J. Ma, et al. (2014). "The placenta harbors a unique microbiome." Sci Transl Med **6**(237): 237ra265.

Achtman, M. (1995). "Epidemic spread and antigenic variability of Neisseria meningitidis." Trends in Microbiology **3 (5)**: 186-192.

Achtman, M., M. Neibert, et al. (1988). "Purification and characterization of eight class 5 outer membrane protein variants from a clone of Neisseria meningitidis serogroup A." J Exp Med **168**(2): 507-525.

Adrian, J., P. Bonsignore, et al. (2019). "Adaptation to Host-Specific Bacterial Pathogens Drives Rapid Evolution of a Human Innate Immune Receptor." Curr Biol **29**(4): 616-630 e615.

Agrawal, A. and T. F. Murphy (2011). "Haemophilus influenzae infections in the H. influenzae type b conjugate vaccine era." J Clin Microbiol **49**(11): 3728-3732.

Allison, A. C. (1954). "Notes on sickel-cell polymorphism." Annal Human Genetics **19**(1): 29-51.

Altmeyer, R. M., J. K. McNern, et al. (1993). "Cloning and molecular characterization of a gene involved in Salmonella adherence and invasion of cultured epithelial cells." Mol Microbiol **7**(1): 89-98.

Antachopoulos, C. and E. Roilides (2005). "Cytokines and fungal infections." Br J Haematol **129**(5): 583-596.

Apicella, M. A. (2010). *Neisseria meningitidis*. Principles and Practice of Infectious Diseases. G. L. Mandell, J. E. Bennett and R. Dolin. Philadelphia, Churchill Livingstone. **2:** 2737-2752.

Arpaia, N., C. Campbell, et al. (2013). "Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation." Nature **504**(7480): 451-455.

Barnes, A. M. T., J. L. Dale, et al. (2017). "Enterococcus faecalis readily colonizes the entire gastrointestinal tract and forms biofilms in a germ-free mouse model." Virulence **8**(3): 282-296.

Barnich, N., F. A. Carvalho, et al. (2007). "CEACAM6 acts as a receptor for adherent-invasive E. coli, supporting ileal mucosa colonization in Crohn disease." J Clin Invest **117**(6): 1566-1574.

Barreiro, L. B., M. Ben-Ali, et al. (2009). "Evolutionary dynamics of human Toll-like receptors and their different contributions to host defense." PLoS Genet **5**(7): e1000562.

Bartosch, S., A. Fite, et al. (2004). "Characterization of bacterial communities in feces from healthy elderly volunteers and hospitalized elderly patients by using real-time PCR and effects of antibiotic treatment on the fecal microbiota." Appl Environ Microbiol **70**(6): 3575-3581.

Batt, C. A. and M. L. Tortorello (2015). "Encyclopedia of food microbiology." Choice: Current Reviews for Academic Libraries **52**(8): 1292-1292.

Baumler, A. and F. C. Fang (2013). "Host specificity of bacterial pathogens." Cold Spring Harb Perspect Med **3**(12): a010041.

Beauchemin, N., P. Draber, et al. (1999). "Redefined nomenclature for members of the carcinoembryonic antigen family." Experimental Cell Research **252**(2): 243-249.

Belambri, S. A., L. Rolas, et al. (2018). "NADPH oxidase activation in neutrophils: Role of the phosphorylation of its subunits." Eur J Clin Invest **48 Suppl 2**: e12951.

Belkaid, Y. and T. W. Hand (2014). "Role of the microbiota in immunity and inflammation." Cell **157**(1): 121-141.

Belogolova, E., B. Bauer, et al. (2013). "Helicobacter pylori outer membrane protein HopQ identified as a novel T4SS-associated virulence factor." Cell Microbiol **15**(11): 1896-1912.

Benchimol, S., A. Fuks, et al. (1989). "Carcinoembryonic antigen, a human tumor marker, functions as an intercellular adhesion molecule." Cell **57**: 327-334.

Bennett, J. S., D. T. Griffiths, et al. (2005). "Genetic diversity and carriage dynamics of Neisseria lactamica in infants." Infect Immun **73**(4): 2424-2432.

Benson, A. K., S. A. Kelly, et al. (2010). "Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors." Proc Natl Acad Sci U S A **107**(44): 18933-18938.

Berger, C. N., O. Billker, et al. (2004). "Differential recognition of members of the carcinoembryonic antigen family by Afa/Dr adhesins of diffusely adhering *Escherichia coli* (Afa/Dr DAEC)." Mol Microbiol **52**(4): 963-983.

Bhat, K. S., C. P. Gibbs, et al. (1992). "The opacity proteins of *Neisseria gonorrhoeae* strain MS11 are encoded by a family of 11 complete genes." Molecular Microbiology **6**: 1073-1076.

Bilek, N., C. A. Ison, et al. (2009). "Relative contributions of recombination and mutation to the diversification of the opa gene repertoire of Neisseria gonorrhoeae." J Bacteriol **191**(6): 1878-1890.

Billker, O., A. Popp, et al. (2002). "Distinct mechanisms of internalization of *Neisseria gonorrhoeae* by members of the CEACAM receptor family involving Rac1- and Cdc42- dependent and -independent pathways." The EMBO Journal **21**(4): 560-571.

Boamah, D. K., G. Zhou, et al. (2017). "From Many Hosts, One Accidental Pathogen: The Diverse Protozoan Hosts of Legionella." Front Cell Infect Microbiol **7**: 477.

Bond, M., M. F. Tejedor, et al. (2015). "Eocene primates of South America and the African origins of New World monkeys." Nature **520**(7548): 538-541.

Bonsignore, P., J. W. P. Kuiper, et al. (2019). "CEACAM3-A Prim(at)e Invention for Opsonin-Independent Phagocytosis of Bacteria." Front Immunol **10**: 3160.

Bonsor, D. A., Q. Zhao, et al. (2018). "The *Helicobacter pylori* adhesin protein HopQ exploits the dimer interface of human CEACAMs to facilitate translocation of the oncoprotein CagA." EMBO J **37**(13).

Bookwalter, J. E., J. A. Jurcisek, et al. (2007). "A CEACAM1 homologue plays a pivotal role in nontypeable *Haemophilus influenzae* colonization of the chinchilla nasopharynx via the OMP P5-homologous adhesin." Infect Immun.

Booth, J. W., D. Telio, et al. (2003). "Phosphatidylinositol 3-kinases in carcinoembryonic antigen-related cellular adhesion molecule-mediated internalization of *Neisseria gonorrhoeae*." J Biol Chem **278**(16): 14037-14045.

Bos, M. P., F. Grunert, et al. (1997). "Differential recognition of members of the carcinoembryonic antigen family by Opa variants of *Neisseria gonorrhoeae*." Infection & Immunity **65**: 2353-2361.

Bos, M. P., M. Kuroki, et al. (1998). "CD66 receptor specificity exhibited by neisserial Opa variants is controlled by protein determinants in CD66 N-domains." Proc Natl Acad Sci U S A **95**(16): 9584-9589.

Brandt, E., J. Van Damme, et al. (1991). "Neutrophils can generate their activator neutrophil-activating peptide 2 by proteolytic cleavage of platelet-derived connective tissue-activating peptide III." Cytokine **3**: 311-321.

Brauer, M. J., C. Huttenhower, et al. (2008). "Coordination of growth rate, cell cycle, stress response, and metabolic activity in yeast." Mol Biol Cell **19**(1): 352-367.

Brewer, M. L., D. Dymock, et al. (2019). "Fusobacterium spp. target human CEACAM1 via the trimeric autotransporter adhesin CbpF." J Oral Microbiol **11**(1): 1565043.

Brouwer, S., T. C. Barnett, et al. (2016). "Streptococcus pyogenes adhesion and colonization." FEBS Lett **590**(21): 3739-3757.

Buntru, A., K. Kopp, et al. (2011). "Phosphatidylinositol-3' kinase activity is critical for initiating the oxidative burst and bacterial destruction during CEACAM3-mediated phagocytosis." J Biol Chem **286**(11): 9555-9566.

Buntru, A., A. Roth, et al. (2012). "HemITAM signaling by CEACAM3, a human granulocyte receptor recognizing bacterial pathogens." Arch Biochem Biophys **524**: 77-83.

Buntru, A., T. Zimmermann, et al. (2009). "FRET-based subcellular visualization of pathogen-induced host receptor signalling." BMC Biol **7**: 81.

Cameron, C. E., E. L. Brown, et al. (2004). "Treponema pallidum fibronectin-binding proteins." J Bacteriol **186**(20): 7019-7022.

Carbone, L., R. A. Harris, et al. (2014). "Gibbon genome and the fast karyotype evolution of small apes." Nature **513**(7517): 195-201.

Cario, E. (2005). "Bacterial interactions with cells of the intestinal mucosa: Toll-like receptors and NOD2." Gut **54**(8): 1182-1193.

Carvalho, F. A., N. Barnich, et al. (2009). "Crohn's disease adherent-invasive Escherichia coli colonize and induce strong gut inflammation in transgenic mice expressing human CEACAM." J Exp Med.

Castrillo, J. I., L. A. Zeef, et al. (2007). "Growth control of the eukaryote cell: a systems biology study in yeast." J Biol **6**(2): 4.

Cates, W., Jr., R. T. Rolfs, Jr., et al. (1990). "Sexually transmitted diseases, pelvic inflammatory disease, and infertility: an epidemiologic update." Epidemiol Rev **12**: 199-220.

Chan, C. H. and C. P. Stanners (2004). "Novel mouse model for carcinoembryonic antigen-based therapy." Mol Ther **9**(6): 775-785.

Chan, P. J., I. M. Seraj, et al. (1996). "Prevalence of mycoplasma conserved DNA in malignant ovarian cancer detected using sensitive PCR-ELISA." Gynecol Oncol **63**(2): 258-260.

Chang, A. H. and J. Parsonnet (2010). "Role of bacteria in oncogenesis." Clin Microbiol Rev **23**(4): 837-857.

Chang, C. L., J. Semyonov, et al. (2013). "Widespread divergence of the CEACAM/PSG genes in vertebrates and humans suggests sensitivity to selection." PLoS One **8**(4): e61701.

Chen, A., I. C. Boulton, et al. (2003). "Induction of HIV-1 long terminal repeat-mediated transcription by Neisseria gonorrhoeae." AIDS **17**(4): 625-628.

Chen, T., S. Bolland, et al. (2001). "The CGM1a (CEACAM3/CD66d)-mediated phagocytic pathway of *Neisseria gonorrhoeae* expressing opacity proteins is also the pathway to cell death." Journal of Biological Chemistry **276**(20): 17413-17419.

Chen, T. and E. C. Gotschlich (1996). "CGM1a antigen of neutrophils, a receptor of gonococcal opacity proteins." Proceedings of the National Academy of Sciences of the United States of America **93**: 14851-14856.

Chen, T., F. Grunert, et al. (1997). "Several carcinoembryonic antigens (CD66) serve as receptors for gonococcal opacity proteins." Journal of Experimental Medicine **185**: 1557-1564.

Clarke, G., R. M. Stilling, et al. (2014). "Minireview: Gut microbiota: the neglected endocrine organ." Mol Endocrinol **28**(8): 1221-1238.

Collado, M. C., S. Rautava, et al. (2016). "Human gut colonisation may be initiated in utero by distinct microbial communities in the placenta and amniotic fluid." Sci Rep **6**: 23129.

Connell, T. D., D. Shaffer, et al. (1990). "Characterization of the repertoire of hypervariable regions in the Protein II (opa) gene family of Neisseria gonorrhoeae." Mol Microbiol **4**(3): 439-449.

Conners, R., D. J. Hill, et al. (2008). "The *Moraxella* adhesin UspA1 binds to its human CEACAM1 receptor by a deformable trimeric coiled-coil." EMBO J **27**(12): 1779-1789.

Coyne, M. J., B. Reinap, et al. (2005). "Human symbionts use a host-like pathway for surface fucosylation." Science **307**(5716): 1778-1781.

Crocker, P. R., J. C. Paulson, et al. (2007). "Siglecs and their roles in the immune system." Nat Rev Immunol **7**(4): 255-266.

Crouch, L. I., M. V. Liberato, et al. (2020). "Prominent members of the human gut microbiota express endo-acting O-glycanases to initiate mucin breakdown." Nat Commun **11**(1): 4017.

Crowe, J. D., I. K. Sievwright, et al. (2003). "Candida albicans binds human plasminogen: identification of eight plasminogen-binding proteins." Mol Microbiol **47**(6): 1637-1651.

Crowley, M. T., P. S. Costello, et al. (1997). "A critical role for Syk in signal transduction and phagocytosis mediated by Fcgamma receptors on macrophages." Journal of Experimental Medicine **186**: 1027-1039.

Crowther, G. S., S. D. Baines, et al. (2013). "Evaluation of NVB302 versus vancomycin activity in an in vitro human gut model of Clostridium difficile infection." J Antimicrob Chemother **68**(1): 168-176.

Danby, C. S., L. A. Cosentino, et al. (2016). "Patterns of Extragenital Chlamydia and Gonorrhea in Women and Men Who Have Sex With Men Reporting a History of Receptive Anal Intercourse." Sex Transm Dis **43**(2): 105-109.

Das, P., P. Babaei, et al. (2019). "Metagenomic analysis of microbe-mediated vitamin metabolism in the human gut microbiome." BMC Genomics **20**(1): 208.

David, L. A., C. F. Maurice, et al. (2014). "Diet rapidly and reproducibly alters the human gut microbiome." Nature **505**(7484): 559-563.

De Filippo, C., D. Cavalieri, et al. (2010). "Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa." Proc Natl Acad Sci U S A **107**(33): 14691-14696.

de Jonge, M. I., M. P. Bos, et al. (2002). "Conformational analysis of opacity proteins from Neisseria meningitidis." Eur J Biochem **269**(21): 5215-5223.

de Jonge, M. I., H. J. Hamstra, et al. (2003). "Mapping the binding domains on meningococcal Opa proteins for CEACAM1 and CEA receptors." Mol Microbiol **50**(3): 1005-1015.

Dearlove, B. L., A. J. Cody, et al. (2016). "Rapid host switching in generalist Campylobacter strains erodes the signal for tracing human infections." ISME J **10**(3): 721-729.

Degn, S. E. and S. Thiel (2013). "Humoral pattern recognition and the complement system." Scand J Immunol **78**(2): 181-193.

Delgado Tascon, J., J. Adrian, et al. (2015). "The granulocyte orphan receptor CEACAM4 is able to trigger phagocytosis of bacteria." J Leukoc Biol **97**(3): 521-531.

Dempsey, J. A., W. Litaker, et al. (1991). "Physical map of the chromosome of Neisseria gonorrhoeae FA1090 with locations of genetic markers, including opa and pil genes." J Bacteriol **173**(17): 5476-5486.

DeVoe, I. W. (1982). "The meningococcus and mechanisms of pathogenicity." Microbiol Rev **46**(2): 162-190.

Dreux, N., J. Denizot, et al. (2013). "Point mutations in FimH adhesin of Crohn's disease-associated adherent-invasive Escherichia coli enhance intestinal inflammatory response." PLoS Pathog **9**(1): e1003141.

Dubin, K. and E. G. Pamer (2014). "Enterococci and Their Interactions with the Intestinal Microbiome." Microbiol Spectr **5**(6).

Eckburg, P. B., E. M. Bik, et al. (2005). "Diversity of the human intestinal microbial flora." Science **308**(5728): 1635-1638.

Eggimann, P. and D. Pittet (2014). "Candida colonization index and subsequent infection in critically ill surgical patients: 20 years later." Intensive Care Med **40**(10): 1429-1448.

English, B. K., S. L. Orlicek, et al. (1997). "Bacterial LPS and IFN-gamma trigger the tyrosine phosphorylation of vav in macrophages: evidence for involvement of the hck tyrosine kinase." J Leukoc Biol **62**(6): 859-864.

Evans, C. M., C. B. Pratt, et al. (2011). "Nasopharyngeal colonization by Neisseria lactamica and induction of protective immunity against Neisseria meningitidis." Clin Infect Dis **52**(1): 70-77.

Evans, J. M., L. S. Morris, et al. (2013). "The gut microbiome: the role of a virtual organ in the endocrinology of the host." J Endocrinol **218**(3): R37-47.

Eyre, D. W., N. D. Sanderson, et al. (2018). "Gonorrhoeae treatment failure caused by a *Neisseria gonorrhoeae* strain with combined ceftriaxone and high-level azithromycin resistance." Euro Surveillance **23**(27): 1800323.

Faith, J. J., N. P. McNulty, et al. (2011). "Predicting a human gut microbiota's response to diet in gnotobiotic mice." Science **333**(6038): 101-104.

Fazio, A., M. C. Jewett, et al. (2008). "Transcription factor control of growth rate dependent genes in Saccharomyces cerevisiae: a three factor design." BMC Genomics **9**: 341.

FitzGerald, G., D. Botstein, et al. (2018). "The future of humans as model organisms." Science **361**(6402): 552-553.

Forsythe, P., N. Sudo, et al. (2010). "Mood and gut feelings." Brain Behav Immun **24**(1): 9-16.

Frank, D. N., A. L. St Amand, et al. (2007). "Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases." Proc Natl Acad Sci U S A **104**(34): 13780-13785.

Frickey, T. and A. Lupas (2004). "CLANS: a Java application for visualizing protein families based on pairwise similarity." Bioinformatics **20**(18): 3702-3704.

Fukata, M., K. S. Michelsen, et al. (2005). "Toll-like receptor-4 is required for intestinal response to epithelial injury and limiting bacterial translocation in a murine model of acute colitis." Am J Physiol Gastrointest Liver Physiol **288**(5): G1055-1065.

Fuller, G. L., J. A. Williams, et al. (2007). "The C-type lectin receptors CLEC-2 and Dectin-1, but not DC-SIGN, signal via a novel YXXL-dependent signaling cascade." J Biol Chem **282**(17): 12397-12409.

Fumagalli, L., H. Zhang, et al. (2007). "The Src family kinases Hck and Fgr regulate neutrophil responses to N-formyl-methionyl-leucyl-phenylalanine." J Immunol **178**(6): 3874-3885.

Fumagalli, M., M. Sironi, et al. (2011). "Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution." PLoS Genet **7**(11): e1002355.

Fung, T. C., D. Artis, et al. (2014). "Anatomical localization of commensal bacteria in immune cell homeostasis and disease." Immunol Rev **260**(1): 35-49.

Funkhouser, L. J. and S. R. Bordenstein (2013). "Mom knows best: the universality of maternal microbial transmission." PLoS Biol **11**(8): e1001631.

Furusawa, Y., Y. Obata, et al. (2013). "Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells." Nature **504**(7480): 446-450.

Gales, A., A. Conduche, et al. (2010). "PPARgamma controls Dectin-1 expression required for host antifungal defense against Candida albicans." PLoS Pathog **6**(1): e1000714.

Garcia-Gutierrez, E., M. J. Mayer, et al. (2019). "Gut microbiota as a source of novel antimicrobials." Gut Microbes **10**(1): 1-21.

Garlanda, C., F. Riva, et al. (2004). "Intestinal inflammation in mice deficient in Tir8, an inhibitory member of the IL-1 receptor family." Proc Natl Acad Sci U S A **101**(10): 3522-3526.

Gibbs, R. A., J. Rogers, et al. (2007). "Evolutionary and Biomedical Insights from the Rhesus Macaque Genome." Science **316**(5822): 222 - 234.

Gilfillan, G. D., D. J. Sullivan, et al. (1998). "Candida dubliniensis: phylogeny and putative virulence factors." Microbiology (Reading) **144 (Pt 4)**: 829-838.

Gill, S. R., M. Pop, et al. (2006). "Metagenomic analysis of the human distal gut microbiome." Science **312**(5778): 1355-1359.

Gold, P. and S. O. Freedman (1965). "Specific carcinoembryonic antigens of the human digestive system." J Exp Med **122**(3): 467-481.

Goleski, C. A., A. H. Rule, et al. (1972). "Cea-Like Antigens - Presence in Tumor, Normal Colon and Meconium Extracts." Federation Proceedings **31**(2): A639-&.

Goodrich, J. K., E. R. Davenport, et al. (2017). "The Relationship Between the Human Genome and Microbiome Comes into View." Annu Rev Genet **51**: 413-433.

Gordon, J. H. and R. Dubos (1970). "The anaerobic bacterial flora of the mouse cecum." J Exp Med **132**(2): 251-260.

Goyal, S., J. C. Castrillon-Betancur, et al. (2018). "The Interaction of Human Pathogenic Fungi With C-Type Lectin Receptors." Front Immunol **9**: 1261.

Gray-Owen, S. D. and R. S. Blumberg (2006). "CEACAM1: contact-dependent control of immunity." Nat Rev Immunol **6**(6): 433-446.

Gray-Owen, S. D., C. Dehio, et al. (1997). "CD66 carcinoembryonic antigens mediate interactions between Opa-expressing *Neisseria gonorrhoeae* and human polymorphonuclear phagocytes." The EMBO Journal **16**: 3435-3445.

Gray-Owen, S. D., D. R. Lorenzen, et al. (1997). "Differential Opa specificities for CD66 receptors influence tissue interactions and cellular response to *Neisseria gonorrhoeae*." Molecular Microbiology **26**: 971-980.

Grenham, S., G. Clarke, et al. (2011). "Brain-gut-microbe communication in health and disease." Front Physiol **2**: 94.

Groussin, M., F. Mazel, et al. (2017). "Unraveling the processes shaping mammalian gut microbiomes over evolutionary time." Nat Commun **8**: 14319.

Guignot, J., S. Hudault, et al. (2009). "Human decay-accelerating factor and CEACAM receptor-mediated internalization and intracellular lifestyle of Afa/Dr diffusely adhering Escherichia coli in epithelial cells." Infect Immun **77**(1): 517-531.

Guignot, J., I. Peiffer, et al. (2000). "Recruitment of CD55 and CD66e brush border-associated glycosylphosphatidylinositol-anchored proteins by members of the Afa/Dr diffusely adhering family of *Escherichia coli* that infect the human polarized intestinal Caco-2/TC7 cells." Infect Immun **68**(6): 3554-3563.

Hagemans, D., I. A. van Belzen, et al. (2015). "A script to highlight hydrophobicity and charge on protein surfaces." Front Mol Biosci **2**: 56.

Hammami, R., B. Fernandez, et al. (2013). "Anti-infective properties of bacteriocins: an update." Cell Mol Life Sci **70**(16): 2947-2967.

Hammerum, A. M. (2012). "Enterococci of animal origin and their significance for public health." Clin Microbiol Infect **18**(7): 619-625.

Harari, A., M. Ooms, et al. (2009). "Polymorphisms and splice variants influence the antiretroviral activity of human APOBEC3H." J Virol **83**(1): 295-303.

Hauck, C. R., E. Gulbins, et al. (1999). "Tyrosine phosphatase SHP-1 is involved in the opsonin-independent phagocytosis of Opa52-expressing *Neisseria gonorrhoeae*." Infection & Immunity **67**: 5490-5494.

Hauck, C. R., T. Hunter, et al. (2001). "The v-Src SH3 domain facilitates a cell adhesion-independent association with FAK." Journal of Biological Chemistry **276**(21): 17653-17662.

Hauck, C. R., T. F. Meyer, et al. (1998). "CD66-mediated phagocytosis of Opa_{52} *Neisseria gonorrhoeae* requires a Src-like tyrosine kinase- and Rac1-dependent signalling pathway." The EMBO Journal **17**: 443-454.

Hehemann, J. H., G. Correc, et al. (2010). "Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota." Nature **464**(7290): 908-912.

Heinrich, A., K. A. Heyl, et al. (2016). "Moraxella catarrhalis induces CEACAM3-Syk-CARD9-dependent activation of human granulocytes." Cell Microbiol **18**(11): 1570-1582.

Heinsbroek, S. E., P. R. Taylor, et al. (2008). "Stage-specific sampling by pattern recognition receptors during Candida albicans phagocytosis." PLoS Pathog **4**(11): e1000218.

Hill, D. J., A. M. Edwards, et al. (2005). "Carcinoembryonic antigen-related cell adhesion molecule (CEACAM)-binding recombinant polypeptide confers protection against infection by respiratory and urogenital pathogens." Mol Microbiol **55**(5): 1515-1527.

Hill, D. J. and M. Virji (2003). "A novel cell-binding mechanism of *Moraxella catarrhalis* ubiquitous surface protein UspA: specific targeting of the N-domain of carcinoembryonic antigen-related cell adhesion molecules by UspA1." Mol Microbiol **48**(1): 117-129.

Hill, M. J. (1997). "Intestinal flora and endogenous vitamin synthesis." Eur J Cancer Prev **6 Suppl 1**: S43-45.

Hooi, J. K. Y., W. Y. Lai, et al. (2017). "Global Prevalence of Helicobacter pylori Infection: Systematic Review and Meta-Analysis." Gastroenterology **153**(2): 420-429.

Huang, J. Y., S. M. Lee, et al. (2011). "The human commensal Bacteroides fragilis binds intestinal mucin." Anaerobe **17**(4): 137-141.

Iliev, I. D., V. A. Funari, et al. (2012). "Interactions between commensal fungi and the C-type lectin receptor Dectin-1 influence colitis." Science **336**(6086): 1314-1317.

Islam, E. A., V. C. Anipindi, et al. (2018). "Specific binding to differentially expressed human Carcinoembryonic Antigen-related cell adhesion molecules determines the outcome of *Neisseria gonorrhoeae* infections along the female reproductive tract." Infect Immun **86**(8).

Jauch, A., J. Wienberg, et al. (1992). "Reconstruction of genomic rearrangements in great apes and gibbons by chromosome painting." Proc Natl Acad Sci U S A **89**(18): 8611-8615.

Javaheri, A., T. Kruse, et al. (2016). "Helicobacter pylori adhesin HopQ engages in a virulence-enhancing interaction with human CEACAMs." Nat Microbiol **2**: 16189.

Jeong, H., B. Arif, et al. (2019). "Horizontal gene transfer in human-associated microorganisms inferred by phylogenetic reconstruction and reconciliation." Sci Rep **9**(1): 5953.

Jerse, A. E., M. S. Cohen, et al. (1994). "Multiple gonococcal opacity proteins are expressed during experimental urethral infection in the male." Journal of Experimental Medicine **179**: 911-920.

Jimenez, E., M. L. Marin, et al. (2008). "Is meconium from healthy newborns actually sterile?" Res Microbiol **159**(3): 187-193.

Johansson, M. E., J. K. Gustafsson, et al. (2010). "Bacteria penetrate the inner mucus layer before inflammation in the dextran sulfate colitis model." PLoS One **5**(8): e12238.

Johswich, K. O., S. E. McCaw, et al. (2013). "In Vivo Adaptation and Persistence of *Neisseria meningitidis* within the Nasopharyngeal Mucosa." PLoS Pathog **9**(7): e1003509.

Jones, B. W. and M. K. Nishiguchi (2004). "Counterillumination in the Hawaiian bobtail squid, Euprymna scolopes Berry (Mollusca: Cephalopoda)." Marine Biology **144**(6): 1151-1155.

Jong, A. Y., S. H. Chen, et al. (2003). "Binding of Candida albicans enolase to plasmin(ogen) results in enhanced invasion of human brain microvascular endothelial cells." J Med Microbiol **52**(Pt 8): 615-622.

Joshi, T., J. P. Butchar, et al. (2006). "Fcgamma receptor signaling in phagocytes." Int J Hematol **84**(3): 210-216.

Kamada, N., Y. G. Kim, et al. (2012). "Regulated virulence controls the ability of a pathogen to compete with the gut microbiota." Science **336**(6086): 1325-1329.

Kammerer, R., M. Mansfeld, et al. (2017). "Recent expansion and adaptive evolution of the carcinoembryonic antigen family in bats of the Yangochiroptera subgroup." BMC Genomics **18**(1): 717.

Kammerer, R., T. Popp, et al. (2007). "Species-specific evolution of immune receptor tyrosine based activation motif-containing CEACAM1-related immune receptors in the dog." BMC Evol Biol **7**: 196.

Kammerer, R. and W. Zimmermann (2010). "Coevolution of activating and inhibitory receptors within mammalian carcinoembryonic antigen (CEA) families." BMC Biology **8**: 12.

Kelly, D., J. I. Campbell, et al. (2004). "Commensal anaerobic gut bacteria attenuate inflammation by regulating nuclear-cytoplasmic shuttling of PPAR-gamma and RelA." Nat Immunol **5**(1): 104-112.

Kennedy, M. J. and P. A. Volz (1985). "Ecology of Candida albicans gut colonization: inhibition of Candida adhesion, colonization, and dissemination from the gastrointestinal tract by bacterial antagonism." Infect Immun **49**(3): 654-663.

Kim, J. and P. Sudbery (2011). "Candida albicans, a major human fungal pathogen." J Microbiol **49**(2): 171-177.

Kimbrell, D. A. and B. Beutler (2001). "The evolution and genetics of innate immunity." Nat Rev Genet **2**(4): 256-267.

Klaile, E., M. M. Muller, et al. (2017). "Binding of Candida albicans to Human CEACAM1 and CEACAM6 Modulates the Inflammatory Response of Intestinal Epithelial Cells." MBio **8**(2).

Kondelkova, K., D. Vokurkova, et al. (2010). "Regulatory T cells (TREG) and their roles in immune system with respect to immunopathological disorders." Acta Medica (Hradec Kralove) **53**(2): 73-77.

Koniger, V., L. Holsten, et al. (2016). "Helicobacter pylori exploits human CEACAMs via HopQ for adherence and translocation of CagA." Nat Microbiol **2**: 16188.

Kopp, K., A. Buntru, et al. (2012). "GRB14 is a negative regulator of ceacam3-mediated phagocytosis of pathogenic bacteria." J Biol Chem **287**: 39158-39170.

Korber, B., M. Muldoon, et al. (2000). "Timing the ancestor of the HIV-1 pandemic strains." Science **288**(5472): 1789-1796.

Koretke, K. K., P. Szczesny, et al. (2006). "Model structure of the prototypical non-fimbrial adhesin YadA of Yersinia enterocolitica." J Struct Biol **155**(2): 154-161.

Korotkova, N., Y. Yang, et al. (2008). "Binding of Dr adhesins of *Escherichia coli* to carcinoembryonic antigen triggers receptor dissociation." Mol Microbiol **67**(2): 420-434.

Kupsch, E.-M., B. Knepper, et al. (1993). "Variable opacity (Opa) outer membrane proteins account for the cell tropisms displayed by *Neisseria gonorrhoeae* for human leukocytes and epithelial cells." The EMBO Journal **12**(2): 641-650.

Kuroki, M., F. Arakawa, et al. (1991). "Molecular cloning of nonspecific cross-reacting antigens in human granulocytes." J Biol Chem **266**(18): 11810-11817.

Kuroki, M., Y. Matsuo, et al. (1990). "Nonspecific cross-reacting antigen (NCA) expressed by human granulocytes: six species with different peptide sizes and membrane anchoring forms." Biochemical & Biophysical Research Communications **166**: 701-708.

Kuroki, M., T. Yamanaka, et al. (1995). "Immunochemical analysis of carcinoembryonic antigen (CEA)-related antigens differentially localized in intracellular granules of human neutrophils." Immunological Investigations **24**: 829-843.

Laga, M., A. Meheus, et al. (1989). "Epidemiology and control of gonococcal ophthalmia neonatorum." Bull World Health Organ **67**(5): 471-477.

Landig, C. S., A. Hazel, et al. (2019). "Evolution of the exclusively human pathogen Neisseria gonorrhoeae: Human-specific engagement of immunoregulatory Siglecs." Evol Appl **12**(2): 337-349.

Lebensohn, A. M. and M. W. Kirschner (2009). "Activation of the WAVE complex by coincident signals controls actin assembly." Mol Cell **36**(3): 512-524.

Lee, J. S., H. Y. Choi, et al. (2002). "Gonococcal keratoconjunctivitis in adults." Eye (Lond) **16**(5): 646-649.

Lee, S. M., G. P. Donaldson, et al. (2013). "Bacterial colonization factors control specificity and stability of the gut microbiota." Nature **501**(7467): 426-429.

Lerner, A., T. Matthias, et al. (2017). "Potential Effects of Horizontal Gene Exchange in the Human Gut." Front Immunol **8**: 1630.

Ley, R. E., M. Hamady, et al. (2008). "Evolution of mammals and their gut microbes." Science **320**(5883): 1647-1651.

Ley, R. E., D. A. Peterson, et al. (2006). "Ecological and evolutionary forces shaping microbial diversity in the human intestine." Cell **124**(4): 837-848.

Ley, R. E., P. J. Turnbaugh, et al. (2006). "Microbial ecology: human gut microbes associated with obesity." Nature **444**(7122): 1022-1023.

Li, Y., R. Li, et al. (2020). "Enterococcus faecalis alpha1-2-mannosidase (EfMan-I): an efficient catalyst for glycoprotein N-glycan modification." FEBS Lett **594**(3): 439-451.

Liu, X., Y. Wu, et al. (2015). "Dietary fiber intake reduces risk of inflammatory bowel disease: result from a meta-analysis." Nutr Res **35**(9): 753-758.

Lo, H. J., J. R. Kohler, et al. (1997). "Nonfilamentous C. albicans mutants are avirulent." Cell **90**(5): 939-949.

Locke, D. P., L. W. Hillier, et al. (2011). "Comparative and demographic analysis of orang-utan genomes." Nature **469**(7331): 529-533.

Macfarlane, G. T. and G. R. Gibson (1991). "Formation of glycoprotein degrading enzymes by Bacteroides fragilis." FEMS Microbiol Lett **61**(2-3): 289-293.

Mackie, R. I., A. Sghir, et al. (1999). "Developmental microbial ecology of the neonatal gastrointestinal tract." Am J Clin Nutr **69**(5): 1035S-1045S.

Maclean, C. A., N. P. Chue Hong, et al. (2015). "hapbin: An Efficient Program for Performing Haplotype-Based Scans for Positive Selection in Large Genomic Datasets." Mol Biol Evol **32**(11): 3027-3029.

Macpherson, A. J. and T. Uhr (2004). "Induction of protective IgA by intestinal dendritic cells carrying commensal bacteria." Science **303**(5664): 1662-1665.

Makino, S., J. P. M. van Putten, et al. (1991). "Phase variation of the opacity outer membrane protein controls invasion by *Neisseria gonorrhoeae* into human epithelial cells." The EMBO Journal **10**(6): 1307-1315.

Malorny, B., G. Morelli, et al. (1998). "Sequence diversity, predicted two-dimensional protein structure, and epitope mapping of neisserial Opa proteins." J Bacteriol **180**(5): 1323-1330.

Marcobal, A., M. Barboza, et al. (2010). "Consumption of human milk oligosaccharides by gut-related microbes." J Agric Food Chem **58**(9): 5334-5340.

Marcobal, A. and J. L. Sonnenburg (2012). "Human milk oligosaccharide consumption by intestinal microbiota." Clin Microbiol Infect **18 Suppl 4**: 12-15.

Mathewson, N. D., R. Jenq, et al. (2016). "Gut microbiome-derived metabolites modulate intestinal epithelial cell damage and mitigate graft-versus-host disease." Nat Immunol **17**(5): 505-513.

Mayer, L. W. (1982). "Rates in vitro changes of gonococcal colony opacity phenotypes." Infect Immun **37**(2): 481-485.

McCaw, S. E., E. H. Liao, et al. (2004). "Engulfment of *Neisseria gonorrhoeae*: revealing distinct processes of bacterial entry by individual carcinoembryonic antigen-related cellular adhesion molecule family receptors." Infect Immun **72**(5): 2742-2752.

McCaw, S. E., J. Schneider, et al. (2003). "Immunoreceptor tyrosine-based activation motif phosphorylation during engulfment of *Neisseria gonorrhoeae* by the neutrophil-restricted CEACAM3 (CD66d) receptor." Mol Microbiol **49**(3): 623-637.

McFall-Ngai, M. J. and E. G. Ruby (1991). "Symbiont recognition and subsequent morphogenesis as early events in an animal-bacterial mutualism." Science **254**(5037): 1491-1494.

Mikkelsen, T. S. (2005). "Initial sequence of the chimpanzee genome and comparison with the human genome." Nature **437**(7055): 69-87.

Miller, M. G. and A. D. Johnson (2002). "White-opaque switching in Candida albicans is controlled by mating-type locus homeodomain proteins and allows efficient mating." Cell **110**(3): 293-302.

Millette, M., G. Cornut, et al. (2008). "Capacity of human nisin- and pediocin-producing lactic Acid bacteria to reduce intestinal colonization by vancomycin-resistant enterococci." Appl Environ Microbiol **74**(7): 1997-2003.

Missbach, S., D. Aleksic, et al. (2018). "Alternative splicing after gene duplication drives CEACAM1-paralog diversification in the horse." BMC Evol Biol **18**(1): 32.

Mitchell, C. and M. Prabhu (2013). "Pelvic inflammatory disease: current concepts in pathogenesis, diagnosis and treatment." Infect Dis Clin North Am **27**(4): 793-809.

Miyashiro, T. and E. G. Ruby (2012). "Shedding light on bioluminescence regulation in Vibrio fischeri." Mol Microbiol **84**(5): 795-806.

Moonens, K., Y. Hamway, et al. (2018). "Helicobacter pylori adhesin HopQ disrupts trans dimerization in human CEACAMs." EMBO J **37**(13).

Muenzner, P., V. Bachmann, et al. (2010). "Human-restricted bacterial pathogens block shedding of epithelial cells by stimulating integrin activation." Science **329**(5996): 1197-1201.

Muenzner, P., V. Bachmann, et al. (2008). "The CEACAM1 transmembrane domain, but not the cytoplasmic domain, directs internalization of human pathogens via membrane-microdomains." Cellular Microbiology **10**: 1074-1092.

Muenzner, P., C. Dehio, et al. (2000). "Carcinoembryonic antigen family receptor specificity of *Neisseria meningitidis* Opa variants influences adherence to and invasion of proinflammatory cytokine-activated endothelial cells." Infect Immun **68**(6): 3601-3607.

Muenzner, P. and C. R. Hauck (2020). "Neisseria gonorrhoeae Blocks Epithelial Exfoliation by Nitric-Oxide-Mediated Metabolic Cross Talk to Promote Colonization in Mice." Cell Host Microbe **27**(5): 793-808 e795.

Muenzner, P., A. Kengmo Tchoupa, et al. (2016). "Uropathogenic E. coli Exploit CEA to Promote Colonization of the Urogenital Tract Mucosa." PLoS Pathog **12**(5): e1005608.

Muenzner, P., M. Rohde, et al. (2005). "CEACAM engagement by human pathogens enhances cell adhesion and counteracts bacteria-induced detachment of epithelial cells." Journal of Cell Biology **170**: 825-836.

Mugisha, L., S. Kondgen, et al. (2014). "Nasopharyngeal colonization by potentially pathogenic bacteria found in healthy semi-captive wild-born chimpanzees in Uganda." Am J Primatol **76**(2): 103-110.

Mukherjee, S., D. Ganguli, et al. (2014). "Global footprints of purifying selection on Toll-like receptor genes primarily associated with response to bacterial infections in humans." Genome Biol Evol **6**(3): 551-558.

Nagaishi, T., Z. Chen, et al. (2008). "CEACAM1 and the regulation of mucosal inflammation." Mucosal Immunol **1 Suppl 1**: S39-42.

Nagel, G., F. Grunert, et al. (1993). "Genomic organization, splice variants and expression of CGM1, a CD66-related member of the carcinoembryonic antigen gene family." European Journal of Biochemistry **214**: 27-35.

Nakano, Y., H. Aso, et al. (2017). "A conflict of interest: the evolutionary arms race between mammalian APOBEC3 and lentiviral Vif." Retrovirology **14**(1): 31.

Nantel, A., D. Dignard, et al. (2002). "Transcription profiling of Candida albicans cells undergoing the yeast-to-hyphal transition." Mol Biol Cell **13**(10): 3452-3465.

Nathan, C. (2006). "Neutrophils and immunity: challenges and opportunities." Nat Rev Immunol **6**(3): 173-182.

Neish, A. S. (2014). "Mucosal immunity and the microbiome." Ann Am Thorac Soc **11 Suppl 1**: S28-32.

Newman, L., J. Rowley, et al. (2015). "Global Estimates of the Prevalence and Incidence of Four Curable Sexually Transmitted Infections in 2012 Based on Systematic Review and Global Reporting." PLoS One **10**(12): e0143304.

Nimmerjahn, F. and J. V. Ravetch (2008). "Fcgamma receptors as regulators of immune responses." Nat Rev Immunol **8**(1): 34-47.

Nishida, A. H. and H. Ochman (2018). "Rates of gut microbiome divergence in mammals." Mol Ecol **27**(8): 1884-1897.

Noble, R. C., R. M. Cooper, et al. (1979). "Pharyngeal colonisation by Neisseria gonorrhoeae and Neisseria meningitidis in black and white patients attending a venereal disease clinic." Br J Vener Dis **55**(1): 14-19.

Nord, C. E., A. Heimdahl, et al. (1986). "Antimicrobial induced alterations of the human oropharyngeal and intestinal microflora." Scand J Infect Dis Suppl **49**: 64-72.

Norskov-Lauritsen, N. (2014). "Classification, identification, and clinical significance of haemophilus and aggregatibacter species with host specificity for humans." Clin Microbiol Rev **27**(2): 214-240.

Nyholm, S. V. and M. J. McFall-Ngai (2004). "The winnowing: establishing the squid-vibrio symbiosis." Nat Rev Microbiol **2**(8): 632-642.

O'Hara, A. M. and F. Shanahan (2006). "The gut flora as a forgotten organ." EMBO Rep **7**(7): 688-693.

Ochman, H., M. Worobey, et al. (2010). "Evolutionary relationships of wild hominids recapitulated by gut microbial communities." PLoS Biol **8**(11): e1000546.

Oikawa, S., C. Inuzuka, et al. (1991). "A specific heterotypic cell adhesion activity between members of carcinoembryonic antigen family, W272 and NCA, is mediated by N-domains." Journal of Biological Chemistry **266**(13): 7995-8001.

Oikawa, S., C. Inuzuka, et al. (1989). "Cell adhesion activity of non-specific cross-reacting antigen (NCA) and carcinoembryonic antigen (CEA) expressed on CHO cell surface: homophilic and heterophilic adhesion." Biochemical & Biophysical Research Communications **164**(1): 39-45.

Otte, J. M., E. Cario, et al. (2004). "Mechanisms of cross hyporesponsiveness to Toll-like receptor bacterial ligands in intestinal epithelial cells." Gastroenterology **126**(4): 1054-1070.

Ouwerkerk, J. P., W. M. de Vos, et al. (2013). "Glycobiome: bacteria and mucus at the epithelial interface." Best Pract Res Clin Gastroenterol **27**(1): 25-38.

Pan, X., Y. Yang, et al. (2014). "Molecular basis of host specificity in human pathogenic bacteria." Emerg Microbes Infect **3**(3): e23.

Park, Y. S., W. S. Shin, et al. (2008). "In vitro and in vivo activities of echinomycin against clinical isolates of Staphylococcus aureus." J Antimicrob Chemother **61**(1): 163-168.

Patti, J. M., B. L. Allen, et al. (1994). "MSCRAMM-mediated adherence of microorganisms to host tissues." Annu Rev Microbiol **48**: 585-617.

Peitsch, M. C. (1996). "ProMod and Swiss-Model: Internet-based tools for automated comparative protein modelling." Biochem Soc Trans **24**(1): 274-279.

Peterson, K. M., J. B. Baseman, et al. (1983). "Treponema pallidum receptor binding proteins interact with fibronectin." J Exp Med **157**(6): 1958-1970.

Pickard, J. M. and G. Nunez (2019). "Pathogen Colonization Resistance in the Gut and Its Manipulation for Improved Health." Am J Pathol **189**(7): 1300-1310.

Pier, G. B., M. Grout, et al. (1998). "Salmonella typhi uses CFTR to enter intestinal epithelial cells." Nature **393**(6680): 79-82.

Pils, S., D. Gerrard, et al. (2008). "CEACAM3: an innate immune receptor directed against human-resticted bacterial pathogens." International Journal of Medical Microbiology **298**(7-8): 553-560.

Pils, S., K. Kopp, et al. (2012). "The adaptor molecule Nck localizes the WAVE complex to promote actin polymerization during CEACAM3-mediated phagocytosis of bacteria." PLoS One **7**(3): e32808.

Pils, S., T. Schmitter, et al. (2006). "Quantification of bacterial invasion into adherent cells by flow cytometry." Journal of Microbiological Methods **65**: 301-310.

Powell, J. E., S. P. Leonard, et al. (2016). "Genome-wide screen identifies host colonization determinants in a bacterial gut symbiont." Proc Natl Acad Sci U S A **113**(48): 13887-13892.

Qin, Y., L. Zhang, et al. (2016). "Innate immune cell response upon Candida albicans infection." Virulence **7**(5): 512-526.

Quast, C., E. Pruesse, et al. (2013). "The SILVA ribosomal RNA gene database project: improved data processing and web-based tools." Nucleic Acids Res **41**(Database issue): D590-596.

Quigley, E. M. (2013). "Gut bacteria in health and disease." Gastroenterol Hepatol (N Y) **9**(9): 560-569.

Rader, B. A. and S. V. Nyholm (2012). "Host/microbe interactions revealed through "omics" in the symbiosis between the Hawaiian bobtail squid Euprymna scolopes and the bioluminescent bacterium Vibrio fischeri." Biol Bull **223**(1): 103-111.

Rakoff-Nahoum, S., J. Paglino, et al. (2004). "Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis." Cell **118**(2): 229-241.

Ramanan, N. and Y. Wang (2000). "A high-affinity iron permease essential for Candida albicans virulence." Science **288**(5468): 1062-1064.

Rao, V. K., G. P. Krasan, et al. (1999). "Molecular determinants of the pathogenesis of disease due to non-typable Haemophilus influenzae." FEMS Microbiol Rev **23**(2): 99-129.

Reddy, M. S., T. F. Murphy, et al. (1997). "Middle ear mucin glycoprotein: purification and interaction with nontypable Haemophilus influenzae and Moraxella catarrhalis." Otolaryngol Head Neck Surg **116**(2): 175-180.

Regenberg, B., T. Grotkjaer, et al. (2006). "Growth-rate regulated genes have profound impact on interpretation of transcriptome profiling in Saccharomyces cerevisiae." Genome Biol **7**(11): R107.

Rejman, J., V. Oberle, et al. (2004). "Size-dependent internalization of particles via the pathways of clathrin- and caveolae-mediated endocytosis." Biochem J **377**(Pt 1): 159-169.

Rescigno, M., M. Urbano, et al. (2001). "Dendritic cells express tight junction proteins and penetrate gut epithelial monolayers to sample bacteria." Nat Immunol **2**(4): 361-367.

Rest, R. F. and W. M. Shafer (1989). "Interactions of *Neisseria gonorrhoeae* with human neutrophils." Clinical Microbiology Reviews **2 (Suppl.)**: S83-S91.

Rhesus Macaque Genome Sequencing and Analysis Consortium, R. A. Gibbs, et al. (2007). "Evolutionary and Biomedical Insights from the Rhesus Macaque Genome 10.1126/science.1139247." Science **316**(5822): 222-234.

Richard, M. L. and H. Sokol (2019). "The gut mycobiota: insights into analysis, environmental interactions and role in gastrointestinal diseases." Nat Rev Gastroenterol Hepatol **16**(6): 331-345.

Rikkerink, E. H., B. B. Magee, et al. (1988). "Opaque-white phenotype transition: a programmed morphological transition in Candida albicans." J Bacteriol **170**(2): 895-899.

Rizzatti, G., L. R. Lopetuso, et al. (2017). "Proteobacteria: A Common Factor in Human Diseases." Biomed Res Int **2017**: 9351507.

Roberton, A. M. and R. A. Stanley (1982). "In vitro utilization of mucin by Bacteroides fragilis." Appl Environ Microbiol **43**(2): 325-330.

Rogers, N. C., E. C. Slack, et al. (2005). "Syk-dependent cytokine induction by Dectin-1 reveals a novel pattern recognition pathway for C type lectins." Immunity **22**(4): 507-517.

Rook, G. A. and L. R. Brunet (2005). "Microbes, immunoregulation, and the gut." Gut **54**(3): 317-320.

Roth, A., C. Mattheis, et al. (2013). "Innate recognition by neutrophil granulocytes differs between *Neisseria gonorrhoeae* strains causing local or disseminating infections." Infect Immun **81(7)**(Juli): 2358-2370.

Rougerie, P., V. Miskolci, et al. (2013). "Generation of membrane structures during phagocytosis and chemotaxis of macrophages: role and regulation of the actin cytoskeleton." Immunol Rev **256**(1): 222-239.

Rushmore, J., A. B. Allison, et al. (2017). "Screening wild and semi-free ranging great apes for putative sexually transmitted diseases: Evidence of Trichomonadidae infections." Am J Primatol **79**(4): 1-2.

Saijo, S., N. Fujikado, et al. (2007). "Dectin-1 is required for host defense against Pneumocystis carinii but not against Candida albicans." Nat Immunol **8**(1): 39-46.

Sakwinska, O., M. Giddey, et al. (2011). "Staphylococcus aureus host range and human-bovine host shift." Appl Environ Microbiol **77**(17): 5908-5915.

Salama, N. R., M. L. Hartung, et al. (2013). "Life in the human stomach: persistence strategies of the bacterial pathogen Helicobacter pylori." Nat Rev Microbiol **11**(6): 385-399.

Saleheen, D., P. Natarajan, et al. (2017). "Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity." Nature **544**(7649): 235-239.

Sandstrom, I. (1987). "Etiology and diagnosis of neonatal conjunctivitis." Acta Paediatr Scand **76**(2): 221-227.

Sarantis, H. and S. D. Gray-Owen (2007). "The specific innate immune receptor CEACAM3 triggers neutrophil bactericidal activities via a Syk kinase-dependent pathway." Cell Microbiol **9**(9): 2167-2180.

Sarantis, H. and S. D. Gray-Owen (2012). "Defining the roles of human carcinoembryonic antigen-related cellular adhesion molecules during neutrophil responses to Neisseria gonorrhoeae." Infect Immun **80**(1): 345-358.

Sasse, C., M. Hasenberg, et al. (2013). "White-opaque switching of Candida albicans allows immune evasion in an environment-dependent fashion." Eukaryot Cell **12**(1): 50-58.

Sasson, G., S. Kruger Ben-Shabat, et al. (2017). "Heritable Bovine Rumen Bacteria Are Phylogenetically Related and Correlated with the Cow's Capacity To Harvest Energy from Its Feed." mBio **8**(4).

Sato, K., Y. Kuroki, et al. (2015). "Resequencing of the common marmoset genome improves genome assemblies and gene-coding sequence analysis." Sci Rep **5**: 16894.

Savage, D. C. (1970). "Associations of indigenous microorganisms with gastrointestinal mucosal epithelia." Am J Clin Nutr **23**(11): 1495-1501.

Savage, D. C. (1977). "Microbial ecology of the gastrointestinal tract." Annu Rev Microbiol **31**: 107-133.

Scally, A., J. Y. Dutheil, et al. (2012). "Insights into hominid evolution from the gorilla genome sequence." Nature **483**(7388): 169-175.

Schloss, P. D. and J. Handelsman (2004). "Status of the microbial census." Microbiol Mol Biol Rev **68**(4): 686-691.

Schmitter, T., F. Agerer, et al. (2004). "Granulocyte CEACAM3 is a phagocytic receptor of the innate immune system that mediates recognition and elimination of human-specific pathogens." Journal of Experimental Medicine **199**: 35-46.

Schmitter, T., S. Pils, et al. (2007). "The granulocyte receptor CEACAM3 directly associates with Vav to promote phagocytosis of human pathogens." Journal of Immunology **178**(8): 3797-3805.

Schmitter, T., S. Pils, et al. (2007). "Opa proteins of pathogenic *Neisseriae* initiate Src-kinase-dependent or lipid raft-mediated uptake via distinct human CEACAM isoforms." Infection & Immunity **75**(8): 4116-4126.

Schorey, J. S. and C. Lawrence (2008). "The pattern recognition receptor Dectin-1: from fungi to mycobacteria." Curr Drug Targets **9**(2): 123-129.

Sekirov, I., S. L. Russell, et al. (2010). "Gut microbiota in health and disease." Physiol Rev **90**(3): 859-904.

Shanahan, F. (2005). "Physiological basis for novel drug therapies used to treat the inflammatory bowel diseases I. Pathophysiological basis and prospects for probiotic therapy in inflammatory bowel disease." Am J Physiol Gastrointest Liver Physiol **288**(3): G417-421.

Shepherd, E. S., W. C. DeLoache, et al. (2018). "An exclusive metabolic niche enables strain engraftment in the gut microbiota." Nature **557**(7705): 434-438.

Si, H., A. D. Hernday, et al. (2013). "Candida albicans white and opaque cells undergo distinct programs of filamentous growth." PLoS Pathog **9**(3): e1003210.

Simren, M., G. Barbara, et al. (2013). "Intestinal microbiota in functional bowel disorders: a Rome foundation report." Gut **62**(1): 159-176.

Singh, R. K., H. W. Chang, et al. (2017). "Influence of diet on the gut microbiome and implications for human health." J Transl Med **15**(1): 73.

Sintsova, A., H. Sarantis, et al. (2014). "Global analysis of neutrophil responses to Neisseria gonorrhoeae reveals a self-propagating inflammatory program." PLoS Pathog **10**(9): e1004341.

Sintsova, A., H. Wong, et al. (2015). "Selection for a CEACAM receptor-specific binding phenotype during Neisseria gonorrhoeae infection of the human genital tract." Infect Immun **83**(4): 1372-1383.

Sit, S. T. and E. Manser (2011). "Rho GTPases and their role in organizing the actin cytoskeleton." J Cell Sci **124**(Pt 5): 679-683.

Skubitz, K. M. and A. P. Skubitz (2008). "Interdependency of CEACAM-1, -3, -6, and -8 induced human neutrophil adhesion to endothelial cells." J Transl Med **6**: 78.

Slevogt, H., S. Zabel, et al. (2008). "CEACAM1 inhibits Toll-like receptor 2-triggered antibacterial responses of human pulmonary epithelial cells." Nat Immunol **9**(11): 1270-1278.

Slutsky, B., M. Staebell, et al. (1987). ""White-opaque transition": a second high-frequency switching system in Candida albicans." J Bacteriol **169**(1): 189-197.

Snoeck, V., I. R. Peters, et al. (2006). "The IgA system: a comparison of structure and function in different species." Vet Res **37**(3): 455-467.

Soto, A., V. Martin, et al. (2014). "Lactobacilli and bifidobacteria in human breast milk: influence of antibiotherapy and other host and clinical factors." J Pediatr Gastroenterol Nutr **59**(1): 78-88.

Stark, P. L. and A. Lee (1982). "The microbial ecology of the large bowel of breast-fed and formula-fed infants during the first year of life." J Med Microbiol **15**(2): 189-203.

Stern, A., M. Brown, et al. (1986). "Opacity genes in *Neisseria gonorrhoeae*: Control of phase and antigenic variation." Cell **47**: 61-71.

Stevens, J. S. and A. K. Criss (2018). "Pathogenesis of Neisseria gonorrhoeae in the female reproductive tract: neutrophilic host response, sustained infection, and clinical sequelae." Curr Opin Hematol **25**(1): 13-21.

Streichert, T., A. Ebrahimnejad, et al. (2001). "The microbial receptor CEACAM3 is linked to the calprotectin complex in granulocytes." Biochemical & Biophysical Research Communications **289**(1): 191-197.

Svobodova, E., P. Staib, et al. (2012). "Differential interaction of the two related fungal species Candida albicans and Candida dubliniensis with human neutrophils." J Immunol **189**(5): 2502-2511.

Swanson, J., O. Barrera, et al. (1988). "Expression of outer membrane protein II by gonococci in experimental gonorrhea." Journal of Experimental Medicine **168**: 2121-2129.

Swanson, J., K. Robbins, et al. (1987). "Gene conversion variations generate structurally distinct pilin polypeptides in Neisseria gonorrhoeae." Journal of Experimental Medicine **165**: 1016-1025.

Swidsinski, A., V. Loening-Baucke, et al. (2005). "Spatial organization of bacterial flora in normal and inflamed intestine: a fluorescence in situ hybridization study in mice." World J Gastroenterol **11**(8): 1131-1140.

Tchoupa, A. K., S. Lichtenegger, et al. (2015). "Outer membrane protein P1 is the CEACAM-binding adhesin of *Haemophilus influenzae*." Mol Microbiol **98**(3): 440-455.

Tchoupa, A. K., T. Schuhmacher, et al. (2014). "Signaling by epithelial members of the CEACAM family - mucosal docking sites for pathogenic bacteria." Cell Commun Signal **12**: 27.

Teglund, S., A. Olsen, et al. (1994). "The pregnancy-specific glycoprotein (PSG) gene cluster on human chromosome 19: fine structure of the 11 PSG genes and identification of 6 new genes forming a third subgroup within the carcinoembryonic antigen (CEA) family." Genomics **23**(3): 669-684.

Teixeira, A. M., J. Fawcett, et al. (1994). "The N-domain of the biliary glycoprotein (BGP) adhesion molecule mediates homotypic binding: domain interactions and epitope analysis of BGPc." Blood **84**(1): 211-219.

Thomas, C. J. and K. Schroder (2013). "Pattern recognition receptor function in neutrophils." Trends Immunol **34**(7): 317-328.

Thompson, J. and W. Zimmermann (1988). "The carcinoembryonic antigen gene family: structure, expression and evolution." Tumour Biol **9**(2-3): 63-83.

Thompson, J. A., F. Grunert, et al. (1991). "Carcinoembryonic antigen gene family: molecular biology and clinical perspectives." Journal of Clinical Laboratory Analysis **5**: 344-366.

Toleman, M., E. Aho, et al. (2001). "Expression of pathogen-like Opa adhesins in commensal *Neisseria*: genetic and functional analysis." Cell Microbiol **3**(1): 33-44.

Trompette, A., E. S. Gollwitzer, et al. (2014). "Gut microbiota metabolism of dietary fiber influences allergic airway disease and hematopoiesis." Nat Med **20**(2): 159-166.

Tsai, H. H., D. Sunderland, et al. (1992). "A novel mucin sulphatase from human faeces: its identification, purification and characterization." Clin Sci (Lond) **82**(4): 447-454.

Turnbaugh, P. J., F. Backhed, et al. (2008). "Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome." Cell Host Microbe **3**(4): 213-223.

Underhill, D. M. and A. Ozinsky (2002). "Phagocytosis of microbes: complexity in action." Annu Rev Immunol **20**: 825-852.

Underhill, D. M., E. Rossnagle, et al. (2005). "Dectin-1 activates Syk tyrosine kinase in a dynamic subset of macrophages for reactive oxygen production." Blood **106**(7): 2543-2550.

Vaca, D. J., A. Thibau, et al. (2020). "Interaction with the host: the role of fibronectin and extracellular matrix proteins in the adhesion of Gram-negative bacteria." Med Microbiol Immunol **209**(3): 277-299.

Valdes-Varela, L., M. Alonso-Guervos, et al. (2016). "Screening of Bifidobacteria and Lactobacilli Able to Antagonize the Cytotoxic Effect of Clostridium difficile upon Intestinal Epithelial HT29 Monolayer." Front Microbiol **7**: 577.

Van Eldere, J., M. P. Slack, et al. (2014). "Non-typeable Haemophilus influenzae, an under-recognised pathogen." Lancet Infect Dis **14**(12): 1281-1292.

van Tongeren, S. P., J. P. Slaets, et al. (2005). "Fecal microbiota composition and frailty." Appl Environ Microbiol **71**(10): 6438-6442.

van Valen, L. (1973). A new evolutionary law. In: Evolutionary Theory.

Vasseur, E., E. Patin, et al. (2011). "The selective footprints of viral pressures at the human RIG-I-like receptor family." Hum Mol Genet **20**(22): 4462-4474.

Villullas, S., D. J. Hill, et al. (2007). "Mutational analysis of human CEACAM1: the potential of receptor polymorphism in increasing host susceptibility to bacterial infection." Cell Microbiol **9**(2): 329-346.

Virji, M. (1996). "Meningococcal disease: epidemiology and pathogenesis." Trends in Microbiology **4**: 466-469.

Virji, M. (2009). "Pathogenic neisseriae: surface modulation, pathogenesis and infection control." Nat Rev Microbiol **7**(4): 274-286.

Virji, M., D. Evans, et al. (2000). "Carcinoembryonic antigens are targeted by diverse strains of typable and non-typable *Haemophilus influenzae*." Molecular Microbiology **36**(4): 784-795.

Virji, M., K. Makepeace, et al. (1996). "Carcinoembryonic antigens (CD66) on epithelial cells and neutrophils are receptors for Opa proteins of pathogenic *Neisseriae*." Molecular Microbiology **22**: 941-950.

Virji, M., S. M. Watt, et al. (1996). "The N-domain of the human CD66a adhesion molecule is a target for Opa proteins of *Neisseria meningitidis* and *Neisseria gonorrhoeae*." Molecular Microbiology **22**: 929-939.

Voges, M., V. Bachmann, et al. (2012). "Extracellular IgC2-like domains of CEACAMS mediate PI3K sensitivity during uptake of pathogens." PloS One accepted **7 e39908**.

Voight, B. F., S. Kudaravalli, et al. (2006). "A map of recent positive selection in the human genome." PLoS Biol **4**(3): e72.

Wall, R., R. P. Ross, et al. (2009). "Role of gut microbiota in early infant development." Clin Med Pediatr **3**: 45-54.

Wang, B. and L. Li (2015). "Who determines the outcomes of HBV exposure?" Trends Microbiol **23**(6): 328-329.

Wang, J., S. D. Gray-Owen, et al. (1998). "Opa binding to cellular CD66 receptors mediates the transcellular traversal of *Neisseria gonorrhoeae* across polarized T84 epithelial cell monolayers." Molecular Microbiology **30**(3): 657-671.

Waterhouse, A., M. Bertoni, et al. (2018). "SWISS-MODEL: homology modelling of protein structures and complexes." Nucleic Acids Res **46**(W1): W296-W303.

Watt, S. M., A. M. Teixeira, et al. (2001). "Homophilic adhesion of human CEACAM1 involves N-terminal domain interactions: structural analysis of the binding site." Blood **98**(5): 1469-1479.

Wei, M., R. Shinkura, et al. (2011). "Mice carrying a knock-in mutation of Aicda resulting in a defect in somatic hypermutation have impaired gut homeostasis and compromised mucosal defense." Nat Immunol **12**(3): 264-270.

Weiss, A. and D. R. Littman (1994). "Signal transduction by lymphocyte antigen receptors." Cell **76**(2): 263-274.

Wenger, J. D., A. W. Hightower, et al. (1990). "Bacterial meningitis in the United States, 1986: report of a multistate surveillance study. The Bacterial Meningitis Study Group." J Infect Dis **162**(6): 1316-1323.

Whitman, W. B., D. C. Coleman, et al. (1998). "Prokaryotes: the unseen majority." Proc Natl Acad Sci U S A **95**(12): 6578-6583.

Wikstrom, K., G. Kjellstrom, et al. (1996). "Homophilic intercellular adhesion mediated by C-CAM is due to a domain 1-domain 1 reciprocal binding." Exp Cell Res **227**(2): 360-366.

Wolf, J. B., A. Kunstner, et al. (2009). "Nonlinear dynamics of nonsynonymous (dN) and synonymous (dS) substitution rates affects inference of selection." Genome Biol Evol **1**: 308-319.

Wu, G. D., J. Chen, et al. (2011). "Linking long-term dietary patterns with gut microbial enterotypes." Science **334**(6052): 105-108.

Xu, J. and J. I. Gordon (2003). "Honor thy symbionts." Proc Natl Acad Sci U S A **100**(18): 10452-10459.

Yamaguchi, M., Y. Terao, et al. (2013). "Pleiotropic virulence factor - Streptococcus pyogenes fibronectin-binding proteins." Cell Microbiol **15**(4): 503-511.

Zebhauser, R., R. Kammerer, et al. (2005). "Identification of a novel group of evolutionarily conserved members within the rapidly diverging murine Cea family." Genomics **86**(5): 566-580.

Zhan, Y., J. V. Virbasius, et al. (2002). "The p40phox and p47phox PX domains of NADPH oxidase target cell membranes via direct and indirect recruitment by phosphoinositides." J Biol Chem **277**(6): 4512-4518.

Zheng, J., K. K. Miller, et al. (2011). "Carcinoembryonic antigen-related cell adhesion molecule 16 interacts with alpha-tectorin and is mutated in autosomal dominant hearing loss (DFNA4)." Proc Natl Acad Sci U S A **108**(10): 4218-4223.

Zhou, G. Q., Y. Zhang, et al. (2001). "The carcinoembryonic antigen (CEA) gene family in non-human primates." Gene **264**(1): 105-112.

Zhou, H., A. Fuks, et al. (1993). "Homophilic adhesion between Ig superfamily carcinoembryonic antigen molecules involves double reciprocal bonds." J Cell Biol **122**(4): 951-960.

Zhu, Y., K. J. Kwiatkowski, et al. (2015). "Outer membrane proteins related to SusC and SusD are not required for Cytophaga hutchinsonii cellulose utilization." Appl Microbiol Biotechnol **99**(15): 6339-6350.

Zimin, A. V., A. S. Cornish, et al. (2014). "A new rhesus macaque assembly and annotation for next-generation sequencing analyses." Biol Direct **9**(1): 20.