A gut commensal microbiome-host protein network map reveals bacterial modulation of human immune signaling

Dissertation der Fakultät für Biologie

der Ludwig-Maximilians-Universität München

Veronika Young

München, 2023

Diese Dissertation wurde angefertigt

unter der Leitung von Prof. Dr. Pascal Falter-Braun

am Institut für Netzwerkbiologie (INET)

am Helmholtz Zentrum München.

Erstgutachter: Prof. Dr. Pascal Falter-Braun

Zweitgutachter: Prof. Dr. Simon Heilbronner

Tag der Abgabe: 20.12.2023

Tag der mündlichen Prüfung: 11.06.2024

Eidesstattliche Erklärung

Ich versichere hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbstständig und ohne unerlaubte Hilfe angefertigt wurde. Des Weiteren erkläre ich, dass ich nicht anderweitig ohne Erfolg versucht habe, eine Dissertation einzureichen oder mich der Doktorprüfung zu unterziehen. Die folgende Dissertation liegt weder ganz noch in wesentlichen Teilen einer anderen Prüfungskommission vor.

München, den 25.06.2024

Veronika Young

Abstract

Complex diseases such as cardiovascular illnesses, cancer, respiratory diseases, and diabetes have been on the rise worldwide contributing to 70% of all deaths [1]. While these illnesses are superficially not associated with microbial organisms accumulating research links bacteria of the gut microbiome to a diverse range of complex diseases [2]. Especially Pseudomonadota, the third most abundant phylum in the gut, has been associated with several of those illnesses e.g., metabolic sicknesses and cancer [3]. However, the underlying mechanisms mediating the bacterial impact on host health remain largely unknown.

Pathogenic representatives of the Pseudomonadota phylum are well-known for employing a type three secretion system (T3SS) to manipulate host cells and thereby mediate infectious diseases. Prominent examples are *Salmonella, Shigella*, enterohemorrhagic *Escherichia coli (EHEC), and* enteropathogenic *Escherichia coli (*EPEC) as well as *Yersinia pestis* responsible for wiping out one-third of the European population during the plague [4,5]. Yet, T3SSs have also been detected in plant mutualists e.g., rhizobia strains of the rhizobia-legume symbiosis [6]. Furthermore, bacteria that do not exhibit a pathogenic or mutualistic lifestyle also seem to encode for T3SSs [5]. Work outside of this thesis as part of the same project detected T3SSs also in commensal Pseudomonadota of the human gut microbiome [7]. However, the impact of the T3SS effectors on the host cell and subsequently host health is not known. Therefore, this thesis aimed to elucidate the impact of T3SS effectors expressed by commensal gut Pseudomonadota on host functions in the context of human health and disease.

To assess the impact of bacterial effectors on the human host a network map of protein-protein interactions (PPIs) between gut commensal bacterial effectors and human proteins was generated. To this end, an ORFeome collection of bacterial effectors was established by cloning 959 T3SS effectors from known, culturable strains as well as from metagenomic data of the human gut. Testing these bacterial effectors against the human ORFeome v9.1 collection consisting of 17,408 protein-coding genes with a systematic, high-throughput yeast two-hybrid (Y2H) pipeline gave rise to the human-microbiome meta-interactome map (HuMMI). The network consists of 1,263 interactions mediated by 289 effectors and 430 human proteins. HuMMI was subjected to a detailed quality control to assess the reliability of the used pipeline and the quality of the interactions. For this purpose, reference sets were assembled to benchmark the Y2H and an orthogonal assay, which was employed to re-test a subset of HuMMI. In addition, the saturation of HuMMI i.e., the percentage of discovered interactions compared to all detectable interactions, was assessed by a Y2H repeat screen.

After demonstrating the reliability of the employed Y2H pipeline and the comparability of HuMMI to well-documented, literature-curated-interactions, validation experiments *in vitro* were conducted based on the functional analysis of the effector targets. As the bacterial effectors targeted human proteins involved in immune signaling their impact on the transcription factor nuclear factor kappa B (NF-κB) was assessed. A cell-based reporter assay was employed testing the ability of the effectors to modulate NF-κB activity. Five effectors significantly activated NF-κB, while three effectors seemed to inhibit the transcription factor significantly. Further impacts on human immune signaling by T3SS effectors were shown by collaborators reporting increased ICAM1 expression as well as up- and downregulation of pro-inflammatory cytokine secretion from a colon cell line upon

effector transfection. The opposing effects of bacterial effectors on immune signaling pathways suggest different influences of gut commensal T3SS effectors on the human host.

In conclusion, this study introduced a novel mechanism by which gut commensals might impact the human host. T3SS effectors potentially affect human immune signaling locally and systemically via cytokine secretion potentially affecting the risk of complex disease. Thereby, this works launches the investigation into gut commensal T3SS effectors and their impacts on host health and disease.

Zusammenfassung

Komplexe Krankheiten z.B. Herz-Kreislauf-Erkrankungen, Krebs, Atemwegserkrankungen und Diabetes sind weltweit für 70% aller Todesfälle verantwortlich. Während diese Krankheiten oberflächlich betrachtet nicht mit mikrobiellen Organismen in Verbindung stehen, gibt es zunehmend mehr Forschungsergebnisse, die eine Assoziation zwischen dem Darmmikrobiom und einer Vielzahl komplexer Krankheiten zeigen. Insbesondere Pseudomonadota, das dritthäufigste Phylum im Darmmikrobiom, wurde mit mehreren dieser Krankheiten in Verbindung gebracht, z. B. mit Stoffwechselkrankheiten und Krebs. Die zugrunde liegenden Mechanismen hinter dem Einfluss der Bakterien auf die Gesundheit des Wirts, sind jedoch noch weitgehend unbekannt.

Pathogene Vertreter der Pseudomonadota sind für ihr Typ-3-Sekretionssystem (T3SS) bekannt, mit dem sie Wirtzellen manipulieren, um den Organismus zu infizieren. Prominente Beispiele sind Salmonellen, Shigellen, enterohämorrhagische *Escherichia coli* (EHEC) und enteropathogene *Escherichia coli* (EPEC) sowie *Yersinia pestis*. T3SS wurden jedoch auch in Pflanzen-Symbionten nachgewiesen, z. B. in Rhizobien-Stämmen der Rhizobien-Leguminosen-Symbiose. Darüber hinaus scheinen auch Bakterien, die keine pathogene oder symbiotische Lebensweise aufweisen, T3SS zu kodieren. Im Rahmen eines EU-Projekts, zu der auch diese Arbeit gehört, wurden T3SS ebenfalls in kommensalen Pseudomonadota des menschlichen Darmmikrobioms nachgewiesen. Die Auswirkungen der T3SS Effektoren auf die Wirtszelle und damit auf die Gesundheit des Wirts sind jedoch nicht bekannt.

Um die Auswirkungen bakterieller Effektoren von Darmkommensalen auf den menschlichen Wirt zu bewerten, wurde eine Netzwerkkarte der Protein-Protein-Interaktionen (PPIs) zwischen T3SS Effektoren und menschlichen Proteinen erstellt. Zu diesem Zweck wurde eine ORFeome-Sammlung von bakteriellen Effektoren angefertigt, indem 959 T3SS Effektoren aus bekannten, kultivierbaren Stämmen sowie aus metagenomischen Daten des menschlichen Darms kloniert wurden. Diese Effektoren wurden auf Interaktionen mit 17.408 proteinkodierenden menschlichen Genen getestet, mit Hilfe einer systematischen Hochdurchsatz-Hefe-Zwei-Hybrid-Pipeline (Y2H) um die human-microbiome meta-interactome (HuMMI) Netzwerkkarte zu erstellen. HuMMI besteht aus 1.263 Interaktionen zwischen 289 Effektoren und 430 menschlichen Proteinen. HuMMI wurde einer detaillierten Qualitätskontrolle unterzogen, um die Zuverlässigkeit der Pipeline und die Qualität der Interaktionen zu bewerten. Dafür wurden Referenzdatensätze zusammengestellt um die Y2H-Pipeline und einen orthogonaler Assay zu prüfen, mit dem ein Teil des HuMMI Datensatzes validiert wurde. Darüber hinaus wurde die Sättigung von HuMMI, d. h. der Prozentsatz der entdeckten Interaktionen im Vergleich zu allen nachweisbaren Interaktionen, durch Wiederholungen eines Teils der Y2H Experimente bewertet.

Nachdem die Zuverlässigkeit der verwendeten Y2H-Pipeline nachgewiesen und die Vergleichbarkeit der Interaktionen in HuMMI mit in der Literatur gut dokumentierten Interaktionen gezeigt worden war, wurden Validierungsexperimente *in vitro* auf Grundlage der funktionellen Analyse der interagierenden Humanproteinen durchgeführt. Da die bakteriellen Effektoren mit menschlichen Proteinen interagierten, die an Signalwegen des Immunsystems beteiligt sind, wurde ihr Einfluss auf den Transkriptionsfaktor Nuklearfaktor kappa B (NF-κB) untersucht. Mit einem zellbasierten Reporter-Assay wurde getestet, ob die

Effektoren die Aktivität von NF-κB beeinflussen können. Fünf Effektoren aktivierten den Transkriptionsfaktor signifikant, während drei Effektoren den Transkriptionsfaktor signifikant zu hemmen schienen. Weitere Auswirkungen der T3SS Effektoren auf das menschliche Immunsystem wurden von Kooperationspartnern nachgewiesen, die über eine erhöhte ICAM1-Expression sowie eine Hoch- und Herunterregulierung der Sekretion proinflammatorischer Zytokine in einer Darmzelllinie berichteten. Der gegensätzliche Effekt bakterieller Effektoren auf Immunsignalwege lassen auf unterschiedliche Einflüsse kommensaler T3SS Effektoren auf den menschlichen Wirt schließen.

Diese Arbeit zeigt einen neuen Mechanismus auf, mit dem Darmkommensale den menschlichen Wirt beeinflussen können. T3SS Effektoren beeinträchtigen möglicherweise Signalwege des menschlichen Immunsystems auf lokaler und systemischer Ebene über die Sekretion von Zytokinen, was sich möglicherweise auf das Risiko komplexer Krankheiten auswirkt. Diese Arbeit ist somit der Auftakt zur Erforschung der T3SS Effektoren im Darmmikrobiom und ihrer Auswirkungen auf die Gesundheit des Wirts.

List of Publications

Young V, Dohai B, Hitch TCA, Hyden P, Weller B, van Heusden NS, Saha D., Fernandez Macgregor J, Maseko SB, Lin CW, Boujeant M, Choteau S.A, Ober F, Schwehn P, Rothballer ST, Altmann M, Altmann S, Strobel A, Rothballer M, Tofaute MJ, Heinig M, Clavel T, Twizere JC, Vincentelli R, Boes M, Krappmann D, Falter C, Rattei T, Brun C, Zanzoni A, Falter-Braun P. A gut meta-interactome map reveals modulation of human immunity by microbiome effectors. bioRxiv 2023.09.25.559292 (2023)

Weller B, Lin C, Pogoutse O, Sauer M, Marín-de la Rosa N, Strobel A, **Young V**, Knapp JJ, Rayhan A, Falter C, Kim DK, Roth FP, Falter-Braun P. A resource of human coronavirus protein-coding sequences in a flexible, multipurpose Gateway Entry clone collection. G3 Bethesda Md 13, jkad105 (2023)

Kim DK, Weller B, Lin CW, Sheykhkarimli D, Knapp JJ, Dugied G, Zanzoni A, Pons C, Tofaute MJ, Maseko SB, Spirohn K, Laval F, Lambourne L, Kishore N, Rayhan A, Sauer M, **Young V**, Halder H, la Rosa NM, Pogoutse O, Strobel A, Schwehn P, Li R, Rothballer ST, Altmann M, Cassonnet P, Coté AG, Vergara LE, Hazelwood I, Liu BB, Nguyen M, Pandiarajan R, Dohai B, Coloma PAR, Poirson J, Giuliana P, Willems L, Taipale M, Jacob Y, Hao T, Hill DE, Brun C, Twizere JC, Krappmann D, Heinig M, Falter C, Aloy P, Demeret C, Vidal M, Calderwood MA, Roth FP, Falter-Braun P. A proteome-scale map of the SARS-CoV-2–human contactome. Nat. Biotechnol. 41, 140–149 (2023)

Altmann M, Altmann S, Rodriguez PA, Weller B, Elorduy Vergara L, Palme J, Marín-de la Rosa N, Sauer M, Wenig M, Villaécija-Aguilar JA, Sales J, Lin CW, Pandiarajan R, **Young V**, Strobel A, Gross L, Carbonnel S, Kugler KG, Garcia-Molina A, Bassel GW, Falter C, Mayer KFX, Gutjahr C, Vlot AC, Grill E, Falter-Braun P. Extensive signal integration by the phytohormone protein network. Nature 583, 271–276 (2020)

Brei C, Stecher L, Meyer DM, **Young V**, Much D, Brunner S, Hauner H. Impact of Dietary Macronutrient Intake during Early and Late Gestation on Offspring Body Composition at Birth, 1, 3, and 5 Years of Age. Nutrients 10, 579 (2018)

Contents

List of Figures

List of Tables

Abbreviations

3-AT	3-amino-1,2,4 triazole
5-HT	5-hydroxytryptamine (serotonin)
abbr	abbreviation
AD	activation domain
AMP	adenosine monophosphate
AMR	antimicrobial resistance
AP-MS	affinity purification-mass spectrometry
bhLit_BM-v1	bacterial human literature binary multiple
bhRRS-v1	bacterial host random reference set
bp	base pair
CCK	cholecystokinin
CD	Crohn's disease
CHX	cycloheximide
DB	DNA-binding domain
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen
EHEC	enterohemorrhagic *Escherichia coli*
EPEC	enteropathogenic *Escherichia coli*
EPS	exopolymeric substances
ERK	extracellular signal-regulated kinase
ESCRT	endosomal sorting complexes required for transport
F/R	value of the firefly luciferase/value of the renilla luciferase
FMT	fecal microbiota transplantation
GALT	gut-associated lymphoid tissue
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
GFP	green fluorescent protein
GLP-1	glucagon-like peptide-1
GO	Gene Ontology
GPR	G-protein coupled receptor
GWAS	genome-wide association studies
His	histidine
HMP	Human Microbiome Project
hsPRS-v1/2	*Homo sapiens* Positive Reference Set version 1 or 2

hsRRS-v1/2	*Homo sapiens* Random Reference Set version 1 or 2
HSV-1	herpes simplex virus type 1
HuMEOme_v1	human microbiome effector ORFeome v1
HuMMI	human-microbiome meta-interactome map
HuRI	reference map of the human binary protein interactome
IKBKG	Inhibitor Of Nuclear Factor Kappa B Kinase Regulatory Subunit Gamma
IKK	inhibitory kappa B kinase
IL	interleukin
IQR	interquartile range
IκB	inhibitory kappa B
JNK	Jun N-terminal Kinase
kbp	kilo base pair
KEGG	Kyoto Encyclopedia of Genes and Genomes
Leu	leucin
LPS	lipopolysaccharides
M cells	membranous/microfold cells
MAGs	metagenome-assembled genomes
MAPK	mitogen-activated protein kinases
MetaHIT	METAgenomics of the Human Intestinal Tract
NF	nodulation factor
NF-κB	nuclear factor kappa B
NGS	next-generation sequencing
NLR	normalized luminescence ratio
OF	Old Friends
OMIM	Online Mendelian Inheritance in Man
ORF	open reading frame
PCR	polymerase chain reaction
PMA	phorbol 12-myristate 13-acetate
PPI	protein-protein-interactions
PRS	positive reference set
PYY	peptide tyrosine tyrosine
rRNA	ribosomal RNA
RRS	random reference set

S_A	assay sensitivity
SARS-CoV-2	severe acute respiratory syndrome coronavirus 2
SCFAs	short-chain fatty acids
SDS	sodium dodecyl sulfate
SE	standard error
SNPs	single nucleotide polymorphisms
S_o	overall sensitivity
S_s	sampling sensitivity
T3SS	type 3 secretion system
T_a	annealing temperature
TK	thymidine kinase
TLRs	toll-like receptors
T_m	melting temperature
TMA	trimethylamine
TMAO	trimethylamine N-oxide
TNFR	TNF receptor
Trp	tryptophan
UC	ulcerative colitis
Ura	uracil
V	volt
Y2H	yeast two-hybrid
Y3H	yeast three-hybrid
yN2H	yeast-based nanoluciferase complementation assay

1. Introduction

1.1 Shifts in perspectives on microbe-host interactions

For most of human history, the existence of microbial organisms and their various impacts on the human host remained undetected. In the 17th century, advancements in technology facilitated the microscopic observation of microbial organisms allowing their exploration. With increasing knowledge about microbial organisms, the perspectives on them became more differentiated. While microbes were mostly viewed as pathogenic during the 19th- and 20th-century, in the 21st century they are perceived on a spectrum ranging from parasitic over commensal to mutualistic [8].

1.1.1 The discovery of microbes

Even though microbial organisms impact numerous aspects of human life, missing technologies prevented the recognition of their existence until the 17th century. In 1609, Galileo Galilei (1564-1642) was among the first to modify a telescope into an early version of a microscope to facilitate the observation of insects [4,8]. Gradually, microscopes were modified and equipped with stronger magnifying lenses allowing the study of smaller insects or fungi yet were unsuitable for observing small organisms like microbes. The first microscope that fulfilled the requirements for detecting microbes was built by the Dutch merchant Antoni van Leeuwenhoek (1632-1723) who tailored his microscopes perfectly to this undertaking [4,8]. For instance, he ground powerful lenses that exceeded most commercially available ones at the time. Furthermore, his microscopes offered the stability that is needed for high-magnification microscopy and could be operated using transmitted light instead of incident light, which is more suitable for translucent samples [8]. Due to these adaptations, van Leeuwenhoek was the first human to observe microbial organisms, which he termed "animalcules" [4,9].

The existence of these animalcules was common knowledge by the 18th century and sparked heated debates about their origins among scientists and clergy. This gave rise to the theory of spontaneous generation i.e., that microbes can arise spontaneously without a parent generation. Disputes about this theory persisted stubbornly and only deliberate experiments by e.g., Lazzaro Spallanzani (1729-1799) and Louis Pasteur (1822-1895) could convince most contemporaries of the theory's falseness. In one of those experiments, Spallanzani demonstrated that microbial growth was not observed in a sealed flask containing bouillon that had been sterilized through boiling. His experiment

Figure 1 | Pasteur's swan neck flask. *Adapted from Slonczewski & Foster [9].*

contradicted the theory of spontaneous generation, yet his critics believed that the microbes inside the flask could not grow due to a lack of oxygen. Pasteur conducted a subsequent experiment employing a specially designed swan neck flask (Fig. 1), demonstrating that fermentation or microbial growth did not occur in the flask when the contained bouillon was boiled, and the air was filtered through a tube "heated to redness" [10]. However, when the

cooled-off flask was tilted, the bouillon encountered dust and microbes that had gathered inside the curve of the tube, which subsequently led to microbial growth inside the flask. Thereby, Pasteur disproved the theory of spontaneous generation and demonstrated that microbial organisms are present in the "atmosphere" [4,10].

Pasteur's experiments on the theory of spontaneous generation were greatly influential on e.g., the food industry and various fields of medicine. The former, of course, owes Pasteur the process of "pasteurization", a method to conserve foods and liquids by heating them to a specific temperature to destroy germs [11]. In medicine, Pasteur's work laid the ground for many practices and understandings that we depend on today. For instance, inspired by Pasteur's findings, the English surgeon Joseph Lister (1827-1912) hypothesized that airborne microbes provoked wound sepsis in patients after surgery. The number of deaths following sepsis was enormous at the time, which is not surprising from today's perspective as surgeon's coats were rarely washed and surgical tools remained mostly uncleaned between operations. Lister not only urged surgeons to work under clean conditions but also invented a spray apparatus to distribute phenol (carbolic acid) as a means of disinfection during surgery. Even though phenol is not considered safe for wound treatment anymore, Lister's antiseptic methods were powerful in his day and earned him the title "father of modern surgery" [12]. Lister's work is a good example of the application of Pasteur's findings and not the only area of medicine that greatly profited from his discoveries.

1.1.2 Microbes as human pathogens

Based on the work of early microbiologists such as van Leeuwenhoek and Pasteur, physicians began to understand the impact of microbes on human diseases. One prominent example is the German physician Robert Koch (1843-1910) who demonstrated that a particular disease is caused by a specific bacterium. His observations and empiric efforts together with those of his colleagues gave rise to the germ theory of disease i.e., the concept that human diseases are caused by microbes. This new concept gradually replaced the previous miasma theory which made bad smells and pollution responsible for human and animal diseases [13].

Koch had started his research on microbes as a physician in a small village at the German-Polish border. His first investigations focused on anthrax as local farmers were suffering the loss of infected sheep and cattle. He started his experimental series by injecting the blood of a deceased animal into a healthy rabbit upon which the rabbit died as well. Injecting the blood of the dead rabbit into another healthy rodent led to the same outcome. When observing the blood of the dead animals under the microscope Koch discovered the same rod-shaped bacteria in huge amounts in all samples. This suggested that the rod-shaped bacteria were responsible for the disease and that infections can be passed on from one organism to another establishing a chain of infection [4].

Koch's experiments on anthrax were relatively uncomplicated as *Bacillus anthracis* is present in large amounts in the blood of the infected organism which eases its microscopical detection [4,14]. It furthermore causes symptoms very quickly and remains infectious for a long time outside of a host [4]. Research on tuberculosis, Koch's next undertaking, was more tedious as the bacteria are small and difficult to spot under the microscope. Additionally, the disease develops slowly with periods of inactivity [4]. To identify the underlying pathogen *Mycobacterium tuberculosis* Koch needed to take a slightly different, more complex approach compared to his research on anthrax. The protocol that led to the successful identification of the pathogen

consisted of Koch isolating the bacterium in pure culture from several patients, transferring it to guinea pigs, which developed the disease, and finally re-isolating the pathogen from the animals again [14]. This approach fulfills Koch's postulates, which Koch formalized in a publication eight years later. For quite some time, these postulates were attributed to Koch alone, however, it is assumed that several researchers were involved in their conceptualization among which are Jacob Henle (1809-1885) [15], Edwin Klebs (1834-1913) and Friedrich Loeffler (1852-1915) [9,14]. The postulates are not exempt from exceptions as has been encountered by Koch himself during his research on cholera. Koch had isolated *Vibrio cholerae* in 1884 from a stool sample collected in Egypt and described its existence in 1886. He hypothesized that the bacterium secretes a "poison" that leads to the symptoms of cholera [16]. However, as *Vibrio cholerae* does not colonize the intestinal tract of other adult mammals besides humans the lack of suitable animal models was a hindrance in fulfilling his postulates [14,17]. Acceptable solutions were only much later identified: around 1959, Indian researcher Sambhu Nath De experimented on ligated ileal loops of adult rabbits. This model helped advance research on cholera and experiments on its underlying molecular mechanisms. De observed that the cell-free filtrate from a *Vibrio cholerae* culture could induce massive fluid loss from the ileal loops. This observation suggested the presence of a toxin, which was successfully isolated and purified in 1964 [16]. As De's model of rabbit ligated ileal loops requires surgery prior to the experimental procedures, researchers nowadays rely more often on infant mice as cholera models. However, as these mice do not develop diarrhea, a typical symptom of cholera, researchers are still looking for and suggesting new cholera models [17].

The obstacle of working with suitable animal models when fulfilling Koch's postulates has been encountered by many scientists since Koch's work on cholera. Further challenges in achieving all postulates arise when researching polymicrobial causes of disease or when diseases are inflicted by viruses that cannot be grown in isolation [14]. Nevertheless, at the time of Koch and his colleagues, these postulates provided clear guidelines for identifying specific bacterial pathogens causing diseases [4,14]. Thereby, our understanding of disease etiology increased enabling the prevention of such diseases by vaccinations and promoting the search for antibiotics to successfully treat sick patients.

1.1.3 The golden age of antibiotics

The usage of antibiotics dates back to ancient times, when antimicrobial products came in the form of herbs, soil or moldy bread and communities lacked the proper knowledge of the underlying mechanisms facilitating recovery from diseases [18,19]. The modern era of antibiotic discoveries launched at the beginning of the 20th century when Paul Ehrlich (1854-1915) [4] experimented with synthetic dyes as a means to stain certain microbes but not others. This inspired his idea of a "magic bullet" i.e., a substance specifically targeting a particular disease-causing pathogen similarly to dyes staining only certain bacteria. Subsequently, Ehrlich and his colleagues systematically synthesized and screened hundreds of different compounds to cure syphilis-infected rabbits. The screened compounds were derivatives of a highly toxic drug and Ehrlich hoped to identify an altered version that displayed toxicity towards bacteria but not the human organism. After several years of testing, compound number 606 could effectively cure the animals and was subsequently marketed under the name Salvarsan. The drug became the standard treatment for syphilis, even though it was not ideal as its injection procedure was rather laborious and patients suffered from several side effects [18,20].

Encouraged by Ehrlich's findings, Gerhard Domagk (1895-1964) [4] tested a dye for its potential antimicrobial activity, which had previously been synthesized by the Bayer company. His experiments proved successful and the compound sulfonamidochrysoidine was marketed as Prontosil from 1935 onward. Prontosil and its derivates, referred to as sulfa drugs, were the first broad-spectrum antibiotics in clinical use and are still on the market today [19]. As the sulfa drugs and Salvarsan were derived from synthetically produced dyes, they were classified as *synthetic* antibiotics. *Natural* antibiotics were first discovered by Alexander Fleming (1881-1955) [4] in 1928. According to the famous story of the Penicillin discovery, Fleming returned from his summer vacation to find a forgotten petri dish with *Staphylococcus aureus* contaminated by the fungus *Penicillium notatum*. Upon closer observation, Fleming noticed a specific pattern suggesting bacterial lysis upon encountering the fungus. This phenomenon had been observed by other scientists as well, but Fleming conducted further experiments and hypothesized that a fungal substance, which he termed penicillin, was responsible for the observed bacterial growth inhibition. His attempts to isolate and purify the substance in sufficient amounts for clinical testing failed and it took another twelve years to find collaborators in support of this project. In 1940, Howard Florey (1898-1968) [4] and Ernst Chain (1906-1979) [4] were able to purify penicillin and successfully demonstrate its antibacterial effects in a mouse model as well as in humans. In light of the Second World War, mass penicillin production was made a priority to cure infections of the Allied forces and eventually became available for the public in 1945 [18,20].

Fleming's experimental methods using agar-medium plates for determining antimicrobial substances that inhibit pathogenic growth displayed an inexpensive tool for the search for new natural antibiotics. Together with Ehrlich's systematic screening approach, these methods triggered a wave of discovering new synthetic as well as natural antibiotics in the 1940s up to the 1970s. Today, not all of these drugs are potent remedies in the treatment of bacterial infections anymore as drug resistance has become a serious issue. Bacteria have been involved in the war of antibiotics longer than us as we are not the first to use antimicrobial products against them. For instance, other bacteria and fungi have produced antibiotics to fight against bacterial organisms which led to the development of antimicrobial resistance (AMR) by targeted bacteria. AMR can be achieved by various mechanisms e.g., by alterations of bacterial targets, neutralization of the antibiotic using degradation enzymes, or by modifications of antibiotics via e.g., acetylation. The underlying genetic changes can be shared with other bacteria through horizontal gene transfer enabling the quick spread of AMR [18,20]. It is therefore estimated that by 2050, antimicrobial-resistant infections will result in the deaths of 10 million people every year [21]. However, while the "development of resistance is inevitable, its spreading is not" [20]. Measures to fight AMR are widely discussed and include cessation of antibiotic use in animals as growth hormones, increased use of vaccines as disease prevention, proper differentiation between viral and bacterial infections, and funding for research on new antibiotics [20].

The era of antibiotic discoveries is characterized by a very "pathogen-dominated view of human-associated microorganisms" [22]. As uncurable, infectious diseases were one of the biggest threats to human life it was of utmost importance to gain the ability to protect and fight against them. Naturally, this fostered a perspective focused on the pathogenic potential of bacteria. Furthermore, important technologies for in-depth studies of bacteria were missing and traditional methods of cultivating microbes limited the investigations to specific bacteria. With the advent of molecular biology and new technologies, our understanding of microbes

shifted towards a more differentiated perspective including commensal and even mutualistic bacteria.

1.1.4 Microbiome studies

Part of Koch's heritage to all microbiologists was the practice of growing microbes in pure culture to study their nature. This method, however, only allows for those microbes to be cultured that grow under the specific conditions applied and commonly misses the ones that need e.g., anaerobic environments or a precise nutrient composition. In 1959, the marine biologist Holger Jannasch (1927-1998) pointed out that many more bacteria from the same aquatic samples could be observed under the microscope compared to the number of bacteria successfully grown on agar plates. He concluded that most marine microbes are not culturable in the laboratory which stirred debates about their aliveness given the fact that they did not grow [4]. With the advent of gene sequencing, this challenge could be addressed as it allowed the identification of bacteria by their genomic information without bacterial cultivation. In the early days, gene sequencing was quite laborious and time-consuming encouraging researchers to develop faster and more efficient ways to sequence DNA. Several sequencing techniques were practiced when, in 1977, Frederick Sanger proposed a more accurate and faster method which is still widely used today. The novelty of his approach was the usage of four terminators, analogous to the four nucleotides, that were incorporated into the new DNA strand and inhibited the DNA polymerase resulting in numerous DNA fragments of different lengths. The four mixtures of DNA fragments corresponding to the incubations with either of the four terminators were analyzed by gel electrophoresis. Comparing the length of the different fragments from the four mixtures gave information on which terminator was incorporated at which location. This method was easier to perform than contemporary methods as it contained fewer incubation steps, was more accurate due to the easy determination of fragments on the gels, and was faster as the four mixtures could be run in parallel [23].

Despite Sanger's efforts, it was not possible to obtain sequences longer than ~ 50 kilobase pairs (kbp) as the longest continuous single reads were only about 200-300 nucleotides, and assembling them into a consecutive DNA sequence was challenging [23,24]. Therefore, as whole bacterial genomes could not be sequenced at the time, a smaller genetic entity was required to distinguish and classify bacterial species. In 1977, Carl Woese and George Fox proposed ribosomal RNA (rRNA) as a phylogenetic marker to identify different domains, phyla, and species. Ribosomal RNA is suitable for this purpose as it is present in all "self-replicating systems", can be isolated from an organism, and changes only slowly over time (but enough to differentiate between species) [25]. Specifically, the 16S rRNA was determined for the analysis of distinct species due to its presence in a great variety of organisms [25]. However, even though its isolation from a bacterial cell was possible, it was tedious, and therefore, the discovery of unknown bacteria progressed only slowly (Fig. 2). This improved in 1985 when a more efficient 16S rRNA sequencing protocol was published relying on synthetic primers binding to a conserved region within the 16S rRNA [26]. Thereby, no isolation and purification of the 16S rRNA was required enabling time-efficient sequencing of 16S rRNA. Even though 16S rRNA sequencing presents a huge milestone in bacterial classification, it is subject to potential errors. For instance, organisms with 97% sequence identity are typically assigned to the same species, which can lead to grouping organisms even though they might be dissimilar in their physiology, biochemistry, and genome content [27]. Alternatively, sequencing an organism's

entire genome can provide more information on genetic and biochemical diversity [27]. This was first achieved in 1995, when – due to progress in computational methods – Craig Venter and his team were able to sequence the first bacterial genome of 1,830 kbp (*Haemophilus influenzae* Rd) [24]. Venter and his team relied on automated DNA sequencing machines using fluorescent dyes to distinguish between each nucleotide and a computational approach to identify repeat regions and assemble the whole genome [24]. In 2003-2004, Venter and his team extended their sequencing efforts to analyzing microbial metagenomes i.e., large numbers of nucleic acids collected from a particular ecosystem [28]. In a nine-month sailing expedition in the Atlantic and Pacific Oceans, they gathered 44 surface (mostly marine) aquatic samples yielding six million base pairs identifying approximately 25,000 different microbial species per litre [4,27]. This was the first systematic whole genome sequencing project which drastically increased the known protein sequences of all living organisms at the time (Fig. 2) [4].
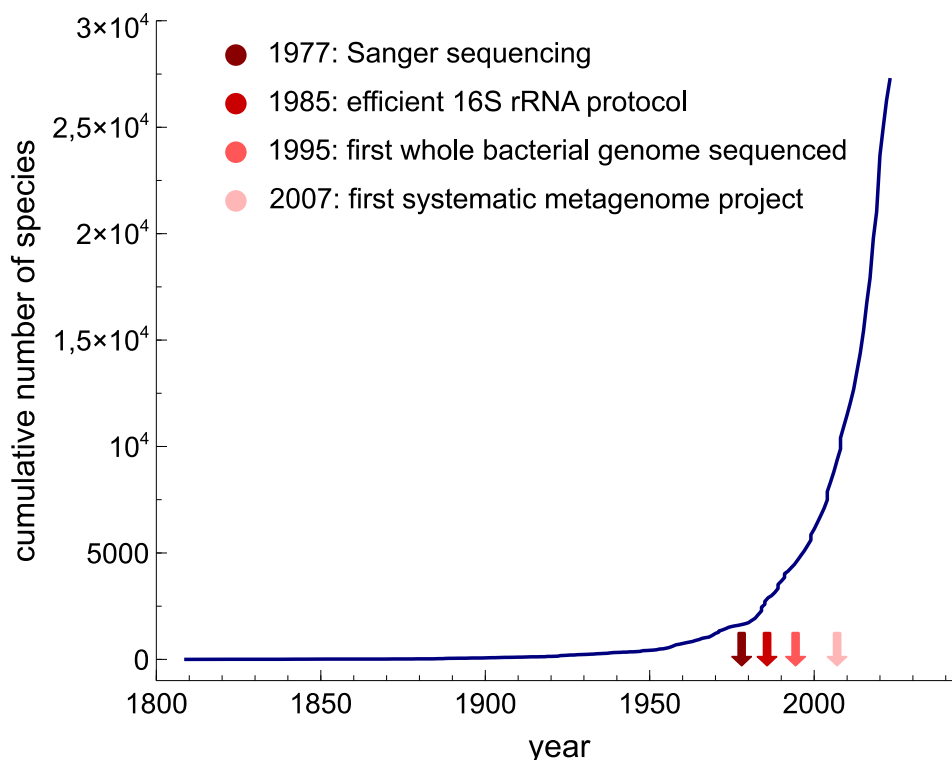
Figure 2 | Number of prokaryotic species in publications from 1800 to 2022. *Data based on the "List of Prokaryotic names with Standing in Nomenclature" (LPSN) [29]. Bacterial species were selected from the LPSN and the number of species was determined according to publication date.*

Besides detecting new bacterial species, our knowledge also expanded concerning bacterial habitats and microbial communities. The latter is termed microbiota i.e., a community consisting of numerous different species which interact with each other within their shared environment [28]. Historically, studies focused on the relationship between pairs of bacterial species to understand dynamics within microbiotas [30]. This reductionistic approach has been criticized by a recent study demonstrating that the coexistence of two bacterial species often requires further community members [30]. The authors co-cultured different combinations of two bacterial species from 13 different microbiotas of "stably coexisting species" showing that most pairs could not coexist when grown in isolation from their community members [30]. This demonstrates that microbial communities require to be viewed more holistically e.g., as the collection of genomes and with a focus on their functions and interactions within their environment [28]. This is reflected in the term microbiome which comprises the organisms of a

community, their metabolites, and structural elements (e.g., proteins, lipids, nucleic acids, etc.) as well as impacts of the environment [28].

The interactions between the members of the microbiome can have positive, negative, or neutral effects on the "microbial fitness, population dynamics, and functional capacities within the microbiome" [28]. For instance, some microbes produce reactive oxygen species or antibiotics to inhibit the growth of their neighbors. Microbial growth inhibition can also be the consequence of bacterial metabolic by-products such as lactate or short-chain fatty acids (SCFAs) [22]. Other bacteria enable each other's growth due to complementary energy metabolisms but can also compete for specific nutrients such as nitrogen and/or binding sites. Furthermore, gut bacteria cooperate in the generation of biofilms which consist of exopolymeric substances (EPS) that are typically attached to a surface and include the EPS-producing bacteria and other microbial species [22,31]. Biofilms are often formed via quorum sensing, a bacterial cell-cell communication via extracellular signaling molecules to achieve collaboration e.g., to adjust gene expression [31]. The advantages of such biofilms lie in their protective characteristics from antibiotics, disinfectants, and other stressors [32]. Some bacteria in these biofilms exhibit so-called social traits e.g., the capacity to produce "public goods", meaning they secrete molecules that other cells in the community benefit from [32]. Such public goods can be binding sites offered to other bacteria or enzymes that fractionate complex molecules facilitating their uptake by the surrounding microbial cells [32].

Furthermore, the microbiome is impacted by factors from its environment such as pH, temperature, available substrates, and other parameters that are specific to the niche. This becomes evident when looking at the results of the Human Microbiome Project (HMP) which started collecting samples from different human body sites in the fall of 2007 [33]. Determining the microbial species in the different samples by sequencing revealed "distinct microbial community compositions" for each body part dependent on the specific ecological conditions [33]. For instance, in the case of the human gut microbiota, the host's nutrition can greatly impact the composition of the gut bacterial community and even modulate microbial gene expression [34]. Naturally, this microbiome is particularly vulnerable to alterations due to antibiotics as it comes in direct contact with the medication after ingestion by the host. Furthermore, the microbiome changes in its composition and density along the intestinal tract owing to changes in e.g., acids, oxygen availability, and antimicrobials [35].

The second half of the 20th century is characterized by a more inclusive perspective on microbes acknowledging their involvement in pathogenic, commensal, and mutualistic symbioses. Additionally, awareness of microbial omnipresence increased, and microbes were understood as part of large communities that are influenced by the interactions among their members and by their habitat. Naturally, this shift in perspective impacted human behavior concerning microbial exposure and hygienic protocols.

1.1.5 Hygienic practices in response to shifting views on microbes

While the germ theory of disease encouraged the avoidance of all microbes as they were viewed as solely pathogenic, the Hygiene Hypothesis proposed by David Strachan in 1989 argued for increased exposure to pathogens during childhood. Strachan suggested that infections during childhood could be protective against allergic diseases such as asthma and food allergies which were on the rise during the 20th century in westernized countries. As infectious diseases during childhood declined in these regions Strachan argued for a

correlation between a decreased exposure to pathogens during childhood and an increase in allergic diseases [36]. However, with more research on the interplay between the immune system and microbes and more findings on body-associated microbiotas, more differentiated concepts were suggested. In 2003, the Old Friends (OF) hypothesis was proposed advocating for the "vital microbial exposures" being OF-microbes instead of pathogens during childhood. OF-microbes are characterized by their presence during early human evolution "when the human immune system was evolving" [37]. They inhabit various indoor and outdoor environments positively impacting human immunoregulatory mechanisms which prevents overreactions of the immune system to harmless agents [37]. Especially during early life, exposure to these OF is essential for increased tolerance of the immune system towards diverse microbes, microbial products, and non-microbial agents [37]. From 2004 to 2008 three studies were published conducted in different European countries demonstrating that infections during childhood were not protective of allergic diseases ultimately providing evidence for the flaws in the Hygiene Hypothesis [37]. The increase in allergies during the 20th century is now considered to be a result of combined changes in lifestyle, nutrition, medical, and public health decreasing the exposure to "potentially beneficial microbial agents" [36]. For instance, through sanitary measurements, the exposure to not only pathogenic but also commensal and mutualistic microbes was reduced with negative effects for immunoregulatory mechanisms [37].

To harness the benefits of exposure to harmless microbes but avoid infectious diseases and antibiotic administration, a targeted hygiene approach was proposed with measures to decrease contact with pathogens and foster relations with non-pathogenic microbes [36]. Actions to support such an approach lie in the promotion of natural childbirth, breastfeeding, exposure to outdoor environments, a more traditional diet with unprocessed foods, and deliberate use of antibiotics [37]. Unfortunately, the idea of the population living in a "too clean" environment has persisted in the media and the minds of the public fostering the neglect of hygienic protocols [36]. Therefore, stressing the importance of hygienic practices, for instance, applied to high-risk surfaces (e.g., technical devices or door handles) and situations (e.g., public transport) is an essential task in implementing the targeted hygiene approach [36].

The full implications of hygienic practices and groundbreaking discoveries, such as antibiotics, on host health are often not immediately discernible. While avoiding all microbes is sensible in light of infectious diseases, this measure also contributed to the rise of allergic illnesses. Similarly, while the usage of antibiotics cures infectious diseases, it promotes the development of AMR increasing antimicrobial-resistant infections. Furthermore, while antibiotics prolonged human lifespan, they thereby elevated the risk for complex diseases, which now predominate as the leading causes of mortality on a global scale [1,20,38].

1.2 Complex diseases

Even though AMR poses a threat, infectious bacterial diseases are no longer as threatening as they were prior to the discovery of antibiotics. This does not implicate the resolution of the hazard of illnesses altogether as new diseases emerged and became prevalent. Today, complex diseases have superseded infectious diseases as the deadliest illnesses worldwide. It is estimated that 70% of all deaths can be attributed to complex diseases such as cardiovascular diseases, cancer, respiratory diseases, and diabetes [1]. Complex diseases are labeled "complex" as they result from a combination of genetic, environmental, and lifestyle factors. The extent to which the different factors contribute to disease susceptibility and how to weigh them in relation to each other is, however, difficult to assess [39].

1.2.1 Risk factors of complex diseases

Besides the demographic changes in the human population, altered environments and economic modifications further promoted complex diseases by elevating risk factors (Table 1). Processes like mechanization and modernization led to changes in lifestyle and diet in Westernized countries over the last centuries and more recently in the Global South. For instance, globalization and modernization of society lead to the production and consumption of convenient foods, thereby increasing amounts of fat and sugar in the diet with decreases in fiber compared to traditional diets. The mechanization of production and life, as well as urbanization encourages a sedentary lifestyle with reduced physical activity. Additionally, leisure activities are increasingly sedentary as well, further negatively impacting body weight and composition. Other lifestyle changes such as the use of tobacco and alcohol, increased stress, depression, and anxiety additionally elevate the risk for complex diseases [40]. Moreover, environmental factors also negatively impact disease susceptibility such as air pollution or elevated UV radiation which increased through mechanization and urbanization [38].

Table 1 | Risk factors of complex diseases. Examples of different factors (e.g., genetics, environmental, etc.) contributing to non-communicable diseases according to Budreviciute et al. [38]. Non-communicable diseases form "the basis for complex diseases" involving intricate interactions between various factors [41]. BP, blood pressure.

genetics	environmental	demographic	lifestyle	medical
• family disease history	• air pollution	• age	• tobacco use	• medication
• genetic inheritance	• weather change	• gender	• alcohol use	• BP
• epigenetic changes	• UV radiation	• race	• physical activity	• lipids
• mutations		• ethnicity	• body weight	• glucose
- environment (radiation)		• education	• nutrition	• viruses
- toxic material-based		• income	• dental health	• obesity
				• stress

Genetic factors also play a role in complex diseases, however, they do not adhere to the standard Mendelian patterns of inheritance [39]. In complex diseases, genetic predisposition and/or epigenetic changes can increase disease susceptibility, however, for an individual to develop the disease further risk factors need to be present [38,39]. Moreover, compared to the Mendelian principle of single-gene diseases, several genetic loci are often contributing to complex disease susceptibility. The variations in the genetic loci between individuals are often single DNA base-pair changes called single nucleotide polymorphisms (SNPs) [39]. Genome-wide association studies (GWAS) investigate associations between individual genetic variations such as SNPs and a particular disease or phenotype in large populations. GWAS

aim at identifying genetic variations that are associated with a specific phenotype to understand the genetic contributions to disease risk [39]. However, several challenges are encountered when trying to interpret GWAS data. For instance, even if a genetic locus does not show statistical significance, it could still be relevant to the phenotype [42]. Moreover, as genetic loci can include up to tens of genes identifying the gene responsible for the phenotypic trait is difficult [42]. Therefore, further experiments are required to link genetic variation to phenotypic traits and to understand the underlying mechanisms. Network biology (Chapter 1.3) provides a valuable tool to integrate GWAS data by linking genetic variation to biological networks [42].

Although at first glance, complex diseases seem unaffected by microbes, accumulating research suggests microbial involvement in this context as well. While causal mechanisms remain mostly elusive, manipulating the gut microbiota has been effective in alleviating the symptoms of some of those diseases [43]. Uncovering the underlying effects and potentially identifying effective treatments requires a thorough understanding of various aspects of the human gut microbiota in health and disease.

1.2.2 The gut microbiome

The human gut is densely populated by a vast number of microorganisms, including bacteria, archaea, viruses, and eukarya, consisting of trillions of cells [35,44]. As bacteria contribute 99.1% of genes to the gut metagenome, researching gut bacteria to understand the gut microbiome and its impact on the human host is at the focus [45]. Therefore, when subsequently referring to the gut microbiota, the focus lies on the community of bacterial residents.

As only an estimated 10-15% of these bacterial residents are culturable, most culture-based techniques are unsuited for capturing the majority of gut bacteria. Despite new, sophisticated culturing methods, which can be worthwhile attempts for studying difficult-to-grow bacteria, only the advances in sequencing technologies allowed identifying a great proportion of the members of the gut community. High-throughput sequencing is the method of choice for a broad overview of gut bacterial species in a given sample and typically targets the 16S rRNA [43]. Alternatively, sequencing of the whole genome (shotgun sequencing) is employed when a higher taxonomic resolution and/or more detail on the potential functional capacities of the microbes is desired [35].

Both kinds of high-throughput sequencing methods have been used in the two major projects providing us with the most comprehensive datasets about the gut microbiota: the European METAgenomics of the Human Intestinal Tract (MetaHIT) and the US HMP. The MetaHIT contributed gut metagenomic data from a cohort of 124 healthy, overweight, and obese European individuals as well as IBD patients [45]. The HMP focused on bacterial communities of 15 (male) or 18 (female) body sites from 242 healthy subjects offering 16S rRNA as well as whole-genome shotgun data. Since the first analyses of the HMP, the consortium has launched additional projects to look into different microbiotas under varying disease conditions [46]. All of those studies contributed greatly to our understanding of the composition of the gut microbiota: in a healthy individual, the bacterial community is comprised of mostly Bacteroidota (formerly Bacteroidetes) and Bacillota (formerly Firmicutes) whereas Pseudomonadota (formerly Proteobacteria), Actinomycetota (formerly Actinobacteria), and Verrucomicrobiota (formerly Verrucomicrobia) constitute smaller portions (Fig. 3) [3].
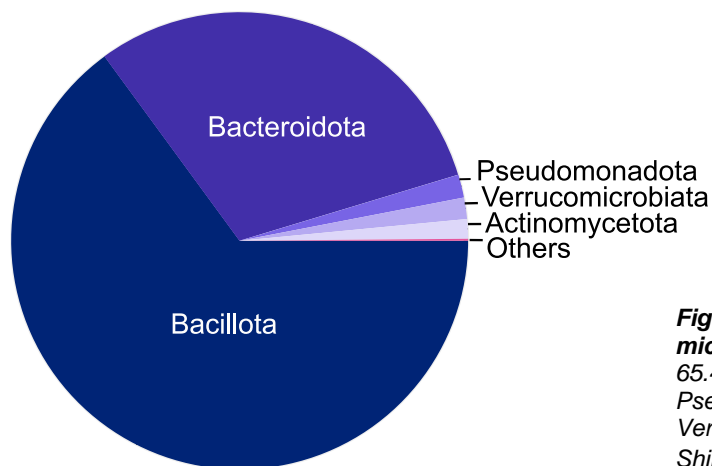
Figure 3 | Composition of the healthy human gut microbiota. *Percentages in relative abundance: 65.4% Bacteroidota, 24.4% Bacillota, 4.5% Pseudomonadota, 2.2% Actinomycetota, 0.7% Verrucomicrobiota, 2.8% others. Adapted from Shin et al.* [3].

Since the MetaHIT and HMP projects, research on the gut microbiota has increased and advanced our understanding of how microbes can interact with their host (Fig. 4b) and which factors influence its composition (Fig. 4a). For instance, whether a newborn is delivered vaginally or via C-section has a great impact on the early gut microbiota composition as well as whether it is breast- or bottle-fed. Of importance are also environmental factors and geography i.e., whether the individual grows up in rural or populated spaces, western or non-westernized societies. Furthermore, diet greatly influences the gut microbiota [35,47]. For instance, when comparing the effects of a plant-based versus an animal-based diet on the gut microbiota of human volunteers, researchers identified differences in microbial gene expression. This observation suggests "regulatory and taxonomic shifts within the microbiome" in response to food intake by the host [34]. Furthermore, foodborne microbes e.g., from fermented foods, temporarily colonize the gut potentially being metabolically active [34]. Besides such dietary influences, other lifestyle choices e.g., regular exercise or living arrangements further impact the gut microbiota [47]. Moreover, gut bacteria composition may change when humans fall ill, age, or are under medication [47]. Especially, in the case of medication and diseases, influences between microbes and host can be bidirectional. For example, bacteria have the potential to manipulate host cells, including immune cells, and thereby impact host health [47]. On the other hand, a variety of host factors that influence host health also impact gut bacteria e.g., host genetics, inflammatory mediators, stress, etc. [35,47]. Regarding medications, drugs can alter gut microbiota composition e.g., by inhibiting or promoting bacterial growth and by altering bacterial metabolism [48]. But bacteria can also impact drugs' activity and toxicity by e.g., modification of the drug (biotransformation) or by intracellular accumulation inside the bacterial cell (bioaccumulation). Subsequently, gut microbiota-dependent biotransformation and bioaccumulation can be responsible for interpersonal differences in drug response. These observations regarding drug-bacteria interactions are not limited to drugs but encompass various xenobiotics e.g., environmental contaminants, food contact material, natural toxins, food additives, etc. [48].

As mentioned before, studies not only provided us with knowledge on factors shaping the gut microbiota but also described some mechanisms used by bacteria to influence their host (Fig. 4b). Even though more understanding is required, research has shown that the gut microbiota can impact human metabolism. This can be achieved via the release of bacterial metabolites which can trigger human enteroendocrine cells in the gut lining to release hormones impacting host processes. Examples are SCFAs, secondary bile acids, and structural components of the bacterial cell wall (flagella, lipopolysaccharides (LPS)), which can bind to different receptors

expressed on enteroendocrine cells. These receptors include free fatty acid receptors, G-protein coupled receptor TGR5, and toll-like receptors (TLRs) respectively. Hormones released from enteroendocrine cells are e.g., glucagon-like peptide 1 (GLP-1), 5-hydroxytryptamine (5-HT), peptide tyrosine-tyrosine (PYY), or cholecystokinin (CCK). GLP-1 augments insulin while inhibiting glucagon secretion and influences satiety and food intake. Besides its neurological functions, 5-HT (serotonin) positively correlates with body mass index and poor glycemic control. PYY is known for its regulation of neurons in the hypothalamic arcuate nucleus, thereby controlling food intake and satiety. CCK-containing cells are mainly found in the small intestine and regulate appetite and gastric motility, amongst others [49].
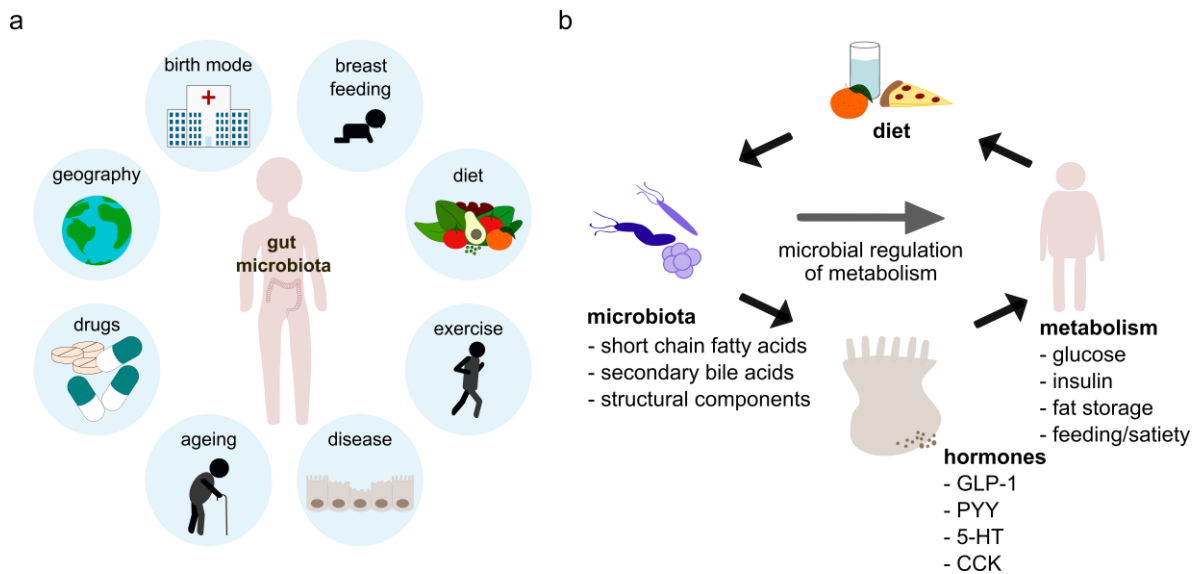
Figure 4 | Interplay between the gut microbiome and the human host. a | *Factors shaping the gut microbiota composition: methods of delivery and infant nutrition, diet and exercise, diseases, aging, medications, and geography i.e., place of living. Adapted from Quigley [47].* **b |** *Gut microbial regulation of host metabolism: microbial metabolites impact enteroendocrine cells, which in turn release hormones impacting human metabolism. Human influences on gut microbiota composition are mediated via the diet. Adapted from Martin et al. [49].*

Impact on host metabolic processes can also take place via the *de novo* synthesis of essential vitamins by the gut microbiota. The host profits especially from the supply of vitamin K and several B vitamins which are produced by different gut bacteria. A healthy gut microbiota provides further benefits for the host including the strengthening of the host epithelial barrier and regulation of host immunity. The former is achieved by e.g., microbial stimulation of mucus production as well as of antimicrobial peptide secretion by the host [50,51]. Regulation of host immunity is important to create a tolerogenic environment towards gut microbes by e.g., inhibiting the activation of NF-κB or by blockage of the transcription of its regulated, pro-inflammatory genes [51]. Furthermore, certain microbes such as *Lactobacillus acidophilus* NCFM can directly interact with dendritic cells promoting the release of anti-inflammatory interleukin-10 (IL-10) [52]. Other benefits for the host offered by gut commensal microbes include protection against pathogens and the harvest of energy. Whether the latter is beneficial or harmful is, of course, context-dependent: in obese people, an increase in energy provision by gut bacteria is rather detrimental, whereas it can be helpful when energy supply is scarce. Protection against pathogens by gut commensals is offered by occupying the physical space in the lumen. Subsequently, pathogens have to compete for attachment sites and nutrient sources as well as protect themselves from antimicrobial substances produced by the gut microbiota [35]. Even protection against viruses is fostered by gut bacteria through enhancement

of the epithelial barrier and by supporting host defenses during viral infection. Viruses in turn have evolved to exploit gut bacteria in order to evade the host's immune system and to facilitate host invasion [53]. Furthermore, it seems likely that viruses compete with bacteria for pathway regulation, as both organisms can target the same human pathways such as NF-κB signaling, even via interactions with the same molecules [54].

The described benefits offered by the gut bacteria are mainly attributed to a healthy microbiota, even though the exact nature of this healthy community is the subject of debate. In the most basic sense, "healthy", of course, refers to "the absence of any overt disease" [55]. More precisely, a healthy microbiome is often characterized as a community of bacterial species that provide a functional core of metabolic pathways [55]. A large diversity of bacterial species in this community ensures substantial functional redundancy of those metabolic pathways which might explain why microbial diversity is typically linked to greater health [55]. Further characteristics of a healthy microbiome are stability over time, resistance to perturbations as well as resilience i.e., the ability to recover from these perturbations and return to homeostasis [55]. A dysbiotic microbiota is therefore simply characterized as disrupted [43], showing a weakened "resistance to colonization" by disadvantageous bacteria [3] and diverging from the normal microbiota or failing "to provide the host with (…) beneficial properties" [35].
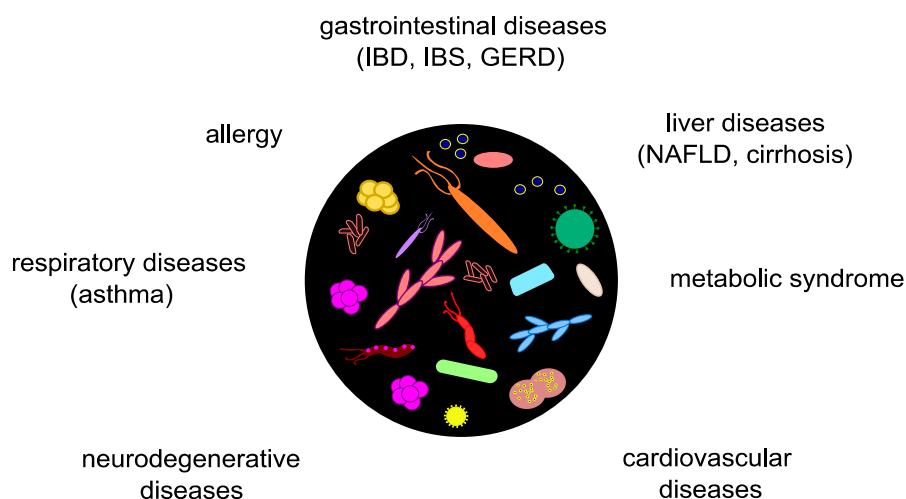
Figure 5 | Associations between a dysbiotic microbiota and diseases. Overview of diseases that have been associated with disturbances in the gut microbiota. Adapted from Catinean et al. [2]. IBD, inflammatory bowel disease; IBS, irritable bowel syndrome; GERD, gastroesophageal reflux disease; NAFLD, nonalcoholic fatty liver disease.

A plethora of studies have shown that a dysbiotic microbiota is associated with several diseases ranging from intestinal ones over metabolic illnesses to even neurodegenerative abnormalities (Fig. 5) [2]. Surely, this raises the question of causality: does the disorder in the gut microbiota result from the disease or is the illness a consequence of changes in the gut microbiota? In the overwhelming majority of cases, this question remains unanswered as establishing causality is aggravated by several challenges. Firstly, discovering the microbial causes for disease etiology is difficult as the gut microbiota forms a complex community, in which several microbes might be disease-causing and potentially solely in combination with certain other co-habitants. Furthermore, the bacterial impact on host health might be context dependent as other factors such as host lifestyle or genetics exert important effects on disease manifestation as well. Designing a suitable experiment when investigating the gut microbiome

presents further challenges including the selection of appropriate gut microbiome representatives and animal models that most closely resemble the human host [56].

Despite the challenge of proving causality, studies could demonstrate a causal link in some instances. For example, it is well established that a disturbed microbiota exerts a causative role in *Clostridium difficile* infection. Hosts are susceptible to this infection after major disruptions in the gut bacterial community due to e.g., antibiotic therapy. Thereby, the dysbiotic microbiota no longer offers resistance to colonization by disadvantageous bacteria such as *Clostridium difficile*. Generally, patients are subsequently treated with a broad-spectrum antibiotic aiming at *Clostridium difficile*, but simultaneously further depleting the healthy microbiota. Relapse followed by morbidity and mortality is quite common and chronic *Clostridium difficile* infection additionally reduces bacterial species diversity resulting in a vicious cycle. Studies demonstrated that fecal microbiota transplantation (FMT) from a healthy donor to an infected patient is an effective therapy with a success rate of 90%. FMT allows the re-establishment of a healthy microbiota through the administration of a complete and stable bacterial community which protects against recurring *Clostridium difficile* infection. So far, FMT is only practiced in a few centers around the world and, as only several hundred patients have received FMT, larger studies are needed to develop reliable protocols and standardized procedures [57].

Suitable therapies, such as FMT, are of course best developed when the etiology and progression of a disease are well understood. Therefore, a lot of research has been conducted over the past years in the pursuit of unraveling the associations between several complex illnesses and their dysbiotic microbiota. During these investigations, studies detected that gut Pseudomonadota have been "a common factor" [58], especially in illnesses characterized by inflammation [58].

1.2.3 Associations between Pseudomonadota and complex diseases

As mentioned before, Pseudomonadota were formerly known as Proteobacteria named after the Greek god Proteus who was allegedly able to assume different shapes similar to the heterogeneity of Proteobacteria [58]. Due to recent sequencing analyses and to create more consistency in phyla names, "Pseudomonadota" replaced "Proteobacteria" based on one of its genera according to the new guidelines [59]. Next to the eponymous genus *Pseudomonas* [59], other well-known Pseudomonadota genera include *Klebsiella*, *Escherichia,* and *Enterobacter* (all three belonging to the *Enterobacteriaceae* family) as well as *Bilophila* and *Proteus* [3].

Pseudomonadota are the third-most abundant phylum in the gut microbiome and the most unstable one out of the four prominent gut phyla over time and under different conditions. For instance, Pseudomonadota incrase in response to a diet high in calories, fat, artificial sweeteners, and emulsifiers as well as a diet low in fiber [3]. Furthermore, they show increased abundance in patients with e.g., metabolic disorders, inflammation, and cancer or gastric bypass (Fig. 6) [3]. For instance, a study demonstrated that, in a healthy host, *Enterobacteriaceae* are present in low numbers and without any obvious adverse health outcomes. However, during inflammation, the bacteria utilize host-derived nitrate, which is part of the inflammatory response, and colonize inflamed gut sections. This, of course, generates a vicious cycle of furthering inflammation within the host. Yet, not only host inflammation or diet can foster the propagation of certain Pseudomonadota but also other factors such as host genetics. For instance, IL-10-deficient mice exhibited increases in Pseudomonadota and

spontaneous colitis suggesting that a dysregulated host immune response can enable the growth of Pseudomonadota [3].
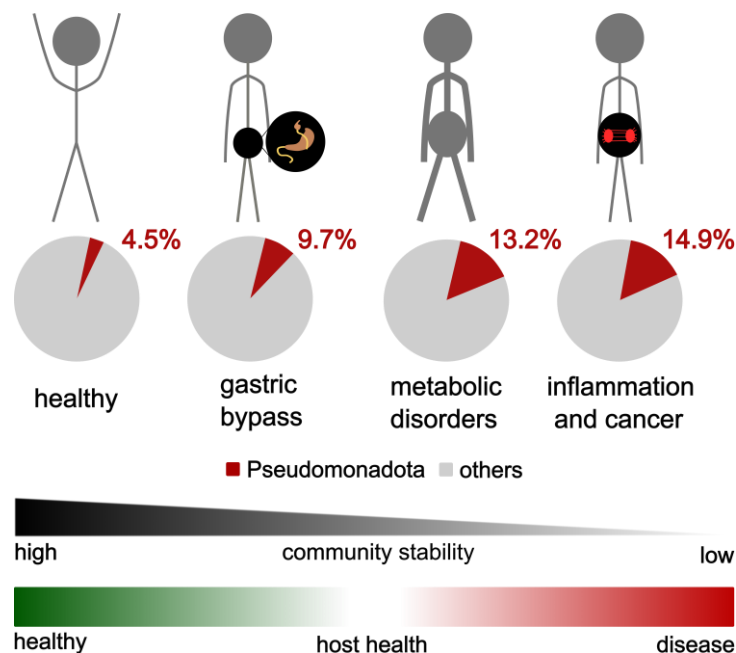
Figure 6 | *Pseudomonadota abundance under different health conditions.* *In a healthy gut, Pseudomonadota represent only a small portion of the gut microbiota. A bloom in Pseudomonadota has been observed in patients with e.g., gastric bypass, metabolic disorders, inflammation, and cancer. Adapted from Shin et al. [3].*

As Pseudomonadota are associated with adverse health conditions, they can be seen as a marker of gut dysbiosis and microbiota instability [58]. Besides inflammatory diseases, evidence is emerging of increased Pseudomonadota numbers in patients with neurodegenerative disease as well. For instance, Pseudomonadota increased in patients with Alzheimer's disease and major depressive disorder compared to healthy controls. *Gammaproteobacteria* and specifically *Enterobacteriaceae* were gradually elevated in samples from healthy controls over patients with amnestic mild cognitive impairment, a predementia stage, to patients with Alzheimer's disease. In patients suffering from Alzheimer's, *Enterobacteriaceae* correlated significantly with disease severity and a decline in cognitive functioning [60]. A different study showed that *Escherichia* and *Shigella* genera were enriched in microbiotas of cognitively impaired Alzheimer's disease patients with brain amyloidosis compared to patients with no brain amyloidosis or controls [61]. The study showed that the abundance of the two genera was positively correlated with pro-inflammatory cytokines in the bloodstream suggesting a link between gut Pseudomonadota abundance and neurodegenerative diseases via the immune system. Potentially, members of the Pseudomonadota phylum can increase peripheral inflammation promoting brain amyloidosis, neurodegeneration, and cognitive symptoms [61].

Besides inflammatory mechanisms, researchers hypothesize that Pseudomonadota may impact neurodegenerative diseases via the gut-brain axis. This axis bidirectionally connects the brain with the gut through neuronal pathways like the vagus nerve but also through messaging systems employing small molecules released from e.g., enteroendocrine cells. The axis facilitates bidirectional gut-brain communication concerning digestive functions and satiety as well as normal brain function and behavior. Studies now suggest a microbiota-gut-brain axis highlighting the fact that gut bacteria might be able to hijack the gut-brain axis to impact the host's brain and behavior [62]. For example, microbial metabolites can interact with

enteric neurons that regulate local gut homeostasis regarding motility, blood flow, and secretion. Interactions are also possible with vagal nerve terminals which transmit signals to the central nervous system. Moreover, bacterially produced neuroactive metabolites can reach the brain via the systemic circulation. These neuroactive metabolites are e.g., modified tryptophan from the diet such as 5-HT or molecules fully synthesized by the bacteria e.g., indoles. As seen in the previous chapter, SCFAs can also bind to receptors on enteroendocrine cells which in turn release hormones to act on afferent neurons inducing e.g., satiety [63]. Furthermore, SCFAs have been observed to cross the blood-brain barrier potentially binding to receptors that regulate reward, appetite, mood, memory, and stress responses [64,65].

Even though Pseudomonadota are mainly associated with gut dysbiosis the phylum can also offer some benefits to the human host. For instance, due to their facultative anaerobic nature, Pseudomonadota are able to colonize the oxygen-rich neonatal gut preparing the habitat for strict anaerobes by oxygen consumption, pH-alterations, and the production of nutrients and carbon dioxide [3]. Furthermore, the Pseudomonadota strain *Escherichia coli* Nissle 1917 is especially protective against pathogens by e.g., limiting the iron availability to pathogens and secreting antimicrobials. It is therefore often administered as part of probiotic treatments. Another Pseudomonadota strain, *Escherichia coli* HS has been shown to compete with pathogens for carbon sources thereby inhibiting pathogen colonization of the host [66].

Hence, gut Pseudomonadota as a phylum are linked to gut dysbiosis, however, as Pseudomonadota comprise a huge variety of different bacterial strains, some of them also exhibit beneficial properties for the human host. Several mechanisms of gut bacterial impact on host health are known such as the release of bacterial metabolites or food-related metabolites which exert their impacts not only locally but also systemically via the circulation. Furthermore, gut bacteria can directly act on neurons signaling to the central nervous system or increase peripheral inflammation negatively impacting brain health. Different from other microbial phyla is the presence of T3SSs in Pseudomonadota strains to deliver effector proteins into the host cell. So far, T3SSs have mainly been studied in human pathogens, however, the secretion system has also been detected in 5-20% of gut commensal Pseudomonadota, yet, their impact remains unresearched [7].

1.2.4 The type three secretion system

Secretion systems are widespread tools employed by bacteria for various purposes [67]. At least six (type I-VI) different secretion systems are known from gram-negative bacteria and one (type VII) from a gram-positive bacterium. These secretion systems are used to release proteins into the environment or directly into a host cell. Differences between the secretion systems can be found in structure (depending on e.g., the number of membranes over which the cargo needs to be transported), regulation (involving e.g., secretion system assembly, protein translocation, and feedback regulation), and substrate specificity (each secretion system transports specific proteins). As T3SSs can transport effectors across the inner and outer bacterial membrane as well as across an additional third membrane, this secretion system is often employed on eukaryotic cells by gram-negative bacteria. T3SSs are well-researched in human pathogens but are also known from plant mutualists and commensals [5,68]. Examples of human-pathogenic bacteria using a T3SS are mainly found in the Pseudomonadota phylum e.g., *Yersinia, Salmonella, Shigella*, EHEC, and EPEC. Also, plant pathogens employ the T3SS such as *Xanthomonas* spp. and *Pseudomonas syringae* [5]. In plants, T3SS effectors are also important in mutualistic relationships such as in the rhizobia-

legume symbiosis [6]. Furthermore, genes of the T3SS were also detected in bacteria for which no pathogenic or beneficial activity is known [5].
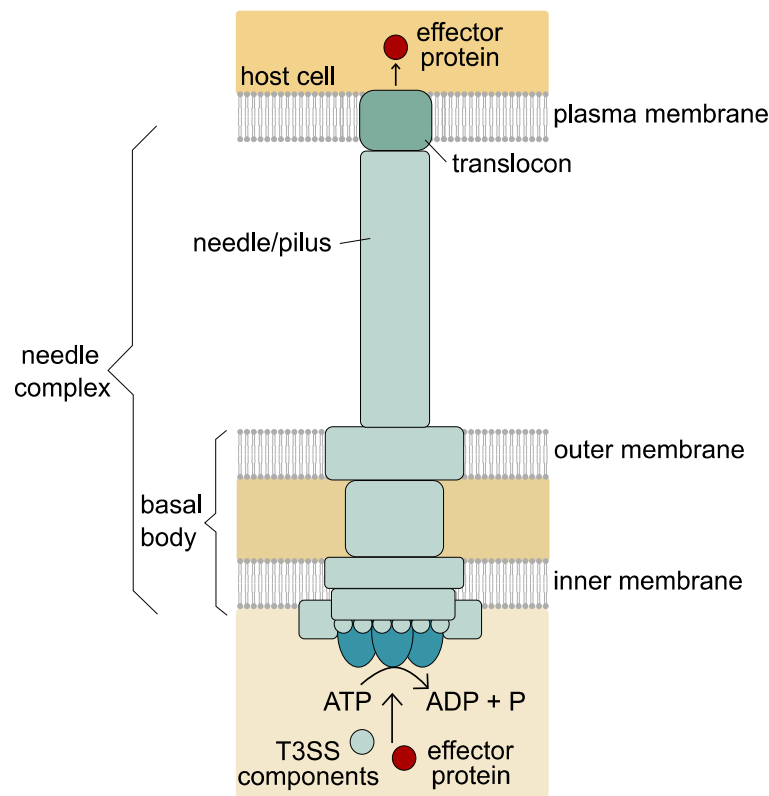
Figure 7 | The type three secretion system. *The basal body is incorporated into the bacterial membranes and attached to a needle which transports effector proteins into the host cell via the translocon. The docking and unfolding of the extracellular components of the T3SS and the effector proteins are energy-dependent, which is provided by an ATPase (dark turquoise). Adapted from Buttner et al.* [5]*. ATP, adenosine triphosphate; ADP, adenosine diphosphate; P, phosphate.*

Evolutionarily related to flagellin, the T3SS is a highly conserved protein-secretion apparatus consisting of more than 20 components forming a "needle and syringe"-like instrument which earned it the name "injectisome" (Fig. 7) [68]. Essentially, the T3SS consists of a basal body, the needle complex, and the translocon, which provides enough space for an unfolded protein to translocate from one cell to the other [68]. The assembly of those components occurs sequentially, however, we lack a detailed understanding of its hierarchical nature. The corresponding genes of the components are either located on pathogenicity islands, in the bacterial genome, or on plasmids. T3SSs are typically shared with other bacteria via horizontal gene transfer as is demonstrated by *Shigella* and *Escherichia coli* sharing highly homologous genomes but more similarities with other bacteria in their T3SSs (e.g. *Shigella* with *Salmonella*) [68]. This variety in T3SS components is reflected in different T3SS families which probably arose due to adaptation to different host organisms or extracellular environments [5]. T3SS families are e.g., the Inv/Mxi-Spa T3SS family or the Hrc-Hrp T3SS family based on sequence identity among structural components of the basal body and named according to the genetic loci encoding the T3SS [69,70]. Bacteria differ not only in T3SS components but also in function and amount of their effectors e.g., EHEC encodes for around 40 effectors whereas *Pseudomonas syringae* secretes 190 effectors [67]. Secretion of the effector proteins is also subject to a certain hierarchy just as the assembly of the T3SS components. This hierarchical secretion can be triggered e.g., by a certain extracellular pH as in *Salmonella* spp. In other

cases, contact with the host cell signals the expression of effectors as has been observed in *Yersinia* spp. and *Shigella flexneri* [5].

After injection of the effectors into the host cell they engage in different assignments supporting bacterial adherence, replication, and/or dissemination [67]. In the human host, studies detected that effectors often impact the host cytoskeleton for the bacterial cell to gain access to the human cell and enable movement within it. For instance, by altering the cellular regulators, actin filament polymerization is impacted, or via direct interaction, microtubules are destabilized. These and other manipulations of the host cell are often mediated by bacterial interference with host phosphorylation cascades by e.g., binding of host kinases or imitating host phosphatases [67]. Mimicry of host proteins is a common theme of effectors, and other examples can be found in the imitation of GTPase-activating proteins or E3 ubiquitin ligases. The former allows disrupting host signaling cascades whereas the latter can change the fate and activity of host proteins. Modification of host proteins is also accomplished by e.g., acetylation of members of the mitogen-activated protein kinases (MAPK) pathway resulting in the inability of the host to activate these proteins. Similarly, effectors can add adenosine monophosphate (AMP) to Rho, Rac, or Cdc42 (AMPylation) inhibiting their ability to "communicate" with their host interaction partners [71]. These manipulations of host proteins are often incompletely understood especially regarding their ultimate function in the host cell.

In the mutualistic rhizobia-legume symbiosis, T3SS effectors are involved in different steps of the nodule symbiosis depending on the bacterial strain and the host. Typically, bacteria do not rely on T3SSs to initiate the rhizobia-legume symbiosis but on nodulation factors (NFs): NFs are released in response to flavonoids secreted by the plant into the soil and stimulate early steps in nodule formation and bacterial colonization. T3SS effectors can initiate nodulation in the absence of NF by activating the required signaling pathways in plants [6]. Notably, while an effector protein can induce nodulation in certain legumes, it can be harmful to the same process in a different legume species [72]. While some strains employ effectors to form root nodules on legumes, others use effectors to suppress proteins of the MAPK pathway to downregulate host immune responses against the bacteria [6]. The latter is achieved via e.g., inhibiting the expression of defense genes or by repressing the generation of reactive oxygen species [72]. Furthermore, effectors can impact the host's ubiquitin system possibly marking plant defense proteins for proteasomal degradation [72]. Suppressing the host's immune system is not always required as some plant hosts have evolved to tolerate certain rhizobia strains resulting in particular host-strain combinations that are compatible. In other cases, suppression of the host immune system is ineffectual as some legumes developed an immune response upon the recognition of bacterial effectors to inhibit infection by the strain.

As T3SS effectors in humans have only been investigated in pathogens, nothing is known about potential human host tolerance towards effector proteins. Mainly, mechanisms of the bacteria to evade the host immune system have been studied. For example, evasion of the immune system can be achieved by inhibiting NF-κB and thereby preventing the expression of pro-inflammatory cytokines, anti-apoptotic factors, and defensins. NF-κB-signaling is an important part of the host's innate immune system which is activated after bacteria are recognized by intra- or extracellular receptors. Therefore, manipulating this pathway at various points seems to be an effective and therefore popular strategy of bacteria to alter the host's defenses [67]. In plants, T3SS effectors suppressing the host immune system have not only been detected in rhizobia strains but also in non-pathogenic and beneficial *Pseudomonas*

strains in the rhizosphere [72]. Similarly to rhizobia strains, *Pseudomonas* strains do so by inhibiting the generation of oxidative reagents or the expression of defense genes [72].

Plant-beneficial *Pseudomonas* strains use T3SSs to e.g., promote mycorrhizal symbioses i.e., mutualistic relationships between a plant and a root-colonizing fungus. Findings show that T3SS-positive *Pseudomonas* were enriched in the mycorrhizosphere of specific fungi, and knock-out of T3SS genes left the bacterial strains unable to support fungal colonization [72]. Whether this support is mediated via interactions with plant proteins or fungi proteins requires further research. Other plant-beneficial *Pseudomonas* strains are involved in the biocontrol activities of their plant host via T3SSs. This has been observed in cucumber plants in which biocontrol activities against an oomycete were reduced when T3SS genes of root-colonizing *Pseudomonas* strains were mutated [72]. Notably, expression of the bacterial T3SS genes remained unchanged when in contact with the plant but increased upon encountering the oomycete with the ability to inhibit the production of virulence factors of the pathogen [72]. Even though diverse research on T3SS effectors in plant mutualists is missing, the secretion system seems to support multifaceted functions in plant mutualistic symbioses.

As pointed out by several examples from plant mutualists and human pathogens, effector proteins impact the host in myriad ways with some mechanisms being shared between bacteria whereas others seem to be unique to a specific bacterium or group of bacteria. To this complexity, the spatial and temporal regulation of the effector library is added. The latter can be ensured by e.g., a hierarchical interaction of effectors with the T3SS-compounds prior to their secretion. Alternatively, effectors with different half-lives can be employed allowing their simultaneous secretion while ensuring their often antagonistic effects. For instance, while an effector with a shorter half-life can disturb the host cytoskeleton to enable bacterial uptake, an effector with a longer half-life can return the actin filaments to homeostasis after successful bacterial invasion [71]. Galàn *et al.* suggest that bacteria might co-secrete proteins that support these temporal regulations of effectors e.g., proteins imitating host E3 ubiquitin ligases marking effector proteins for degradation [71]. Spatial regulations are managed e.g., by mimicking host domains to target specific cellular proteins, for instance, by a nuclear localization signal as a "ticket" for the nuclear import machinery, or by the acquisition of modifications through the host adding localization information such as ubiquitination [71].

Bacterial effectors exhibit further mechanisms of host manipulation than mentioned so far e.g., effectors from human pathogens can be involved in the inhibition of autophagy, cell cycle arrest, Golgi fragmentation, obstruction of inflammatory cell death, impediment of immune cell migration, fostering of immune cell apoptosis, etc. [73]. While many gaps in the knowledge of T3SSs in pathogenesis remain, comprehension of T3SSs employed in mutualism or commensalism exhibits greater holes. Sufficient examples of T3SS-positive commensals and plant mutualists employing T3SS effectors exist to support research on T3SSs used in microbe-host interactions other than pathogenic. This requires consideration of the complexity of T3SSs and their effectors (e.g., the timing of effector secretion, spatial and temporal regulation of effectors, effector targets and functions) in combination with the host cellular intricacy that the effectors encounter.

1.3 A systems biology approach to understanding complexity

Biological systems ranging from the human cell to the entire organism are complex i.e., multiple components are involved that are often interdependent, function simultaneously, and sometimes work together [74]. Changes in these biological systems can result in various phenotypic outcomes including illnesses such as complex diseases. Understanding the underlying molecular processes of such complexities requires a holistic approach.

1.3.1 Complex biological systems

The French philosopher René Descartes (1596-1650) greatly impacted various aspects of our Western society, for instance, through his concept of reductionism. This approach suggests that complex systems consist of several parts and by understanding each part one can grasp the whole system [74]. This view has greatly influenced science and how researchers conduct their work, namely typically by focusing on single entities of complex systems [74]. In most cases, and in biological systems especially, this approach does not consider all the different components involved which function together, at the same time, with various interdependencies. The human cell, for instance, consists of thousands of different genes and their proteins that carry out different functions in parallel and by working together [74]. Adding the regulatory and feedback mechanisms that exist in the cell, even small disturbances can lead to several, unpredictable consequences [75].

In 1941, George Beadle and Edward Tatum drew attention to the "complex ways" in which components of the organism are interdependent [76]. This is especially remarkable as the two scientists provided great support for the reductionistic "one gene – one enzyme" hypothesis i.e., that one gene is responsible for one step in a metabolic pathway [76,77]. Beadle and Tatum's work revolutionized genetics in those days as it connected biochemical reactions in the cell with genetics for the first time. Naturally, through more research, we now understand that the "one gene – one enzyme" hypothesis does not capture the complexity that is found in the relationship between genetics and cell biology. Since 1941, many more concepts of gene expression, variation, and regulation have been detected. For instance, in the mid-1950s and early 1960s, Seymour Benzer drew attention to gene recombination and gene mutation indicating that genes can be subject to change. In the mid-1960s concepts of gene regulation received great attention including e.g., repressor genes, promoters (binding sites of the RNA polymerase), and operators (binding sites of repressor proteins) [77]. Besides gene regulatory mechanisms, gene expression can also be impacted by epigenetic modification e.g., methylation of nucleotides and histone alterations [77]. Other concepts that add to the complexity between genetics and cell biology are gene duplication and gene pleiotropy. Gene duplication results in more than one copy of a gene in the genome which can be functionally redundant i.e., they perform identically with respect to their function [78]. Alternatively, one of the duplicates can acquire a function different from the original gene, or the copies can each perform one part of the original function [78]. Gene pleiotropy describes the concept that one gene affects several, unrelated phenotypic traits [79]. This is evident in some human diseases such as phenylketonuria where the defect in only one gene leads to multiple phenotypic outcomes such as mental retardation, eczema, and pigment defects [79].

To understand the consequences of these genetic variations, and dynamics in gene expression it is essential to observe the behavior of the gene products within the human cell. As gene products interact with other cellular components, altered gene products not only

exhibit altered functions themselves but also impact the functions of their interaction partners [42,80]. The interactions of the gene products with each other or with other cellular components can be mapped as a network. This aids in comprehending how genetic variations can "spread along the links" formed between interacting molecules inside the cell also affecting the behavior of components that are not directly interacting with the altered protein [80]. Such a network approach abandons the reductionistic view and has proven helpful in assuming a more holistic perspective.

1.3.2 Networks as integrators

As genes and their products "function not in isolation but as components of complex networks" [81], dysfunctions in these networks assist in explaining genotype-phenotype relationships. The components (nodes) in such network models can represent e.g., macromolecules (DNA, RNA, and proteins) or metabolites. The connections between nodes (edges) typically depict the biochemical or physical interactions linking these nodes together. As the edges in the network are dependent on functional interaction interfaces, mutations in nodes can lead to the deletion of edges (Fig. 8). Depending on the type of mutation, different edges might be affected resulting in different phenotypes. For instance, mutations leading to the stop of transcription or greatly changed protein structures can result in the loss of all interactions with this node (node removal). Other mutations can perturb specific binding sites or truncate a protein which might impact not all interactions of the node (edgetic perturbation) and can even lead to the acquisition of new binding partners. The resulting disturbances in the network following the mutation of a node or several nodes can help explain phenotypic outcomes [81].
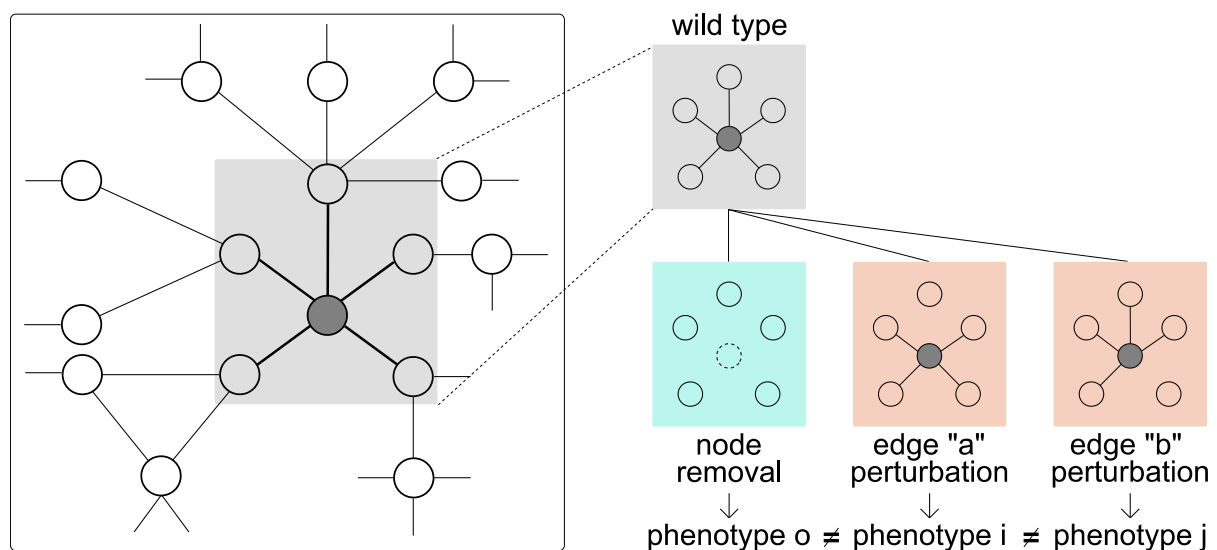
Figure 8 | Different network perturbations and resulting phenotypes. *Nodes represent proteins and edges the interactions between them. The blue box illustrates a node removal i.e., all interactions of a protein are lost due to greatly altered protein interfaces or loss of them. The red boxes indicate edge removals i.e., specific interactions between proteins are lost due to different node mutations affecting different interfaces. Adapted from Zhong et al.* [81].

When trying to understand phenotypic outcomes it can be valuable to consider the interdependencies between different molecular layers inside a cell [82]. For instance, the activity of a gene is often regulated by other genes via PPIs or protein-DNA binding thereby linking two different molecular layers. If several layers are analyzed together, multilayer biological molecular networks can be constructed. These contain e.g., a gene regulatory network (genes

are linked by regulatory relations), a PPI network (proteins are linked by physical interactions), and a metabolic network (metabolites are linked by chemical-chemical interactions) [82]. To analyze the functionality of such a multilayer biological network the interdependencies between the different layers are assessed. For instance, the perturbation of one gene affects all genes that are under its regulatory control leading to their dysfunction as well. This impacts the proteins of all perturbed genes which will stop functioning properly also affecting the function of other proteins in the network that were previously interacting with the now dysfunctional proteins [82]. Additionally, as the metabolites depend on the proteins in the PPI network e.g., as enzymes, dysfunctional proteins can negatively affect the activity of these molecules as well [82].

Besides connecting different molecular layers by mapping e.g., protein-DNA interactions, networks can also join layers of different cells, tissues, and organ systems [80]. Yet, due to little systematic data, most studies focus on the intracellular networks [80]. These can already offer great value when trying to understand phenotypic manifestations, as cellular network perturbations can have systemic effects that contribute to certain phenotypic outcomes such as complex diseases. Disturbances of the cellular networks can arise through e.g., environmental by-products that affect PPIs, microbial interference with host molecules, or altered gene products due to genetic variations [39,83,84]. For instance, to understand network disturbances in ataxia, Lim *et al.* mapped a PPI network to analyze the functions of involved disease genes [83]. They detected that disease genes of different types of inherited ataxias are linked by direct and indirect (via a third protein) interactions between their proteins suggesting that these illnesses share certain disease processes. The authors identified genetic modifiers i.e., gene products that impact gene expression [85], as direct interactors of several disease-causing proteins suggesting that alterations in genetic modifiers can impact disease risk. Given the strong connectivity between ataxia genes and that several genetic modifiers physically interacted with ataxia-causing proteins the network offers the possibility of identifying causative genes for those ataxias of which disease genes are unknown [83].
To assess the nature of the proteins that are associated with ataxias Lim *et al.* annotated the proteins in the network with information from the Gene Ontology (GO) database. This database describes the molecular functions of a protein, the biological process it is involved in, and its cellular location. Lim *et al.* detected that interacting proteins are localized in the same cellular compartments and that specific biological functions are enriched in the network. The latter allows the formation of hypotheses concerning the mechanisms underlying disease manifestation [83].

A year after Lim *et al.*'s work, Goh *et al.* published a study analyzing relations between 1,284 disorders and 1,777 disease genes [86]. What Lim *et al.* had demonstrated for a particular disorder Goh *et al.* confirmed on a much bigger scale: that proteins expressed from genes associated with the same disease are much more likely (ten times in Goh *et al.*'s study) to interact with each other compared to random expectation [86]. This means that components linked with a particular disease phenotype often cluster in the same network neighborhood forming subnetworks i.e., "groups of nodes that link to each other" [80]. Subnetworks that are associated with a particular disease phenotype via e.g., biochemical processes contributing to altered cellular functions, are referred to as disease modules [80]. Identifying and analyzing disease modules can help determine altered pathway regulations during disease, disease genes, or markers for prognostic procedures [87]. Furthermore, drug targets can be identified to manipulate protein functions to reach a cellular network state closer to homeostasis [87].

Cellular networks cannot only be perturbed from the inside i.e., via genetics, but also from the outside, for instance, by viral proteins [39,84]. Rozenblatt-Rosen *et al.* demonstrated that cancer viruses such as the Human Papillomavirus, Epstein-Barr Virus, Adenovirus, and Polyomavirus are capable of rewiring the host's cellular PPI network. They observed that viral proteins interacted with several human proteins altering their degree i.e., the number of edges that are linked to the protein. Analysis of the viral targets displayed their involvement in pathways typically perturbed during cancer and showed an overrepresentation of tumor suppressors among them. Furthermore, the viruses targeted candidate cancer genes suggesting that the network can promote the identification of cancer-associated genes [84]. The fact that perturbation of cellular networks by exogenous factors can give rise to disease phenotypes raises the question of whether bacterial T3SS effectors might have similar capacities.

The examples by Lim *et al.* and Rozenblatt-Rosen *et al.* reflect the importance of proteins in disease manifestation. Through interactions with each other, proteins mediate host signaling and metabolic pathways as well as cellular processes up to organismal systems [87]. Proteins are the "main agents of biological function" and "determine the phenotype of all organisms" [87]. Ultimately, proteins convey the mutations in disease-causing genes through e.g., alterations in their interfaces. Given their importance in biological functions, protein interactions "control the mechanisms leading to healthy and diseased states in organisms" [87]. Therefore, when trying to understand the molecular basis of diseases and comparing differences between healthy and diseased states it is important to consider changes in the cellular PPIs [87].

1.3.3 Protein-protein interaction maps

In order to map any type of network, three different approaches are typically used to obtain the necessary data: 1) assembly of existing data from the literature, 2) computational predictions, and 3) experimental data collection. Data compiled from the literature typically stem from heterogeneous experimental procedures and assays [88]. This aggravates the collation into one network due to variable quality, lack of systematization, and no information on negative data [88,89]. Computational predictions use e.g., conserved sequences or sequence similarities to predict network data and while this method is fast and can provide huge datasets, it is unknown how well the data reflect biological systems [88]. Finally, experimental data can be prepared utilizing different laboratory assays, which, in the case of PPI networks, are often either of two high-throughput methods: affinity purification followed by mass spectrometry (AP-MS) or the Y2H (Fig. 9). AP-MS yields direct and indirect PPIs: proteins interacting with a bait protein (direct) are detected as well as all proteins that are associated with the prey's protein complex (indirect). The Y2H on the other hand catches only direct PPIs. Therefore, datasets from either of these methods can vary greatly and provide different perspectives on the interactome i.e., the entirety of all interactions [42,88].

The Y2H system was first described by Fields and Song in 1989 and has since been modified and improved rendering it "one of the most reliable protein interaction detection methods" [42] when including all necessary controls [42,89]. As the Y2H is relatively fast, inexpensive, and scalable, it has become the most widely used large-scale experiment to map PPI networks [42,89]. Direct PPI interactions from Y2H experiments allow the deduction of hypotheses on molecular mechanisms more readily compared to indirect PPIs [90]. Another advantage of this *in vivo* system is its suitability for the expression of proteins from varying (model) organisms like *Caenorhabditis elegans* and *Drosophila melanogaster*, as well as from humans and plants [88].
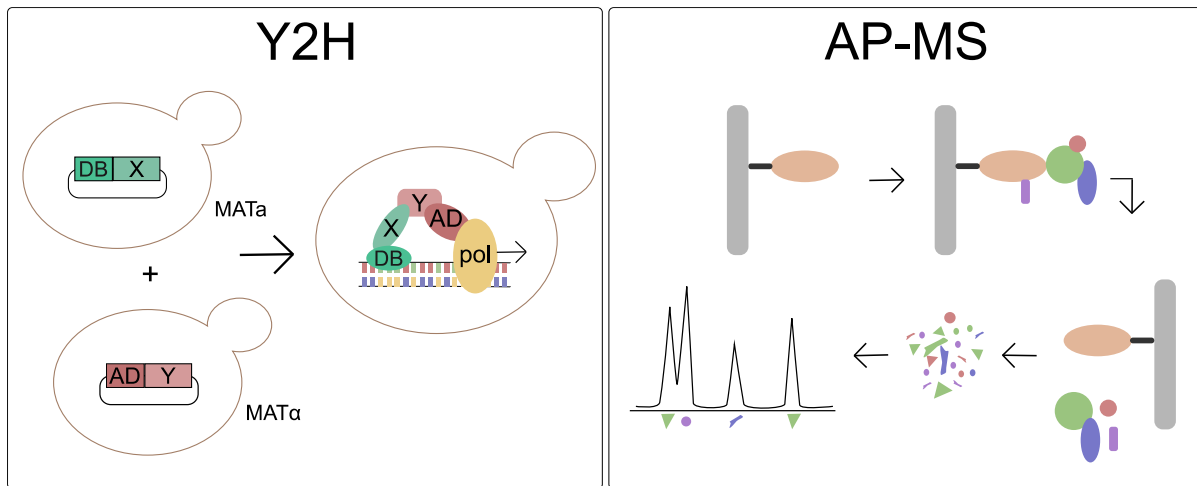
Figure 9 | Schematic representations of the mechanisms of the Y2H and AP-MS. Y2H | *The bait protein is fused to the DNA-binding domain (DB), whereas the prey protein is bound to the activation domain (AD) of a transcription factor. Both constructs are transformed into different yeast cells. After mating, the bait and prey protein are present in the same yeast cell, and upon interaction, the transcription factor is reconstituted, and the reporter gene is transcribed by the polymerase (pol).* ***AP-MS |*** *The bait protein is bound to a matrix, to which a protein mixture is added. By interacting with the bait protein, prey proteins are captured. The prey proteins can be analyzed using mass spectrometry. Adapted from Koh et al.* [89].

Luck *et al.* employed different high-throughput versions of the Y2H to map a very comprehensive human PPI network [90]. The authors screened 17,408 protein-coding genes "all-by-all" representing ~ 94% of all expressed human genes (based on three individual transcriptome sequencing studies). The resulting protein interactome called the "reference map of the human binary protein interactome" (HuRI) published in 2020 consists of approximately 53,000 PPIs constituting four times more interactions than PPIs available from small-scale studies at the time [90]. Around 15-20 years ago, PPIs from small-scale studies were considered of higher quality compared to PPIs from large-scale studies. However, several publications have demonstrated that by benchmarking and validating the utilized assays, large-scale experiments are just as reliable [42]. Small-scale approaches are suitable for testing well-formulated hypotheses, whereas large-scale experiments are conducted when knowledge in a given field is scarce. The latter allows a more unbiased approach to gathering information on an unknown topic. More precise hypotheses can be drawn in a subsequent step and analyzed in smaller-scale experiments [42].

While Luck *et al.* used the Y2H to provide a reference interactome map as a general resource, Mukhtar *et al.* and Weßling *et al.* employed the Y2H to map a network in the approach of a specific research question [91,92]. The researchers wondered how different plant pathogens manipulate host cellular processes resulting in host disease phenotypes. Mukthar *et al.* investigated interactions between proteins of the host *Arabidopsis thaliana* and effector proteins from two plant pathogens, the bacterium *Pseudomonas syringae* and the oomycete *Hyaloperonospora arabidopsidis* [91]. The resulting network was extended by Weßling *et al.* adding data on PPIs between effector proteins from the fungus *Golovinomyces orontii* and proteins from *Arabidopsis thaliana*. Analyzing this network map, the researchers identified a conserved host-pathogen interface across the three kingdoms of life. The plant pathogens seemed to have evolved independently targeting the same, limited set of host proteins (convergence). Subsequently, the authors showed that this convergence was biologically meaningful by employing host mutant lines created by altering host genes that were subject to convergence. Upon colonization of the mutant lines with the pathogens, changes in disease

susceptibility or disease resistance were assessed. A positive correlation between the degree of convergence on host proteins and the likelihood of an altered infection phenotype was observed. This demonstrates that the higher the intra- and interspecies convergence on a host protein the more likely its relevance for a certain host phenotype [92].

Importantly, Weßling *et al.* demonstrated that proteins relevant to disease severity and fitness, encoded by highly variant genes, interacted with host proteins that were targeted by the plant pathogens. Thereby, genetic variation located in the network neighborhood of pathogen targets impacted disease outcomes by modifying downstream effects [92,93]. This has also been shown by Kim *et al.* who published "a systematic contactome map of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) with the human host" in 2022 [93]. As infection is mediated via direct contacts between viral and host proteins using a Y2H approach as opposed to AP-MS was essential in detecting vital PPIs [93]. Examining the neighborhood of the targeted host proteins, the authors detected proteins "encoded from a critical illness-associated locus" interacting with targeted host proteins. Repeating the same analysis with interaction data from AP-MS studies detected no statistically relevant critical illness proteins demonstrating the advantage of the Y2H when studying circumstances where direct contacts are essential. Thereby, Kim *et al.* verified Weßling *et al.*'s findings in a human host demonstrating that "clinically relevant genetic variation acts in the local network neighborhood of viral contact proteins" [93]. Kim *et al.* looked even further into the host target's neighborhood analyzing subnetworks and communities (dense subnetworks that often exhibit a common function [80]) for human genetic variation that can impact disease susceptibility [93]. The authors identified several virus-targeted communities exhibiting genetic variations that were linked to SARS-CoV-2-induced COVID-19 disease severity [93]. Thereby, proteins in the host target's neighborhood reflecting genetic variation may impact COVID-19 progression and severity.

Information about genetic variation and association to diseases can be obtained from GWAS data. As mentioned in Chapter 1.2.1 network approaches aid in integrating GWAS data by linking genetic variation to biological networks. For instance, Duan *et al.* used a PPI network integrated with GWAS data to detect susceptibility genes for coronary artery disease [94]. They built a PPI network employing different databases with a focus on genes associated with the disease. After annotating the network with GWAS data, it was analyzed for susceptibility modules (modules represent highly connected regions in the network [80]) based on enrichment with differentially expressed genes. These modules were subjected to a functional analysis using the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database. Most of the modules were annotated with functions relevant to the disease i.e., they were directly or indirectly involved with vascular endothelial growth and inflammation, which contribute to coronary artery disease. In one of those modules, MAPK10 was identified as a susceptibility gene by three statistical gene-based association tests and two independent GWAS datasets. This is plausible considering that other studies connect MAPK10 to vascular endothelial dysfunction and pathogenesis of atherosclerosis. Thereby, the study demonstrates that PPI network analysis can help integrate GWAS data to identify disease susceptibility genes [94].

Hence, PPI networks are valuable in understanding mechanisms of diseases via e.g., the identification of disease modules and susceptibility genes and their functional analysis informing hypotheses formulation concerning underlying mechanisms of diseases.

1.4 Study objective

Research on the gut microbiota has detected associations between gut bacterial composition and various complex diseases, however, causality has only been established for a few cases. Pseudomonadota, the third most abundant phylum in the gut, has been linked to several complex illnesses e.g., metabolic sicknesses and cancer. This phylum differentiates itself from other phyla by encoding for T3SSs used to inject effector proteins into host cells. While T3SSs have been predominantly researched in human pathogens, they are also essential in mediating mutualistic relationships in plants. After injection into the host cell, effectors interact with various host proteins manipulating diverse processes such as host immune responses, and activities specific to the microbe-host interaction such as nodule formation in the rhizobia-legume relationship. This thesis aims to elucidate the role of T3SS effectors expressed by commensal gut Pseudomonadota concerning their impacts on host functions in the context of human health and disease.

Inside the cell, the T3SS effectors impact the cellular network potentially mediating effector functions along the network's links. Given that research on gut commensal T3SS effectors presents a completely new undertaking, a large-scale network approach is most suitable to provide an unbiased overview of the interactions. As PPIs moderate the interactions underlying healthy and diseased phenotypes, the effectors' impact on the host protein interactome will be investigated. To this end, the Y2H assay will be employed considering the interactions between effectors and host proteins are mediated via direct contacts. The bacterial ORFeome will be generated by cloning predicted effector open reading frames (ORFs) from commensal gut Pseudomonadota genomes as well as from metagenomic data. Screening this collection against the human ORFeome collection v9.1 consisting of 17,408 human genes will provide a PPI network map with insights into targeted host functions and signaling pathways. After quality control of the network map including benchmarking of the Y2H assay, assessing the saturation of the data, and re-testing a subset of the interactions in an orthogonal assay, further investigations in cell culture assays will be performed to functionally validate the findings.

2. Results

To elucidate the functions of T3SS effectors employed by human gut commensal Pseudomonadota a protein-protein meta-interactome is mapped and the effector targets are analyzed. This requires the selection of T3SS effectors of gut commensal Pseudomonadota (Chapter 2.1) and the generation of a bacterial effector ORFeome (Chapter 2.2) that can be screened against the human ORFeome collection v9.1. Interactions between gut commensal effectors and human proteins are detected employing a Y2H pipeline (Chapter 2.3) and the resulting network is subjected to a thorough quality control (Chapter 2.4). As protein homology is often used to infer functional similarity, interaction patterns of homologous effectors are analyzed (Chapter 2.5). To determine whether bacterial effectors can hijack human protein-protein interfaces, the ability of effectors to disrupt human-human PPIs is assessed (Chapter 2.6). Finally, a functional enrichment analysis of the effector targets is performed (Chapter 2.7). Based on the analysis of the functions of the effector targets, two hypotheses concerning effector impacts are proposed and tested in experiments *in vitro* (Chapter 2.8 and Chapter 2.9).

This thesis was part of a bigger project involving several molecular biologists and bioinformaticians who informed each other's work. Therefore, some contributions to my work were obtained: Patrick Hyden predicted the effector sequences from gut commensal Pseudomonadota strains and identified effectors with SfiI sites and low-frequency start-codons. Furthermore, he clustered the homologous effectors according to sequence similarity. Dr. Stefan Altmann identified transmembrane effectors and clustered the effectors phylogenetically. Dr. Chung-Wen Lin designed the primers and demultiplexed the next-generation sequencing (NGS) data. Dr. Benjamin Weller performed the yN2H. Bushra Dohai conducted the random selection of effectors for the RRS and for the HuMMI-subset tested by the yN2H as well as the convergence analysis of the effector targets. She also identified two negative controls for the experiments on the effectors' impact on apoptosis using the shortest path analysis. Dr. Andreas Zanzoni gathered the publications for the assembly of the PRS. The team assembling the bacterial-human PRS consisted of Dr. Andreas Zanzoni, Prof. Dr. Pascal Falter-Braun, Dr. Claudia Falter, Dr. Melina Altmann, and myself. Patrick Schwehn supported the calculation of the sampling saturation curve. The cloning and transformation in preparation for the Y3H experiments were supported by Katharina Frey, a master's student performing an internship at INET. Dr. Thomas Hitch analyzed the metagenomic data of IBD patients for the presence of the effectors investigated in this study. Niels van Heusden performed the assays analyzing cytokine secretion from colonocytes and ICAM1 expression.

2.1 Selection of gut bacterial T3SS effectors

For the mapping of effector-human PPIs, a bacterial ORFeome was required. To this end, gut commensal Pseudomonadota strains available from culture collections were analyzed for the presence of T3SSs and effector sequences. Culture collections are susceptible to culturing bias i.e., bacterial strains that grow well under the applied conditions or were identified due to continuous efforts are often overrepresented among identified bacteria [95]. Culturing bias can be reduced by including metagenomic data which are obtained independent of culturing. Hence, metagenome-assembled genomes (MAGs) were also analyzed for the presence of T3SS effectors.

2.1.1 Effectors from cultured gut bacterial strains

The Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ) and the HMP were analyzed for Pseudomonadota strains meeting three criteria (Fig. 10a): 1) isolation from the human intestine to obtain Pseudomonadota strains from the human gut. 2) availability of a reference genome as the prediction of effector sequences depended on it. 3) accessibility of the strains' DNA required for the cloning of bacterial effectors. This yielded 77 strains which were screened for the presence of T3SSs using EffectiveDB, a widely used annotation tool of bacterial secretion systems [96]. Of those, 44 strains encoded complete T3SSs (Table S1). To reliably identify effector proteins in these T3SS-positive strains, the overlapping results of three complementary machine learning models predicting putative T3SS effectors were considered (EffectiveT3 [97], DeepT3 [98], and pEffect [99]).

As the cloning strategy involved the introduction of effectors into an entry vector by SfiI digestion, all effector sequences with an inherent SfiI site were excluded. Furthermore, effectors containing a start codon with < 3% frequency were eliminated to secure translation initiation (leaving "ATG", "GTG", and "TTG"). This was determined based on the work of Hecht *et al.* who quantified translation initiation from 64 start codons [100]. Moreover, 480 genes containing a transmembrane domain were removed due to their decreased relevance as effector proteins and lessened compatibility with the Y2H system. This resulted in 3002 effector proteins of 44 T3SS-positive strains (Fig. 10b).
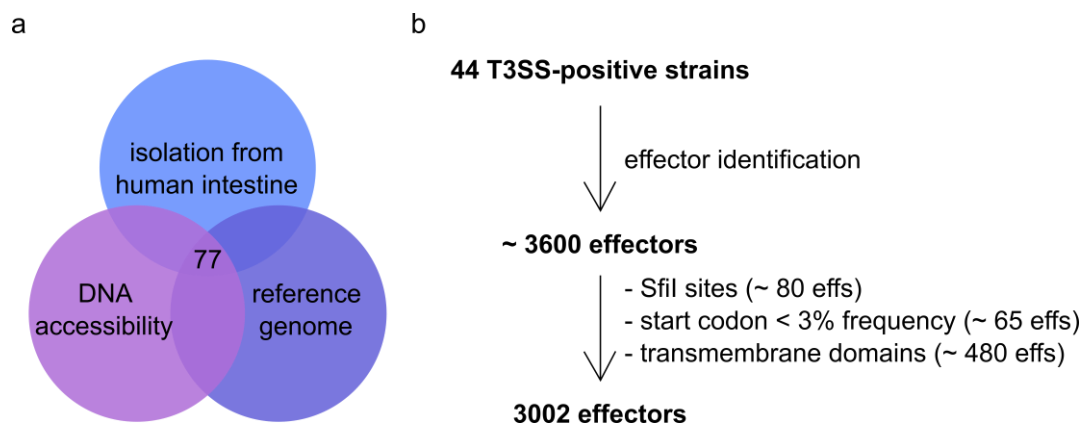
Figure 10 | Effector identification. a | *77 Pseudomonadota strains were detected following three selection criteria.* **b |** *Number of effectors (effs) after effector identification from T3SS-positive strains and after removal of effectors incompatible with the experimental setup. Some effectors were incompatible due to more than one factor e.g., due to transmembrane domains and the presence of an SfiI site.*

To keep the cloning process manageable concerning time and resources, the number of strains included in the study was decreased aiming for a selection of diverse, diet-dependent strains (Table 2). Selecting diverse strains gives a broader overview of the effectors' capabilities and functions. Focusing on strains that are affected by diet is relevant due to the links between diet, Pseudomonadota abundance, and disease: Pseudomonadota abundance fluctuates in response to dietary intake and bacterial abundance, in turn, is associated with certain diseases [3]. To select diverse, diet-dependent strains from the 44 T3SS-positive gut commensal Pseudomonadota, the characteristics "association of strain abundance to nutrition" and "diversity of strains" were determined.

Table 2 | Overview of the 18 gut bacterial strains selected for this study. *Color-code describes the basis on which strains were selected: ■, bacterial abundance is impacted by nutrition according to previous studies (Table 3); ■, contribution to strain diversity within the Pseudomonadota collection (Fig. 11). Abbreviations (abbr.) of the strains used throughout this work are stated. The number of predicted effectors (# eff) is indicated per strain.*

	species	strain	abbr.	# eff
■	*Aeromonas jandaei*	*Aeromonas jandaei CECT 4228*	Aja	59
■	*Cedecea davisae*	*Cedecea davisae Grimont et al. 1981*	Cda	81
■	*Citrobacter youngae*	*Citrobacter pasteurii Clermont et al. 2015*	Cyo	83
■	*Escherichia coli*	*Escherichia coli MS 200-1*	Ec2	40
■	*Escherichia coli*	*Escherichia coli MS 69-1*	Ec6	75
■	*Escherichia fergusonii*	*Escherichia fergusonii Farmer et al. 1985*	Efe	58
■	*Edwardsiella tarda*	*Edwardsiella tarda ATCC 23685*	Eta	29
■	*Klebsiella pneumoniae*	*Klebsiella sp. MS 92-3*	Kpn	57
■	*Morganella morganii*	*Morganella morganii subsp. morganii NBRC 3848*	Mmo	69
■	*Pseudomonas_E massiliensis*	*Pseudomonas sp.*	Pem	73
■	*Pseudocitrobacter faecalis*	*Pseudocitrobacter faecalis Kämpfer et al. 2014*	Pfa	72
■	*Phytobacter massiliensis*	*Enterobacter massiliensis Lagier et al. 2014*	Pma	54
■	*Providencia rettgeri_D*	*Providencia rettgeri DSM 1131*	Pre	119
■	*Pantoea septica*	*Pantoea septica*	Pse	54
■	*Providencia stuartii*	*Providencia stuartii ATCC 25827*	Pst	107
■	*Vibrio furnissii*	*Vibrio furnissii NCTC 13120*	Vfu	101
■	*Yersinia enterocolitica*	*Yersinia enterocolitica subsp. palearctica Y11*	Yen	118
■	*Yokenella regensburgei*	*Yokenella regensburgei ATCC 43003*	Yre	58

<u>Association of strain abundance to nutrition:</u> In the third quarter of 2019 a scientific literature search was performed assessing nutritional impacts on gut bacterial abundance or gut bacterial metabolism (Table 3). Information on the strain level was scarce given the specificity of the search criteria. Furthermore, studies often focused on a particular subject such as the bacteria's ability to degrade choline to trimethylamine (TMA), which was the most common information available. Choline is present in eggs, dairy products, and red meat in the human nutrition and some gut bacteria metabolize the compound to TMA [101]. TMA is absorbed by the host and travels to the liver where it is oxidized to trimethylamine N-oxide (TMAO) by a hepatic enzyme. Elevated levels of TMAO in the blood circulation are strongly associated with increased inflammation in the human host leading to, e.g. higher risks of cardiovascular diseases [102]. Notably, the metabolic capacity of the microbes to degrade choline to TMA can be manipulated by dietary measures: Panyod *et al.* showed that raw garlic juice inhibited TMA production of the gut microbiota *in vitro* [101]. Given their malleable capacity to degrade choline to TMA, both *Escherichia coli* strains were selected, as well as *Edwardsiella tarda*, *Klebsiella pneumoniae*, and *Providencia rettgeri_D*.

Furthermore, a study conducted in pigs demonstrated a decreased abundance of some species of the *Providencia* genus in response to a diet supplemented with prebiotics (dietary fiber), probiotics (living microorganisms), or synbiotics (pre- and probiotics) [103]. However, the resolution of the effect on the strain level was missing. Therefore, information concerning the diversity of effector genes of the T3SS-positive *Providencia* strains was consulted to decide which strains to include in the study.

Table 3 | Associations of T3SS-positive strains to nutrition. *Association to nutrition of the Providencia genus is described on the species level, for all other findings the association was stated on the strain level.*

strain	association to nutrition	study type
Ec2	degraded choline to TMA	*in vitro* [104]
Ec6	degraded choline to TMA	*in vitro* [105]
Eta	degraded choline to TMA	*in vitro* [102]
Kpn	degraded choline to TMA	*in vitro* [105]
Pre	degraded choline to TMA	*in vitro* [102]
Providencia stuartii	reduction with pre-, pro- and synbiotics	*in vivo* (pigs) [103]
Providencia rettgeri	reduction with pre-, pro- and synbiotics	*in vivo* (pigs) [103]
Providencia alcalifaciens	reduction with pre-, pro- and synbiotics	*in vivo* (pigs) [103]

Diverse strains: The effector genes of the 44 T3SS-positive strains were clustered according to 90% sequence identity over 90% sequence length (Fig. 11). This analysis mostly coincided with the clustering of the strains according to genera revealing similarities between effectors of the same genus. The number of similar effectors varied between strains of the same genus with some strains sharing 90% sequence similarity in almost half of their effectors (*Escherichia coli MS 69-1* and *Escherichia coli MS 198-1*), while others exhibited similarities only in < 5% of their effectors (*Providencia alcalifaciens* and *Providencia stuartii*). Only the genus *Pseudomonas* stood as an exception, as the strains exhibited no similarities within their respective effector complements. To reduce redundancy between the cloned effectors, only one strain per genus was chosen besides the strains that had already been selected based on their association with nutrition. Per genus, the strain was selected for which genomic DNA or living cultures were accessible or which encoded the highest numbers of predicted effectors.

As mentioned earlier, three species of the genus *Providencia* were detected to decrease in abundance in response to pre-, pro-, and synbiotics [103] (Table 3), with missing resolution on the strain level. *Providencia stuartii* was selected due to its phylogenetically unique effector complement compared to the other strains in the *Providencia* cluster (Fig. 11).

Figure 11 | Phylogeny matrix of effector genes between the 44 T3SS-positive strains. *Numbers of shared genes with 90% sequence identity over 90% sequence length are listed. Darker blue indicates > 5 similar effector genes that are shared between strains, whereas lighter blues marks < 5 similar effector genes shared between strains within the same genus.*

Column/row index key (same order for rows and columns):

1 Aeromonas dhakensis CIP 107500 · 2 Aeromonas enteropelogenes CECT 4255T · 3 Aeromonas enteropelogenes CECT 4487 · 4 Aeromonas jandaei CECT 4228 · 5 Aeromonas tecta CECT 7082 · 6 Cedecea davisae Grimont et al. 1981 · 7 Citrobacter europaeus Ribeiro et al. 2017 · 8 Citrobacter pasteurii Clermont et al. 2015 · 9 Edwardsiella tarda ATCC 23685 · 10 Edwardsiella tarda Ewing and McWhorter 1965 · 11 Enterobacter cloacae (Jordan 1890) Hormaeche and Edwards 1960 · 12 Enterobacter massiliensis Lagier et al. 2014 · 13 Escherichia albertii 19982 · 14 Escherichia coli D9 · 15 Escherichia coli MS 110-3 · 16 Escherichia coli MS 115-1 · 17 Escherichia coli MS 16-3 · 18 Escherichia coli MS 196-1 · 19 Escherichia coli MS 198-1 · 20 Escherichia coli MS 200-1 · 21 Escherichia coli MS 21-1 · 22 Escherichia coli MS 57-2 · 23 Escherichia coli MS 69-1 · 24 Escherichia fergusonii Farmer et al. 1985 · 25 Escherichia sp. · 26 Klebsiella pneumoniae subsp. pneumoniae WGLW3 · 27 Klebsiella pneumoniae subsp. pneumoniae WGLW5 · 28 Klebsiella sp. MS 92-3 · 29 Morganella morganii subsp. morganii NBRC 3848 · 30 Pantoea septica · 31 Providencia alcalifaciens DSM 30120 · 32 Providencia rettgeri DSM 1131 · 33 Providencia rustigianii DSM 4541 · 34 Providencia stuartii ATCC 25827 · 35 Pseudocitrobacter faecalis Kämpfer et al. 2014 · 36 Pseudomonas sp. · 37 Pseudomonas sp. 2_1_26 · 38 Pseudomonas sp. HPB0071 · 39 Vibrio fluvialis ATCC 33809 · 40 Vibrio furnissii NCTC 13120 · 41 Yersinia aleksiciae 159 · 42 Yersinia bercovieri Wauters et al. 1988 · 43 Yersinia enterocolitica subsp. palearctica Y11 · 44 Yokenella regensburgei ATCC 43003

Row	Shared-gene values (column: value)
1	1:45 2:6 3:4 4:7 5:4
2	1:6 2:28 3:15 4:7 5:3
3	1:4 2:15 3:34 4:4 5:2
4	1:7 2:7 3:4 4:45 5:4
5	1:4 2:3 3:2 4:4 5:47
6	6:66 7:2 8:2 11:2 12:2 13:2 15:1 16:1 17:1 20:1 21:1 22:1 24:2 25:2 35:2 44:2
7	6:2 7:62 8:27 11:2 12:2 13:4 14:4 16:3 17:2 18:1 19:1 20:2 21:3 22:1 23:1 24:4 25:4 35:1 36:2 39:2 44:4
8	6:2 7:27 8:69 11:2 12:2 13:5 14:3 16:2 17:2 18:2 19:2 20:2 21:3 22:1 23:2 24:4 25:4 35:1 36:2 39:2 44:5
9	9:20 10:5 42:1 43:1
10	9:5 10:18
11	6:2 7:2 8:2 11:50 12:2 13:2 14:1 17:1 24:2 25:1 35:1 36:3 44:2
12	6:2 7:2 8:2 11:2 12:46 13:2 14:1 17:1 24:2 25:1 35:1 36:1 38:3 44:2
13	6:2 7:4 8:5 11:2 12:2 13:68 14:9 15:1 16:5 17:7 18:7 19:6 20:9 21:7 22:7 23:9 24:9 25:13 35:1 36:2 39:2 44:3
14	6:2 7:4 8:3 13:9 14:55 15:2 16:22 17:12 18:21 19:23 20:12 21:17 22:10 23:23 24:12 25:21 36:1 39:1 44:2
15	14:1 15:2 16:9 17:0 18:1 19:0 20:0 21:2 22:1 23:2 24:3 25:1 26:7
16	6:1 7:3 8:2 13:1 14:1 15:5 16:22 17:0 18:49 19:10 20:15 21:15 22:7 23:10 24:9 25:15 26:6 27:12 35:1 36:2 39:1 44:1
17	6:1 7:2 8:2 13:7 14:12 15:1 16:10 17:37 18:12 19:14 20:13 21:11 22:14 23:14 24:8 25:19 35:1 39:1 44:1
18	6:1 7:1 8:2 13:7 14:21 15:0 16:15 17:12 18:57 19:20 20:7 21:12 22:8 23:20 24:7 25:18 35:1 39:2 44:1
19	8:1 12:2 13:6 14:23 15:0 16:15 17:14 18:20 19:97 20:10 21:32 22:12 23:34 24:11 25:26 35:1 39:2
20	6:1 7:2 8:2 13:9 14:12 15:2 16:7 17:13 18:7 19:10 20:34 21:12 22:12 23:13 24:7 25:19 35:1 39:1 44:1
21	6:1 7:3 8:3 13:7 14:17 15:1 16:10 17:11 18:12 19:32 20:12 21:72 22:9 23:26 24:11 25:24 35:1 39:1 44:1
22	6:1 7:1 8:1 13:7 14:10 15:2 16:9 17:14 18:8 19:12 20:12 21:9 22:34 23:12 24:6 25:19 35:2 39:1 44:1
23	8:1 12:2 13:9 14:23 15:3 16:15 17:14 18:20 19:34 20:13 21:26 22:12 23:68 24:9 25:24 35:1 39:1
24	6:2 7:4 8:4 11:2 12:2 13:9 14:12 15:1 16:6 17:8 18:7 19:11 20:7 21:11 22:6 23:9 24:51 25:13 35:1 36:2 39:2 44:3
25	6:2 7:4 8:4 13:1 14:1 15:13 16:21 17:7 18:12 19:19 20:18 21:26 22:19 23:24 24:19 25:24 13:... 25:78 35:1 39:1 44:2
26	13:1 16:1 17:1 18:1 19:1 22:1 23:1 26:33 27:9 28:18 32:1 39:1
27	7:1 8:1 14:1 16:1 17:1 19:1 26:9 27:35 28:11 32:1 39:3 44:1
28	6:2 7:2 8:2 13:3 14:3 15:2 16:1 18:2 19:1 20:2 21:2 23:1 24:1 25:1 26:2 27:1 28:18 ... 26:18 27:11 28:47 32:1 39:3 44:2
29	29:52 32:1
30	30:50
31	31:98 32:6 33:12 34:3
32	26:1 27:1 28:1 31:6 32:100 33:5 34:7
33	31:12 32:5 33:69 34:1
34	28:1 31:3 32:7 33:3 34:81
35	7:2 8:2 11:2 12:1 13:2 14:1 16:1 17:1 18:1 19:1 20:1 21:2 22:1 23:2 24:1 25:1 26:3 28:3 35:59 36:0 37:0 38:0 44:1
36	35:0 36:62 37:0 38:0
37	35:0 36:0 37:71 38:0
38	35:0 36:0 37:0 38:85
39	39:66 40:25
40	39:25 40:82
41	41:64 42:20 43:12
42	13:1 41:20 42:71 43:16
43	13:1 41:12 42:16 43:99
44	6:2 7:4 8:5 11:2 12:2 13:3 14:2 16:1 17:1 18:1 20:1 21:1 22:1 24:3 25:2 35:1 36:2 39:1 44:45

2.1.2 Effectors from gut MAGs

Besides effectors from known bacterial strains, T3SS effectors from MAGs were included in the bacterial ORFeome to reduce culturing bias. Initially, the goal was to detect effectors from MAGs responding in abundance to diet due to the links between diet, Pseudomonadota abundance, and disease as described in the previous chapter. However, data on metagenomes in relation to diet were sparse and the few dietary intervention studies available provided very little data (February 2020). Therefore, two large meta-studies without associations to diet were consulted for metagenomic data of the human gut microbiota. Pasolli *et al.* acquired metagenomic data from 7,783 stool samples of differently-aged individuals from four continents following a westernized or non-westernized lifestyle [106]. Almeida *et al.*, acquired metagenomic data from 11,850 human gut microbiomes of mainly diseased, adult patients from 75 studies predominantly carried out in North America and Europe [107]. Data from both studies taken together offered the largest set of gut metagenomic data available.

To predict effector sequences from the MAGs, the three complementary machine learning models used for the prediction of T3SS effectors in the known strains were employed. The overlapping results predicted 186 effectors within the MAGs, which are referred to as "metagenomic effectors" throughout this work.

2.2 Bacterial effector ORFeome collection

In total, 1,307 effectors of the 18 strains and 186 effectors from MAGs were identified. If available, genomic DNA was obtained from the 18 selected strains, or alternatively, their live cultures. Effectors from MAGs were obtained as chemically synthesized nucleotide sequences. Depending on the effector length, they were acquired as gene fragments (< 1800 bp) or as clonal genes in an entry vector (> 1800 bp). The linker regions located up and downstream of the metagenomic effector ORFs were deliberately designed to be identical to the linker regions of the primers used to clone effectors from genomic DNA. A cloning pipeline was developed to systematically generate a bacterial ORFeome collection. Depending on the starting material (genomic DNA, gene fragment, or clonal gene) the ORFs were channeled through different stages (Fig. 12).
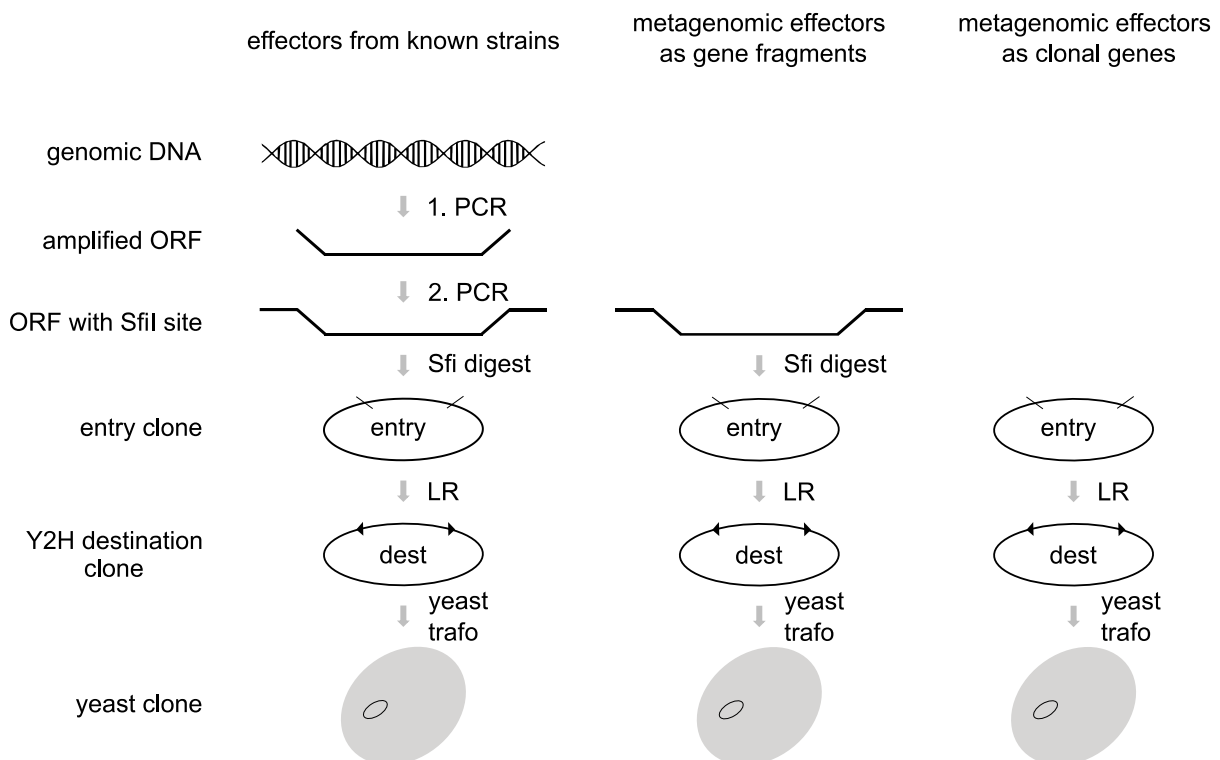
Figure 12 | Cloning strategy according to starting material. *The full cloning pipeline, which was entered at different stages, was designed as follows: first, ORFs from genomic DNA are amplified by polymerase chain reaction (PCR), and part of a SfiI side is added by the primer. The second PCR completes the SfiI sites up and downstream of the effector ORF. PCR products and entry vector are digested with the SfiI enzyme and sticky ends of the vector and the ORF are ligated. ORFs are transferred into the Y2H destination vector by an LR Gateway™ reaction and the plasmids are transformed into yeast.*

Genomic DNA could be obtained for 13 of the 18 strains and live bacterial cultures were acquired for the remaining five for which genomic DNA was unavailable (*Escherichia coli MS 200-1, Escherichia coli MS 69-1, Edwardsiella tarda, Klebsiella pneumoniae* and *Providencia stuartii)*. Prior to growing the bacteria in culture, experiments on a lab strain of *Escherichia coli* were performed to determine the best method to isolate high-quality DNA for effector amplification from bacterial cultures. First, genomic DNA extraction was attempted using Invitrogen™ TRIzol™ Reagent. The compound contains phenol which induces cell lysis and in combination with chloroform separates proteins and DNA allowing the isolation of genomic DNA. However, this method did not yield sufficient amounts of high-quality genomic DNA. Therefore, the DNeasy® UltraClean® Microbial Kit was employed relying on mechanical force

against specialized beads to induce cell lysis subsequently binding DNA on a silica column. This method was most successful employing the FastPrep-24™ for cell lysis, a high-speed homogenizer. Genomic DNA yield was relatively low but ORF amplification via PCR provided good results. Due to the low DNA concentration and operational challenges regarding the FastPrep-24™, a protocol based on alkaline lysis was tested. The detergent sodium dodecyl sulfate (SDS) was used for cell lysis and protein denaturation, followed by a NaOH-rich buffer to denature genomic DNA into single strands. Decreasing the alkalinity in the sample the DNA was reconstituted into double strands which were more soluble in the solution compared to the denatured proteins and could be isolated from the mixture [108]. While DNA concentration in the sample was high (> 3 µg/µL), the success of ORF amplification via PCR was diminished, and if ORFs could be amplified successfully, large DNA amounts per reaction (10 ng/µL) were required. This was presumably due to the incomplete recovery of the genomic DNA into double strands. Hence, a fourth method was employed testing the NucleoSpin® Plasmid kit with two vortexing steps after the addition of the lysis buffer and the neutralization buffer to shear genomic DNA. The kit is also based on SDS and alkaline lysis, after which DNA is bound to a silica membrane for its purification and isolation. Similarly to the alkaline lysis, ORF amplification via PCR was not satisfactory. Despite the relatively low DNA yield, the DNeasy® UltraClean® Microbial Kit demonstrated the best results concerning ORF amplification and was employed for isolating genomic DNA of the five strains for which live bacterial cultures were obtained.

The primers for the first PCR were designed using the Primer3 program with melting temperatures (T_m) i.e., "the dissociation temperature of the primer/template duplex" [109], between 50-64 °C selecting higher T_m if possible. This is important to provide "a sufficient thermal window for efficient annealing" as a high annealing temperature (T_a) of the primer to the DNA template is necessary to avoid unspecific DNA binding and as the T_a should be close to the T_m to further support sequence specificity [109,110]. The average primer length varied between 15-24 base pairs (bp) with an additional 12 bp for the forward linker and 15 bp for the reverse linker which is in line with recommended primer length [109].

Most genomic DNA of the 13 strains was obtained in sufficient amounts ranging from 2.0-7.3 µg per strain. However, the genomic DNA of three strains (*Providencia rettgeri_D*, *Phytobacter massiliensis,* and *Citrobacter youngae*) only reached a total of 0.4-0.5 µg and for *Pseudomonas_E massiliensis* only 0.125 µg could be acquired. Therefore, experiments were performed to determine the lowest DNA amount possible for successful ORF amplification by PCR, as well as to identify the optimal DNA polymerase. To this end, different polymerases such as KOD, Phusion, and Taq were combined with different amounts of DNA (0.1 ng-10 ng per reaction). On average, 0.2 ng purified genomic DNA per PCR reaction was sufficient for successful ORF amplification using the KOD polymerase, a high-fidelity polymerase due to its proof-reading activity. These conditions constituted the standard protocol to amplify the effector ORFs from the bacterial genomic DNA. Optimal primer T_a required to be determined for every strain separately (Table 4). Thus, PCR reactions for the amplification of a few ORFs per strain were set up in batches using varying T_as which allowed selecting the optimal T_a per strain (Fig. 13a). For both *Escherichia coli* strains and *Klebsiella pneumoniae* this protocol was not successful questioning the standard PCR conditions as described above. Hence, varying DNA concentrations per sample, as well as different polymerases (Fig. 13b), were combined in separate PCR reactions to determine optimal PCR conditions. 5 ng genomic DNA and amplification by the high-fidelity Phusion polymerase proved successful and were employed as the basis for determining the optimal T_a (Table 4). As differently sized ORFs require

different elongation times during PCR, the effectors were grouped according to their sequence length: "small" effectors under 1 kbp, "middle" effectors between 1 kbp and 2.5 kbp, and "large" effectors over 2.5 kbp (Fig. 13c). As the same primers were used in the second PCR to complete the SfiI sites for all ORFs, the same conditions could be employed for all strains.

Table 4 | PCR conditions per strain. *The amount of DNA, annealing temperature (T$_a$), and polymerase used for the first PCR per strain are stated. The second PCR was conducted using the same condition for each strain.*

		1. PCR		2. PCR
strain	DNA (ng)	T$_a$ (°C)	polymerase	T$_a$ + polymerase
Aja	0.2	52.5	KOD	
Cda	0.2	50.5	KOD	
Cyo	0.2	52.5	KOD	
Ec2	5.0	50.0	Phusion	
Ec6	5.0	50.0	Phusion	
Efe	0.2	50.5	KOD	
Eta	0.2	54.0	KOD	
Kpn	5.0	50.0	Phusion	
Mmo	0.2	56.0	KOD	51 °C + KOD
Pem	0.2	56.0	KOD	
Pfa	0.2	50.5	KOD	
Pma	0.2	56.6	KOD	
Pre	0.2	50.5	KOD	
Pse	0.2	52.5	KOD	
Pst	0.2	52.5	KOD	
Vfu	0.2	54.0	KOD	
Yen	0.2	52.5	KOD	
Yre	0.2	56.0	KOD	

a
Phytobacter massiliensis

T$_a$ = 51.1 °C

T$_a$ = 56.6 °C

c

b
Klebsiella pneumoniae

KOD

Taq

Phusion

After optimal PCR conditions were determined per strain, a nested PCR was conducted to amplify effector ORFs adding a stop codon after the effector sequence as well as SfiI-sites up and downstream of the ORF. PCR products were purified using magnetic beads to remove remnants of the PCR reaction. Subsequently, PCR products were cut by SfiI restriction employing a standardized protocol according to the manufacturer's instructions. This did not yield the expected results which was determined by gel electrophoresis analyzing the digested PCR. As the SfiI enzyme was originally intended to cut exclusively between recognition sites at both ends of the PCR product, thereby generating sticky ends for subsequent ligation, only a singular band was anticipated on the agarose gel. However, the enzyme's star activity resulted in multiple bands on the gel due to the digestion of ORF sequences with similarities, but not exact matches to the SfiI's recognition sequence (Fig. 14). Measures to reduce the enzyme's star activity included the reduction of the enzyme's concentration per reaction and diminishing the buffer concentration. Furthermore, adjustments were made to the incubation temperature and duration. The combinations of these refinements resulted in a tailored SfiI-digestion protocol, which, in most instances, provided precise digestion of PCR products.

Gel electrophoresis of digested PCR products of Cedecea davisae

- 0.25 µl Sfi / well
- 5 µl Cut smart buffer / well
- 50°C
- 2 h incubation

- 0.0125 µl Sfi / well
- 4 µl Cut smart buffer / well
- 48°C
- 7 min incubation
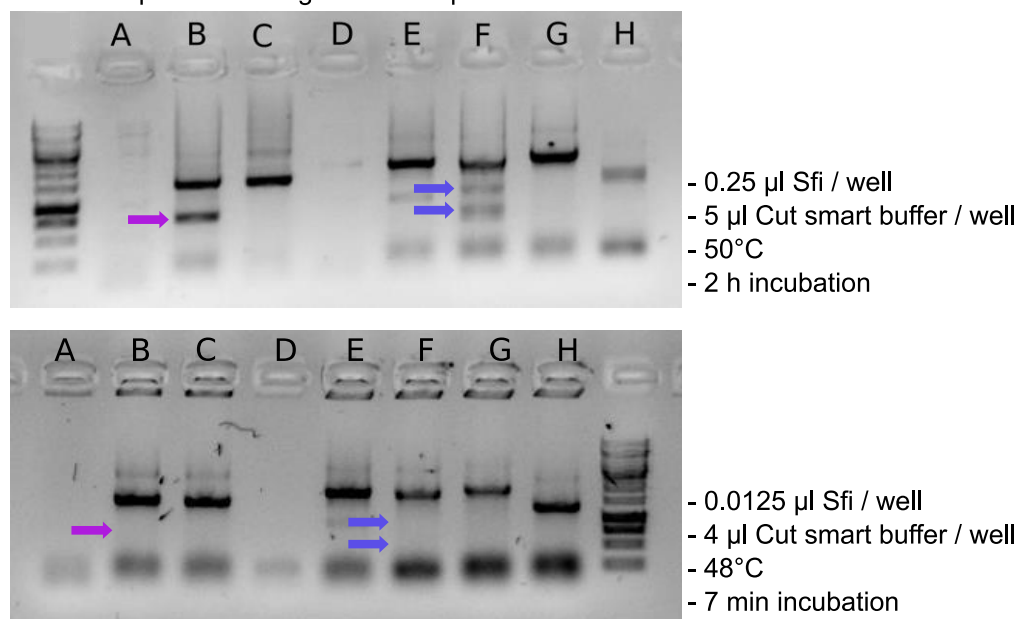
Figure 14 | Differences between SfiI digestion protocols. *Gel electrophoresis of digested PCR products of the strain Cedecea davisae. The first image shows the result of the SfiI restriction performed according to the manufacturer's protocol. The image below depicts the results after several adjustments to the SfiI-restriction protocol. Arrows indicate the bands that vanished after protocol improvement.*

After SfiI-digestion, PCR products were subjected to another cleaning step with magnetic beads to remove remains from the digestion mixture. The ORFs were ligated into an entry vector displaying sticky ends after digestion by the SfiI enzyme. Subsequently, ORFs were transferred from the entry vectors into the destination Y2H expression vectors via Gateway™ LR reaction. The destination vectors were transformed into the *Saccharomyces cerevisiae* Y8930 strain.

Not all effector ORFs were successfully processed in the different stages of the cloning pipeline. Hence, the collection of cloned effector ORFs required several consolidation steps for more convenient processing of the collection in the following cloning step. Figure 15 demonstrates the amount and percentage of successfully processed effector ORFs after each step. For instance, ~ 14% of effectors from genomic DNA could not be amplified in the first PCR. Further ~ 11% of bacterial effectors were not transferred from the entry to the destination vector. To confirm the identity of the ORFs and ensure that PCR amplification had not halted prematurely, forward and reverse end-reads using Sanger-sequencing were obtained. The number of effectors with incorrect sequences was ~ 2%. Overall, ~ 60% of effectors (786 ORFs) from known strains were cloned into yeast and successfully sequence-identified by Sanger Sequencing (Table S3). For the metagenomic effectors, the success was ~ 95% (173 ORFs) (Table S3).
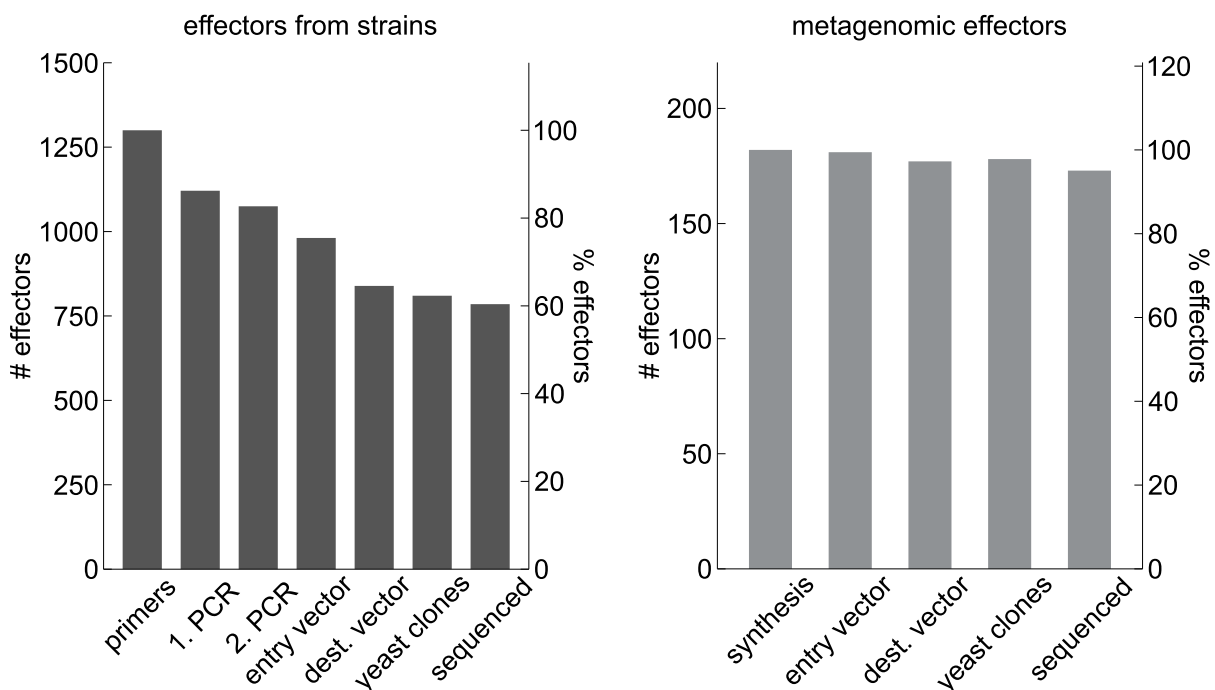
Figure 15 | Number of cloned ORFs per cloning stage. *As both types of effectors (effectors from strains and metagenomic effectors) entered the cloning process at different stages, different amounts of stages are depicted in the right versus the left panel. On the left are the number and percentages of effectors cloned from the strains at every step of the cloning process. "primers" refers to the primers obtained, and "sequenced" to correct end reads of the effectors by Sanger-sequencing. On the right are the number and percentages of the metagenomic effectors cloned after different stages. "synthesis" refers to the number of metagenomic effectors that were chemically synthesized (182 ORFs were synthesized of the 186 identified metagenomic effectors due to challenges during synthesis and export). Gene fragments were sequenced by Sanger-sequencing whereas the clonal genes had already been correctly sequenced directly after chemical synthesis by the company providing the synthesized ORFs.*

Figure 16 depicts the successfully cloned ORFs of the human microbiome effector ORFeome version 1 (HuMEOme_v1) consisting of 959 yeast clones containing different T3SS effectors. The different bacterial strains contribute with different amounts of effectors to HuMEOme_v1. This is due to differences in the number of predicted effectors and variations in cloning success per strain. For instance, 118 effectors were predicted for *Yersinia enterocolitica*, whereas only 29 were predicted for *Edwardsiella tarda*. The cloning success exhibited great differences between the strains with ~ 80% of effectors cloned for *Providencia stuartii* and only ~ 17% for *Pantoea septica*. The elevated GC content of Pantoea septica's genomic DNA, approaching 60%, could account for this phenomenon. Other strains with a similar genomic GC content are *Aeromonas jandaei* and *Pseudomonas_E massiliensis*, which also exhibited a relatively low cloning success.

In conclusion, HuMEOme_v1 constitutes a valuable resource of diverse gut commensal Pseudomonadota effectors from known strains and MAGs, not only for this study but also for future endeavors investigating gut commensal T3SS effectors. This is facilitated by the availability of the effector ORFs in Gatway™-compatible entry vectors which enables a convenient transfer into a variety of destination vectors allowing their usage in a multitude of assays. In this work, HuMEOme_v1 was employed for the meta-interactome mapping of PPIs between gut commensal effectors and human host proteins.
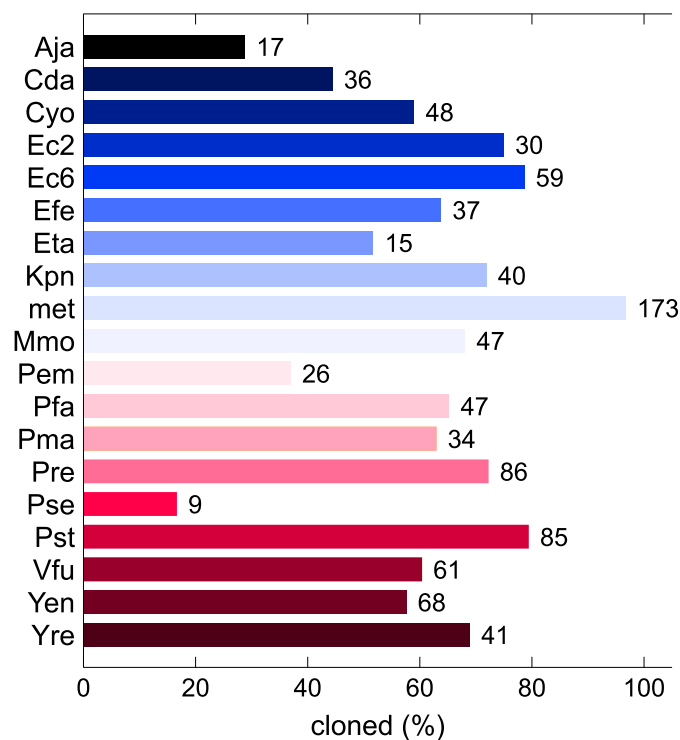
Figure 16 | Gut commensal Pseudomonadota effectors in yeast. *Percentages of cloned effector ORFs per strain based on the number of predicted effectors. Numbers of cloned effector ORFs on the right side of the bars. Metagenomic effectors are referred to as "met".*

2.3 Gut commensal microbiome-host protein interactome mapping

To generate the bacterial-human interactome, a Y2H mapping pipeline was employed to detect direct PPIs. Crucial to the Y2H is the split of the yeast transcription factor Gal4 into the DB and the AD. The cDNA encoding a protein of interest (bait) is fused to the DB (DB-X construct), while the cDNA encoding a different protein (prey) is fused to the AD (AD-Y construct) [111]. The two proteins encounter each other after co-transformation into the same yeast cell, or upon mating of haploid yeast strains of opposite mating types where each strain carries either the DB-X or the AD-Y construct [111]. Interaction of the two proteins reconstitutes the transcription factor subsequently expressing the reporter gene (e.g. HIS3) under its control [112].

The Y2H system is typically employed to test whether two known proteins interact or to identify novel interaction partners of a protein of interest by screening it against a large library of prey proteins. When dealing with large libraries consisting of thousands of genes, mating yeast cells is more advantageous compared to co-transforming all possible combinations of DB-X and AD-Y constructs into yeast. For one, yeast mating circumvents the challenge of low transfection efficiency during co-transformation [108]. Furthermore, it provides more flexibility and renders screening against large libraries manageable. The throughput can be further increased by pooling yeast strains expressing prey proteins and mating yeast pools against the protein of interest [41]. Successful mating of two haploid strains of opposing mating types is usually assessed by auxotrophic selection [109]. To that end, the two haploid yeast strains are auxotrophs for leucine and tryptophan respectively whereas the Y2H vectors contain the complementing genes LEU2 and TRP1 respectively [109]. Mated yeast cells have both vectors enabling yeast cell growth on selection plates. Interaction of the bait and prey proteins can be identified using selection plates that require the expression of the reporter gene for yeast cell growth [109]. Growth strength of the yeast clones varies which "may not reflect the actual affinity of protein-protein interactions as they take place in their native environment" [110]. Guiding the identification of yeast growth indicating PPIs and to control for appropriate selection plates, six Y2H controls can be added at the bottom of each selection plate [110,111]. The diploid Y2H control strains contain different combinations of DB-X and AD-Y pairs displaying varying interaction strengths on different selection plates (Table 10 in Chapter 4) [110].

A common artifact of the Y2H is the autoactivation of the reporter gene by the DB-X construct in the absence of an interaction with an AD-Y construct. This can be controlled for by identifying constitutive autoactivators that contain a protein structure similar to the AD and by detecting spontaneous *de novo* autoactivators arising by mutation during the Y2H assay [113]. The former are detected by a pre-screen testing the DB-X against the empty prey vector, whereas the latter are identified during the Y2H assay by mating against empty plasmids and by employing the counter-selectable marker CYH2 and cycloheximide (CHX) [113]. The AD-Y coding plasmids contain the CYH2 gene conferring sensitivity to CHX [113]. Growing on selection plates containing CHX, yeast clones are identified that have segregated out the AD-Y construct and in which the DB-X activated the reporter gene expression.
Besides autoactivation, mutations of the bait or prey proteins during the Y2H assay can also result in false-positive or false-negative interactions. Therefore, candidate interactions are retested multiple times to verify them by selecting the haploid yeast clones containing either of the proteins of the candidate interaction pair from their archival stocks and repeating the yeast mating [113].

Three separate Y2H experiments are performed in this study to answer three different questions. To obtain insights into effector targets and their functions, a main screen is conducted screening all bacterial effectors of the HuMEOme_v1 against all human proteins of the human ORFeome v9.1. The resulting network map is called the human-microbiome meta-interactome map (HuMMI)$_{main}$. For information about the saturation of HuMMI$_{main}$ and thereby about the completeness of the map, a repeat screen is performed. It consists of three repeat screens of a subset of the bacterial effectors against a subset of the human proteins (HuMMI$_{repeat}$). In addition to overlaps of interactions between HuMMI$_{main}$ and HuMMI$_{repeat}$ new interactions are detected informing about the sampling sensitivity of the one repeat constituting HuMMI$_{main}$. Lastly, the interaction patterns of homologous bacterial effectors are investigated with the help of HuMMI$_{hom}$. Mapping this network is informed by HuMMI$_{main}$ revealing interaction partners of effectors with homologs in HuMEOme_v1. Subsequently, these homologs are tested against the human interaction partners of similar effectors identified in HuMMI$_{main}$.

a

primary screening	→ primary positives
secondary phenotyping	→ secondary positives
sequence identification	→ candidate interaction partners
4-fold independent verification	→ Y2H interaction partners

b

Y2H stage	pairs (#)
primary positives	11,800
secondary positives	5,200
interaction candidates	4,100
Y2H interaction partners	1,263
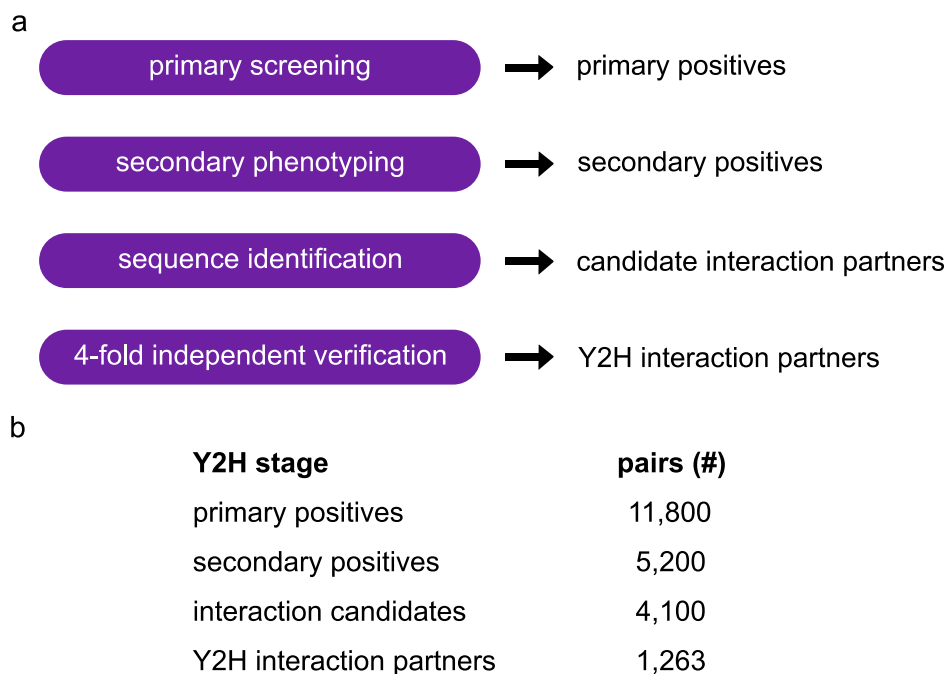
Figure 17 | Y2H pipeline and number of pairs detected at every step. a | *The Y2H pipeline used to generate HuMMI as previously described* [114]. *Effectors and human proteins are mated during the primary screening and yeast spots growing on selection plates are picked as primary positives. During the secondary phenotyping primary positives are grown again on selection plates and growing yeast spots are picked (secondary positives). The secondary positives are identified by sequencing and the interaction candidates are tested one-on-one in a four-fold independent verification step. Growth of the respective yeast clones on selection plates indicates bona fide Y2H interaction partners.* *b |* *The approximate number of candidate interaction pairs for each step of the Y2H pipeline during the mapping of HuMMI.*

Before screening the bacterial effectors against human proteins to generate the human-microbial interactome, constitutive autoactivators were identified by mating the yeast clones containing the bacterial effectors (DB-X yeast clones) against yeast clones with empty prey vector (AD-empty yeast clones). Diploid yeast clones exhibiting growth stronger than Y2H control number 1, which carries no insert, were considered autoactivators [113]. 59 bacterial ORFs accounting for ~ 6% of all cloned effectors exhibited autoactivation and were excluded from the screen. Subsequently, a Y2H mapping pipeline (Fig. 17) which has been used multiple times in our lab, recently to establish a SARS-CoV-2-human contactome [93], was

employed to generate the human-microbial interactome. The first step constitutes the primary screening i.e., haploid yeast clones containing one DB-X construct from the HuMEOme_v1 were screened against a pool of haploid yeast clones containing a combined ~ 188 human ORFs from the human ORFeome v9.1. For HuMMI$_{main}$ this was repeated until the complete 900 bacterial effector ORFs of HuMEOme_v1 were screened against pools containing the combined 17,408 human ORFs of the human ORFeome v9.1. To obtain HuMMI$_{repeat}$ only a subset of each ORFeome was screened consisting of 288 effector proteins and 1475 human proteins. Diploid yeast cells were grown on selection plates lacking histidine selecting for cells that express the reporter gene HIS3 indicating the interaction of the bacterial effector with a human protein. Comparison with the growth phenotype of diploid yeast cells containing the respective DB-X and AD-empty construct on selection plates allowed the identification of spontaneous autoactivators: similar or stronger growth of yeast clones with a DB-X and AD-empty construct compared to growth of yeast clones containing a DB-X and AD-Y construct indicated autoactivation (Fig. 18). While autoactivators were excluded, three colonies per yeast spot growing on selection plates indicating interaction between the bacterial and human ORF (primary positive) were picked.
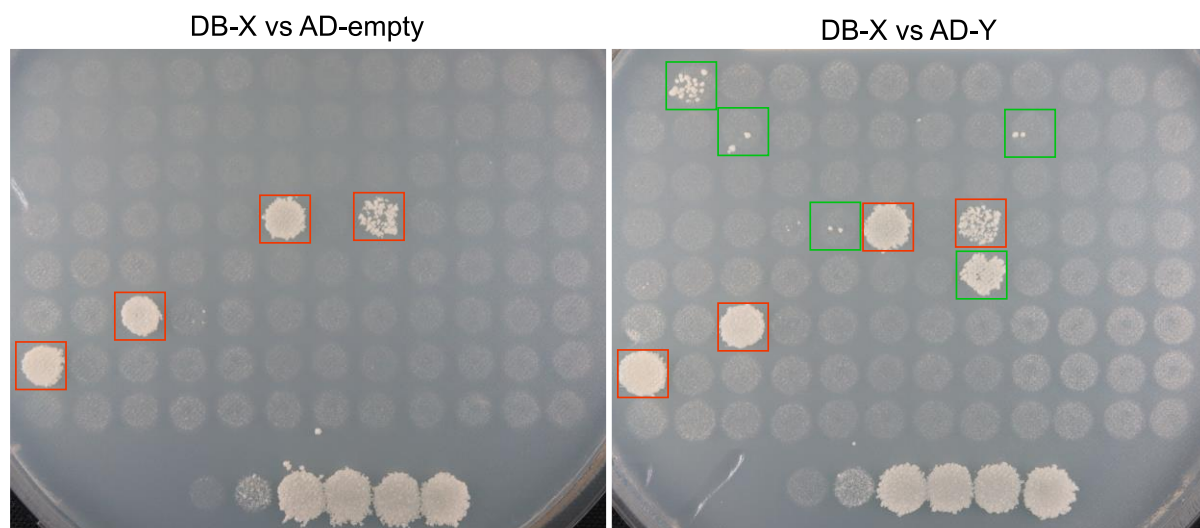
Figure 18 | Selection plates of the primary screening. *The plate on the left shows the yeast growth on selection plates of DB-X yeast clones mated with AD-empty yeast clones indicating autoactivators (red). The plate on the right shows the corresponding DB-X yeast clones mated against a pool of AD-Y yeast clones. Comparison between the two plates allows the identification of primary positives (green). The six Y2H controls are spotted at the bottom of the plate.*

Primary positives were grown again on selection plates during the secondary phenotyping to ensure the robustness of the interaction. Additionally, CHX-containing selection plates were employed to control for spontaneous autoactivators. Comparison of yeast growth phenotypes on selection plates with and without CHX indicated autoactivators and interactions between effectors and human ORFs (secondary positives) as described above. The secondary positives were picked, and the candidate interaction partners were identified by NGS. The last step in the pipeline was performed to verify the candidate interactions (4-fold independent verification). This step was also used to map HuMMI$_{hom}$. Each DB-X/AD-Y yeast clone was obtained from the archival stocks and mated against yeast clones with its candidate interaction partner four separate times. Furthermore, the DB-X yeast clones were mated against AD-empty yeast clones, and AD-Y yeast clones against DB-empty yeast clones. Yeast growth of yeast clones containing a DB-X and AD-Y construct was compared to yeast clones with a DB-

X or AD-Y construct and an empty plasmid to identify autoactivators. Yeast growth indicating an interaction between an effector and human protein was scored using an in-house developed scoring tool based on machine learning assessing the growth strength of each yeast clone. Pairs scoring positive at least three out of the four repeats were considered bona fide Y2H interactors. To confirm interaction partners after the 4-fold independent verification, all interaction partners were sequence-verified once again with the effector ORFs undergoing full-length sequencing. Deviations in eleven effector sequences were detected mainly consisting of changes in single amino acids (Table S3). In total, 1263 interactions were detected between 289 bacterial effectors and 430 human proteins (Table 5 and Fig. 19).

Table 5 | Overview of the number of nodes and edges per network. *Each network and the number of corresponding human proteins and bacterial effectors are listed as well as the number of interactions (edges). HuMMI total encompasses all nodes and edges. As some interactions are detected in more than one Y2H experiment, overlaps between the networks exist. All interactions are listed in Table S4.*

network	human proteins	bacterial effectors	edges
HuMMI$_{main}$	406	255	1071
HuMMI$_{hom}$	165	117	398
HuMMI$_{repeat}$	17	25	39
HuMMI all	430	289	1263
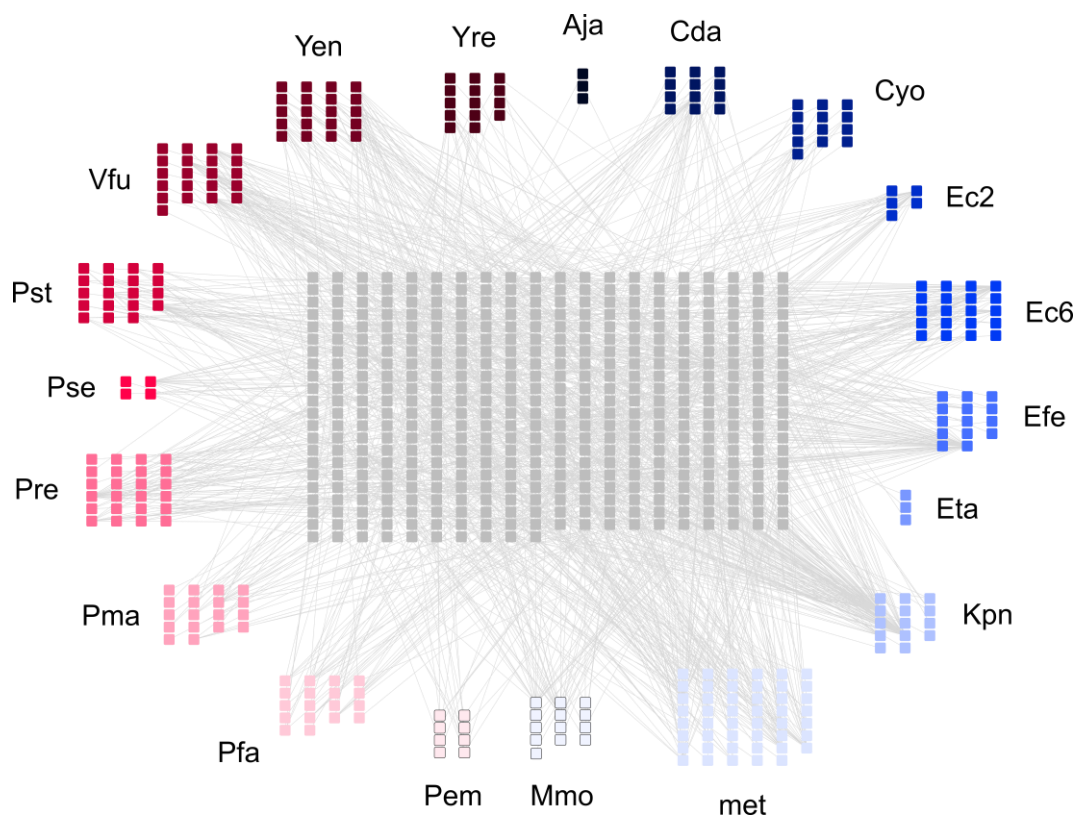
Figure 19 | HuMMI network. *In the combined network of all interactions human proteins are depicted in grey whereas effectors are shown in color according to the respective strain following the color code in Figure 16. Metagenomic effectors are referred to as "met".*

Given the rigorous exclusion of autoactivators and the 4-fold verification retest to verify candidate interactions, HuMMI consists of robust interactions. Whether the network map is a good representation of the interactome still needs to be determined by subjecting the network to a thorough quality control.

2.4 Quality assessment

Before the generated network map can be employed to approach biological questions the quality and coverage of the map need to be assessed i.e., the biophysical quality of the interactions (precision) and the map's comprehensiveness. These parameters can be evaluated by considering the sources of false negatives and false positives during the map generation [42,115].

Comprehensiveness: The sensitivity of an assay employed to map a network and the sampling sensitivity of a screen contribute to the comprehensiveness of the network map. Assay sensitivity is the fraction of interactions that can be detected by a given assay, while sampling sensitivity is the saturation of a screen i.e., the fraction of interactions identified with the repeat(s) performed using a certain assay [42,115].

Assay sensitivity: Limitations in the sensitivity of the assay used to map the network negatively impact the comprehensiveness of the map (false negatives), while an enhanced sensitivity increases the background (false positives) [41]. Assay sensitivity is influenced by the restrictions of the assay to detect any interaction e.g., the Y2H is unlikely to detect interactions involving a protein that requires post-translational modifications that are absent in yeast [112]. Furthermore, the applied experimental conditions influence the sensitivity of the assay, for instance, using high-copy-number plasmids can increase sensitivity but also background [41]. To estimate the assay sensitivity a positive reference set (PRS) and a random reference set (RRS) are employed. The reference sets are used to benchmark the assay i.e., assessing the sensitivity and background of the assay used to map the network [41,112].

Reference sets need to be "appropriate for the class of interactions evaluated" e.g., reference sets including interactions detected in protein complexes might not be scoring positive in binary assays [41]. This is due to the "different nature of protein complexes and binary interactions" as proteins in complexes might not interact directly with each other and therefore, binary assays cannot detect them [41]. In 2009, Venkatesan *et al.* assembled a binary *Homo sapiens* PRS and RRS version 1 for benchmarking assays used to detect human interactome maps [112]. A second-generation *Homo sapiens* PRS and RRS (hsPRS-v2 and hsRRS-v2) has been published in 2019 consisting of "fully sequence-verified, full-length" ORFs based on updates in the literature and annotation databases of binary PPIs [113]. These two updated reference sets have been used to benchmark ten different versions of four binary assays representing suitable standards against which new datasets and reference sets can be compared [113].

Depending on the generated network map, novel reference sets might be required to best reflect the new dataset in terms of the nature of the investigated interactions and the involved interaction partners [112]. For instance, if plant proteins of *Arabidopsis thaliana* are investigated using a Y2H assay, binary interactions between *Arabidopsis thaliana* proteins should be employed for the reference sets. Such interactions can be compiled employing different databases and publications gathering literature-curated interactions. Thereby, the quality of the reported interaction needs to be considered e.g., interactions relying on solitary evidence such as a single method in one publication exhibit lower quality compared to interactions reported by multiple lines of evidence [41].

Sampling sensitivity: Sampling sensitivity is the fraction of identifiable interactions that can be detected in the context of a specific pipeline [112]. This allows estimation of the degree of

saturation that is reached with the mapping experiment and thereby of the comprehensiveness of the map [41]. Sampling sensitivity is assessed by repeat screens e.g., by re-screening a subset of the employed ORFeomes [112]. Extrapolating the number of newly detected interactions with every additional repeat permits the estimation of the degree of saturation achieved by the experiment [114].

Precision: Precision refers to the fraction of true positives in a dataset i.e., true biophysical interactions. The opposite, biophysical false positives are "unidentified artifacts of the assay" i.e., technical false positives [42]. The precision can be assessed by employing an orthogonal assay to validate the detected interactions. Interpreting the resulting retest data requires benchmarking of the orthogonal assay. Assay benchmarking is achieved by employing the PRS and RRS to define the signal-to-noise ratio [42]. Hence, reference sets are tested together with a subset of the new dataset in the orthogonal assay. Subsequently, a scoring threshold is set reflecting a well-thought-out compromise between high sensitivity and low background [42]. In high-quality datasets, the fraction of positive pairs is similar to the fraction scoring positive of the PRS, whereas low-quality datasets show more similarity with the RRS in their fraction of positively scoring pairs [42].

Assembling a microbe-host reference set: To assess the quality of HuMMI, reference sets were gathered to benchmark the Y2H assay and the orthogonal yeast-based nanoluciferase complementation assay (yN2H), which was employed to estimate the precision. For a most suitable PRS, a selection of literature-curated binary PPIs between gut commensal Pseudomonadota T3SS effectors and human proteins was aimed at. As investigating gut commensal T3SS effectors is a completely new undertaking, no studies were available in that regard. Therefore, publications describing PPIs between gut pathogenic Pseudomonadota T3SS effectors and human proteins were analyzed. Studies were collected by programmatically searching the IMEx consortium protein interaction databases using the UniProtKB accession numbers of T3SS effectors. From these publications, information on the interaction partners, the type of assay performed to detect the interaction, employed controls, and protein length (fragment or full-length) were extracted. To ensure high-quality PPIs each paper was assessed by two individuals and only PPIs documented by two methods, or two papers were considered. As the Y2H used to map HuMMI detects binary interactions, only PPIs from binary assays were regarded e.g., Y2H, pulldown assays, protein microarray, and competition assays. This resulted in 67 binary interactions consisting of 29 bacterial effector proteins and 64 human proteins constituting the bacterial human literature binary multiple (bhLit_BM-v1) (Table S5). The bacterial host random reference set (bhRRS-v1) was assembled by randomly picking 100 interaction pairs from all bacterial effector proteins (bhLit_BM-v1 and HuMEOme_v1) and all human proteins (human ORFeome v9.1) (Table S6). In addition, hsPRS-v2 (60 pairs) and hsRRS-v2 (78 pairs) were employed as standard reference sets to benchmark the assay which were kindly provided by the Center for Cancer Systems Biology, Dana-Farber Cancer Institute in Boston (Table S7).

Assay sensitivity was evaluated by pairwise testing the four reference sets according to the 4-fold verification protocol. To that end, ORFs of the bhLit_BM-v1 and the bhRRS-v1 were processed according to the cloning strategy applied to the bacterial effector ORFs. Human ORFs were obtained from the human ORFeome collection v9.1 and verified by sequencing. Non-verified clones were excluded from the experiment as were yeast clones with impaired yeast growth, and autoactivators. Hence, 54 pairs of the bhLit_BM-v1 were tested, 73 of the bhRRS-v1, 60 pairs of the hsPRS-v2, and 78 of the hsRRS-v2. Yeast clones containing the

protein pairs of the PRS and the RRS were alternately distributed on the plates. The human reference sets were tested in both configurations (AD-Y and DB-X) as configuration affects detection rate [116]. To compare the results to the one configuration of the bacterial-host reference sets, the results of the human reference sets were averaged. None of the RRS pairs scored positive, whereas seven unique interactions were detected from the bhLit_BM-v1 and nine and twelve pairs from the hsPRS-v2 in both configurations. This corresponds to 13% pairs scoring positive of the bhLit_BM-v1 and 17.5% pairs scoring positive of the hsPRS-v2 (Fig. 20a). The fact that no interactions of the RRS were detected by the Y2H demonstrates the low background of the assay. Thereby, the Y2H pipeline used for mapping the interactome detected interactions with a very low false positive rate.
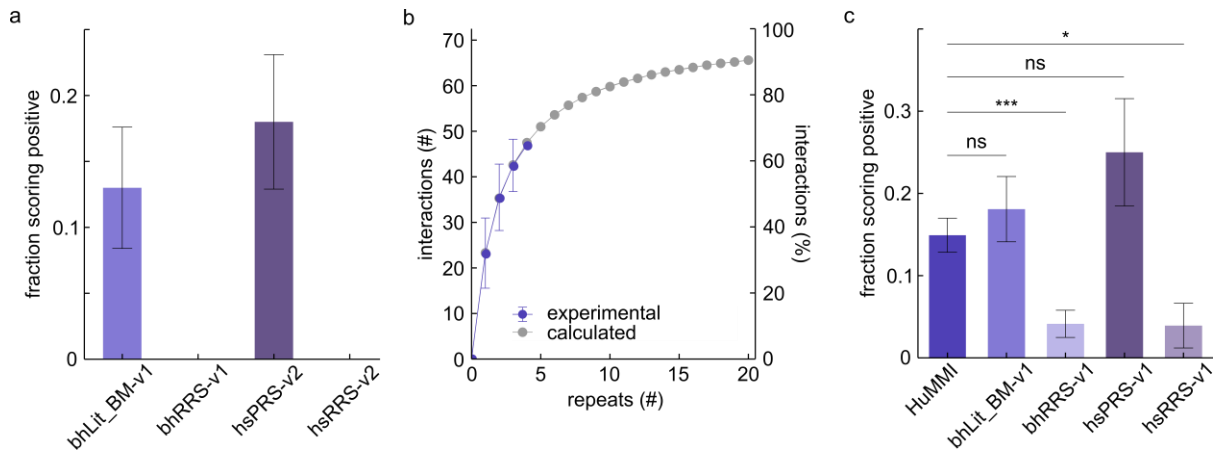
Figure 20 | Quality control of HuMMI. a | Assay sensitivity based on interactions of the reference sets identified using the Y2H pipeline. Details are presented in Tables S5-S7. Error bars represent the standard error (SE) of proportion. *b |* Experimental values stem from the four repeats screening subsets of the two ORFeomes. The sampling sensitivity curve (calculated) was predicted from the experimental values. *c |* Validation rate of a subset of HuMMI and the four reference sets assessed in the yN2H. Tested pairs and pairs scoring positive in Table S8 and Table S9. * P-value = 0.04; *** P-value = 0.0006; ns, "no significant difference" (Fisher's exact test). Error bars present SE of proportion. P-values are shown in Table S10.

To investigate the sampling sensitivity, a subset of the main screen consisting of 288 effector proteins and 1475 human proteins, was re-screened three additional times following the Y2H mapping pipeline. The resulting HuMMI$_{repeat}$ consists of 39 interactions between 17 effectors and 25 human proteins. Sampling sensitivity was calculated as follows according to a previously applied method [114,115]: the number of unique interactions of the four repeats (subset from the main screen and the three additional repeats) was determined, excluding homologs, and a hyperbolic curve was estimated using an adapted Michaelis-Menten equation to determine maximum saturation (Fig. 20b). As the Michaelis-Menten equation describes the asymptotic approach of the maximum reaction velocity with increasing substrate concentrations, it can be adapted to determine the maximum saturation approached by repeat screens [114]. To account for "discontinuities in the accumulation of interactions", the detected interactions during the four repeats were combined in all possible variations [114]. From these, the variables of the Michaelis-Menten equation can be calculated, and by using an adapted function f(x) of the Michaelis-Menten equation with x representing the repeats, a saturation curve can be calculated. As HuMMI was mapped with one screen, the fraction of identifiable interactions detected was 32%. This represents a reasonable fraction of identifiable interactions rendering HuMMI a relevant network map of representative interactions.

The overall sensitivity (S_o) of a screen can be calculated as the product of the assay sensitivity (S_A) and the sampling sensitivity (S_s) [93]. In this study, the S_o is 4.18% demonstrating that HuMMI represents only a fraction of the gut meta-interactome. However, since this is the first dataset of interactions between gut commensal T3SS effectors and human proteins it offers valuable insights into gut commensal effector targets and functions.

Precision was assessed employing the yN2H. Briefly, parts of the bioluminescence NanoLuc protein are fused to the proteins of interest (bait and prey). Upon their interaction, the NanoLuc protein is reconstituted emitting a light signal, which can be detected using a plate reader [116]. 173 interactions were randomly selected from HuMMI (Table S8) and the respective ORFs together with the ORFs of the four reference sets were cloned into suitable vectors and yeast strains. Cloning success was assessed by PCR and unsuccessfully processed ORFs were excluded from the experiment. Yeast clones containing corresponding protein pairs were mated and to control for background signal yeast clones were also mated with yeast clones containing the respective empty plasmid. All protein pairs were tested in two orientations (NanoLuc fragments F1 (1-65 amino acid) and F2 (66-171 amino acid) linked to the N-terminus) as orientation can affect detection rates [116]. Yeast clones containing the protein pairs of the PRS and the RRS were alternately distributed on the plates. The retest rates of the bacterial-host reference sets were generally lower compared to the retest rates of the human reference sets (Fig. 20c and Fig. S1). This might be due to the good detectability of the human-human interactions enabled by their well-researched nature and by their congruency with the eukaryotic assay system putting prokaryotic proteins at a disadvantage. The fraction scoring positive of the HuMMI dataset was statistically indistinguishable from those of the positive reference sets but significantly deviated from those of the RRSs. Thereby, the biophysical quality of the HuMMI interactions is comparable to high-quality, literature-curated interactions.

In conclusion, the Y2H pipeline used to map HuMMI reliably detected 32% of all identifiable interactions with a very low false positive rate. Furthermore, the biophysical quality of the interactions in HuMMI is comparable to those of literature-curated interactions. Therefore, HuMMI represents a reliable dataset that can be employed to approach biological questions.

2.5 Interaction patterns of homologous effectors

As obtaining experimental data is resource-intensive, properties of well-researched proteins are often assumed to be transferrable to homologous proteins of different species. Especially in computational biology, many study designs are based on the link between sequence homology and functional similarity [117]. For instance, protein sequence similarity is frequently used to infer similarity in protein functionality or PPIs [117]. This concept suggests the similarity of PPI partners between homologous T3SS effectors in HuMMI. To test this hypothesis, the relation between sequence-similar T3SS effectors and similarity in interaction partners in HuMMI$_{hom}$ was analyzed.

Effector similarity over 90% sequence length was assessed and effectors sharing ≥ 30% sequence similarity were assigned to the same homology cluster. This resulted in 122 clusters consisting of 2-13 effectors (Table S11). Human interaction partners of homologous effectors were tested against all effectors in the shared homology cluster (Fig. 21). Thereby 1,470 pairs were assessed according to the 4-fold verification protocol. The results are gathered in HuMMI$_{hom}$ which consists of 399 interacting pairs of 117 homologous effectors and 165 human proteins.
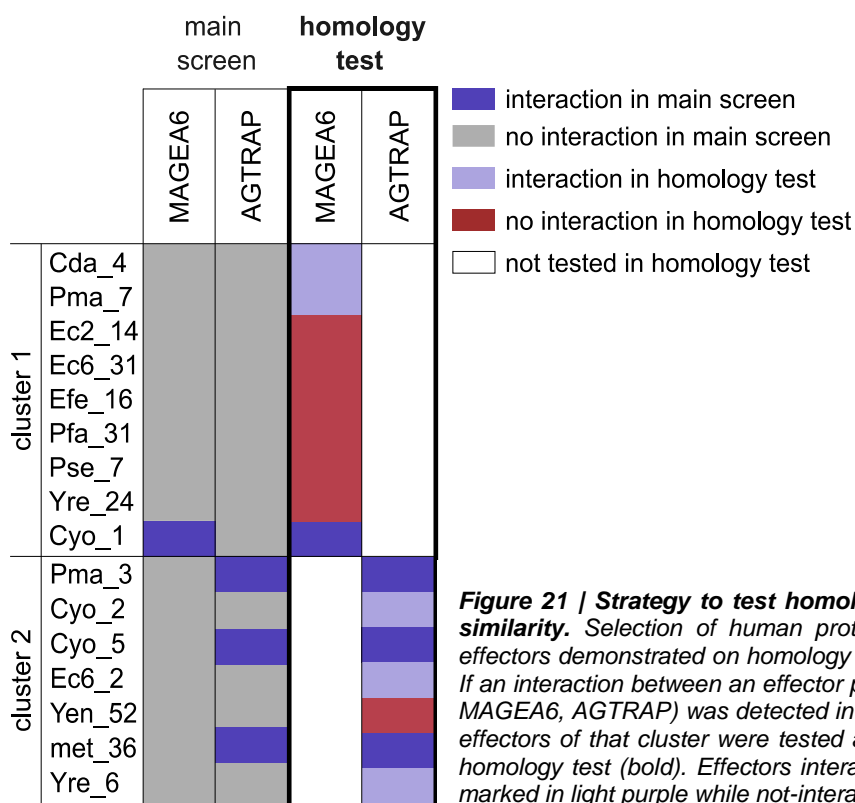
Figure 21 | Strategy to test homologous effectors for interaction similarity. *Selection of human proteins tested against homologous effectors demonstrated on homology cluster 1 and homology cluster 2. If an interaction between an effector protein and a human protein (e.g., MAGEA6, AGTRAP) was detected in the main screen (dark purple), all effectors of that cluster were tested against that human protein in the homology test (bold). Effectors interacting with the human protein are marked in light purple while not-interacting effectors are marked in red.*

To determine whether the sequence similarity of two effectors correlated with their interaction patterns, their sequence similarity over 90% sequence length was plotted against their Jaccard index (Fig. S2). The Jaccard index is used to determine the similarity between two sets of data by calculating the ratio of intersection of the sets (human proteins targeted by both effectors) to their union (all human proteins targeted by either of the two effectors) [118]. To assess the relationship between the sequence similarity of an effector pair and their interaction similarity, the Spearman correlation coefficient ρ was calculated based on the non-linear distribution of both variables. When considering a union of at least three human interaction partners of an

effector-pair to render the Jaccard index meaningful, ρ (0.536) shows a moderate positive correlation. This indicates that, while the two variables are connected, sequence similarity between two proteins does not necessarily indicate similar interaction patterns [119].

Figure 22a exhibits the interaction similarity versus the sequence similarity of effector pairs of clusters with at least three effector proteins and with a union of at least three human proteins. The restrictions are inferred to present meaningful data more concisely. Low sequence similarity (< 60%) correlated with zero shared interaction partners, while a tendency between high sequence similarity and higher interaction similarity was observed. Complete similarity in the interaction pattern (Jaccard index of 1) was only observed in effector-pairs sharing > 87% sequence similarity. Yet, very high sequence similarities i.e., close to 100%, as seen in cluster 3 (bright pink in Fig. 22a) did not always lead to the same interaction profiles. The diverging interaction patterns of the homologous effectors of cluster 3 can be seen in Figure 22b which depicts the areas on Y2H selection plates corresponding to mated yeast containing the respective effector and human protein. Whereas some homologous effectors shared all targets based on the respective yeast growth (Efe_5 and Pma_2), others had only partially overlapping interaction profiles (Cyo_3 and Efe_5) while a few did not share any targets at all (Yen_7 and Cyo_3).
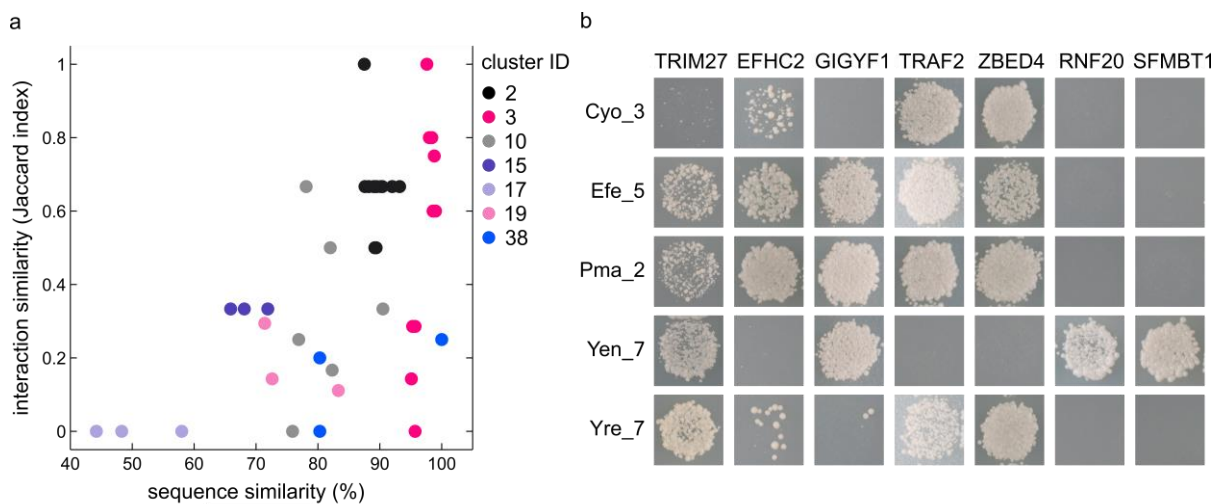
***Figure 22 | Sequence similarity and interaction similarity of homologous effectors. a** | Interaction similarity calculated by Jaccard index and sequence similarity over 90% sequence length for effector-pairs having a union of at least three human interactors and a cluster of at least three effectors. **b** | Y2H selection plates with/without yeast growth indicating whether an effector of homology cluster 3 (bright pink in a) interacted with a human protein.*

In conclusion, although sequence similarity and interaction similarity are connected, PPIs cannot be reliably predicted based on protein similarity of commensal Pseudomonadota T3SS effectors. This is especially true for low sequence similarity, whereas the probability for similar interaction patterns increased with higher sequence similarity. Nevertheless, even highly related effectors do not necessarily share any or all interaction partners. Therefore, the degree of T3SS effector interaction similarity needs to be experimentally determined.

2.6 Effect of T3SS effectors on human PPIs

T3SS effectors mediate their impacts on the host via directly interacting with host proteins. This is often mediated via effector interface mimicry i.e., effectors imitate host protein motifs to facilitate the interaction with host targets [120]. Thereby, effectors can "hijack host protein-protein interfaces" competing for binding sites with host proteins [120]. This prompted the question of whether gut Pseudomonadota T3SS effectors disturb interactions between human host proteins. To investigate this, the effector's impact on human-human PPIs was assessed. To this end, one relevant effector target was selected that interacted with several human proteins which were also targeted by T3SS effectors. REL was selected due to its high connectedness in the human protein interactome HuRI as well as in HuMMI. REL is an NF-κB subunit rendering the protein functionally relevant as several gut commensals were shown to alter NF-κB signaling [51,121].

Figure 23 | Interaction partners of REL. *Human proteins are depicted in grey whereas effectors are shown in color according to the respective strain following the color code in Figure 16. 67 bacterial effectors targeted REL of which 14 also targeted eight of the 190 human interaction partners of REL. Effectors that were tested in the Y3H are circled in green and shown in Figure 24.*

To explore whether the effectors can impede human-human interactions a version of the yeast three-hybrid (Y3H) assay was employed. Originally, the Y3H was developed by Licitra and Liu in 1996 to investigate protein-small molecule interactions [122]. The third party consists of a well-known ligand that binds a known small molecule-binding protein (here LBP) and is covalently attached to a small molecule of interest (here X) [123]. The LBP is fused to the DB domain while potential binding partners of X are fused to the AD domain [123]. Reconstitution of the transcription factor is dependent on the DB-LBP interacting with the ligand-X which needs to bind the AD-Y construct [123]. This method can be used for the analysis of trimeric complexes,

and Y3H modifications are employed to assess the effects that a third component exerts on two interacting proteins [42]. Hence, to evaluate the effect of T3SS effectors on human-human PPIs, a version of the Y3H was employed with which the impact of a third protein on PPIs can be assessed. While the human proteins were fused to either the DB-domain or the AD-domain the effector proteins were added as a third component on an additional plasmid. Yeast growth on selection plates was assessed as a reflection of the effector impact on the human PPIs.

Figure 23 depicts the 67 bacterial effectors in HuMMI targeting REL of which 14 also targeted eight of the 190 human interaction partners of REL. Five effectors were chosen according to two selection criteria: 1) difference in genus and 2) difference in human interaction partners of REL. Initially, the effectors were transformed into yeast strains together with REL fused to the AD-domain. However, the success rates of the co-transformations were low. Thus, a sequential transformation was performed which successfully resulted in yeast strains containing REL as well as one of the selected effectors. Thereby, five different yeast strains carrying REL as AD-Y construct and one of the five selected effectors were generated. These yeast clones were mated against yeast clones containing one of the five interaction partners of both the effectors and REL as DB-X constructs. This was conducted according to the four-fold verification protocol with modified selection plates due to the presence of three plasmids in the mated yeast cells. As a control, yeast clones containing REL were mated with yeast clones expressing one of REL's human interaction partners. Furthermore, yeast clones with the AD-Y or DB-X construct were mated against yeast clones containing the respective empty plasmid to detect autoactivators.
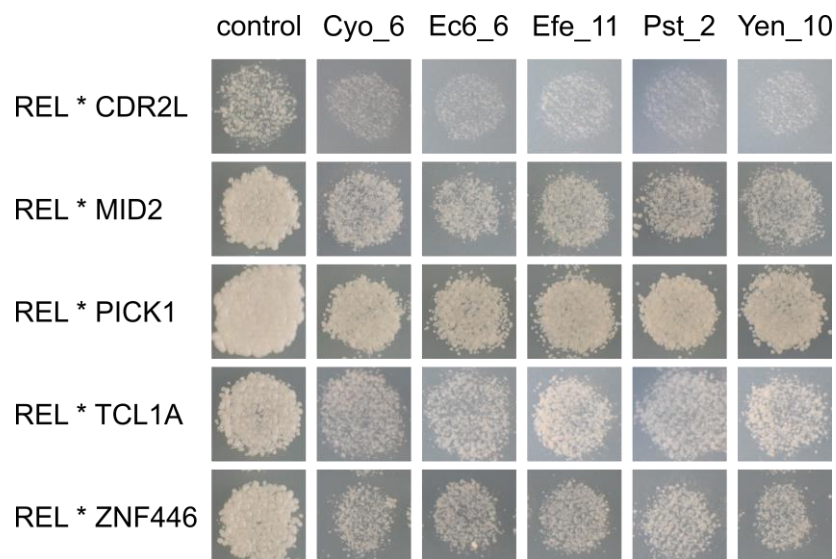
Figure 24 | Yeast growth during the Y3H. *Yeast growth of yeast clones containing REL and one of its human interaction partners is depicted as control. All other columns show the yeast growth in the presence of one of the selected effectors.*

Yeast growth on selection plates was compared between the control yeast clones and the yeast clones containing REL, one of its human interactors, and one of the five effectors. As Figure 24 shows, yeast growth in the presence of an effector is only marginally lighter compared to the controls. This is more likely to be attributed to factors inherent to the experiment than to impacted human PPIs. For instance, the presence of a third plasmid in the yeast clones demands more intracellular resources and potentially decreases yeast growth. Furthermore, differences in plate processing can lead to slight variations in yeast growth on different plates. The fact that all effectors impact the human PPIs in a similar manner i.e.,

slightly decreasing yeast growth, supports the notion that the effectors do not impact the human PPIs as it is unlikely that all effectors affect the human PPIs to a similar degree. In conclusion, the experiment conducted suggests that the investigated gut commensal T3SS effectors do not disturb the interactions between REL and its explored interaction partners. Potentially, the effectors bind different protein interfaces of REL than the human interactors.

2.7 Functional analysis

As most networks consisting of binary PPIs are mapped using an assay system different from the *in vivo* environment of the proteins, biological false positives can occur. These are protein pairs that are capable of interacting biophysically, but due to spatial or temporal separation never interact *in vivo* [42]. For instance, protein pairs might not be expressed in the same cell or never simultaneously. Analyzing the "enrichment in protein pairs annotated with common GO terms" can give a first impression of the functions of the targets and suggest biological relevance [42]. Such a functional enrichment analysis is an over-representation analysis annotating genes with known information on gene functions and detecting significantly enriched terms [124]. Despite inherent biases of the employed database due to incompleteness and biases in the reported studies that inform database records, this approach is useful to assess whether the interactions of a mapped network reflect biological information [42]. Besides the assessment of the biological quality of a network map, functional enrichment analyses enable the formulation of hypotheses concerning the impact of the detected PPIs.

Biologically relevant functions are often exhibited by host proteins that are targeted by several effectors (convergence). This was demonstrated by Weßling *et al.* using an effector-host network map of interactions between *Arabidopsis thaliana* and three of its pathogens [92]. Employing host mutant lines with alterations in convergence proteins and evaluating their disease phenotype upon colonization with the pathogens demonstrated that a higher interspecies convergence on host proteins increased the probability of observing an infection phenotype. Therefore, effector targets subject to interspecies convergence in HuMMI were selected to undergo functional analysis.
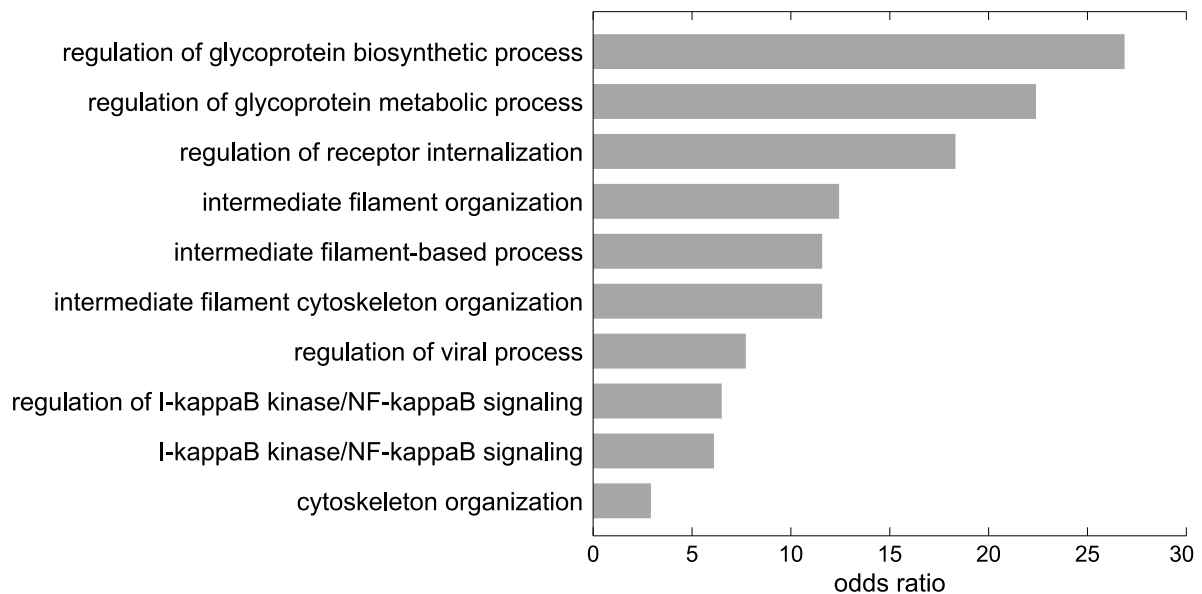
Figure 25 | Significantly enriched functions among the effector targets. *Of the 64 human proteins subject to convergence (Table S12) 60 were present in HuRI which was used as background. Hence, the 60 proteins were annotated using the g:Profiler web server [124]. Annotations were detected using the GO Biological Process database. HuRI [90] was used as background. Benjamini-Hochberg FDR was selected as significance threshold. Odds ratios were calculated with HuRI as background. GO Accession numbers and respective genes are listed in Table S13.*

In HuMMI$_{MAIN}$, 64 host proteins are subject to significant interspecies convergence i.e., that these proteins are targeted by the strains more often than expected by chance (≥ 4 strains

targeting the same human protein). To identify meaningful annotations, the functional enrichment analysis was performed with the reference interactome map of human PPIs, HuRI [90], as background. As only 60 of the 64 human proteins subject to convergence were present in HuRI, the functional enrichment analysis was conducted including these 60 proteins (Table S12). Databases that were queried included GO, KEGG, and Reactome. Significant annotations were obtained from the GO Biological Process database and are shown in Figure 25.

To assess the biological relevance of the interactions in HuMMI, the enriched functions among the effector targets are compared to known impacts of gut Pseudomonadota or gut commensals in general. As T3SS effectors of gut commensals have not been studied so far comparisons to existing literature can only be conducted based on pathogenic T3SS effectors, or gut commensals employing mechanisms different from the T3SS. As gut commensal Pseudomonadota are not classified as pathogens, a comparison to gut commensal functions seems more suitable. Potentially, gut commensal Pseudomonadota mediate similar functions by T3SS effectors. An extensive discussion of the terms enriched among the effector targets including hypotheses concerning the impacts of gut commensal Pseudomonadota effectors on the host in the context of host health is presented in Chapter 3.

The two biological processes most enriched among the effector targets are related terms namely the "regulation of glycoprotein biosynthetic process" and the "regulation of glycoprotein metabolic process". Glycoproteins are present in the gut e.g., as mucin glycans or receptors on host immune cells [125–127]. Gut commensals are known to impact mucus production to strengthen the mucus barrier facilitating the symbiosis between gut commensals and the human host [50]. Furthermore, a gut commensal *Escherichia coli* strain has been shown to stimulate glycoprotein expression on dendritic cells rendering them highly active thereby contributing to colitis [127].

The term "regulation of receptor internalization" refers to the endocytosis of receptors from the plasma membrane to the cytoplasm to downregulate receptor signaling [128]. The proteins annotated with this function can be linked to regulating or interacting with G-protein coupled receptors (GPRs). These are present on various human cells including intestinal epithelial cells where they regulate e.g., innate immune responses or apoptosis, and are linked to intestinal diseases such as IBD [129]. Gut commensals have been shown to impact GPR signaling including receptor internalization via bacterial metabolites potentially impacting gut functions e.g., gut hormone secretion, and signaling processes concerning appetite regulation [64].

Further processes that were significantly targeted by the effectors regard intermediate filament organization. Intermediate filaments are mainly comprised of keratins, and form a dense barrier in intestinal epithelial cells underneath the microvillar brush border to protect the cells from environmental stressors or microbes [130,131]. Manipulation of intermediate filaments has been demonstrated by the gut commensal *Bifidobacterium breve* increasing keratin expression in host cells of the colon potentially promoting host protection against environmental stressors [130].

"Regulation of viral process" was another function enriched among the functions of the effector targets. Gut commensals have been shown to support host defenses during viral infection e.g., by increasing host expression of pro-inflammatory cytokines [53]. In contrast, gut commensals can indirectly contribute to viral infection e.g., by viruses using LPS to induce host immunosuppressive pathways [53].

Furthermore, NF-κB signaling was significantly enriched among the functions of the effector targets. Gut commensals can downregulate the host's immune system via SCFAs suppressing NF-κB signaling to circumvent the transcription of pro-inflammatory molecules thereby enabling the symbiosis with the host [51,121]. In contrast, some gut commensal *Escherichia coli* strains activate NF-κB via microbe-associated molecular patterns such as LPS or flagellin [132–134]. Potentially, NF-κB activation, as part of the innate immune response, is triggered by the bacterial presence in the gut without gut commensals explicitly activating NF-κB signaling.

Lastly, "cytoskeleton organization" was enriched among the functions of the effector targets. A gut commensal *Escherichia coli* strain was shown to disrupt actin filaments decreasing cell stability and thereby negatively impacting the epithelial barrier [135]. Furthermore, a gut commensal *Enterococcus faecalis* strain produces superoxide that impacts microtubules disrupting the mitotic spindle [136].

In conclusion, the investigated commensal Pseudomonadota strains target host proteins involved in functions that are known to be impacted by gut commensals as shown by previous studies. The plausibility of the results of the functional enrichment analysis suggests biological relevance of the dataset offering first insights into the functions of gut commensal T3SS effector targets. This encourages investigations to validate the findings of the functional enrichment analysis and uncover details of the underlying mechanisms.

2.8 Impact of bacterial effectors on apoptosis

An early functional enrichment analysis of proteins subject to convergence in HuMMI, identified the regulation of apoptotic processes as significantly enriched among the effector targets. Apoptosis is known as programmed cell death, a process employed to dispose of specific cells to protect the organism [137]. Disturbances in the apoptotic process are associated with multiple illnesses such as cancer, neurodegenerative diseases, heart diseases, autoimmune diseases as well as bacterial and viral illnesses [137]. As commensal Pseudomonadota are associated with several of these diseases they may manipulate apoptotic processes to influence human illnesses. This hypothesis is supported by research on commensal gut bacteria showing that their impact on host apoptosis affected host illnesses. For instance, moderate apoptosis of gut epithelial cells is mediated by gut commensals with protective effects against colon cancer [138]. Altonsy *et al.* demonstrated that different species have varying capacities of apoptosis induction [138]. They compared commensals such as *Escherichia coli* K-12, the probiotic bacteria *Lactobacillus rhamnosus,* and *Bifidobacterium latis* as well as pathogenic EHEC. EHEC exhibited the most pronounced apoptotic effect in colonic epithelial Caco-2 cells potentially facilitating host infection. Probiotic strains induced cell apoptosis to a weaker degree and *Escherichia coli* showed no signs of apoptosis induction [138]. Notably, a different study observed that some commensal *Escherichia coli* strains increased in abundance in colon cancer patients and could drive metastasis via their inhibitory effects on apoptosis [139].
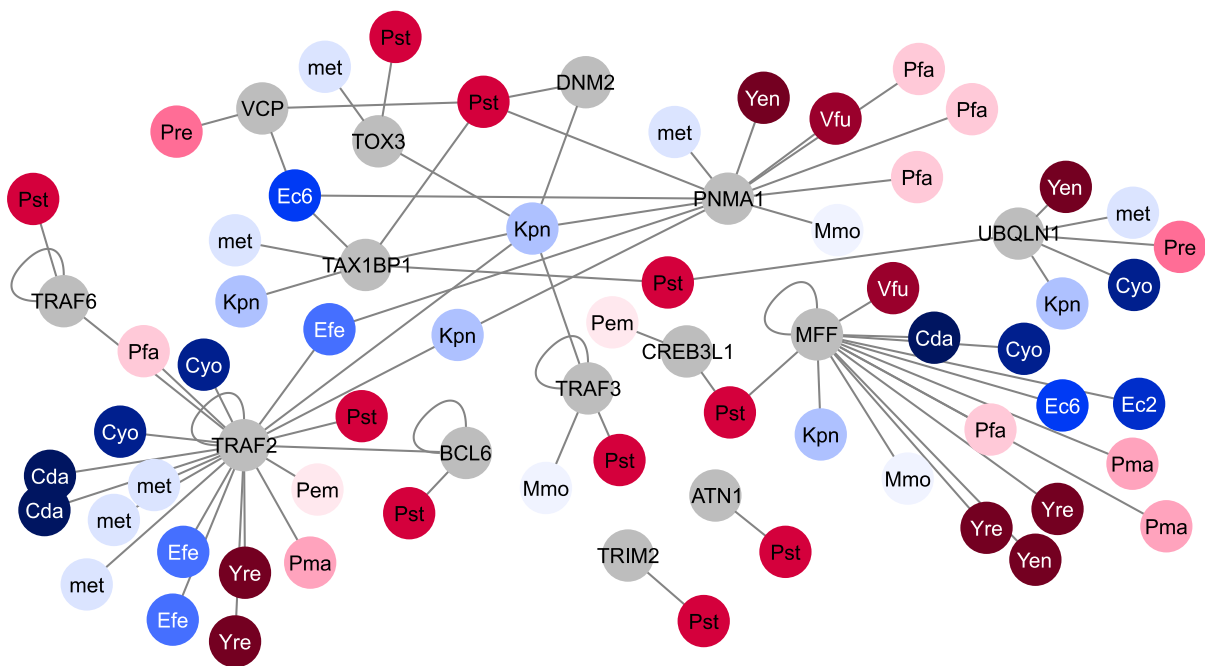
Figure 26 | Apoptosis subnetwork. *Human proteins are depicted in grey, whereas effector proteins are shown in color according to the respective strain following the color code in Figure 16.*

To assess the impact of Pseudomonadota T3SS effectors on apoptosis, human proteins involved in apoptosis were obtained from the GO database [140]. The apoptosis subnetwork was extracted from a merged network of HuMMI and HuRI demonstrating the links between human proteins involved in apoptotic processes with each other and with the bacterial effectors (Fig. 26). As 59 effectors (Table S14) were part of this network a two-step experimental setup was developed to enable the assessment of all effectors concerning their apoptotic effect. First,

effectors were screened for their potential to decrease cell viability and subsequently, were analyzed for their effect on cell apoptosis assessed by the TUNEL assay. While the first approach is more general and can easily be performed on a large number of samples, the latter is more specific and effortful requiring smaller sample sizes. For both experiments, effectors were transfected into HEK293 cells with a transfection efficiency ranging from 18-38% depending on the effector. As under physiological conditions, injection of T3SS effectors is energy-dependent [5], only small amounts of effectors may be present in the human cell. Therefore, the plasmid containing the effector ORF constituted only a small fraction of the DNA content of the transfection mix with the remainder comprising the empty vector and a plasmid containing the green fluorescent protein (GFP) ORF serving as transfection control.

The assay employed to detect cell viability determines the amount of ATP as a signal of metabolically active cells; this is quantified by measuring the light signal of luciferin oxygenated by a luciferase in the presence of ATP. Three controls were included: 1) untransfected cells, 2) cells containing only the empty vector and a plasmid encoding GFP, and 3) cells transfected with two effectors that showed the greatest distance in the network from proteins involved in apoptosis (met_26 and Vfu_17), which was analyzed using the shortest path analysis. As apoptosis is a process consisting of several stages and onset and progression of which depend on the cell cycle phase, impacts of early and late apoptosis onset on cell viability were considered [141]. Therefore, the cell viability assay was performed at three different time points (24h, 48h, and 72h after transfection). The results are presented in Figure 27 showing that no effector exhibited a clear decrease in cell viability at any of the three time points.

Despite no consistent effect of the effectors on cell viability, six effectors (Pst_2, Pst_5, Pst_6, Pst_9, Pst_17, and Pst_14), which showed a tendency to decrease cell viability in some experiments, were analyzed for their effect on cell apoptosis. To this end, the TUNEL assay was employed which detects fragmented DNA, a hallmark of apoptosis, by adding nucleotides labeled with a fluorophore to the free ends of the fragmented DNA. Subsequently, the fluorophore can be detected by a light signal under the microscope. As a positive control, DNase was employed to induce DNA strand breaks, whereas the effector Vfu_12 was used as a negative control due to its great distance in the network from proteins involved in apoptosis (see control 3 for the cell viability assay). Images of the samples were analyzed for an overlap between a GFP signal indicating successful transfection and a signal from the fluorophore indicating DNA-strand breaks (Fig. 28). No noteworthy difference was detected between the negative control and the effectors tested. Therefore, no evidence for apoptosis-induction by the bacterial effectors could be detected in this experimental setup. Nevertheless, this does not completely exclude the possibility of apoptosis induction by the effectors as the assay used here may not have been sensitive enough. DNA fragmentation is one of the last steps of apoptosis and the commensal Pseudomonadota effectors could affect apoptotic signaling at earlier stages. In the study conducted by Altonsy *et al.,* probiotic strains did not induce DNA fragmentation but caused earlier signs of apoptosis such as cytochrome c release from mitochondria as well as the activation of caspases 9 and 3 which are important mediators of cell death [138]. Yet, as Pseudomonadota are in most cases not probiotic strains the results of this study might be more comparable to Altonsy *et al.*'s observation of *Escherichia coli* K-12 which did not show any signs of apoptosis induction [138]. This is supported by the disappearance of the terms regarding the regulation of apoptotic processes from the functional enrichment analysis after the compilation of the entire HuMMI$_{main}$ dataset.

Figure 28 | Apoptosis induction in cells expressing T3SS effectors. *Yellow signal indicates DNA strand breaks. Red borders mark GFP-positive cells indicating successful transfection. Positive control: DNase-treated cells. Negative control: effector absent of the network neighborhoods of human proteins involved in apoptotic processes.*

2.9 Impact of bacterial effectors on NF-κB activity

The functional enrichment analysis revealed that NF-κB signaling was significantly enriched among the functions of the effector targets. To visualize all proteins involved in the interactions the NF-κB subnetwork (Fig. 29) was extracted from a merged network between HuMMI and HuRI. It consists of 97 effectors and ten human proteins. The human proteins involved in NF-κB signaling were obtained from the GO (TRAF6, TRAF2, TRAF1, CARD10, IKBKG) and KEGG (TRAF3) databases as well as from the literature (MID2 [142], CARD9 [143] and TRIM27 [144]). This subnetwork demonstrated a strong convergence onto REL by 67 effectors of 16 strains and the metagenomic effectors. REL is one of the five NF-κB subunits, which form homo- and heterodimers among each other resulting in variations regarding specificity for distinct DNA sequences and regarding the ability to activate gene transcription [145]. Notably, REL was shown to be predominantly expressed in the inflamed mucosa of patients suffering from Crohn's disease (CD) but not in patients with ulcerative colitis (UC) – which are the two forms of IBD. In CD patients, REL expression is enhanced upon early inflammatory signals during disease onset. This leads to REL-mediated high proinflammatory IL-12 expression and T-cell activation furthering inflammation in the gut [146]. Due to REL's involvement in inflammation during CD and as Pseudomonadota are especially enriched in CD patients compared to UC patients [147], the presence of the effectors of this study was investigated in CD patients. To this end, metagenomic data from patients suffering from IBD were analyzed. In metagenomes of UC patients, three effectors were less prevalent compared to healthy controls [7]. In the metagenomes of CD patients, 64 effectors were significantly more prevalent than in healthy controls [7]. Of these 64 effectors, 13 targeted REL. The fact that the REL-targeting effectors from gut Pseudomonadota strains were specifically enriched in CD-, but not UC-patients could suggest a potential link between the T3SS effectors and Crohn's disease via REL. Based on these findings and the significant enrichment of NF-κB regulation among the functions of the effector targets the impact of bacterial effectors on NF-κB activity was investigated.

Besides REL, other human proteins of the subnetwork were also targeted by multiple effectors from different strains and metagenomic effectors just not to the extent of REL. Also, not all human proteins are subject to convergence as they are only targeted by very few effectors. To investigate the impact of bacterial effectors on NF-κB activity, strains were selected according to the two criteria: 1) differences in effector targets among the human proteins of the NF-κB subnetwork 2) changes in abundance to diet, and 3) phylogenetical differences between the selected strains. Hence, five strains (*Citrobacter youngae, Escherichia coli MS 69-1, Escherichia fergusonii, Klebsiella pneumoniae, Providencia stuartii*) and the metagenomic effectors were selected resulting in 26 effectors (Table S15). The potential of the 26 effectors to activate NF-κB or inhibit its activation was researched employing a dual luciferase reporter assay. With this assay, the activity of the transcription factor NF-κB can be measured by the expression of a reporter gene, in this case, the firefly luciferase. In addition, the renilla luciferase is used as a transfection control encoded by a second plasmid. To interpret the luciferase signals of the effectors, NF-κB-inhibiting A20 as well as NF-κB-activating IKKβ were included in the experiments besides the empty vector.

The 26 selected effectors and the controls were transfected into HEK293 cells along with the NF-κB reporter plasmid (encoding for the firefly luciferase) and the plasmid to control for the transient transfection (encoding the renilla luciferase). To investigate NF-κB inhibition by the effectors, the transcription factor was activated by TNF treatment of the cells. Luciferases were

measured by a luminometer equipped with two auto-injectors to first add the firefly substrate to a well in a 96-well plate, measure the firefly luciferase signal, and subsequently add a stop reagent to quench the firefly signal and add the renilla substrate. After measuring the renilla signal, the firefly/renilla (F/R) ratio was calculated to normalize the firefly luciferase signal to the transfection efficiency. The expression of the effectors in HEK293 cells was controlled by Western Blot as can be seen in Figure 30c. To assess significance the Kruskal-Wallis test was performed to test multiple samples which are not normally distributed and correction for multiple testing was conducted using Dunn's test [148]. Five effectors (Efe_12, Kpn_3, Kpn_9, Kpn_10, and met_7) significantly activated NF-κB (Fig. 29, and Fig. 30a). Three of these were present in the aforementioned CD cohort (Efe_12, Kpn_3, and Kpn_9), and one of them (Kpn_3) was significantly enriched compared to healthy controls (Efe_12 exhibits an FDR-corrected P-value of 0.0534 barely missing significance).
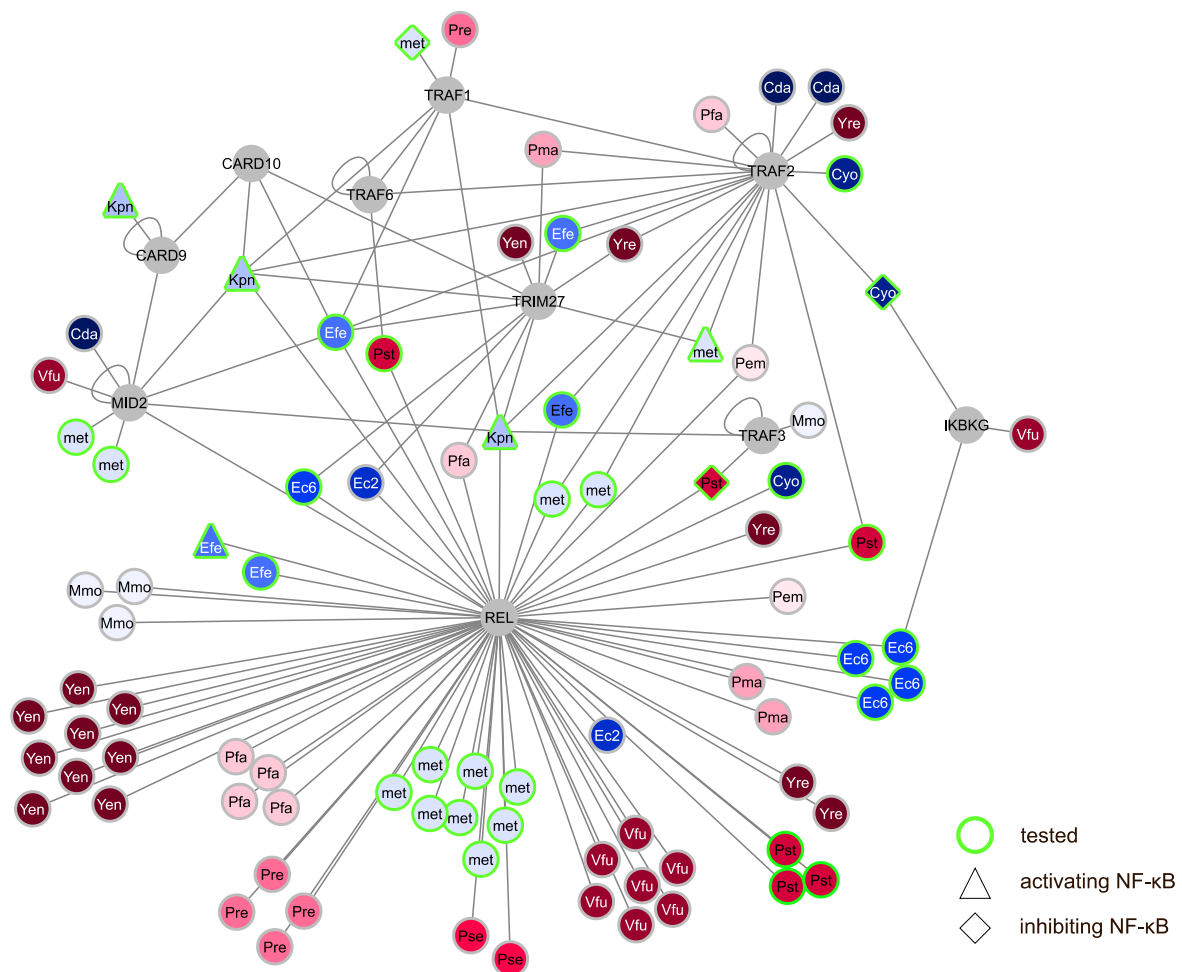
Figure 29 | NF-κB subnetwork. *Human proteins are depicted in grey, whereas effector proteins are shown in red according to the respective strain following the color code in Figure 16. Effectors tested in the NF-κB reporter assay show a green border. Effectors activating NF-κB or potentially inhibiting NF-κB activity are shaped as triangles or diamonds respectively.*
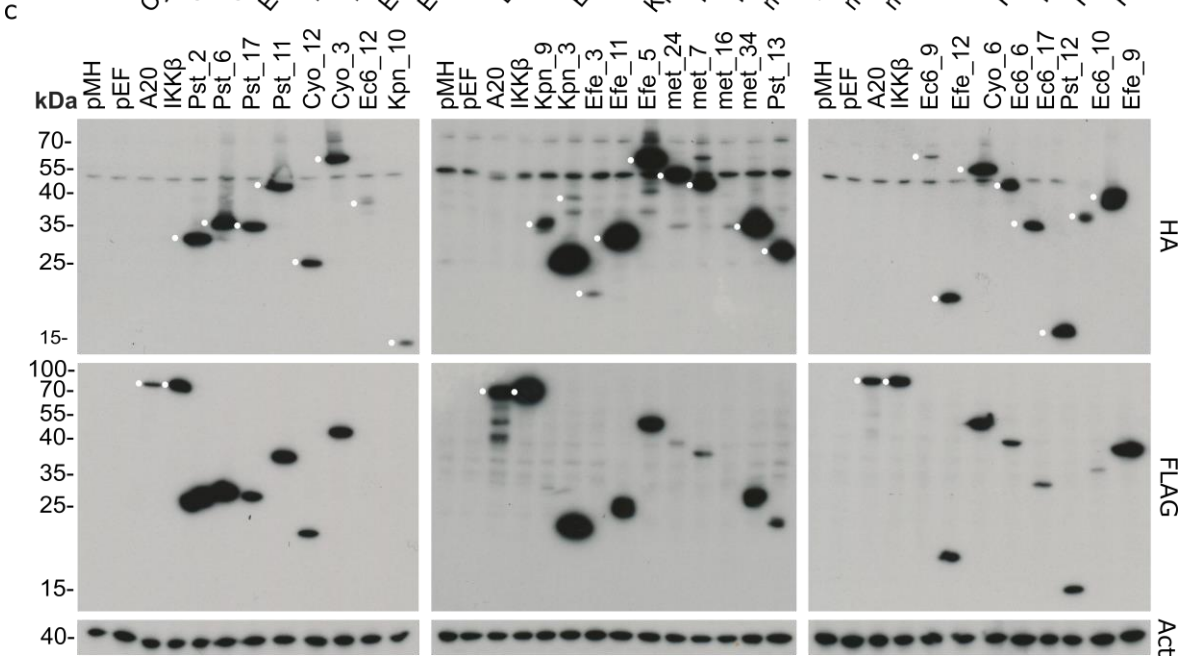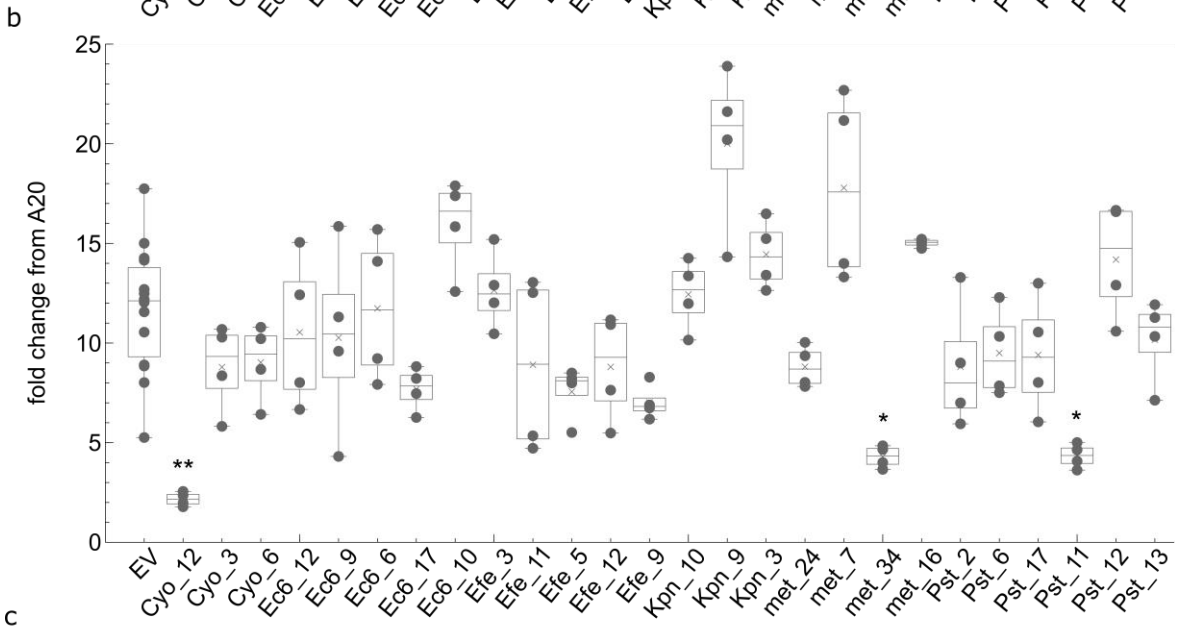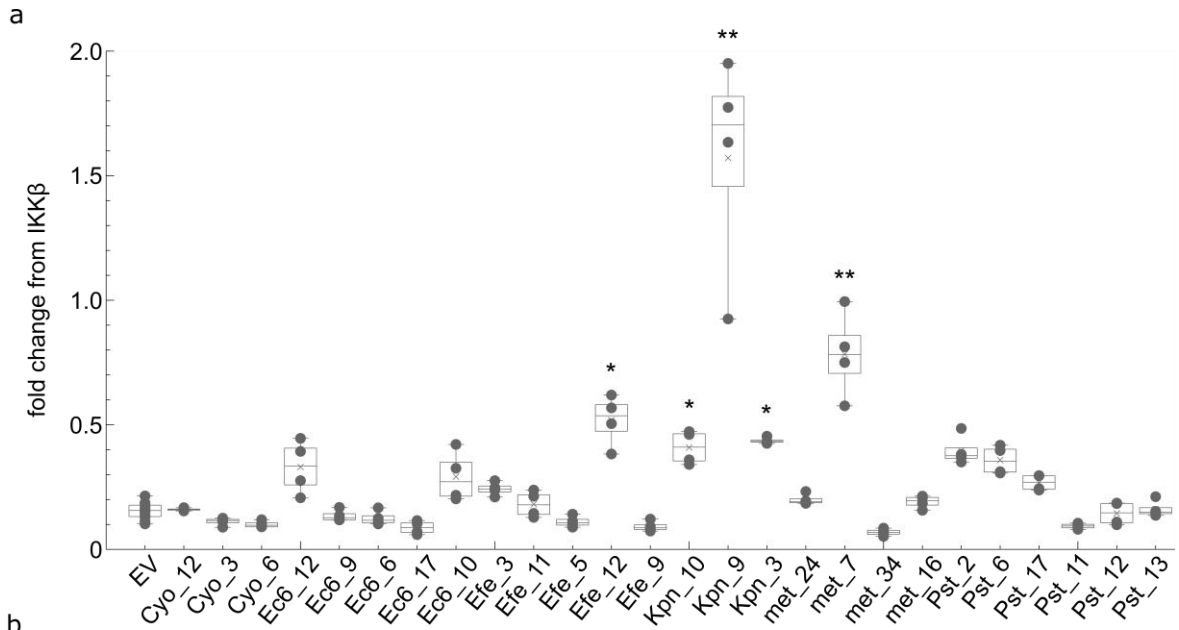
Figure 30 | Effector impact on NF-κB activity. *To be able to compare F/R values from different experiments they were normalized by dividing them through the F/R of the respective controls (A20 for experiments with TNF treatment and IKKβ for untreated cells). A20 and IKKβ were expressed carrying a FLAG-tag, whereas the effectors carried a FLAG and HA tag.* ***a |*** *Fold change from the positive control for each effector as scatter and boxplot (n = 4). Statistically significant differences from the empty vector pMH (n = 14) are indicated with an asterisk (Kruskal-Wallis test with Dunn's correction, * P < 0.05, ** P < 0.01. Details in Table S15.). Boxplots represent the interquartile range (IQR), the black line depicts the mean; whiskers designate the highest and lowest data points within the 1.5 IQR.* ***b |*** *Fold change from the negative control for each effector as scatter and boxplot (n = 4). Statistically significant differences from the empty vector pMH (n = 14) are indicated with an asterisk (Kruskal-Wallis test with Dunn's correction, * P < 0.05, ** P < 0.01. Details in Table S15). Boxplots represent IQR, the black line depicts the mean; whiskers designate the highest and lowest data points within the 1.5 IQR.* ***c |*** *Protein expression by western blot visualized using anti-hemagglutinin (HA) and anti-Flag (FLAG) relative to anti-actin (Act). White dots represent expected bands based on protein size. The fact that the biggest band of Kpn_3 runs lower than expected could be due to e.g., protease degradation or several splice variants* [149]. *pMH, empty pMH-Flag-HA; pEF, empty pEF4.*

The NF-κB-activating effector met_7 was analyzed for a dose-dependent effect on NF-κB activity by transfecting different concentrations ranging from 2-6 µg into HEK293 cells. An increase in NF-κB activity with increased transfected met_7 concentrations could be observed (Fig. 31a). To control for met_7 expression a Western Blot was performed which demonstrated increased protein expression with increased transfection concentrations (Fig. 31b).
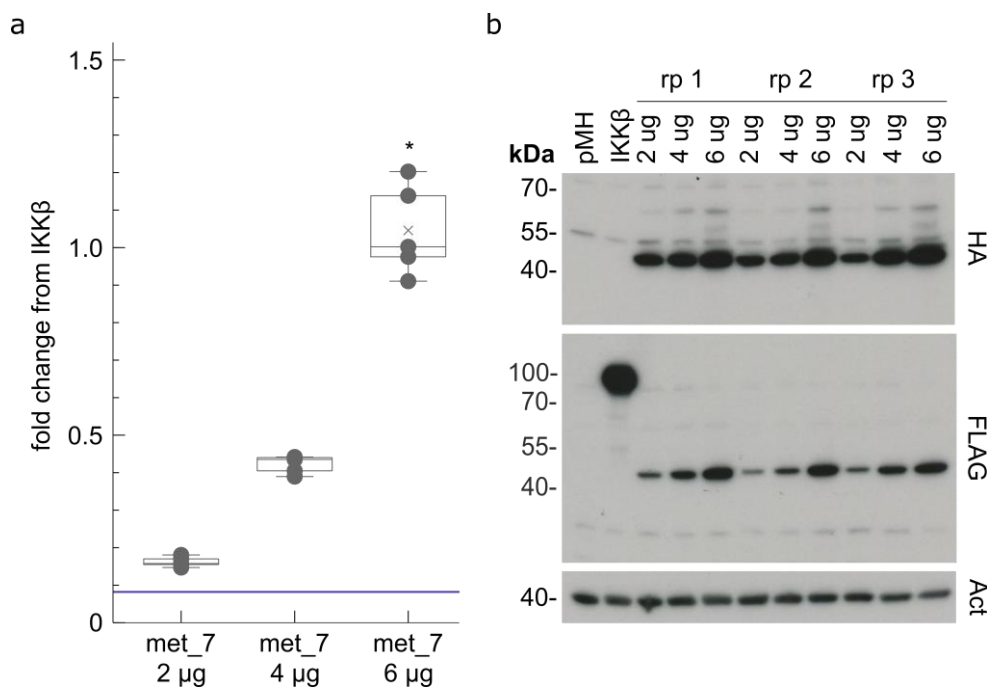
Figure 31 | Dose-dependent effect of met_7 on NF-κB activation. a | *Fold change from IKKβ is depicted for different concentrations. Purple line represents the fold change from IKKβ of the empty vector pMH. Statistically significant differences from the empty vector (n = 1) are indicated with an asterisk (Kruskal-Wallis test with Dunn's correction, * P < 0.05, n = 5. Details in Table S16.). Boxplots represent IQR, the black line depicts the mean; whiskers designate the highest and lowest data points within the 1.5 IQR.* **b |** *IKKβ was expressed carrying a FLAG tag, whereas the effector carried a FLAG and HA tag. Protein expression by western blot was visualized using anti-hemagglutinin (HA) and anti-Flag (FLAG) relative to actin (Act). pMH, empty pMH-Flag-HA.*

At first glance three effectors (Pst_11, Cyo_12, met_34) seemed to inhibit NF-κB activation (Fig. 30b). However, after a closer examination of the individual luciferase values, it became apparent that the firefly values, indicative of NF-κB activity, of the allegedly NF-κB-inhibiting effectors were not lower than the firefly values of the empty vector. Solely through high renilla values, was the F/R ratio indicative of an NF-κB inhibition compared to the empty vector. The high renilla values suggest that the effectors might activate the expression of the renilla luciferase from its herpes simplex virus type 1 (HSV-1) thymidine kinase (TK) promoter.

The increased expression of renilla under the control of a TK promoter has also been reported by others. For instance, Shifera *et al.* noticed high renilla luciferase expression from a TK promoter in HEK293 compared to HeLa cells upon phorbol 12-myristate 13-acetate (PMA) treatment [150]. They showed in subsequent experiments that this was due to the constitutive expression of E1A in HEK293 cells. E1A stems from an adenovirus and was incorporated into the HEK293 genome during the process of immortalization. Shifera *et al.* demonstrated in experiments in HEK293 and HeLa cells that TK activation required the stimulation of the Jun N-terminal Kinase (JNK) and extracellular signal-regulated kinase (ERK) pathways – both induced by PMA – as well as the presence of E1A. The authors suggested a synergistic effect between the activated pathways and E1A, referring to previous studies reporting similar synergism under slightly different conditions [150]. Considering this study, the commensal effectors could potentially activate the JNK or ERK pathway and thereby increase expression from the TK promoter in the presence of E1A. Therefore, the DLA was repeated for the three effectors Cyo_12, met_34, and Pst_11 in HeLa cells. However, high renilla values were observed for all three effectors in the HeLa cells as well suggesting that E1A was not responsible for the observed effect. Therefore, a different promoter controlling the renilla luciferase was tested. The Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) promoter seemed suitable as it controls the expression of GAPDH, a human housekeeping gene that is ubiquitously expressed [151]. Furthermore, this promoter has been used for controlling the expression of the renilla gene in previous studies [151]. However, the promoter change did not alter the high expression of the renilla luciferase gene, neither in HeLa nor in HEK293 cells.

As alterations in the experimental setups showed no improvements, effector impacts on NF-κB activity were assessed by investigating downstream effects of NF-κB. Therefore, collaborators determined the expression of NF-κB controlled human adhesion factor ICAM1 and cytokine secretion from the colon cancer cell line Caco-2 upon effector transfection. ICAM1 is a cell surface glycoprotein that mediates cell adhesion and is involved in leukocyte recruitment to sites of inflammation [152]. The effector met_7 significantly increased ICAM1 expression in Caco-2 cells stimulated with a pro-inflammatory cocktail [7]. Furthermore, met_7 as well as Cyo_12 were capable of altering cytokine release from colonocytes. While Cyo_12 reduced cytokines such as IL-6 and IL-8 in unstimulated cells, met_7 increased IL-1β secretion. Furthermore Cyo_12 reduced cytokine secretion (IL-1β, IL-6, IL-8, IL-18, and IL-23) from Caco-2 cells after proinflammatory stimulation [7].

In conclusion, the reporter assay demonstrated that some effectors are capable of increasing NF-κB activity. Furthermore, effectors can increase or decrease NF-κB-controlled cytokine release from colonic epithelial cells. This suggests that gut commensal Pseudomonadota can up- or downregulate NF-κB activity with potential effects on host inflammation. While effectors might mediate local host immune signaling, they potentially also impact host immune signaling systemically via cytokine secretion. Altering host inflammation might impact host disease risk of e.g., inflammatory diseases such as IBD. Several effectors of the NF-κB subnetwork have been detected in the metagenomes of CD patients, and as some of them induced an increase in NF-κB activity, gut commensal effectors might be able to influence CD risk via the manipulation of NF-κB activity.

3. Discussion

As Pseudomonadota are associated with several complex diseases [3], this thesis aimed to elucidate the role of T3SS effectors expressed by commensal gut Pseudomonadota concerning their impacts on host functions in the context of human health and disease. To this end, the meta-interactome HuMMI (Chapter 3.1) was mapped, which offers information on the functions of the effector targets. This allows the formulation of hypotheses regarding gut commensal effector functions (Chapter 3.2). Subsequently, the impact of gut commensal Pseudomonadota on human health and disease can be assessed (Chapter 3.3). Considering the effects of T3SS effectors on human health raises questions concerning T3SS effector translocation into host cells under physiological conditions (Chapter 3.3). Lastly, the findings of this work contribute to a reassessment of prevalent perspectives on T3SSs (Chapter 3.4).

3.1 Interactions between gut commensal effectors and human proteins

Investigating gut commensal T3SS effectors was approached by assessing the physical interactions between effectors and human proteins. This required the generation of an effector ORFeome (Chapter 3.1.1) which was screened against a library of human proteins to map the first gut meta-interactome map (Chapter 3.1.2).

3.1.1 Generation of a gut microbiome effector ORFeome

With HuMEOme_v1, the first gut microbiome T3SS effector ORFeome was compiled, encompassing diverse effector ORFs from various bacterial strains as well as from gut metagenomes including effectors of strains susceptible to culturing bias. Naturally, HuMEOme_v1 represents only part of the entirety of T3SS effectors in the human gut as not every gut T3SS effectors was identified, and the selection and cloning process inevitably led to a reduction in effector numbers. As described in Chapter 2.1.1, to render the cloning process manageable, the number of strains for PCR amplification was reduced while strain diversity was maintained with the applied selection process. Even more extensive optimization of experimental protocols might have enhanced cloning efficiency. For instance, given that a high genomic GC content ($\geq$ 60%) and large effector ORFs reduced PCR amplification success, additives (such as DMSO and betaine) and a long-range DNA polymerase could have been used [153]. Ultimately, 75% of the T3SS effector ORFs from identified bacterial strains were successfully cloned into entry vectors, while nearly 100% of the metagenomic effectors were processed facilitated by the synthetic synthesis of the ORFs.

Cloning success of previously established ORFeomes ranges between 42% (for uncharacterized ORFs that are not listed in any database) to almost 100% (for ORFs with entries in a database) [154,155]. The cloning success of ORFs depends on e.g., the quality of ORF identification, difficulties during PCR amplification, and optimization efforts [156–158]. Typically, ORFeome completion is progressively attained often resulting in multiple ORFeome versions [90,156,157]. Given the uncharacterized nature of the ORFs in HuMEOme_v1, achieving a 75% cloning success rate for effectors in the entry collection signifies a substantial yield of ORFs.

Hence, HuMEOm_v1 provides a diverse and extensive general resource that is already employed in research projects investigating gut T3SS effector impacts. In this study, the effector

ORFeome was used to map a gut meta-interactome enabling the exploration of gut commensal T3SS effector targets.

3.1.2 A gut meta-interactome map

To investigate gut commensal T3SS effectors, HuMEOme_v1 was employed in a Y2H assay to detect physical interactions between T3SS effectors and human proteins generating the gut meta-interactome map HuMMI. The quality of HuMMI was evaluated by assessing assay sensitivity, sampling sensitivity, and precision as described in Chapter 2.4. Employing the hsPRS-v2 as a standard, assay sensitivity and precision can be compared across different studies demonstrating that HuMMI is on par with first maps [116,114]. The stringency of the employed Y2H pipeline, indicated by the experiments determining assay sensitivity, is necessary for large-scale screening approaches since even a 1% RRS detection rate would lead to a substantial amount of false positive interactions within the dataset [42]. S_o is comparable to initial maps generated with one repeat screen [93,159] but could have been increased with additional repeats. According to the sampling sensitivity curve, one additional screen would have detected 1910 interactions, whereas a third screen would have identified 2305 interactions. Considering the extensive interspecies convergence observed within HuMMI, it is likely that newly identified interactions would have frequently involved already detected human proteins. Lastly, the high biophysical quality of the PPIs in HuMMI was demonstrated by employing an orthogonal assay showing the comparability of the interactions in HuMMI to high-quality, literature-curated interactions.

It is important to keep in mind that no assay can catch all interactions and that different assays share only a small fraction of their discovered interactions [42]. The Y2H is not suitable for reliably detecting interactions involving transmembrane proteins, proteins that are toxic to the yeast cell, or proteins relying on post-translational processing that are lacking in yeast [42,89,90]. Furthermore, false positives can occur in the Y2H by yeast proteins bridging the bait and prey protein [89]. However, this is more likely when testing endogenous yeast proteins with the Y2H compared to heterologous proteins such as bacterial or human ones as done in this study [42]. Despite these limitations, the Y2H assay is one of the most reliable protein interaction detection methods when implementing all necessary controls [42]. This is reflected in the high quality of HuMMI demonstrated by the quality assessment and the plausibility of the results of the functional analysis suggesting biological relevance of the dataset. Subsequent cell-based assays validated aspects of the functional analysis showing that gut commensal T3SS effectors can manipulate NF-κB activation as well as ICAM1 expression, and interleukin secretion from colonocytes.

Comparing HuMMI to other network maps is aggravated by the lack of similar PPI networks. For example, some networks with interactions between T3SS effectors and human proteins were mapped using AP-MS, a technique that identifies both direct and indirect PPIs. This results in datasets that vary in the nature of their interactions compared to HuMMI which contains direct PPIs [160,161]. Other networks included several types of bacterial proteins without a focus on T3SS effectors [162–164]. For example, Yang *et al.* screened 153 virulence-associated proteins of *Yersinia pestis* against human proteins in a Y2H, yet, T3SS effectors constituted less than twelve of these proteins [163]. Commonly, research efforts are limited to one or a few bacterial strains, resulting in the analysis of a smaller subset of T3SS effectors compared to the comprehensive approach employed in the generation of HuMMI [91,165,166]. Furthermore, none of these studies investigated T3SS effectors in commensal bacteria of the human gut microbiome.

Hence, HuMMI represents the first systematic gut meta-interactome map offering insights into gut commensal T3SS effector targets. The extensiveness of the map, encompassing effectors from 18 gut bacterial strains and gut metagenomes, provides a wide-ranging overview of T3SS effector interactions. Thereby, functions of gut commensal T3SS effectors can be assessed for the first time providing insights into gut commensal impacts on human host cells.

3.2 Functions of gut commensal T3SS effectors

The topological characteristics of HuMMI (Chapter 3.2.1) and the functions of the effector targets (Chapter 3.2.2 and Chapter 3.2.3) provide insights into common themes of gut commensal T3SS effectors. For instance, the effector targets were involved in cellular structures as well as cellular movements that rely on those structures (internalization of receptors) (Chapter 3.2.2) [167]. Furthermore, effectors interacted with host proteins participating in host immune signaling (Chapter 3.2.3). Notably, several of the targeted cellular structures and host proteins with immunomodulatory functions show disturbances during CD.

3.2.1 Functional redundancy of gut commensal effectors

Analysis of the topological characteristics of HuMMI revealed a substantial inter- and intraspecies convergence of the T3SS effectors. While convergence can emerge as an artifact of the Y2H system, it is more plausible that the observed convergence is biologically relevant based on three arguments. For one, the quality control demonstrated the low false positive rate in HuMMI and the high biophysical quality of the detected interactions. Furthermore, Weßling *et al.* demonstrated that convergence of effector proteins onto host targets is often biologically meaningful showing a positive correlation between the degree of interspecies convergence and the probability of an infection phenotype [92]. Lastly, convergence of T3SS effectors is a common mechanism of bacteria to ensure functional redundancy and thereby reliability in effector impacts as highlighted in a review by J. E. Galán [71]. Functional redundancy was demonstrated by deleting single effector genes and assessing the effect on bacterial virulence [168]. Deletion of single genes had very little impact on bacterial virulence and in *Pseudomonas syringae* the deletion of 18 effector genes was required to restrict bacterial growth in plants [168]. Functional redundancy is realized in the form of homologous effectors, non-related effectors with different biochemical activities but the same cellular targets, and effectors targeting the same host process via different host proteins [71]. Similarities to the functional redundancy of T3SS effectors of pathogenic Pseudomonadota are potentially also found in commensal gut bacteria of that phylum.

Examples of functional redundancy in T3SS effectors have been demonstrated for several Pseudomonadota. For instance, homologous effectors SopE and SopE2 of *Salmonella typhimurium* exhibit the same function and target the same host protein Cdc42 [169]. Yet, the two effectors do not share all interaction partners as SopE interacts with the protein of Rac1, which is not targeted by SopE2 [169]. Functional redundancy can also be mediated via non-homologous effectors targeting the same host proteins. For instance, three effectors of *Yersinia* spp. (YopE, YopT, YpkA) with different biochemical activities interacted with the same host Rho GTPases leading to the same consequences: YopE and YopT interact with the proteins of RhoA, Rac1, and CDC42, while YpkA binds to the protein of Rac1 only. Targeting these Rho GTPases leads to their inhibition preventing phagocytosis of *Yersinia* spp. by the host [71,170]. Besides convergence onto host proteins, functional redundancy can also be conveyed by impacting the same host processes via diverse host proteins using different effectors. For instance, *Shigella flexneri* inhibits host NF-κB signaling by targeting two proteins of the pathway to evade the host's innate immune system: while one effector prevents the degradation of the inhibitor of NF-κB (IκB) whereby NF-κB remains inactive, another effector prevents chromatin remodeling inhibiting the expression of NF-κB-controlled genes [67].

Signs of functional redundancy have also been observed in HuMMI: 15 out of the 18 strains demonstrated significant intraspecies convergence onto the same host proteins [7]. This intraspecies convergence does not automatically imply the same functional outcome yet based on findings of T3SS effectors in the literature, the commensal Pseudomonadota effectors might mediate functional redundancy in several instances. The observed intraspecies convergence was in the majority of cases not due to effector homology. Only ~ 11% of events in which a human protein was targeted by two effectors of the same strain were mediated by homologous effectors. Furthermore, only three homologous effector-pairs of the 19 homologous effector-pairs from the same strain assessed during the homology test, showed similar, yet not identical interaction patterns. Therefore, the strains investigated in this study seem to mainly employ non-related effectors to mediate convergence onto the same host proteins. This suggests that gut commensal Pseudomonadota potentially evolved effectors independently to target the same host proteins. Besides convergence onto human proteins, convergence onto the same host processes was also observed as the NF-κB subnetwork was heavily targeted by several effectors per strain. For instance, *Klebsiella pneumoniae* interacted with seven of the ten human proteins in the subnetwork via its effectors. Two of these effectors activated NF-κB via different host proteins demonstrating functional redundancy of the effectors.

While intraspecies convergence is employed to ensure functional redundancy and with it, a robustly working system for a bacterial cell, in microbiotas, interspecies convergence could provide functional redundancy for a community of bacteria. Comparable metabolic activities are often mediated by different commensals of the healthy gut microbiota within and across individuals providing a functional core of metabolic activities [55]. Similarly, T3SS effectors of different gut commensal Pseudomonadota strains could convey functional redundancy by targeting the same host proteins and processes. In HuMMI$_{MAIN}$, 64 host proteins are subject to significant interspecies convergence participating in > 500 of the detected 1071 interactions. While convergence does not always suggest similarity in effector functions, a significant overlap in the functional outcomes of effector interactions is anticipated. Given that gut commensals often share similar functions as part of a functional core and since the Pseudomonadota strains in this study exhibited high interspecies convergence, other gut Pseudomonadota strains that were not included in this study might target similar host proteins and processes.

A literature search to assess the interspecies convergence of T3SS effectors identified mostly homologous effectors as mediators of interspecies convergence. For instance, the VirA effector of *Shigella* spp. and the EspG effector of EPEC both target tubulin, which disassembles upon interaction with the effectors to facilitate invasion of (*Shigella* spp.) or attachment to (EPEC) the host cell [67]. An early version of the literature-curated bhLit_BM-v1 consists of 92 interactions, of which 22 are mediated by eleven human proteins that are each targeted by two effectors of two different species. Assessing the effector pairs revealed only one non-homologous pair, whereas the remaining ten pairs displayed sequence homology. The non-homologous effectors were EspF variants of EPEC/EHEC and *Shigella* effector IpaB targeting the human MAD2L2 protein [171]. In contrast, only 8% of the > 500 interactions mediated by the 64 human proteins in HuMMI$_{main}$ that are subject to convergence are mediated by homologous effectors. The observation based on previous publications that interspecies convergence is often mediated via homologous effectors might be biased by the limited number of studies investigating interspecies convergence of T3SS effectors and the fact that homology is often used to detect effectors in a different species. Thereby, homologous effectors are more often identified than non-homologous effectors introducing bias into the available data. Furthermore, gut

commensals might differ from human pathogens in the way they mediate functional redundancy by employing more non-homologous effectors.

Although instances of non-homologous effectors from different bacterial species interacting with the same host protein are rarely reported in the literature, it is common to find non-homologous effectors targeting the same host process. For instance, both *Shigella flexneri* and *Salmonella enterica* employ different effectors to inhibit NF-κB activation [67]. This is achieved by preventing the ubiquitination (*Shigella*) or phosphorylation (*Salmonella*) of IκB – two mechanisms that are both required to degrade IκB and thereby release NF-κB into the nucleus [67]. Other host processes subject to interspecies convergence are the MAPK pathway, vesicular trafficking, cell-death pathways, ubiquitination, and several mechanisms to evade the innate and adaptive immune system [67,71,172]. This phenomenon is also reflected in HuMMI, where the NF-κB subnetwork is subject to interspecies convergence as it is heavily targeted by several strains employing various effectors. Furthermore, activation of NF-κB by the five effectors Efe_12, Kpn_3, Kpn_9, Kpn_10, and met_7 was potentially mediated via different host proteins. For instance, while Efe_12 bound the protein of REL, Kpn_3 targeted the proteins of CARD10, MID2, REL, TRAF1, TRAF2, and TRIM27, and Kpn_10 interacted with the protein of CARD9. The effector met_7 interacted with the proteins of TRAF2 and TRIM27.

While homologous effectors can mediate functional redundancy, sequence homology between different effectors does not necessarily indicate that the effectors are contributing to convergence or functional redundancy. The homology test conducted in this study showed only a moderate correlation between effector homology and interaction similarity. One finding that is noteworthy in this context is that T3SS effectors comprise several domains that confer distinct functions rendering effectors a collection of "individual modules" with "different, unrelated functions" [172]. Therefore, effectors might share homology only in some of their domains rendering an effector homology analysis with a focus on domains more meaningful compared to the 90% effector sequence length that was considered in this study. This could result in a higher correlation between sequence similarity and interaction similarity. However, while some effectors encode for homologous motifs, they might not target the same host protein as they differ in their organelle-targeting function e.g., one effector may target endosomes, whereas its homolog remains in the cytoplasm [172]. This explains Galàn *et al.*'s observation that identical biochemical activities of homologous effectors can be directed toward different host targets [71]. Additionally, small changes such as point mutations can alter the organelle-targeting function suggesting that even highly related effectors might differ in their behavior [172]. Therefore, the effector sequence organization is important to understand in order to comprehend effector behavior. To this end, more research is needed to determine the organization of effector sequences to perform the most meaningful homology analysis. For now, comparing sequence similarity based on most of the effector sequence length is appropriate as it includes various modules that could influence effector behavior.

The moderate nature of the detected correlation between effector homology and interaction similarity in this study together with previous findings showing that point mutations can alter effector function [172], suggests deliberation when extrapolating findings to homologous effectors. Nevertheless, the findings of this study can inform hypotheses formulation during effector research in microbiotas of other human body sites. As 26% of the effectors investigated in this study were detected in skin microbiome samples, effectors of skin bacteria might target the same host proteins or proteins involved in similar host functions [7]. Hypotheses about the effector

impact of skin bacteria on the human host can be deducted from effector homology with gut bacteria but require experimental validation of the respective PPIs and their consequences.

Hence, the observed significant intra- and interspecies convergence of gut commensal T3SS effectors on host proteins and processes, suggests a functional redundancy among gut commensal T3SS effectors as is commonly reported for pathogenic T3SS effectors. Based on Weßling *et al.*'s findings, the host proteins and host processes subject to convergence can be expected to give insights into host pathways that are relevant to gut commensal bacteria.

3.2.2 Modulation of cellular structures by commensal effectors

Based on effector convergence, the functions of the effector targets were analyzed. Human proteins were involved in cellular structures whose functions can be linked to host disease phenotypes as well as bacterial lifestyles. For plausibility, only human proteins expressed in the gut were included in the discussion of effector impacts. The functional annotations of the host proteins were interpreted considering existing literature. As mentioned previously, T3SS effectors of gut commensals have not been studied so far which allows comparisons of the functional analysis only to findings of pathogenic effectors and to functions of gut commensals not mediated by effectors.
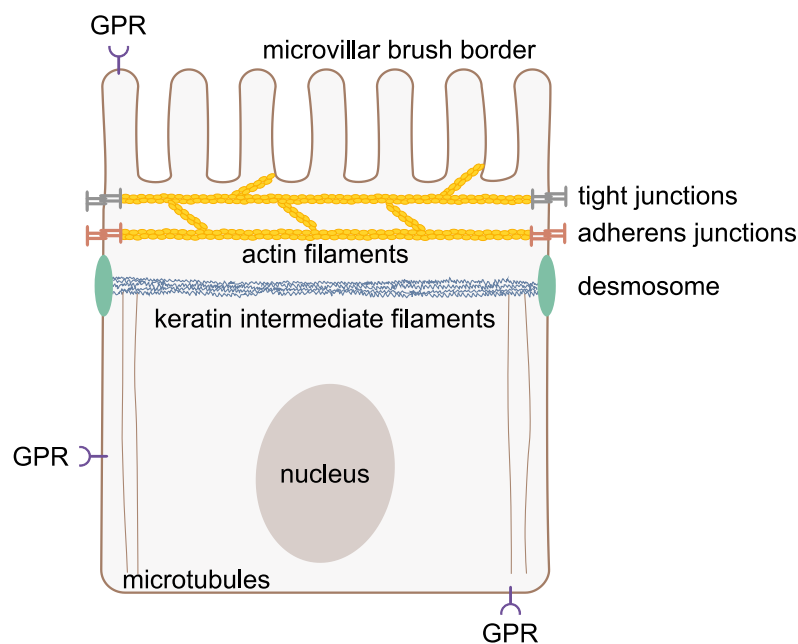
Figure 32 | Structures of intestinal epithelial cells. *Tight junctions and adherens junctions bind to actin filaments in epithelial cells and connect the cells to form the epithelial barrier [173]. Microtubules are linked to intermediate filaments in intestinal cells [174]. Intermediate filaments in intestinal epithelial cells protect against environmental stressors [130]. GPRs are expressed on intestinal epithelial cells to receive various signals from the environment [129]. Adapted from Lechuga and Ivanov [173], Coch and Leube [174], Geisler and Leube [130], and Feng et al. [129].*

One host function subject to convergence was the organization of the cytoskeleton including actin filaments, microtubules, and intermediate filaments, which were all targeted by the effectors.
Actin filaments influence the integrity of the epithelial barrier as they are connected to tight junctions and adherens junctions that join epithelial cells into a tight barrier to protect the organism from the environment (Fig. 32) [173]. Changes in the actin cytoskeleton affect the assembly of the junctions with consequences for the epithelial barrier integrity potentially

resulting in increased barrier permeability [173]. Such disturbances in the intestinal barrier can alter immune reactions in the intestine increasing local inflammation. This, in turn, can promote several diseases such as intestinal inflammatory illnesses, autoimmune diseases as well as metabolic diseases e.g., diabetes and obesity [175]. Gut commensal T3SS effectors converged onto TRIM27 which, among other functions, is involved in the formation of new actin filaments [140]. Additionally, PICK1 was targeted, which negatively regulates actin formation [140]. By interacting with TRIM27 and/or PICK1 bacterial effectors potentially alter actin filament formation and actin dynamics resulting in altered mechanical forces affecting actin-coupled junctions [173]. This could impact epithelial barrier integrity and thereby influence host inflammation and disease risk representing a potential underlying mechanism of the association between gut Pseudomonadota and inflammation as well as metabolic diseases.

Besides influencing the epithelial barrier, actin filaments impact cellular stability. Assessed by electron microscopy, a commensal gut *Escherichia coli* K-12 strain was capable of disrupting actin filaments in Caco-2 cells after prolonged bacterial exposure [135]. Disorganized actin filaments impact the mechanical stability of gut epithelial cells resembling the phenotype of CD characterized by epithelial erosions and crumbling of epithelial villous architecture [135]. Of the 64 effectors investigated in this study that were significantly enriched in metagenomes of CD patients compared to healthy controls, ten effectors (Ec6_9, Ec6_10, Efe_5, Efe_11, Kpn_3, met_18, met_29, Pma_2, Yen_7, Yre_7) targeted either TRIM27 or PICK1, and Efe_11 was targeting both. Potentially, these effectors could impact actin filaments via either of those two proteins contributing to CD manifestation.

Microtubule cytoskeleton (Fig. 32) organization especially during cell division was a function targeted by gut commensal T3SS effectors. Targeted host proteins included SPAG5 involved in "spindle organization" and "establishment of spindle orientation", CCDC102B annotated with "centriole-centriole cohesion", and DYNLT1 implicated in the "establishment of mitotic spindle orientation" [140]. Dalton and Yang described a connection between spindle disturbances and colorectal cancer: colorectal tumors very often demonstrate chromosomal instability i.e., an "unequal segregation of chromosomes" [176]. Aberrations in spindle architecture and spindle dynamics contribute to such chromosomal instability, and spindle disruption can diminish the ability of cells to execute mitotic arrest [176]. A gut commensal *Enterococcus faecalis* strain produces superoxide disrupting the mitotic spindle and promoting chromosome instability [136]. Hence, disrupted spindle organization might be an underlying mechanism of the association between gut commensal Pseudomonadota and colorectal cancer [3] mediated by T3SS effectors.

Intermediate filament organization together with related terms was a significantly targeted function by the investigated gut commensal T3SS effectors in this study. In intestinal epithelial cells, intermediate filaments are located underneath the microvillar brush border to protect the cells from environmental stressors and microbes (Fig. 32) [130]. Keratins represent a huge fraction of intermediate proteins and are involved in the maintenance of the intestinal barrier, especially Keratin 8 and Keratin 18 [131]. Intermediate filaments are divided into six classes (type I, II, II, IV, V, and VI) and keratin members of the same type have highly similar rod domains [177]. Keratin 8 is a member of the keratin type I class, whereas Keratin 18 belongs to the type II group [177]. The keratins subject to convergence that were targeted by the gut commensal T3SS effectors were KRT31, KRT75, KRT76, and KRT27. These keratins are mostly expressed in the skin and hair follicles [178]. KRT27 and KRT31 are members of the type I class whereas KRT75 and KRT76 belong to the type II group [177]. Based on the high domain-similarity shared by keratins of the

same type, effectors might be able to interact with Keratin 8 and Keratin 18 depending on the specific interfaces.

Besides maintaining the intestinal barrier, keratins are involved in the ion transport of colonocytes, their proliferation, and their inflammatory signaling [131]. Studies have shown that gut bacteria are able to impact intermediate filaments: while pathogenic Pseudomonadota use T3SS effectors to disintegrate intermediate filaments to facilitate host colonization, commensal bacteria such as *Bifidobacterium breve* can increase keratin expression in host cells of the colon [130]. Alterations in intermediate filament components have been negatively associated with host health: Keratin 8-knock-out mice exhibited increased barrier defects, inflammation, and tumorigenesis as well as increased levels of pro-inflammatory IL-18 [130]. The mice exhibited a colitis-like phenotype including hyperplastic lesions and damaged colonic epithelium indicating a disease-associated role of keratins [131]. Given the association between defects in epithelial keratins and diseases such as cancer and colitis as well as altered cytokine signaling, impacting keratins could be a potential mechanism underlying the association between Pseudomonadota and human diseases [130]. Further studies are necessary to demonstrate the impacts of Pseudomonadota on intermediate filaments and unravel whether these are beneficial or harmful or require additional factors for the manifestation of a certain phenotype.

Lastly, human proteins involved in the "regulation of receptor internalization" were significantly targeted by gut commensal T3SS effectors. Receptor internalization "serves as a mechanism to downregulate receptor signaling" [128]. In this study, bacterial effectors converged onto PICK1 which can positively regulate receptor internalization and has been identified to interact with GPRs [178,179]. Additionally, the effectors targeted UBQLN2 which is annotated with a "negative regulation of G protein-coupled receptor internalization" [140]. In the human gut, intestinal epithelial cells express GPRs that sense the luminal contents e.g., amino acids, sugars, artificial sweeteners, and bacterial metabolites (Fig. 32) [180]. This information impacts various facets of intestinal functions, such as secretion, motility, and the absorption of nutrients [180]. A study published in 2019 demonstrated that bacterial metabolites of gut commensals can impact GPR signaling including receptor internalization [64]. Potentially, T3SS effectors of gut Pseudomonadota affect the perception of gut luminal contents by modulating GPR expression on intestinal epithelial cells.

GPRs are also involved in the regulation of innate immune responses such as cytokine secretion via the MAPK and NF-κB pathways [129]. Furthermore, intestinal GPRs play a role in apoptosis and the maintenance of epithelial barrier functions and are implicated in intestinal diseases such as IBD [129]. Studies demonstrated that in mouse IBD models and human IBD patients, GPRs have been upregulated in intestinal mucosal tissues [129]. Possibly, modulation of GPR receptor internalization evoked by T3SS effectors can affect host disease risk. While receptor internalization could offer protection against intestinal autoimmune diseases in an inflamed gut, inhibition of receptor internalization by T3SS effectors could increase the risk of intestinal inflammation and diseases.

In conclusion, several of the targeted host functions are relevant in disease phenotypes and have also been shown to be manipulated by gut bacteria. It is therefore plausible, that gut commensal T3SS effectors target the described host functions to modulate cellular structures thereby influencing disease risk. The subsequent effect of this modulation on the host cell needs to be determined in future studies.

3.2.3 Modulation of immune signaling by commensal effectors

Apart from targeting host proteins involved in cellular structures, T3SS effectors converged onto host targets that are part of the host's immune signaling regulating viral processes and glycoprotein processing as well as the NF-κB signaling.

The regulation of viral processes can be supported by gut commensals e.g., via increasing host expression of pro-inflammatory cytokines as well as by impacting the production of reactive oxygen species and defensins [53]. In contrast, gut bacteria are also associated with increased viral survival and host colonization, however, this rarely happens with active bacterial involvement and rather through viral exploitation of bacterial cells [53]. Gut commensal T3SS effector targeted TRIM27, MID2, and TNIP1, which are implicated in the negative regulation of viral transcription, viral entry into host cells, and viral genome replication respectively [140]. Furthermore, effectors converged onto DYNLT1 which binds viral proteins and delivers them to the host nucleus for viral protein production [178]. Lastly, effectors interacted with VPS37B, which is involved in the positive regulation of viral budding [140]. VPS37B is part of the endosomal sorting complexes required for transport (ESCRT) machinery which regulates vesicular trafficking e.g., the fusion of multivesicular bodies with lysosomes or with the plasma membrane releasing macrovesicles from the cell [178,181]. Viruses hijack the ESCRT machinery to overcome the membrane barrier and be released from the cell for further dissemination [181]. Gut commensal T3SS effectors could impact the ESCRT via VPS37B restricting viruses to the cell and diminishing the dissemination rate. Furthermore, effectors potentially target DYNLT1 decreasing the trafficking of viral proteins to the nucleus thereby reducing viral replication. As Pseudomonadota abundance is associated with viral infections such as hepatitis E infection [182], it is plausible that effectors impact TRIM27, MID2, and TNIP1 diminishing their ability to inhibit viral replication and host entry. More research is needed to determine the effect of gut Pseudomonadota effectors on host viral infections investigating whether gut commensal T3SS effectors foster or inhibit viral replication and dissemination.

The regulation of glycoprotein processes was also enriched among the effector targets. Glycoproteins are proteins that carry oligosaccharides of various structures which serve as contact points to facilitate interactions within cells and between cells [183]. The gut commensal T3SS effectors converged onto GOLGA2 which is annotated with "positive regulation of protein glycosylation" at the Golgi apparatus as post-translational modifications [184,140]. Hence, T3SS effectors targeting GOLGA2 could impact the glycosylation of various proteins in the human intestine.

In the human gut, glycoproteins are present e.g., as receptors on membranous/microfold cells (M cells) in the mucosal lining which are part of the gut-associated lymphoid tissue (GALT) [125,126]. M cells express glycoprotein 2, a receptor that binds to a protein of the pilus on gram-negative Enterobacilli e.g., *Escherichia coli* and *Salmonella enterica* [125]. Subsequently, the bacteria are phagocytosed by the M cells and transported to the mesenteric lymph nodes, where they induce bacteria-specific immune responses [125]. Gut commensal Pseudomonadota could use T3SS effectors to alter the glycosylation of glycoproteins on M cells that are involved in bacterial detection. This could decrease bacterial stimulation of the immune system supporting a tolerogenic environment in the gut towards gut commensals.

Another intestinal glycoprotein involved in immune regulation is P-glycoprotein located at the epithelial surface. P-glycoprotein is known to downregulate neutrophil migration, and defects in its expression or function are associated with colonic inflammation [185]. For instance, a reduced expression of P-glycoprotein was observed in biopsy samples of CD patients [186]. Notably, gut

commensals can foster P-glycoprotein expression via bacterial metabolites thereby preventing intestinal inflammation [185]. Given the association between gut commensal Pseudomonadota and IBD [3], T3SS effectors could suppress the glycosylation of P-glycoprotein impacting its function and/or transport to the cell membrane [187] thereby contributing to CD risk.

"Regulation of glycoprotein biosynthetic process" is a parent term to e.g., the regulation of some clusters of differentiation such as CD4, CD80, and CD86, which are glycoproteins expressed on immune cells [140]. A commensal *Escherichia coli* strain, but not a commensal *Bacteroides vulgatus* strain, stimulated dendritic cells to express high levels of CD80 and CD86 resulting in "highly activated and matured dendritic cells" [127]. This could contribute to the colitis induction observed by the *Escherichia coli* strain but not by the *Bacteroides vulgatus* strain in IL-2-deficient mice [127]. The underlying mechanism of *Escherichia coli*-induced activation of dendritic cells remains unknown. As dendritic cells are located in the mucosal lining to sample antigens from the gut lumen, proximity between gut commensals and dendritic cells is facilitated [188]. Thereby, gut commensal T3SS effectors could be injected into the immune cells and increase the expression of CD80 and CD86 leading to activated dendritic cells contributing to colitis risk. Mucus glycans are present in large amounts in the intestine forming the mucus layer [126]. They constitute the mucus barrier protecting the host against gut microbes and thereby preventing activation of the host immune system [50]. Potentially, gut commensal Pseudomonadota effectors impact the glycosylation of mucus glycans resulting in a reduction of the mucus barrier. This would facilitate contact with the host's epithelium and enable effector injection.

Lastly, the functional analysis revealed that NF-κB signaling was significantly targeted by gut commensal T3SS effectors. The transcription factor controls genes that regulate several pathways e.g., immunological and inflammatory responses, cell survival and growth as well as oncogenesis [189]. Consequentially, altered NF-κB gene expression is associated with many diseases e.g., cancer, asthma, IBD, atherosclerosis, diabetes, and viral infections [145]. The investigated gut commensal T3SS effectors targeted several human proteins which are part of the NF-κB subnetwork (Fig. 29). As mentioned in the introduction, NF-κB is a popular target for intestinal bacteria. Pathogenic bacteria typically employ T3SS effectors to downregulate this pathway to evade the host immune system – with some exceptions such as *Salmonella Typhimurium* which activates NF-κB for dissemination purposes prior to NF-κB inhibition [190,191]. Most gut commensals modulating NF-κB inhibit the transcription factor (e.g., several Bacillota species) to avoid the transcription of pro-inflammatory molecules thereby strengthening the symbiosis with the host [121]. In contrast, the gut microbe *Bacteroides fragilis* activates NF-κB via the accumulation of its toxins e.g., during aging [121]. Furthermore, microbe-associated molecular patterns of gut commensal *Escherichia coli* strains can activate the transcription factor [132–134]. Potentially, this is part of the host's innate immune response recognizing bacteria in proximity to the host. Due to the importance of the pathway, the biological validation focused on the impact of gut commensal T3SS effectors on NF-κB activation.

Employing an NF-κB reporter assay, five gut commensal Pseudomonadota effectors (Efe_12, Kpn_3, Kpn_9, Kpn_10, and met_7) increased NF-κB activity with met_7 exhibiting a dose-dependent effect. According to HuMMI, met_7 binds two proteins involved in NF-κB signaling: TRAF2 and TRIM27. TRAF2 participates in the activation of NF-κB (Fig. 33): upon TNF binding to the TNF-receptor a signaling complex forms in the receptor's vicinity comprising several proteins including TRAF2. TRAF2 recruits E3 ubiquitin ligases which trigger a cascade leading to the activation of the trimeric inhibitory κB kinase (IKK) complex consisting of IKKα, IKKβ, and the Inhibitor of Nuclear Factor Kappa B Kinase Regulatory Subunit Gamma (IKBKG). IKK activation leads to the degradation of IκBs which bind NF-κB in the cytosol rendering it inactive.

Upon the release of NF-κB from the IκBs, the transcription factor translocates into the nucleus to activate gene transcription [189]. The involvement of TRAF2 in this process is well-researched, while findings on the role of TRIM27 in NF-κB signaling seem to be less concordant. While TRIM27 was shown to block the IKK complex thereby preventing the release of NF-κB [192], TRIM27 also degraded one of the IκBs (IκBα) and thereby activated NF-κB [144]. Given these opposing results, more research is needed to determine under which conditions TRIM27 contributes to the increase or decrease of NF-κB activity. Hence, the NF-κB activation by met_7 observed in this study could be mediated by either TRAF2 or TRIM27, both, or a so far unidentified protein or mechanism. To determine the component of the NF-κB signaling pathway that exhibits altered function due to interaction with met_7, knock-out cell lines could be employed. This requires consideration of the possibility that met_7 might mediate its effects via several host proteins.
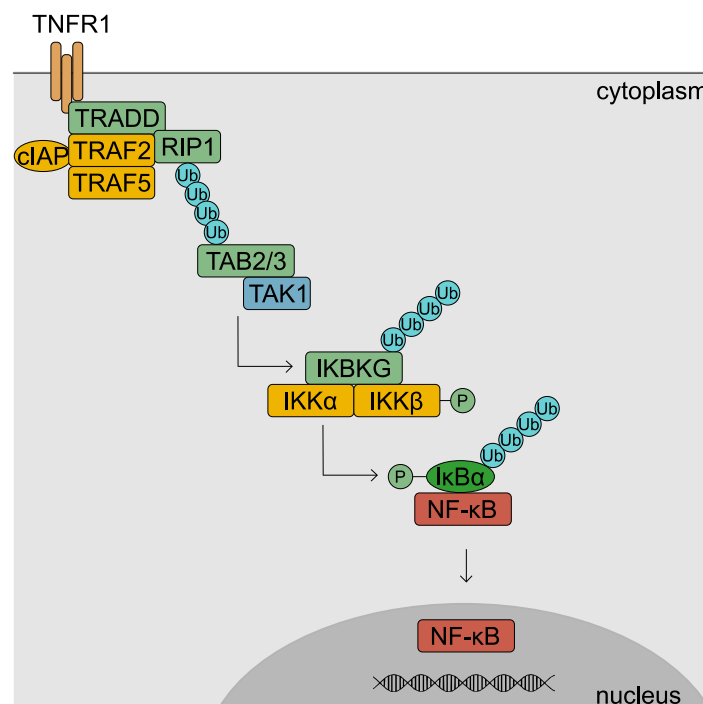
Figure 33 | NF-κB signaling pathway. *After binding of the TNF receptor (TNFR), TRADD is recruited, the E3 ubiquitin ligases cIAP1/2 assemble, and TRAF2 binds the protein kinase RIP1. RIP1 is ubiquitinated and recruited to IKBKG (NEMO) forming the TAK1-IKK complex. TAK1 activates IKKβ by phosphorylation which leads to the phosphorylation and degradation of IκBα releasing NF-κB. The transcription factor translocates into the nucleus for transcription activation. This is just one example of NF-κB activation, and several alternatives mediated via various receptors and pathways exist. Adapted from Yu et al. [193]. Ub, ubiquitin, P, phosphate.*

In addition to modulating NF-κB, collaborators demonstrated that gut commensal T3SS effectors can affect ICAM1 expression on colonocytes and manipulate cytokine secretion from these cells. As overexpression of IL-18 increases the risk for intestinal colitis, the inhibitory effect of Cyo_12 on IL-18 could be protective against the disease [194]. Furthermore, since high levels of IL-1β, IL-6, and IL-8 are associated with irritable bowel syndrome, a reduction of these cytokines by Cyo_12 could decrease disease risk [194]. In contrast, since met_7 increased IL-8 levels, the effector potentially elevates the risk for irritable bowel syndrome. Based on the observation that ICAM1 is upregulated in IBD [152], met_7 could raise IBD risk by increasing ICAM1 expression.

Besides local impacts, alterations of cytokine release from colon cells by commensal effectors could also affect the host systemically. Cytokines released from colonocytes can impact immune cells e.g., dendritic cells, T cells, and B cells in the lamina propria, which is located underneath the epithelium [195]. Activation of these immune cells can trigger further cytokine release or induce immune cell translocation to proximal mesenteric lymph nodes [194,195]. In the mesenteric lymph nodes, the signal potentially induces a specific, local immune response or is further transmitted via cytokines, T cells, or B cells released into the lymphatic system, which is ultimately connected to the blood circulation [195,196]. Thereby, immune cells and cytokines from the gut can be distributed throughout the body impacting immune signaling in distal body sites [196]. The inhibiting impact of Cyo_12 on pro-inflammatory cytokine release could have protective effects on systemic host inflammation, whereas met_7 potentially exerts pro-inflammatory effects.

Although the dual luciferase reporter assay is widely used to detect NF-κB activation or its inhibition, interpreting the results can be difficult as seen in this study. For instance, the expression of the control luciferase can be impacted by the investigated proteins. Furthermore, factors independent of the experiment can lead to reduced or increased control measurements with subsequent over- or underestimation of the reporter luciferase. For instance, binding of endogenous transcription factors to plasmid sequences leads to unintended transcription of plasmid DNA potentially increasing cell stress [197]. Furthermore, genes compete for intracellular resources (polymerases, ribosomes, energy, etc.) which can cause decreased expression levels of exogenous and endogenous genes [197]. For instance, high expression of the firefly luciferase could lead to artificially low renilla luciferase values due to limited resources or *vice versa*. Therefore, the results of such experiments require careful assessment as multiple factors can influence the measured outcomes. This was addressed in this study by deliberately analyzing the measurements of both luciferases independently i.e., assessing the effector impact on the luciferase indicating transfection efficiency as well as the luciferase indicating NF-κB activity.

In conclusion, gut commensal T3SS effectors may modulate host immune signaling via NF-κB activity, cytokine secretion, and ICAM1 expression. While an inhibition of host immune responses may contribute to a tolerogenic environment as has been demonstrated for other gut commensals [121], activation of host immune responses may increase local and potentially systemic host inflammation. As NF-κB activity, cytokine secretion, and ICAM1 expression are relevant for various complex diseases, gut commensal Pseudomonadota might impact disease risk via T3SS effectors. Moreover, gut commensal T3SS effectors potentially affect other immune-relevant processes that are targeted in HuMMI such as viral processes and the regulation of immune-related glycoproteins.

3.3 Implications of commensal T3SS effectors within the gut microbiome

The experimental validation showed that T3SS effectors are capable of altering human immune signaling *in vitro*, and the bioinformatic analysis suggests that the effectors might modulate host immune signaling also via various other processes. This raises questions regarding the impact of T3SS effectors on human health as well as concerning effector translocation into host cells e.g., how gut bacteria overcome the mucus layer to inject effectors into the host's epithelium.

3.3.1 Impacts of gut commensal Pseudomonadota effectors on human health

The gut commensal Pseudomonadota phylum comprises a wide variety of strains with mainly associations to host diseases, even though some strains are employed as probiotics exerting beneficial effects for the host [3,66]. Studies focusing on fecal samples have reported Pseudomonadota to constitute approximately 4-5% of the gut bacterial composition [3]. However, a recent study directly sampling the human intestines revealed that Pseudomonadota exhibit a higher prevalence within the intestinal environment compared to stool samples with some species (*Escherichia/Shigella sp.*) being twice as abundant [198]. Hence, Pseudomonadota are more prevalent in the human intestines than previously assumed, potentially rendering the impact of Pseudomonadota on the host more pronounced than expected. Furthermore, the Pseudomonadota phylum encodes for genes with the greatest variability in abundance between individuals compared to other gut phyla such as Bacteroidota or Bacillota [199]. Thereby, Pseudomonadota species and their encoded genes may explain inter-individual differences in gut microbiomes as well as in microbiome-influenced aspects of host health.

Considering the effector impacts on NF-κB activity representative for effector impacts on immune signaling, various outcomes for host health are conceivable depending on the nature and scope of the effector-mediated impact, and the effect of other bacteria in the gut microbiome: 1) T3SS effectors could be detected within the human cell by intracellular Nod-like receptors [200]. After recognizing bacterial components, NOD1 and NOD2 activate NF-κB resulting in e.g., pro-inflammatory cytokine production, and activation of immune cells [201,202]. Thus, the observed NF-κB activation could be an innate immune response triggered by the presence of bacterial effectors within the host cell. T3SS effector-mediated NF-κB inhibition potentially facilitates the commensal symbiosis between the gut microbiome and the host by downregulating host immune responses. 2) T3SS effectors may induce a moderate activation of NF-κB, possibly exerting a protective effect against host diseases. This hypothesis is based on the finding that not only heightened NF-κB activation is linked to inflammatory diseases, but so is the inhibition of NF-κB, particularly in epithelial cells [203]. Hence, the modulation of NF-κB activity by the effectors, encompassing both activation and inhibition, may play a role in maintaining immune homeostasis. 3) T3SS effectors could impact human health via NF-κB-mediated cytokine release impacting host inflammation locally as well as systemically. Anti-inflammatory effects are potentially protective against host diseases, whereas pro-inflammatory effects could contribute to host disease risk.

Given the association of Pseudomonadota with inflammatory diseases such as IBD and colitis, it is plausible for T3SS effectors to increase local host inflammation and thereby contribute to disease risk. Furthermore, T3SS effector-induced increases in pro-inflammatory cytokine secretion might impact the host systemically, potentially affecting the risk of Pseudomonadota-associated diseases that are not constricted to the human intestine such as metabolic disorders

or cancers. Evidently, this needs to be assessed by further research on gut commensal Pseudomonadota effectors.

The diseases that are linked to Pseudomonadota are characterized as complex diseases whose etiology is impacted by a combination of genetic, environmental, and lifestyle factors affecting the cellular networks resulting in particular phenotypes [39]. For instance, the impact of genetic variation and T3SS effectors on the host cellular PPI network could potentially result in the manifestation of a complex disease phenotype. Furthermore, the observation that effectors can have a dose-dependent effect as demonstrated by met_7 suggests that increased gut T3SS effector concentration can enhance the effector impact. Since Pseudomonadota abundance in the gut can be manipulated by diet [3], nutrition could impact T3SS effector concentration in the intestine affecting the extent of the effector impact. Therefore, in predisposed individuals who adopt a lifestyle leading to a Pseudomonadota bloom, disease manifestation might be more likely compared to individuals preventing an increase in Pseudomonadota abundance via dietary measures.

Several arguments suggest a potential involvement of the effectors in CD based on the participation of the targeted human proteins in actin filament stability, GPR signaling, P-glycoprotein expression, and NF-κB signaling (Fig. 34). For instance, colonocytes of CD patients typically show disruptions of actin filaments which could be impacted by ten (Ec6_9, Ec6_10, Efe_5, Efe_11, Kpn_3, met_18, met_29, Pma_2, Yen_7, Yre_7) of the 64 effectors enriched in metagenomes of CD patients targeting human proteins involved in actin filament formation [135]. Furthermore, CD is associated with a decrease in colonic P-glycoprotein expression [186] potentially mediated by two of the CD-enriched effectors (Efe_11 and Kpn_3) via targeting the protein of GOLGA2. As this human protein is involved in the regulation of protein glycosylation, it could be manipulated by the effectors to alter P-glycoprotein glycosylation. Changes in protein glycosylation can affect protein function and transport to the cell membrane [185,187], reducing functional P-glycoprotein expression at the epithelial surface. Furthermore, CD patients show increases as well as decreases in the expression of GPRs in the ileum and colon which could be influenced by gut commensal T3SS effectors [129]. In total, five of the 64 CD-enriched effectors (Ec6_9, Efe_11, Kpn_3, met_18, met_29) targeted proteins involved in GPR regulation potentially mediating the alterations in GPRs observed in CD. Also, a study on 83 CD patients revealed that 56% exhibited high NF-κB activity, while 44% showed low NF-κB activity [204], which could be mediated via T3SS effector impacts. In HuMMI, 20 of the CD-enriched effectors targeted the NF-κB subnetwork one of which (Kpn_3) significantly activated NF-κB based on the reporter assay. The other two NF-κB-activating effectors (Efe_12, Kpn_9) were also present in the CD metagenomes, however not significantly enriched compared to healthy controls. Cyo_12, which downregulated NF-κB-controlled cytokines, was not found in the metagenomes of CD patients. Potentially, gut commensal T3SS effectors promote CD risk by increasing NF-κB activity. Hence, several of the effectors enriched in CD patients targeted human proteins whose functions can be connected to the CD phenotype raising the possibility that Pseudomonadota effectors contribute to disease manifestation in CD. This is supported by the observation in cell-based assays that met_7 upregulates ICAM1 expression on colonic epithelial cells as increased ICAM1 expression is associated with CD and may be involved in disease pathogenesis [205].
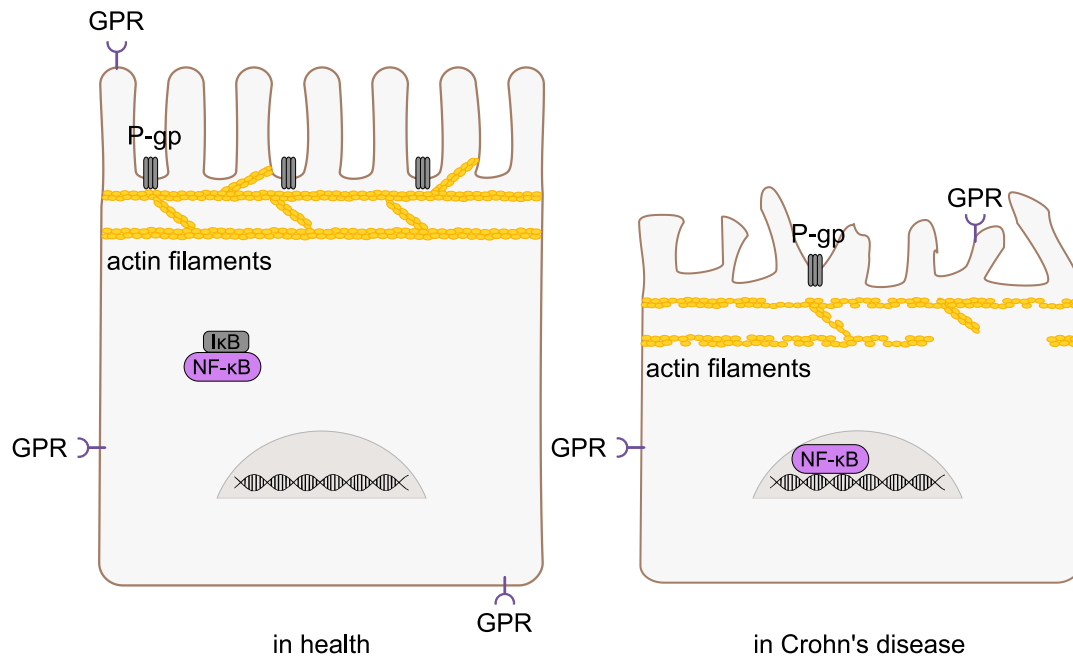
Figure 34 | Colonocytes in healthy controls and Crohn's disease patients. *Schematic representation of some of the differences between colonocytes of healthy controls and CD patients. Based on Wilson et al. [186]. P-gp, P-glycoprotein.*

This thesis aimed to elucidate the impact of T3SS effectors expressed by commensal gut Pseudomonadota on host functions in the context of human health and disease. Based on this work, T3SS effectors potentially have anti- as well as pro-inflammatory effects on host immune signaling. Since gut Pseudomonadota are associated with host inflammation, the hypothesis is proposed that T3SS effectors mediate underlying mechanisms with pro-inflammatory effects. Especially the involvement of T3SS effectors in CD pathology is suggested by several arguments promoting further research in this context. The fact that effectors of gut commensals might exert detrimental or beneficial effects on the human host raises the question of whether the term "commensals" applies to these strains. Commensalism is termed as a form of symbiosis in which "one organism gains" and the other "is affected in neither a positive nor a deleterious manner" [206]. While gut commensals benefit from the relationship with a host since they are provided a protected space to grow as well as nutrient resources [206], the presence of Pseudomonadota might exert positive or deleterious effects on the host. Hence, Pseudomonadota strains with pro-inflammatory effects might be better characterized as pathobionts, while Pseudomonadota strains with anti-inflammatory effects might be commensals in the sense that they contribute to the symbiosis with the host with neither positive nor negative impacts on the individual. Should Pseudomonadota effectors with anti-inflammatory impacts exert protective effects against host inflammation, these strains could be termed mutualists as the host benefits from the effects of T3SS effectors.

3.3.2 Effector translocation into host cells

To elicit the aforementioned effects on host immune signaling, T3SS effectors must be translocated from gut commensals into host epithelial cells via the T3SS. This raises questions concerning the circumstances under which gut bacteria are close enough to host cells to inject effectors.

Direct proximity to host cells can be sufficient to initiate the secretion of effector proteins via a T3SS [5]. To achieve this proximity, bacteria must overcome the mucosal layer with its antimicrobial peptides and secretory IgA. Insights into potential mechanisms for overcoming these barriers can be obtained from examining pathogenic bacteria. Human pathogens gain access to the epithelium by e.g., moving through the mucus using a flagellum or by secreting proteases which degrade mucin glycoproteins [207]. Antimicrobial peptide evasion is mediated by e.g., incorporating positively charged molecules into the bacterial outer membrane thereby altering its negative charge [208]. This is effective as antimicrobial peptides are highly basic and tailored to bind the typically negatively charged phospholipid groups on the bacterial membrane via electrostatic interactions [207]. Coating and clearance by secretory IgA can be escaped by introducing changes in targeted bacterial surface antigens. Alternatively, some pathogens express weakly bound bacterial surface antigens that easily detach from the bacterial cell upon IgA binding/crosslinking, thereby freeing the bacteria from IgA coating [209].

Commensal microbes typically avoid close contact with the host to prevent the activation of the host's immune system. However, in the presence of certain environmental factors, commensals seem to infiltrate the gut mucus layer. For instance, studies in mice have shown that dietary emulsifiers in processed foods such as the widely used polysorbate 80 and carboxymethylcellulose can increase the expression of gut bacterial flagellin [210]. The increased expression of flagellin might enable commensal bacteria to move through the mucus layer as observed in pathogens. Moreover, degradation of the mucus through various mechanisms could further facilitate bacterial contact with the host epithelium. For instance, emulsifiers but also pesticides, and heavy metals negatively affected the mucus layer in rodents including mucus expression, thickness, and penetrability [211]. Additionally, some gut commensals encode for mucin-degrading proteases such as *Bacteroides caccae* [208]. Potentially, mucolytic commensals pave the way to the epithelium enabling close contact between T3SS-expressing bacteria and host cells. Furthermore, in some hosts, the mucus layer might be thinner or less dense due to chronic inflammation and/or genetic predispositions. For example, mice lacking the TLR5, which senses flagellin, exhibited a thinner mucus layer than wild-type mice. TLR5-deficient mice that developed colitis exhibited an even less protective mucus layer with a disordered structure, no solid inner layer, and "with bacteria close to or in contact with the gut epithelium" [212]. Lastly, gut commensals can evade antimicrobial products targeted at their LPS by introducing modifications into these molecules [208]. Thus, multiple mechanisms exist for how gut commensal Pseudomonadota could overcome the mucus barrier and establish close contact with the host epithelium.

The proximity of gut commensal Pseudomonadota to the host's epithelium has indeed been demonstrated by several studies. For instance, in a mouse colitis model, gut Pseudomonadota, especially *Enterobacteriaceae* increased in abundance and resided "very close to, or in direct contact with" the host's epithelium [212]. Furthermore, Pseudomonadota were linked to mucus defects such as increased penetrability in mice resulting in closer contact of the bacteria with the epithelium [213]. The ability to approach host cells was most pronounced in Pseudomonadota

compared to other phyla in the gut microbiome [212,213]. Hence, Pseudomonadota seem to be able to move close to host cells potentially facilitated and/or enhanced by gut inflammation, host genetics, or other factors such as dietary components or environmental toxins. Close contact with host cells probably triggers the secretion of effector proteins from a T3SS. This is presumably exacerbated by increased Pseudomonadota abundance within the gut community in response to e.g., a westernized diet [3]. Given that a westernized diet is a known risk factor for complex diseases [38,214], gut Pseudomonadota abundance may partly mediate the association between dietary habits and the development of complex diseases. Potentially, gut commensal T3SS effectors have evolved to facilitate the relationship with the human host but contribute to disease phenotypes in altered environments influenced by e.g., a westernized diet and environmental toxins.

3.4 The role of T3SSs in microbe-host interactions

The presence of T3SSs in bacterial organisms has typically been correlated with virulence in human and plant pathogens. More recent studies have shown that T3SS are also crucial for mutualists and commensals e.g., plant mutualists using T3SSs to establish the rhizobia-legume symbiosis, and commensal bacteria to colonize the gut of insects [215]. Furthermore, with the advent of genome sequencing, T3SSs have been discovered in several bacteria inhabiting diverse environments e.g., in the soil bacterium *Myxococcus xanthus* as well as in the water bacterium *Verrucomicrobium spinosum*, and in various *Escherichia coli* strains inhabiting the gut of animals [5,7]. Hence, T3SSs are present in multiple bacterial species colonizing several habitats and hosts such as mammals, plants, insects, nematodes, etc. [216]. The identified T3SSs vary in several of their aspects e.g., T3SS components, number and type of effectors, regulation of secretion, interaction with host proteins, etc. [5,67,68,71]. This is reflected in different types of T3SS families as mentioned in Chapter 1.2.4, suggesting that T3SS can be employed for different purposes as observed for *Pantoea stewartia*. This bacterium encodes for two different T3SSs: Hrc-Hrp T3SS, which is important for maize pathogenesis, and Inv-Mxi-Spa T3SS to colonize the gut of a beetle as a gut commensal [215]. Thereby, the bacterium can populate the beetle using it as an insect vector which transmits the bacterium to maize plants in which *Pantoea stewartii* causes Stewart's wilt [215]. By encoding for two different types of T3SSs that are employed for different functions and purposes, the bacterium is able to colonize two hosts of two different kingdoms i.e., plant and insects [215].

As the interpretation of the presence of T3SS in bacterial organisms influences e.g., sequencing data interpretation and bacterial classification into biosafety levels [217], an accurate understanding of T3SS is crucial. Given the fact that T3SSs were identified in non-pathogenic bacteria and that many different types of T3SSs exist, Pallen *et al.* argued for a more inclusive view of the T3SS. They highlighted that this system has evolved as a mechanism to facilitate the "billion-year-old" interactions between bacteria and eukaryotes [216]. The T3SS likely evolved not exclusively for pathogenic purposes but to enable various types of interactions between bacteria and eukaryotic organisms [216]. This thesis supports this notion suggesting that commensal gut bacteria as part of the healthy gut microbiome use T3SS effectors to mediate the interaction with the human host. The effect on the host may be diverse depending on the Pseudomonadota strain and the effectors. While some effectors might assist the commensal relationship with the host by creating a tolerogenic environment, others may increase host inflammation. As these effectors are encoded by the "normal" gut microbiome, effectors are not expected to evoke host diseases by themselves differing from pathogenic bacteria. Rather, diseases may manifest in combination with other risk factors such as diet and/or genetic variations.

In conclusion, this study supports an inclusive perspective on T3SSs based on the observation that non-pathogenic gut commensals employ T3SS effectors to impact host immune signaling. Together with T3SSs of human pathogens and plant mutualists, this suggests that T3SSs are involved in pathogenic, commensal, and mutualistic relationships between bacteria and eukaryotes.

3.5 Limitations of this study

To investigate gut commensal T3SS effectors a high-quality network map was generated pointing to the involvement of the effectors in human immune signaling, which was validated *in vitro* using human cell culture experiments. While this pioneering research offers insights into gut commensal T3SS effector functions for the first time, it is important to keep the limitations of the study in mind. The primary biological constraint of this study lies in the lack of data demonstrating causality between T3SS effector impacts and human health outcomes. This could be shown by experiments investigating the translocation of the effectors from bacterial cells into human cells and exploring the impacts of gut commensal T3SS effectors on host health, which would have exceeded the scope of this thesis.

Investigating effector translocation from the bacterial cell into the human cell via a T3SS requires deliberation during the experimental design. To identify effectors within human cells, the translocated proteins require a distinguishable tag. Tagging effectors with a fluorophore such as GFP is unsuitable as the tightly packed protein cannot pass through the T3SS needle, which requires partially unfolded proteins [218]. Previous studies investigating effector translocation employed e.g., fusion of a reporter enzyme to effector proteins [219], or used self-labeling enzymes i.e., fusing tags to effectors that bind to ligands coupled to a fluorescent dye after translocation into the human cell [220].

While cell culture experiments can demonstrate that gut commensal Pseudomonadota inject effectors into human cells, the experimental setup does not resemble the physiological conditions of the human gut environment, which can be better mimicked using organoids. Subsequently, studies in animal models are commonly the next step when investigating the effects of the gut microbiome on human health. Such studies could explore effector injection into gut epithelial cells and subsequent impacts on the organism. However, it has been argued that experiments in mouse models might not be appropriate when investigating gut microbial impacts on the human host. This is due to fundamental differences between rodents and humans expressed e.g., in the inability of a great proportion of human microbes to colonize mice, and the absence of human-specific lifestyle factors in animal models that greatly affect the gut microbiome [56]. Therefore, animal models may be used that better mimic human physiology such as pigs or humanized mice [56,221]. By transplanting functional human cells or human tissues into mice, thereby humanizing them, the animal models allow investigating the impact of T3SS effectors on human cells or tissues in an *in vivo* context [221]. Furthermore, observational studies in humans could be conducted investigating samples from the gut lumen for the presence of T3SSs and assessing associations to host health.

In conclusion, this work is a starting point for investigating gut commensal T3SS effectors encouraging future studies to test the hypotheses formulated in this thesis and uncover effector impacts on host health.

3.6 Outlook

The observation that T3SS effectors can modulate human immune signaling raises questions concerning potential strategies to prevent host inflammation and enhance beneficial effector impacts. Potentially, therapeutics against effectors with deleterious impacts could be developed. Due to the functional redundancy of T3SS effectors, therapeutic strategies that inhibit the secretion of all effectors are likely to be more effective than those aiming at the neutralization of individual effector proteins. This could be achieved via agents that suppress effector secretion or that inhibit transcription factors responsible for the expression of effectors [222]. Some of these therapeutics are already investigated including natural compounds e.g., flavonoids that have been detected to inhibit the function of T3SSs [222]. It is important to keep in mind that while one therapeutic may inhibit a particular T3SS it might not be effective against a different type of T3SS [222]. A different strategy to manipulate gut effector concentration is to employ synthetic communities of gut commensal T3SS-positive strains with knock-out of T3SS genes in harmful strains but intact T3SS in beneficial Pseudomonadota. These communities could be administered like probiotics maintaining the diversity of gut Pseudomonadota strains without their deleterious effects and potentially inferring benefits in the case of advantageous Pseudomonadota. Notably, the efficacy of probiotics varies greatly between studies demonstrating our lack of detailed understanding in the matter regarding e.g., probiotics colonization and persistence in the gut as well as their effect on the microbiome [223]. Alternatively, since bacteriophages can be used to eliminate pathogenic bacteria from the gut as an alternative to antibiotics [224], similar viruses could also be employed to target deleterious Pseudomonadota. A cheaper approach to reducing harmful Pseudomonadota in the gut is by dietary measures. Pseudomonadota abundance and thereby T3SS effector concentration can be reduced by avoiding a diet high in fats, calories, artificial sweeteners, and emulsifiers [3].

3.7 Conclusion

This study introduced a novel mechanism mediating the relationship between gut commensal Pseudomonadota and the human host. Furthermore, it showed that T3SS effectors of gut commensal Pseudomonadota can impact human immune signaling *in vitro*. Given the opposing effects of T3SS effectors on human immune signaling, the findings reflect the complexity of gut commensal Pseudomonadota as reported by previous findings [121]. While the anti-inflammatory impacts of T3SS effectors could facilitate a tolerogenic environment, the pro-inflammatory effects of gut Pseudomonadota effectors potentially influence the risk of complex diseases that are associated with the phylum [3]. Via altered cytokine secretion T3SS effectors could not only influence local immune responses but also impact host immune signaling on a systemic level. As Pseudomonadota abundance responds to dietary intakes, disease risk could potentially be mediated via an individual's nutrition, especially in genetically predisposed individuals.

This study suggests that T3SSs are not only involved in pathogenic relationships with the human host but also in commensal symbiosis, thereby supporting an inclusive view of T3SSs. The findings of this work encourage further investigations into commensal T3SS effectors to better understand the effect of bacterial effectors on host functions and unravel the contributions of gut Pseudomonadota to disease risk.

4. Material and Methods

4.1 Bacterial strains and material

Bacterial strains were ordered from three different suppliers according to Table 6. Genomic DNA for Yre and bacterial pellets of Eta and Pst were ordered at LGC Standard (Wesel, Germany), which distributes ATCC products in Europe. Ec2, Ec6 and Kpn were obtained as bacterial cultures in glycerol from BEI resources (Manassas, Virginia, USA). Genomic DNA of the remaining strains was ordered from the Leibniz-Institut DSMZ (Braunschweig, Germany). For 15 strains genomic material was shipped ranging from 125 ng to 7300 ng, whereas five strains were delivered as living cultures. The five strains were cultured according to the manufacturer's protocol and genomic DNA was extracted using four different methods: Invitrogen™ TRIzol™ Reagent (Invitrogen cat. no. 15596026) was used as described in the manufacturer's protocol. DNeasy® UltraClean® Microbial Kit (Qiagen cat. no. 12224) was performed according to the manufacturer's protocol employing the FastPrep-24™ to induce cell lysis instead of vortexing using adapter tube holders. Alkalic lysis was performed as follows: 1.8 mL live bacterial culture were centrifuged at 15 g for 1 min and the supernatant was discarded. 150 µL PI buffer (50 mM Glucose, 25 mM Tris HCL pH 8.0, 10 mM EDTA pH 8.0) was added and the pellet was resuspended. 100 µL PII buffer (0.2M NaOH, 1% SDS) was added and the sample was inverted 4-6 times and incubated for 2 min at room temperature to induce cell lysis. After adding 100 µL PIII buffer (3M NaAc, using acetic acid for pH 4.8) and inverting the sample, it was centrifuged for 10 min at maximum speed. The supernatant was transferred to a new tube and 750 µL 100% EtOH was added. After centrifugation for 10 min at maximum speed the supernatant was discarded and 500 µL 70% EtOH was added. After a last centrifugation step for 10 min at maximum speed, the supernatant was discarded, and the pellet dried and resuspended in 50 µL pure H_2O. The NucleoSpin® Plasmid (NoLid) Mini kit (Macherey-Nagel cat. No. 740499) was performed according to the manufacturer's protocol with vortexing after the addition of BufferA2 and BufferA3.

Table 6 | Details of the ordered strains. *Strains with supplier and collection number (no) and material obtained.*

strain	supplier and collection no	material obtained
Aeromonas jandaei CECT 4228	DSM 7311	Genomic DNA
Cedecea davisae Grimont et al. 1981	DSM 4568	Genomic DNA
Citrobacter pasteurii Clermont et al. 2015	DSM 28879	Genomic DNA
Edwardsiella tarda ATCC 23685	ATCC 23685	freeze-dried bacterial pellet
Enterobacter massiliensis Lagier et al. 2014	DSM 26120	Genomic DNA
Escherichia coli MS 200-1	BEI HM-356	bacterial culture in 10% glycerol
Escherichia coli MS 69-1	BEI HM-347	bacterial culture in 10% glycerol
Escherichia fergusonii Farmer et al. 1985	DSM 13698	Genomic DNA
Klebsiella sp. MS 92-3	BEI HM-354	bacterial culture in 10% glycerol
Morganella morganii subsp. morganii NBRC 3848	DSM 30164	Genomic DNA
Pantoea septica	DSM 24604	Genomic DNA
Providencia rettgeri DSM 1131	DSM 1131	Genomic DNA
Providencia stuartii ATCC 25827	ATCC 25827	freeze-dried bacterial pellet
Pseudocitrobacter faecalis Kämpfer et al. 2014	DSM 27453	Genomic DNA
Pseudomonas sp.	DSM 29075	Genomic DNA
Vibrio furnissii NCTC 13120	DSM 19622	Genomic DNA
Yersinia enterocolitica subsp. palearctica Y11	DSM 13030	Genomic DNA
Yokenella regensburgei ATCC 43003	ATCC 43003	Genomic DNA

4.2 PCR for ORF amplification and SfiI restriction site generation

To amplify the ORF and add Sfi restriction sites a nested PCR was performed. The oligonucleotides of the primers were ordered from Eurofins Genomics Germany GmbH and the respective sequences can be found in Table S2. Primers were diluted to a concentration of 2 µM and the KOD Hot Start DNA Polymerase (Merck Millipore cat. no. 71086) was used with 0.4 µL per reaction, or a homemade Phusion High-Fidelity with 0.2 µL per reaction. The effectors of all strains were clustered according to size (> 1 kbp, 1-2.5 kbp, > 2.5 kbp) and for each cluster, a separate PCR was performed with adjusted elongation time. A PCR cycler was programmed as follows for amplification with the KOD (Phusion) polymerase: 2 min at 95°C (1 min at 98°C) followed by 30 cycles of 20 sec at 95°C (10 sec at 98°C), the required elongation time at the T_a, and 95 sec at 70°C (1 min at 72°C). Lastly, the mixture was incubated at 70°C for 5 min (72°C for 5 min), followed by storage at 8°C.

The second primer (2µM) was universal for all first PCR products:
forward 5' GAATTCGGCCGTCAAGGCCAGAAGGAGATATAACCATG 3'
reverse 5' AGTCGACGGCCCATGAGGCCGCCTTA 3'
1 µL of the first PCR was used for the generation of the full Sfi site by the second PCR. For gel electrophoresis, 1% agarose gels were poured with 5 µL Midori Green (Biozym Scientific GmbH cat. no. MG04) per 100 mL gels. GeneRuler 1 kb DNA ladder (Life Technologies GmbH cat. no. SM0312) was used and 5 µL PCR samples were mixed with 7 µL Orange Dye. Gels were run for 10 min at 140 volt (V).

4.3 DNA purification using magnetic beads

To clean PCR products from remnants of enzymatic reactions, a magnetic bead (magtivio cat. no. MDKT00010075) purification was performed after the second PCR and after SfiI-digestion. DNA from PCR or digestion (~ 20-25 µL) was well mixed with 25-30 µL of magnetic beads and incubated for 5 min at room temperature. A 5 min incubation step on a magnetic rack (Alpaqua cat. no. A001322) followed. After removing the supernatant, the beads were washed with 160 µL of 70% Ethanol and incubated for 5 min on the magnetic rack again. Removal of the Ethanol was accompanied by letting the beads dry for 1-2 min to rid them of all remaining Ethanol. 30 µL of an elution buffer (5mM Tris HCl, pH 8.0) was added in which the beads were well resuspended and after 5 min on the bench were put on a magnetic rack for an additional 5 min. The elude was transferred to a new plate or tube.

4.4 Sfi digestion (PCR and plasmid)

To clone the ORF into a plasmid, the second PCR product and the pENTR223 with SfiI restriction site were treated with SfiI recombinant (NEB cat. no. R0123L) to create sticky ends. For reduction of the star activity of the enzyme, the reaction mix for the PCR products was adjusted to 0.0125 µl Sfi and 4 µl CutSmart Buffer (NEB updated cat. no. B6004S) per 50 µL-reaction and a 7 min incubation at 48°C. pENTR223 was digested using 1 µL SfiI enzyme for 1 µg plasmid DNA at 50°C for 3 hours. The mix was run on a 0.8% agarose gel for 70 min at 100 V and the vector backbone was purified using the NucleoSpin Gel and PCR Clean-up Midi kit (Machery-Nagel ca. no. 740986). The backbone was then treated with 1 µL FastAP thermosensitive alkaline phosphatase (ThermoFisher ca. no. EF0651) for 1 µg of DNA according to the manufacturer's protocol for dephosphorylation of sticky ends to prevent recirculation.

4.5 Ligation of PCR products and plasmid backbone

Sfi-digested PCR products and the plasmid backbone were ligated using T4 DNA Ligase (ThermoFisher ca. no. EL0011). A 20 µL-reaction mix consisted of 30-60 ng pENTR223, 0.2 µL ligase, and ~17 µL digested and purified second PCR product.

4.6 Propagation of plasmids in *Escherichia coli* DH5α

Competent *Escherichia coli* DH5α (> 1 * 107 cfu / µg) were transformed in a 96-well format with the plasmids encoding the effector ORFs. To this end, competent cells were thawed on ice and 30 µL distributed per well. 4 µL of the ligation product or 2.5 µL of the LR-reaction was added and incubated for 45 min on ice. The plate was placed in a 42°C-water bath for 1 min to heat-shock the samples, after which they were put on ice for 2 min for cool-down. 100 µL of warm SOB (Super Optimal Broth) medium (0.5% yeast extract, 2% tryptone, 10mM NaCl, 2.5mM KCl, 10mM MgCl$_2$, 10mM MgSO$_4$, pH 7.0) was added per well and incubated for 1-1.5 hours at 37°C. Afterward, the samples were transferred to 1.6 mL TB (Terrific Broth) medium (2.4% yeast extract,1.2% tryptone, 0.4% glycerol, 89mM phosphate buffer) per well in a deep-well plate with the respective antibiotic (Table 7) and incubated for 16-18 hours at 37°C on a plate shaker. Glycerol stocks were prepared by mixing 80 µL bacterial culture with 80 µL 40% Glycerol per well in a skirted 96-well u-bottom plate and stored at -80°C. 50 µL of the remaining culture was centrifuged for 5 min at room temperature at maximum speed, the supernatant was removed, and a PCR mix was added to each well to control for successfully processed clones. The 30 µL-PCR mix consisted of 1 µL (forward and reverse each) 2 µM plasmid-specific primers (Table 7 and Table 8) and 0.5 µL of a homemade Taq polymerase, and the protocol of a standard Taq polymerase PCR reaction was followed. If transformation in a 96-well plate was not successful, transformation in 1.5 mL tubes was performed. To this end, competent *Escherichia coli* DH5α were thawed on ice and 10 µL were distributed per tube. The DNA samples were added as described above as was the heat shock. 1 mL of warm SOB medium was added and incubated for 1-1.5 hours at 37°C on a shaker. 100 µL of the transformation mix were streaked out on LB plates with the respective antibiotics and incubated overnight at 37°C. Grown colonies were picked from the plates and incubated overnight in TB or LB medium. Glycerol stocks were prepared in either 96-well plates as described above or in 2 mL tubes with 900 µL 40% Glycerol and 900 µL sample.

Empty plasmids were propagated in the *Escherichia coli* DB3.1 strain, which is resistant to the toxic effect of the ccdb gene in the empty plasmid. Also, chloramphenicol, located at the region where the ORF is later incorporated, was added to the media in addition to the antibiotic resistance gene in the vector backbone (Table 7).

Table 7 | Plasmids used in this study. Bacterial resistance per plasmid is listed as well as the primers and the annealing temperature.

plasmid	bacterial resistance	primer	Ta [°C]
pENTR223.1	Spectinomycin	M13 forward / M13 reverse	58
pDEST-DB	Ampicillin	Gal4DB forward / term	59
pMH-FLAG-HA	Ampicillin	EF1α-F / IRES reverse	58
pVTU-DEST	Ampicillin	ADH1-F / ADH1-R	59
pDEST-N2H-N1 / -N2	Ampicillin	N2H forward / N2H reverse	57

Table 8 | Plasmid Primer sequences. *Primer names and the according oligonucleotide sequences are listed.*

primer	Sequence (5' to 3')
M13 forward	GTAAAACGACGGCCAGT
M13 reverse	GGAAACAGCTATGACCATG
Gal4DB forward	GGTCAAAGACAGTTGACTGTATCGT
term	GGAGACTTGACCAAACCTCTGGC
EF1α-F	TCAAGCCTCAGACAGTGGTTC
IRES reverse	GCATTCCTTTGGCGAGAG
ADH1-F	AGTTGATTGTATGCTTGGTATAGC
ADH1-R	GCTATACCAAGCATACAATCAACT
N2H forward	ATAAAAGGTGACGCGTGTGG
N2H reverse	GTTCGCTACCTTAGGACCGT

4.7 Plasmid DNA extraction

5 µL of bacterial glycerol stocks were inoculated in 1.6 mL per well in a deep-well plate in TB medium with the respective antibiotic (Table 7). If transformed bacterial cultures were arranged on plates, Plasmid DNA was extracted using the pipetting Bio Robot Universal System (Qiagen cat. no. 9001094) and the QIAprep 96 plus BioRobot kit (Qiagen cat. no. 962241). For smaller batches, the NucleoSpin Plasmid (NoLid) Mini kit (Macherey-Nagel cat. No. 740499) was used. When higher concentrations or transfection-grade plasmid DNA were needed bacterial cultures were grown in 100 mL TB medium and the respective antibiotic, and the NucleoBond Xtra Midi kit (Macherey-Nagel cat. no. 740410) was used. Protocols of all kits were followed according to the manufacturers' recommendations.

Consolidation of the cloned effector ORFs was achieved using a liquid-handling robot which consolidates the desired ORFs into a new plate based on a detailed script and barcode-labelled 96-well plates.

4.8 Sanger Sequencing to verify ORF-sequences

To verify the cloned ORFs in pENTR223, 10 µL of the mini-prepped DNA plus 5 µL H$_2$O was sent to Eurofins Genomics Germany GmbH to be sequenced using the M13 uni (-21) forward primer or the M13 rev (-29) reverse primer. Due to pandemic-related supply shortages sequencing with a forward or a reverse primer was completed after transformation in yeast using the Gal4DB or the term primer.

4.9 Gateway™ cloning

To clone the ORFs from pENTR223 into the Y2H destination plasmid pDEST-DB (pPC97, Cen origin), the Y3H destination plasmid pVTU-DEST, the pDEST-N2H-N1 and -N2, or the mammalian expression vector pMH-FLAG-HA an LR reaction was performed using the Gateway™ System from ThermoFisher. The LR reaction facilitates a recombination reaction between an attL-containing entry clone and an attR-containing destination plasmid to generate an expression clone with the desired ORF. 3.5 µL of mini-prepped pENTR223 containing an ORF (typically < 30 ng/µL), 1 µL of LR Clonase Mix II (ThermoFisher cat. no. 11791100), 1 µL of 100 ng/µL of the respective destination vector and optionally 0.5 µL TE buffer (10mM Tris ph8.0, 1mM EDTA) were mixed well and incubated overnight at 25°C. Before transforming the plasmids in *Escherichia coli* DH5α, the LR reactions were frozen for a couple of hours at -20°C to terminate the reaction.

4.10 Cloning of metagenomic effectors

Metagenomic effectors were ordered at Twist Bioscience (San Francisco, CA, 660 USA): all effectors smaller than 1800 bp could be ordered as gene fragments with a suitable linker for Sfi digestion (Table 9). Twist ships its products dried down; therefore, they were resuspended in 10mM Tris pH 8.0 to a concentration of 100 ng/µL, of which 2 µL were used for Sfi-digestion. Afterward, the sequences were cloned as the effectors, which were isolated from the genome. Metagenomic effectors bigger than 1800bp were ordered as clonal genes in an entry vector with a suitable linker (Table 9). The shipped DNA was resuspended in 10mM Tris pH 8.0 to 50 ng/µL of which 2 µL were used for the LR reaction into pDEST-DB.

Table 9 | Linker sequences for the metagenomic effectors.

type of linker	linker sequence (5' to 3')
fragments forward	ACAAGTTTGTACAAAAAAGCAGGCTGGCCGTCAAGGCCAGAAGGAGAT ATAACCATG
fragments reverse	TAAGGCGGCCTCATGGGCCACCCAGCTTTCTTGTACAAAGTGGTC
clonal genes forward	GGCCGAGGCCAGAAGGAGATATAACCATG
clonal genes reverse	TAAGGCGGCCTGGGCC

4.11 Yeast transformation

To mate the bacterial effector ORFs against human ORFs, *S. cerevisiae* Y8930 (MATα mating type) was transformed with the plasmids containing the effector ORFs as DB-X constructs, whereas the human ORFs were already available from the human ORFeome collection v9.1 provided by the Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Boston, MA as AD-Y constructs in *S. cerevisiae* Y8800 (MATa mating type).

Yeast transformation with the pDEST-DB plasmids was achieved as follows: Y8930 (MATα mating type) was streaked out on 145 mm-YEPD plates and incubated at 30°C for 72 h. Around ten colonies were picked to inoculate 50 mL of YEPD liquid medium (2% bacto peptone, 1% yeast extract, 2% glucose) and incubated at 30°C for 16-18 hours on a shaker at 180 rpm. This pre-culture was grown to OD_{600} 1.0-3.0 and used to inoculate the YEPD-main culture to OD_{600} 0.1 (200 mL main culture per 96-well plate). The main culture was incubated at 30°C on a shaker (~180rpm) until an OD_{600} of 0.4-0.6 was reached. During the incubation, the carrier DNA (Sigma-Aldrich cat. no. D9156) was boiled for 5 min and kept on ice, also the buffers were prepared for the following steps (TE/LiAc und TE/LiAc/PEG, see below). Once the yeast culture reached the recommended density, the culture was centrifuged in 50 mL falcons for 5 min at 800 x g at room temperature. For every 200 mL yeast culture, the cell pellet was resuspended in 10 mL distilled H_2O. After another centrifugation step at 800 x g for 5 min at room temperature, the cell pellet was resuspended in 10 mL TE/LiAc (10mM TRIS pH 8.0, 0.5mM EDTA, 100 mM LiAc), and centrifuged as before. Again, the cell pellet was resuspended in 2 mL TE/LiAc solution and 10 mL of TE/LiAc/PEG (10mM TRIS pH 8.0, 0.5mM EDTA, 100 mM LiAc, 80% of 44% PEG stock solution) plus 200 µL of boiled carrier DNA was added. The solution was mixed gently by inversion and 120 µL of the suspension was pipetted into each well on a 96-well plate. 15 µL plasmid DNA was added (concentration ranged from ~12 ng/µL to ~95 ng/µL) per sample and carefully mixed by pipetting up and down with a liquid-handling robot. The plate was sealed and incubated at 30°C for 30 min without shaking, after which the samples were heat-shocked in a 42 °C water bath for 15 min. The plate was centrifuged at 800 x g for 5 min at room temperature and the supernatant was carefully

removed using a liquid-handling robot. Cell pellets were slowly resuspended in 10 µL of distilled H$_2$O and 5 µL per well were spotted on Sc-Leu plates [114]. The plates were incubated at 30°C for 72 hours after which yeast colonies were picked and transferred to Sc-Leu liquid medium. Again, the plates were incubated at 30°C for 72 hours on a shaker. Two glycerol stocks (stock and working) were prepared using 80 µL of 40% glycerol and 80 µL yeast culture and stored at -80 °C.

4.12 Identification of constitutive autoactivators

Before screening, constitutive autoactivators were identified by mating yeast clones containing the pDEST-DB with a bacterial effector ORF against yeast with an empty pDEST-AD (pPC86+CYH2) plasmid. 5 µL of the respective glycerol stocks were inoculated in 180 µL Sc-Leu medium for DB-X constructs and 180 µL Sc-Trp for empty pDEST-AD plasmid in a 96-well plate. Y2H controls (Table 10) were inoculated in Sc-Leu-Trp. Plates were incubated at 30°C for 72 hours on a shaker (~180 rpm). 5 µL of the yeast culture containing an empty pDEST-AD plasmid were spotted on YEPD plates using a liquid-handling robot. After letting the spots dry for a couple of minutes the yeast cultures containing pDEST-DB with different bacterial ORFs were spotted on top. Y2H controls were spotted at the bottom of the plate. The plates were incubated at 30°C for 24 hours after which the yeast colonies were transferred from the YEPD plates onto Sc-Leu-Trp-His + 1mM 3-amino-1,2,4 triazole (3-AT) (Sigma-Aldrich cat. no. A8056) plates using a replica-plating block and sterile velvets [114]. To remove some of the yeast cells to reduce the background, Sc-Leu-Trp-His + 1mM 3-AT plates were cleaned by pressing them down on a fresh sterile velvet. Plates were incubated at 30°C for 72 hours. If growth was detected on the plate, the respective bacterial ORF was excluded from the following Y2H mapping pipeline.

Table 10 | Y2H controls. *The plasmid pairs per control as well as the expressed protein and the interaction strength are described as they appear on Sc-Leu-Trp-His + 3-AT plates.*

	plasmid pairs	protein	interaction strength
control 1	pDEST-AD pDEST-DB	no inserts	none, background
control 2	pDEST-AD-E2F1 pDEST-DB-CYH2-pRB	human E2F1 aa 342-437 human pRB aa 302-928	weak (control for CHX plates)
control 3	pDEST-AD-Jun mouse pDEST-DB-Fos	mouse Jun aa 250-325 rat Fos aa 132-211	moderately strong
control 4	pDEST-AD pDEST-DB-Gal4	no insert pDEST-DB-Gal4	very strong
control 5	pDEST-AD-dE2F1 pDEST-DB-dDP	Drosophila E2F aa 225-433 Drosophila DP aa 1-377	strong
control 6	pDEST-ADCYH2-dE2F1 pDEST-DB-dDP	Drosophila E2F aa 225-433 Drosophila DP aa 1-377	strong (control for CHX plates)

4.13 Y2H mapping pipeline

The Y2H mapping pipeline is divided into four steps: primary screening, secondary phenotyping, sequence identification, and four-fold independent verification. During the primary screening yeast containing the bacterial effectors were mated against yeast containing a pool of human proteins: yeast containing the pDEST-DB with bacterial effector ORFs were arranged on 10 plates, each well containing effector-specific yeast strains, and

mated against ~ 100 plates, each plate containing ~188 human-ORF-specific yeast clones in each well (one plate contains the same 188 yeast clones in each well). To identify spontaneous autoactivators, DB-X constructs were mated against AD-empty yeast clones. Mating was conducted in the same way as described above for the identification of constitutive autoactivators, except that replica cleaning was done one day after replica plating. After 72 hours at 30°C, three colonies per spot were picked and grown in 180 µL Sc-Leu-Trp medium if the corresponding spot on the plates with the empty plasmids was empty. The plates were incubated at 30°C for 72 hours on a shaker after which glycerol stocks were prepared (80 µL 40% glycerol and 80 µL culture) in 96-well plates and stored at -80°C.

To ensure robust interactions, primary positives were retested during the secondary phenotyping. 5 µL of the glycerol stocks of the primary positives were inoculated in 180 µL Sc-Leu-Trp medium as were the Y2H controls. The plates were incubated at 30°C for 48 hours on a shaker. 5 µL of the primary positives were spotted on Sc-Leu-Trp plates and 5 µL the Y2H controls were spotted at the bottom of the plates. Plates were incubated at 30°C for 48 hours, after which the yeast spots were first replica plated and then cleaned with a fresh velvet onto Sc-Leu-Trp-His + 1mM 3-AT and then onto Sc-Leu-His + 1 mM 3-AT + 1 mg/l CHX (Sigma-Aldrich cat. no. C7698) plates. The pDEST-AD plasmids contain the CYH2 gene which confers sensitivity to CHX. If yeast cells can grow on -His plates, the DB-X was able to autoactivate, and the corresponding yeast clones were excluded from the experiment. After incubation at 30°C for 72 hours, pictures of the plates were taken (Canon 1362C005AA EF-M 28-mm F/3.5 IS STM EU11 IS/Macro) and colonies of spots that showed clear growth on Sc-Leu-Trp-His + 1mM 3-AT but not on Sc-Leu-His + 1 mM 3-AT + 1 mg/l CHX were picked and resuspended in 15 µL yeast lysis buffer on ice in 96-well plates. For the lysis buffer, 45 U of Zymolyase 20T (Amsbio cat. no. 120493) was dissolved in 1 mL 0.1M potassium phosphate buffer pH7.4 (80.2% 1M K_2HPO_4, 19.8% 1M Kh_2PO_4). Afterward, the plates were put in a thermal cycler for 15 min at 37°C followed by 5 min at 95°C and stored at 10 °C. Before freezing the plates at -20 °C, 100 µL of distilled water was added per well. Next, the samples were prepared for NGS to identify interaction candidates as described in the next section.

Identified interaction candidates were re-arrayed in a 96-well format to mate them one-on-one during the four-fold verification step to ensure reproducibility. First, "node plates" were assembled containing the yeast clones with the identified bacterial ORFs in no particular order using a liquid-handling robot (the same was done for the yeast clones containing the identified human ORFs). From the node plates, the yeast clones were re-arrayed in a specific order so that yeast clones with the bacterial interaction candidate could be mated against yeast clones with its potential human interactor. For the assembly and the re-array, 5 µL from the respective glycerol stock was inoculated in either 180 µL Sc-Leu for the DB-X constructs or in Sc-Trp for the AD-Y constructs by a liquid-handling robot. After incubation at 30 °C for 72 hours, glycerol stocks were prepared (80 µL 40% glycerol and 80 µL sample) and stored at -80 °C. The yeast clones from these glycerol stocks were inoculated, incubated, mated and replica plated plus cleaned as described for the identification of the constitutive autoactivators. After incubation at 30°C for 72 hours, pictures of the plates were taken and growth was scored using an in-house developed scoring tool as described by Kim et al [93]. This pairwise testing was done four times in total and interaction candidates that showed robust interactions three out of the four times qualified as bony fide Y2H interactors. All DB-X and AD-Y constructs were verified by NGS or Sanger sequencing and all bacterial ORFs were full-length sequenced. NGS sequencing was performed as described in the next chapter. Sanger sequencing was used for all pairs that could not be identified by NGS sequencing as well as for full-length effector

analysis. For the latter a PCR with 2 µL yeast lysis was prepared using 3 µL buffer/d'NTPs/AD or DB forward primer/term each and 0.2 µL Dreamtaq (ThermoFisher cat. no. EP0702), or KOD for constructs that were difficult to amplify. All incorrectly identified pairs were dismissed and deviations in the effector sequences were documented (Table S3).

4.14 NGS to identify candidate interaction partners

To identify interaction candidates after the secondary phenotyping, a PCR was performed using a primer binding to the Gal4 region and adding a specific taq to the DNA sequence (0.2 µL Dreamtaq DNA Polymerase (ThermoFisher cat. no. EP0702), 3 µL 2 µM term primer, 3 µL µM barcoded primer and 2 µL yeast lysis). Thereby, every ORF in the 96-well plate was assigned a different tag and a TruSeq P7 region. Each PCR plate was then pooled by pipetting 5 µL of each well into the same well of a half-deep-well plate. 30 µL of each PCR pool was purified using 24 µL magnetic beads (see Chapter 4.3) and elution was done with 25 µL elution buffer. DNA concentration of pooled PCRs was assessed using the Quant-iT™ PicoGreen™ dsDNA Assay-Kit (ThermoFisher ca. no. P7589): in dim light, solutions were brought to room temperature, and 1 x TE buffer from 20 x TE buffer was prepared as well as 1 x PicoGreen™ from 200 x stock solution. For the standard curve, lambda DNA was used for a dilution series ranging from 0.390625 ng/µL to 50 ng/µL. 200 µL 1 x TE buffer was distributed per well of a black flat-bottom 96-well plate and 1 µL of the samples and 1 µL of the 16 dilution steps was added to different wells. A plate reader (SPectraMax iD3, Molecular Devices) was used to measure the fluorescence in each well (excitation 480nm, emission 520 nm). A standard curve was generated using the lambda DNA measurements and sample concentration of the PCR pools was calculated. The pooled PCRs were diluted to a concentration of 1-2 ng/µL using an elution buffer (5 mM Tris HCl pH 8.0). Concentrations were controlled by using the PicoGreen™ protocol again. Next, a tagmentation step was performed to fragment the DNA and add adapter sequences (Illumina Tagment DNA TDE1 Enzyme and Buffer Kit cat. no. 20034197). Per well, 0.25 µL TDE enzyme, 2 µL 5 x tagmentation buffer, and 2.75 µL UltraPure DNase/RNase-Free Distilled H_2O was added on ice. 5 µL diluted PCR was pipetted per well, mixed well by pipetting up and down and incubated at 55 °C for 55 min, and stored at 4°C. To amplify the sequences and add plate-specific tags, a PCR with a specific combination of index5 and index7 primers was performed that can bind to the previously added adapter sequence and the TruSeq P7 region respectively: 8 µL PCR-Mix was transferred to the tagmentation mix (0.2 µL Dreamtaq DNA Polymerase (ThermoFisher cat. no. EP0702), 1 µL 10 µM Nextera i7 indexed primer and Nextera i5 indexed primer) and subjected to the following cycler program: 72 °C for 3 min, 10 cycles at 98°C for 10 sec each, 56 °C for 30 sec, 72 °C for 3 min and storage at 4°C. Samples were run on a 1% agarose gel to control for successful tagmentation. 10 µL of the PCRs were pooled and again cleaned with magnetic beads according to Chapter 4.3 (80 µL of beads per 100 µL of PCR and eluted in 30 µL elution buffer). Sequencing was performed by the Genome Analysis Center at Helmholtz Munich using MiSeq Reagent Kits v2 (Illumina cat. no. MS-102-2002). The sequencing data were demultiplexed using bcl2fastq2 (v2.20.0.422) provided by Illumina [93].

4.15 PRS and RRS

For the bhLit_BM-v1, interactions between Pseudomonadota effectors and human proteins were identified from published data. 175 papers were identified by programmatically searching the IMEx consortium protein interaction databases through the PSICQUIC web service using

the T3SS effectors UniprotKB accession numbers. Interactions were identified and selected as described in the results. Pathogenic effector ORFs were ordered at Twist Bioscience (San Francisco, CA, 660 USA) as clonal genes in an entry vector and were treated as the respective metagenomic effectors: ORFs were cloned by LR into pDEST-DB, propagated in *Escherichia coli* DH5α, and mini-prepped. The yeast strain Y8930 was subsequently transformed with the effectors.

For the bhRRS-v1, 100 interaction pairs were randomly selected from the bacterial bhLit_BM-v1 and the HuMEOme_v1 as well as from the human ORFeome9.1.

4.16 Assay sensitivity

Yeast clones containing the human proteins of the bhLit_BM-v1 and bhRRS-v1 were picked from the hORFeome9.1 using a liquid handling robot and sequence-verified by end-reads using Sanger sequencing. The yeast clones containing effector proteins of the bhLit_BM-v1 and bhRRS-v1 were arranged on 96-well plates in a manner that the well-position of each yeast clone matched the well-position of yeast clones with its human counterpart on the respective mating plates. The yeast clones containing the effectors were mated against yeast clones containing the human proteins four times in a pair-wise manner following the four-fold verification step (Chapter 4.13). Additionally, yeast clones with an effector ORF were mated against yeast clones with an empty pDEST-AD plasmid, and yeast clones with a human protein were mated against yeast clones with an empty pDEST-DB plasmid. The human reference sets were tested in two configurations i.e., each protein was tested as AD-Y or DB-X construct. In addition, the hsPRS-v2/hsRRS-v2 were tested following the same protocol. The latter reference sets were provided by the Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Boston, MA. Growth was analyzed using an in-house developed scoring tool. All pairs that exhibited growth at least three times out of the four repeats were considered bona fide Y2H interactors.

4.17 Sampling sensitivity

Sampling sensitivity was assessed using repeat screens. 288 effectors were screened three more times in addition to the main screen, against 1,475 human proteins. To estimate the number of total interactions that can be detected with the employed Y2H pipeline, a saturation curve was calculated as previously described [225]. This approach is based on the Michaelis-Menten equation to determine the number of detected interactions per additional repeat screen and thereby determine saturation.

First, all combinations of all four repeat screens were gathered, and the average and standard deviations were calculated. The reciprocal values of the number of interactions detected during the four repeats were obtained and plotted against the reciprocal values of the repeats. From this linear regression, the slope and intercept were determined. The Lineweaver-Burk double reciprocal plot "$1 / v = 1 / V_{max} + K_m / V_{max} \times 1 / [S]$" with v being the rate of reaction and [S] being the concentration of the substrate was used to calculate V_{max} and K_m [226]. V_{max} was calculated as the reciprocal value of the intercept whereas K_m was obtained from the formula slope/intercept. Using "$f(x) = V_{max} \times x / K_m + x$" with x representing the number of repeats, the number of interactions detected per repeat was calculated.

4.18 Precision determined by the yN2H

A subset of HuMMI was selected by randomly picking 200 interactions. Corresponding ORFs in entry clones were transferred into pDEST-N2H-N1 (LEU2 selection marker) and -N2 (TRP1 selection marker) according to Chapter 4.9. The same was repeated for the ORFs in entry clones of the four reference sets. Cloning success was assessed by PCR and ORFs corresponding to incorrect bands on the agarose gels were excluded from the experiment. *S. cerevisiae* Y8930 (MATα) and Y8800 (MATa) were transformed with the plasmids. The yN2H was performed as previously described [93] testing the HuMMI subset together with the reference sets for assay benchmarking (bhLit_BM-v1, bhRRS-v, sPRS-v2, hsRRS-v2). Briefly, 5 μl of two yeast clones that were to be mated were added to 160 μl YEPD medium and incubated overnight. For background control, each yeast clone was mated against yeast containing the respective empty pDEST-DB or pDEST-AD plasmid. 10 μl of the mated yeast culture was transferred to 160 μL SC-Leu-Trp and grown overnight. Of this culture, 50 μl were inoculated in 1.2 ml Sc-Leu-Trp and grown for 24 h at 1000 rpm. Cell pellets were obtained by centrifugation at 3000 rpm for 15 min and resuspended in 100 μl NanoLuc Assay solution (Promega cat. no. 1120). Incubation occurred for 1 hour at room temperature in the dark in white flat-bottom 96-well plates. Luminescence was measured using a plate reader (SPectraMax iD3, Molecular Devices) with 2 sec integration time, and was normalized by division through the highest luminescence value of the two background measurements. Normalized luminescence ratios (NLRs) were $\log_2$ transformed and the number of hits above the threshold ($\log_2$ NLR = 0) was counted to determine the positive fraction for each dataset. Statistical analysis was performed by Fisher's Exact test.

4.19 Homology test

Next to the main screen and the repeat screen a homology test was performed to assess the interaction similarity of homologous effectors. Therefore, effectors were clustered according to sequence similarity over 90%-sequence-length using the Needleman-Wunsch algorithm. All effectors sharing 30% or more sequence similarity were assigned to the same cluster, which led to 122 clusters consisting of 2-13 effectors. If an effector of a homology cluster interacted with a human protein all other effectors of that homology cluster were tested against that particular human protein. In total, 743 additional interactions were tested during the four-fold verification step. Effector ORFs in yeast were assembled in a 96-well format as were the yeast clones with the human proteins against which the effectors were to be tested. The protocol of the four-fold verification step and subsequent sequencing was followed as for the main screen (Chapter 4.13 and Chapter 4.14). To know which correlation test to use sequence similarity and Jaccard index measurements were tested for normal distribution using the Kolmogorov Smirnov test. Both datasets were not normally distributed, and the Spearman correlation was calculated.

4.20 Y3H assay

HuMMI was merged with HuRI using the Cytoscape software to obtain information on the human interaction partners of the effector targets. By selecting interaction partners of REL, a subnetwork was generated that could be analyzed concerning overlaps between REL interaction partners and effector interaction partners.

Effectors were transferred into pVTU-DEST according to Chapter 4.9, propagated in *Escherichia coli* DH5α (Chapter 4.6), and extracted according to Chapter 4.7 in a small batch. Effector ORFs were sequence-verified by end reads using Sanger-Sequencing. Y8930-yeast strains containing the human interactors of REL as DB-X constructs and Y8800-yeast strains containing REL as AD-Y were picked from the human ORFeome9.1 and were sequence-verified by end-reads using Sanger-Sequencing. Yeast clones containing REL were grown as pre-culture in Sc-Trp medium according to Chapter 4.11. This pre-culture was used to transform yeast cells containing REL as AD-Y with the effectors in pVTU-DEST as described in Chapter 4.11 with some exceptions: yeast cells were incubated 1.5 h at 30°C before the heat shock and Sc-Trp-Ura plates/medium were used as pVTU-DEST contains an uracil metabolic selection marker. Yeast clones containing REL and one effector were mated against yeast clones containing REL's interaction partners according to the four-fold verification step: 5 µL of the yeast clones were inoculated in 180 µL Sc-Leu medium for DB-X constructs and empty pDEST-DB, and Sc-Trp-Ura medium was used for yeast clones with REL + effector ORF. Sc-Trp was used for yeast clones with empty pDEST-AD. Plates were incubated at 30°C for 72 hours on a shaker (~180 rpm). 5 µL of the yeast culture containing an empty pDEST-DB were spotted on YEPD plates using a liquid-handling robot. After letting the spots dry for a couple of minutes the yeast cultures containing REL plus the effectors were spotted on top. Furthermore, yeast clones containing an empty pDEST-AD were mated with yeast clones containing one of REL's interactors. Yeast clones containing REL were mated against yeast clones containing one of its interaction partners. Lastly, yeast clones containing REL and one effector were mated against yeast clones containing one of REL's interactors. The plates were incubated at 30°C for 24 hours after which the yeast colonies were transferred from the YEPD plates onto Sc-Leu-Trp-Ura-His + 1mM 3-AT (Sigma-Aldrich cat. no. A8056) plates using a replica-plating block and sterile velvets [114]. To remove some of the yeast cells to reduce background, Sc-Leu-Trp-Ura-His + 1mM 3-AT plates were cleaned by pressing them down on a fresh sterile velvet. Plates were incubated at 30°C for 72 hours. The growth of yeast clones containing all three plasmids was compared to the growth of yeast clones containing only REL with the respective interactor of REL.

4.21 Functional enrichment analysis of effector targets

Functional enrichment analysis was performed using the g:Profiler web service https://biit.cs.ut.ee/gprofiler/gost. The list of the 60 effector targets subject to convergence and present in HuRI was uploaded to the g:Profiler website. The 8,275 human genes interacting in HuRI [90] were uploaded as "statistical domain scope" and "custom over all known genes" was selected. As a significance threshold "Benjamini Hochberg FDR" was used. Significant terms were only detected in the Gene Ontology "Biological Processes" database. The odds ratio was calculated based on the g:Profiler output file with HuRI as background as follows:

[Intersection size / (effective domain size - term size)] / [term size / (60 effector targets - intersection size)]

The expression of human proteins was assessed using the Human Protein Atlas accessed via https://www.proteinatlas.org/ [178]. The annotations of each human protein were obtained from the GO database via https://www.ebi.ac.uk/QuickGO/ [140].

4.22 Cell culture and transfection

HEK293 cells (BioCat cat. no. PC-002) were grown on a 100 mm cell culture dish in 10 mL DMEM (Sigma-Aldrich cat. no. D6546) + 10% Fetal Bovine Serum (ThermoFisher cat. no. 10437028) + 1% L-Glutamin (ThermoFisher cat. no. 25030081) at 37°C and 5% CO_2. Cells were washed with 5 mL PBS (Sigma-Aldrich cat. no. D8537) and incubated with 0.05% Trypsin/EDTA (Sigma-Aldrich cat. no. 59417C) for 5 min for cell splitting. Typically, a 1:10 split was performed every 3-4 days to maintain healthy cells. Transfection was performed using X-tremeGENE™ HP DNA transfection reagent (Sigma-Aldrich cat. no. 6366236001) with a ratio of 1:3 DNA to transfection reagent.

For a 6-well plate, 300,000 cells per well were seeded 18-24 hours prior to transfection. The transfection reagent was brought to room temperature and 200 µL of serum-free DMEM was pipetted into a 1.5 mL reaction tube. 2 µg of DNA and 6 µL of the transfection reagent were added, carefully vortexed, and incubated at room temperature for 15 min. Meanwhile, the medium of the cells was removed, and 2 mL of fresh medium was added to the wells. The transfection mix was slowly pipetted onto the cells and distributed by carefully swiveling the plate. After 24 hours a subsequent assay was performed.

For a 12-well plate, 100,000 cells were seeded in 1 mL medium and 73 µL serum-free DMEM was used for the transfection mix. Amounts of DNA and transfection reagent can be obtained from the chapters describing the different experiments.

4.23 Cell viability assay

The CellTiter-Glo® Luminescent Cell Viability Assay (Promega cat. no. G7570) was performed as described in the manufacturer's protocol. A standard curve was established to correlate luminescent output with viable cell numbers: four repeats of a twofold dilution of HEK293 cells were made, starting with 50.025 cells per well and performing 14 dilution steps. The standard curve was calculated, and all subsequent experiments relied on it to determine the viable cell number based on the luminescence measured. 59 effector ORFs were cloned into pMH-FLAG-HA (Chapter 4.9 and Chapter 4.6 and Chapter 4.7) and HEK293 cells were transfected with the plasmids in a 96-well format (per well: 20 ng effector, 80 ng empty pMH-FLAG-HA and 100 ng GFP in pMH-FLAG-HA). Two effectors that were far away in the network from human proteins involved in apoptosis were used as negative control (met_26 and Vfu_12) (control 3). Other controls included: cells transfected with 80 ng empty pMH-FLAG-HA and 100 ng GFP in pMH-FLAG-HA (control 2) and untransfected cells (control 1). Luminescence was quantified using a plate reader (SPectraMax iD3, Molecular Devices) in relative light units (RLU). Cell viability percentages were calculated by using the mean of 16 samples of control 2 as 100% viable cells for each experiment. The standard deviation around 100% for every time point is the average of the % standard deviations of both repeats, which are each based on 16 samples of control 2.

4.24 Click-iT™ Plus TUNEL-Assay-Kit for detecting apoptosis

The TUNEL assay (ThermoFisher cat. no. C10618) was performed according to the manufacturer's protocol: HEK293 cells were grown on Collagen I-coated coverslips (VWR cat. no. 734-1009) and transfected with 200 ng of the respective bacterial effector in pMH-FLAG-TAG, 800 ng empty pMH-FLAG-HA plasmid and 1000 ng GFP in pMH-FLAG-TAG in a 6-well

format. As a negative control, the effectors of control 3 of the cell viability assay were transfected in the same manner. As a positive control, cells were treated with ~ 1.8 U of rDNase diluted in 100 µL reaction buffer for rDNase (rDNase set by Machery-Nagel cat. no. 740963) at 37°C for 10 minutes before TdT reaction. The manufacturer's protocol for "cells grown on coverslips" was followed. Afterward, the coverslips were fixed onto microscope slides using nail polish and examined under a Nikon Eclipse Ts 2 microscope. To identify cells that were GFP-positive, indicating successful transfection, and Alexa Fluor-positive, marking apoptotic cells, the Cell-ACDC software was used [227]: GFP-positive cells were segmented, and the picture of the Alexa Fluor channel was laid over the segmentation mask to detect overlaps. This was done with the bright-field image as a background to ensure the fluorescence originated from within the cells.

4.25 NF-κB reporter assay

To analyze the NF-κB subnetwork, HuMMI was merged with HuRI using the Cytoscape software. Human proteins involved in NF-κB signaling (see results) were selected as well as the effectors targeting any of those human proteins.

4.25.1 Initial screen and titration

HEK293 (DSMZ Cat# ACC-305, RRID:CVCL 0045) were grown in DMEM with 10% FBS with 100 U/mL penicillin and 100 U/mL streptomycin (Thermo Fisher cat. no. 15140122) at 37°C and 5% CO_2. 24 hours prior to transfection by calcium phosphate, 1×10^6 cells were seeded in a 60 mm cell culture dish in 3 mL medium. On the day of transfection, 10 ng NF-κB reporter plasmid (6 × NF-κB firefly luciferase pGL2), 50 ng pTK reporter (renilla luciferase), and 2 µg bacterial ORF in pMH-FLAG-HA were mixed with 200 µL 250 mM $CaCl_2$ solution (Carl Roth cat. no. 5239.1). As controls, the respective DNA concentration of NF-κB-activating IKKβ (in pRK5 with a Flag-tag) and NF-κB-inhibiting A20 (in pEF4 with a Flag-tag) were used. For the titration experiments expression vectors of 2 µg, 4 µg, and 6 µg were transfected. The mix was vortexed and added dropwise to 200 µl 2 × HBS (50 mM HEPES (pH 7.0) (Carl Roth cat. no. 9105.4), 280 mM NaCl (Carl Roth cat. no. 3957.2), 1.5 mM $Na_2HPO_4 \times 2 H_2O$ (Carl Roth cat. no. 4984.1), pH 6.93) which was slowly vortexed. After incubating the transfection mix at room temperature for 15 min, it was added dropwise to the HEK293 cells, and the plate was carefully swiveled to distribute the mix evenly. Cells were incubated for ~ 6 hours, after which the medium was changed. 24 hours after transfection, cells were treated with 20 ng/ml TNFα (Sigma-Aldrich cat. no. SRP3177) for 4 hours before they were washed on ice twice with 2 mL PBS and lysed with 350 µL 1 x Passive Lysis Buffer (PLB, see dual luciferase reporter kit) at room temperature for 15 min on a shaker. The lysate was transferred to a tube and after a centrifugation step at full speed at 4°C, the supernatant was transferred to a new tube and stored at 4°C or on ice. 1 µL of the supernatant was added to 9 µL of 1 x PLB on a white flat-bottom 96-well plate. To determine the luciferase activity the dual luciferase reporter kit (Promega cat. no. E1980) was used according to the manufacturer's protocol and the RLU was quantified with a luminometer (Berthold Centro LB960 microplate reader, Software: MikroWin 2010). NF-κB induction was calculated as F/R, the ratio of firefly luminescence to renilla luminescence. The Kruskal-Wallis test with Dunn's correction was performed using the R software.

4.25.2 NF-κB reporter assay in HeLa cells

HeLa (RRID: CVCL_0030; DSMZ) cells were grown in DMEM (Sigma-Aldrich cat. no. D6546) + 10% Fetal Bovine Serum (ThermoFisher cat. no. 10437028) at 37°C and 5% CO_2. 100,000 cells per well were seeded in a 12-well plate 24 hours prior to transfection with the X-tremeGENE™ HP DNA transfection reagent as described in Chapter 4.22. A 1:40 ratio between the NF-κB reporter (with firefly gene) (1.6 ng) and the background plasmid (with renilla gene) (64 ng) (see Chapter 4.25.1) was employed for each sample. Additionally, the cells were transfected with 500 ng of the respective effector in pMH-FLAG-HA. As control, HeLa cells were also transfected with 500 ng empty vector, 500 ng IKKβ, and 500 ng A20. Cells were washed and lysed as described in Chapter 4.25.1 and luminescence in the supernatants was measured using a plate reader (SPectraMax iD3, Molecular Devices) and white flat-bottom 96-well plates (Greiner Bio-One cat. no. 655904). Cells were treated with TNF to examine NF-κB inhibition according to Chapter 4.25.1.

4.25.3 NF-κB reporter assay using the GAPDH promoter

HeLa (RRID: CVCL_0030; DSMZ) or HEK293 cells (BioCat cat. no. PC-002) were transfected and treated as described in Chapter 4.22 and Chapter 4.25.2. However, the pTK renilla plasmid was exchanged for the pGAPDH_PROM_01_Renilla SP Luciferase (Switch Gear Genomics cat. no. S721624) which we thankfully received from Jun. Prof. Konstantin Sparrer from the University of Ulm. The ratio between the NF-κB reporter (1.2 ng) and the background plasmid (48 ng) was 1:40. 380 ng of the effector ORF, empty vector, or controls was used. Cells were treated with TNF to examine NF-κB inhibition according to Chapter 4.25.1.

4.26 Western Blot

To control for protein expression, 12.5% western blot gels were poured to be able to separate proteins by SDS-PAGE. 5 µL of each sample (40 µL supernatant and 15 µL Roti®-Load 1 (Carl Roth cat. no. K929.1)) and 2 µL of PageRuler™ Prestained Protein Ladder (ThermoFisher cat. no. 26616) were loaded and the gels run at 120 V for 2 hours. Proteins were transferred with an electrophoretic semi-dry blotting system on polyvinylidene fluoride (PVDF) membranes (Merck Millipore cat. no. IPVH00010) at 70 mA per gel for 1 hour 50 min. After blotting, the membranes were washed shortly in PBS-T (1 x PBS with 0.1% Tween-20) and then blocked with 5% milk in PBS-T for 1 hour at room temperature. The membranes were then washed twice in PBS-T and incubated in the respective primary antibodies in 2.5% BSA in PBS-T at 4°C overnight. Primary antibodies were: anti-Actin beta (SCBT cat. no. sc-47778, RRID:AB_626632), which was used at a 1:10,000 dilution, anti-FLAG M2 (Sigma Aldrich cat. no. F3165, RRID:AB_259529) at a 1:500 dilution and anti-HA (Sigma-Aldrich cat. no. 11583816001, RRID:AB_514505) at a 1:1,000 dilution. The next morning membranes were washed three times in PBS-T for 15 min each followed by a 1-hour incubation in the anti-mouse secondary antibody in 1.25% BSA in PBS-T (Jackson ImmunoResearch Labs cat. no. 715-035-150, RRID:AB_2340770) at a 1:10,000 dilution. Membranes were washed three times in PBS-T for 15 min each and then incubated for 1 min in LumiGlo reagent (CST cat. no. 7003S) while swiveling. HRP-catalysed enhanced chemiluminescence was detected using a chemiluminescence film (Sigma-Aldrich cat. no. GE28-9068-36).

5. Bibliography

1. Johansson, Å. *et al.* Precision medicine in complex diseases—Molecular subgrouping for improved prediction and treatment stratification. *J. Intern. Med.* **294**, 378–396 (2023).
2. Catinean, A., Neag, M. A., Muntean, D. M., Bocsan, I. C. & Buzoianu, A. D. An overview on the interplay between nutraceuticals and gut microbiota. *PeerJ* **6**, e4465 (2018).
3. Shin, N. R., Whon, T. W. & Bae, J. W. Proteobacteria: Microbial signature of dysbiosis in gut microbiota. *Trends Biotechnol.* **33**, 496–503 (2015).
4. Slonczewski, J. & Foster, J. *Microbiology, An Evolving Science*. (W.W. Norton Company Springer Spektrum, 2011).
5. Buttner, D. Protein Export According to Schedule: Architecture, Assembly, and Regulation of Type III Secretion Systems from Plant- and Animal-Pathogenic Bacteria. *Microbiol. Mol. Biol. Rev.* **76**, 262–310 (2012).
6. Yang, J. *et al.* Mechanisms underlying legume–rhizobium symbioses. *J. Integr. Plant Biol.* **64**, 244–267 (2022).
7. Veronika Young *et al.* A gut meta-interactome map reveals modulation of human immunity by microbiome effectors. *bioRxiv* 2023.09.25.559292 (2023) doi:10.1101/2023.09.25.559292.
8. Cocquyt, T. Positioning Van Leeuwenhoek's microscopes in 17th-century microscopic practice. *FEMS Microbiol. Lett.* **369**, fnac031 (2022).
9. Artenstein, A. W. The discovery of viruses: advancing science and medicine by challenging dogma. *Int. J. Infect. Dis.* **16**, e470–e473 (2012).
10. Br Med J. Spontaneous Generation. *Br. Med. J.* **2**, 311–312 (1862).
11. Watts, S. A mini review on technique of milk pasteurization. *J. Pharmacogn. Phytochem.* **5**, 99–101 (2016).
12. Michaleas, S. N., Laios, K., Charalabopoulos, A., Samonis, G. & Karamanou, M. Joseph Lister (1827-1912): A Pioneer of Antiseptic Surgery. *Cureus* **14**, e32777 (2022).
13. MacDonald, M. A. From Miasma to Fractals: The Epidemiology Revolution and Public Health Nursing. *Public Health Nurs.* **21**, 380–391 (2004).
14. Grimes, D. J. Koch's Postulates – Then and Now. *Microbe Mag.* **1**, 223–228 (2006).
15. Howard-Jones, N. Fracastoro and Henle: a re-appraisal of their contribution to the concept of communicable diseases. *Med. Hist.* **21**, 61–68 (1977).
16. Bharati, K. & Ganguly, N. K. Cholera toxin: A paradigm of a multifunctional protein. *Indian J. Med. Res.* **133**, 179–187 (2011).
17. Ritchie, J. M., Rui, H., Bronson, R. T. & Waldor, M. K. Back to the Future: Studying Cholera Pathogenesis Using Infant Rabbits. *mBio* **1**, e00047-10 (2010).
18. Aminov, R. A Brief History of the Antibiotic Era: Lessons Learned and Challenges for the Future. *Front. Microbiol.* **1**, 134 (2010).
19. Hutchings, M. I., Truman, A. W. & Wilkinson, B. Antibiotics: past, present and future. *Curr. Opin. Microbiol.* **51**, 72–80 (2019).
20. Nicolaou, K. C. & Rigol, S. A brief history of antibiotics and select advances in their synthesis. *J. Antibiot. (Tokyo)* **71**, 153–184 (2018).
21. O'Neill, J. AMR Review Paper - Tackling a crisis for the health and wealth of nations. (2014).
22. Dethlefsen, L., McFall-Ngai, M. & Relman, D. A. An ecological and evolutionary perspective on human-microbe mutualism and disease. *Nature* **449**, 811–818 (2007).
23. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* **74**, 5463–5467 (1977).
24. Fleischmann, R. D. *et al.* Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science* **269**, 496–512 (1995).
25. Woese, C. R. & Fox, G. E. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc. Natl. Acad. Sci.* **74**, 5088–5090 (1977).
26. Lane, D. J. *et al.* Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *PNAS* **82, 6955–6959** (1985).
27. Rusch, D. B. *et al.* The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* **5**, e77 (2007).
28. Berg, G. *et al.* Microbiome definition re-visited: old concepts and new challenges. *Microbiome* **8**, 103 (2020).

29. Parte, A. C., Sardà Carbasse, J., Meier-Kolthoff, J. P., Reimer, L. C. & Göker, M. List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *Int. J. Syst. Evol. Microbiol.* **70**, 5607–5612 (2020).

30. Chang, C.-Y., Bajić, D., Vila, J. C. C., Estrela, S. & Sanchez, A. Emergent coexistence in multispecies microbial communities. *Science* **381**, 343–348 (2023).

31. Rutherford, S. T. & Bassler, B. L. Bacterial Quorum Sensing: Its Role in Virulence and Possibilities for Its Control. *Cold Spring Harb. Perspect. Med.* **2**, a012427 (2012).

32. Dieltjens, L. *et al.* Inhibiting bacterial cooperation is an evolutionarily robust anti-biofilm strategy. *Nat. Commun.* **11**, 107 (2020).

33. Institute of Medicine (US) Food Forum. Study of the Human Microbiome. in *The Human Microbiome, Diet, and Health: Workshop Summary* (National Academies Press (US), 2013).

34. David, L. A. *et al.* Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559–563 (2014)

35. Thursby, E. & Juge, N. Introduction to the human gut microbiota. *Biochem. J.* **474**, 1823–1836 (2017).

36. Scott, E. A., Bruning, E., Nims, R. W., Rubino, J. R. & Ijaz, M. K. A 21st century view of infection control in everyday settings: Moving from the Germ Theory of Disease to the Microbial Theory of Health. *Am. J. Infect. Control* **48**, 1387–1392 (2020).

37. Bloomfield, S. F. *et al.* Time to abandon the hygiene hypothesis: new perspectives on allergic disease, the human microbiome, infectious disease prevention and the role of targeted hygiene. *Perspect. Public Health* **136**, 213–224 (2016).

38. Budreviciute, A. *et al.* Management and Prevention Strategies for Non-communicable Diseases (NCDs) and Their Risk Factors. *Front. Public Health* **8**, 788 (2020).

39. Craig, J. Complex Diseases: Research and Applications. *Nature Education* 1(1),184 (2008).

40. Habib, S. H. & Saha, S. Burden of non-communicable disease: Global overview. *Diabetes Metab. Syndr. Clin. Res. Rev.* **4**, 41–47 (2010).

41. Novelli, G. *et al.* Precision Medicine in Non-Communicable Diseases. *High-Throughput* **9**, 3 (2020).

42. Braun, P. Interactome mapping for analysis of complex phenotypes: Insights from benchmarking binary interaction assays. *Proteomics* **12**, 1499–1518 (2012).

43. Guinane, C. M. & Cotter, P. D. Role of the gut microbiota in health and chronic gastrointestinal disease: understanding a hidden metabolic organ. *Ther. Adv. Gastroenterol.* **6**, 295–308 (2013).

44. Sender, R., Fuchs, S. & Milo, R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLOS Biol.* **14**, e1002533 (2016).

45. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).

46. Human Microbiome Project Consortium. A framework for human microbiome research. *Nature* **486**, 215–221 (2012).

47. Quigley, E. M. M. Gut microbiome as a clinical tool in gastrointestinal disease management: are we there yet? *Nat. Rev. Gastroenterol. Hepatol.* **14**, 315–320 (2017).

48. Lindell, A. E., Zimmermann-Kogadeeva, M. & Patil, K. R. Multimodal interactions of drugs, natural compounds and pollutants with the gut microbiota. *Nat. Rev. Microbiol.* **20**, 431–443 (2022).

49. Martin, A. M., Sun, E. W., Rogers, G. B. & Keating, D. J. The Influence of the Gut Microbiome on Host Metabolism Through the Regulation of Gut Hormone Release. *Front. Physiol.* **10**, 428 (2019).

50. Erturk-Hasdemir, D. & Kasper, D. L. Resident commensals shaping immunity. *Curr. Opin. Immunol.* **25**, 450–455 (2013).

51. Kelly, D., Conway, S. & Aminov, R. Commensal gut bacteria: mechanisms of immune modulation. *Trends Immunol.* **26**, 326–333 (2005).

52. Swiatczak, B. & Rescigno, M. How the interplay between antigen presenting cells and microbiota tunes host immune responses in the gut. *Semin. Immunol.* **24**, 43–49 (2012).

53. Domínguez-Díaz, C., García-Orozco, A., Riera-Leal, A., Padilla-Arellano, J. R. & Fafutis-Morris, M. Microbiota and Its Role on Viral Evasion: Is It With Us or Against Us? *Front. Cell. Infect. Microbiol.* **9**, 256 (2019).

54. Rahman, M. M. & McFadden, G. Modulation of NF-κB signalling by microbial pathogens. *Nat. Rev. Microbiol.* **9**, 291–306 (2011).

55. Lloyd-Price, J., Abu-Ali, G. & Huttenhower, C. The healthy human microbiome. *Genome Med.* **8**, 51 (2016).

56. Walter, J., Armet, A. M., Finlay, B. B. & Shanahan, F. Establishing or Exaggerating Causality for the Gut Microbiome: Lessons from Human Microbiota-Associated Rodents. *Cell* **180**, 221–232 (2020).

57. Borody, T. J. & Khoruts, A. Fecal microbiota transplantation and emerging applications. *Nat. Rev. Gastroenterol. Hepatol.* **9**, 88–96 (2011).

58. Rizzatti, G., Lopetuso, L. R., Gibiino, G., Binda, C. & Gasbarrini, A. Proteobacteria: A Common Factor in Human Diseases. *BioMed Res. Int.* **2017**, e9351507 (2017).

59. Oren, A. & Garrity, G. M. Valid publication of the names of forty-two phyla of prokaryotes. *Int. J. Syst. Evol. Microbiol.* **71**, 005056 (2021).

60. Liu, P. *et al.* Altered microbiomes distinguish Alzheimer's disease from amnestic mild cognitive impairment and health in a Chinese cohort. *Brain. Behav. Immun.* **80**, 633–643 (2019).

61. Cattaneo, A. *et al.* Association of brain amyloidosis with pro-inflammatory gut bacterial taxa and peripheral inflammation markers in cognitively impaired elderly. *Neurobiol. Aging* **49**, 60–68 (2017).

62. Cryan, J.F. *et al.* The Microbiota-Gut-Brain Axis. *Physiol Rev.* **99**, 1877-2013 (2019).

63. Mayer, E. A., Nance, K. & Chen, S. The Gut–Brain Axis. *Rev Med.* **73**, 439-453 (2021).

64. Torres-Fuentes, C. *et al.* Short-chain fatty acids and microbiota metabolites attenuate ghrelin receptor signaling. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **33**, 13546–13559 (2019).

65. Silva, Y. P., Bernardi, A. & Frozza, R. L. The Role of Short-Chain Fatty Acids From Gut Microbiota in Gut-Brain Communication. *Front. Endocrinol.* **11**, 25 (2020).

66. Sassone-Corsi, M. & Raffatellu, M. No Vacancy: How beneficial microbes cooperate with immunity to provide colonization resistance to pathogens. *J. Immunol. Baltim. Md 1950* **194**, 4081–4087 (2015).

67. Bhavsar, A. P., Guttman, J. A. & Finlay, B. B. Manipulation of host-cell pathways by bacterial pathogens. *Nature* **449**, 827–834 (2007).

68. Green, E. R. & Mecsas, J. Bacterial Secretion Systems: An Overview. *Microbiol. Spectr.* **4**, (2016).

69. Klein, J. A., Dave, B. M., Raphenya, A. R., McArthur, A. G. & Knodler, L. A. Functional relatedness in the Inv/Mxi-Spa type III secretion system family. *Mol. Microbiol.* **103**, 973–991 (2017).

70. O'Malley, M. R. & Anderson, J. C. Regulation of the Pseudomonas syringae Type III Secretion System by Host Environment Signals. *Microorganisms* **9**, 1227 (2021).

71. Galán, J. E. Common Themes in the Design and Function of Bacterial Effectors. *Cell Host Microbe* **5**, 571–579 (2009).

72. Zboralski, A., Biessy, A. & Filion, M. Bridging the Gap: Type III Secretion Systems in Plant-Beneficial Bacteria. *Microorganisms* **10**, 187 (2022).

73. Mattock, E. & Blocker, A. J. How Do the Virulence Factors of Shigella Work Together to Cause Disease? *Front. Cell. Infect. Microbiol.* **7**, (2017).

74. Ma'ayan, A. Complex systems biology. *J. R. Soc. Interface* **14**, 20170391 (2017)

75. Mazzocchi, F. Complexity in biology. Exceeding the limits of reductionism and determinism using complexity theory. *EMBO Rep.* **9**, 10–14 (2008).

76. Beadle, G. W. & Tatum, E. L. Genetic Control of Biochemical Reactions in Neurospora. *Proc. Natl. Acad. Sci. U. S. A.* **27**, 499–506 (1941).

77. Gayon, J. From Mendel to epigenetics: History of genetics. *C. R. Biol.* **339**, 225–230 (2016).

78. Baker, E. A., Gilbert, S. P. R., Shimeld, S. M. & Woollard, A. Extensive non-redundancy in a recently duplicated developmental gene family. *BMC Ecol. Evol.* **21**, 33 (2021).

79. Lobo, I. Pleiotropy: One Gene Can Affect Multiple Traits. *Nature Education* **1**(1):10 (2008).

80. Barabási, A.-L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* **12**, 56–68 (2011).

81. Zhong, Q. *et al.* Edgetic perturbation models of human inherited disorders. *Mol. Syst. Biol.* **5**, 321 (2009).

82. Liu, X. *et al.* Robustness and lethality in multilayer biological molecular networks. *Nat. Commun.* **11**, 6043 (2020).

83. Lim, J. *et al.* A Protein–Protein Interaction Network for Human Inherited Ataxias and Disorders of Purkinje Cell Degeneration. *Cell* **125**, 801–814 (2006).

84. Rozenblatt-Rosen, O. *et al.* Interpreting cancer genomes using systematic host perturbations by tumour virus proteins. *Nature* **487**, 491–495 (2012).

85. Nadeau, J. H. Modifier genes in mice and humans. *Nat. Rev. Genet.* **2**, 165–174 (2001).

86. Goh, K.-I. *et al.* The human disease network. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 8685–90 (2007).

87. Gonzalez, M. W. & Kann, M. G. Chapter 4: Protein Interactions and Disease. *PLoS Comput. Biol.* **8**, e1002819 (2012).

88. Vidal, M., Cusick, M. E. & Barabási, A.-L. Interactome networks and human disease. *Cell* **144**, 986–98 (2011).

89. Koh, G. C. K. W., Porras, P., Aranda, B., Hermjakob, H. & Orchard, S. E. Analyzing Protein–Protein Interaction Networks. *J. Proteome Res.* **11**, 2014–2031 (2012).

90. Luck, K. *et al.* A reference map of the human binary protein interactome. *Nature* **580**, 402–408 (2020).

91. Mukhtar, M. S. *et al.* Independently evolved virulence effectors converge onto hubs in a plant immune system network. *Science* **333**, 596–601 (2011).

92. Weßling, R. *et al.* Convergent targeting of a common host protein-network by pathogen effectors from three kingdoms of life. *Cell Host Microbe* **16**, 364–375 (2014).

93. Kim, D.-K. *et al.* A proteome-scale map of the SARS-CoV-2–human contactome. *Nat. Biotechnol.* **41**, 140–149 (2023)

94. Duan, S., Luo, X. & Dong, C. Identification of susceptibility modules for coronary artery disease using a genome wide integrated network analysis. *Gene* **531**, 347–354 (2013).

95. del Campo, J. *et al.* The others: our biased perspective of eukaryotic genomes. *Trends Ecol. Evol.* **29**, 252–259 (2014).

96. Eichinger, V. *et al.* EffectiveDB—updates and novel features for a better annotation of bacterial secreted proteins and Type III, IV, VI secretion systems. *Nucleic Acids Res.* **44**, D669–D674 (2016).

97. Arnold, R. *et al.* Sequence-Based Prediction of Type III Secreted Proteins. *PLOS Pathog.* **5**, e1000376 (2009).

98. Jing, R. *et al.* DeepT3 2.0: improving type III secreted effector predictions by an integrative deep learning framework. *NAR Genomics Bioinforma.* **3**, lqab086 (2021).

99. Goldberg, T., Rost, B. & Bromberg, Y. Computational prediction shines light on type III secretion origins. *Sci. Rep.* **6**, 34516 (2016).

100. Hecht, A. *et al.* Measurements of translation initiation from all 64 codons in E. coli. *Nucleic Acids Res.* **45**, 3615–3626 (2017).

101. Panyod, S. *et al.* Atherosclerosis amelioration by allicin in raw garlic through gut microbiota and trimethylamine-N-oxide modulation. *Npj Biofilms Microbiomes* **8**, 1–13 (2022).

102. Romano, K. A., Vivas, E. I., Amador-Noguez, D. & Rey, F. E. Intestinal microbiota composition modulates choline bioavailability from diet and accumulation of the proatherogenic metabolite trimethylamine-N-oxide. *mBio* **6**, e02481 (2015).

103. Grela, E., Kowalczyk-Pecka, D., E., H. & Matras, J. Effect of inulin and a probiotic supplement in the diet of pigs on selected traits of the gastrointestinal microbiome. *Med. Weter.* **72**, 448–452 (2016).

104. Romano, K. A. *et al.* Metabolic, Epigenetic, and Transgenerational Effects of Gut Bacterial Choline Consumption. *Cell Host Microbe* **22**, 279-290.e7 (2017).

105. Martínez-del Campo, A. *et al.* Characterization and Detection of a Widely Distributed Gene Cluster That Predicts Anaerobic Choline Utilization by Human Gut Bacteria. *mBio* **6**, e00042-15 (2015).

106. Pasolli, E. *et al.* Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell* **176**, 649-662.e20 (2019).

107. Almeida, A. *et al.* A new genomic blueprint of the human gut microbiota. *Nature* **568**, 499–504 (2019).

108. Birnboim, H. C. & Doly, J. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7**, 1513–1523 (1979).

109. Dieffenbach, C. W., Lowe, T. M. & Dveksler, G. S. General concepts for PCR primer design. *PCR Methods Appl.* **3**, S30-37 (1993).

110. Rychlik, W., Spencer, W. J. & Rhoads, R. E. Optimization of the annealing temperature for DNA amplification in vitro. *Nucleic Acids Res.* **18**, 6409–6412 (1990).

111. Paiano, A., Margiotta, A., De Luca, M. & Bucci, C. Yeast Two-Hybrid Assay to Identify Interacting Proteins. *Curr. Protoc. Protein Sci.* **95**, e70 (2019).

112. Mehla, J., Caufield, J. H., Sakhawalkar, N. & Uetz, P. A comparison of two hybrid approaches for detecting protein-protein interactions. *Methods Enzymol.* **586**, 333–358 (2017).

113. Dreze, M. *et al.* High-quality binary interactome mapping. *Methods Enzymol.* **470**, 281–315 (2010).

114. Altmann, M., Altmann, S., Falter, C. & Falter-Braun, P. High-Quality Yeast-2-Hybrid Interaction Network Mapping. *Curr. Protoc. Plant Biol.* **3**, e20067 (2018).

115. Venkatesan, K. *et al.* An empirical framework for binary interactome mapping. *Nat. Methods* **6**, 83–90 (2009).

116. Choi, S. G. *et al.* Maximizing binary interactome mapping with a minimal number of assays. *Nat. Commun.* **10**, 3907 (2019).

117. Lewis, A. C. F., Jones, N. S., Porter, M. A. & Deane, C. M. What evidence is there for the homology of protein-protein interactions? *PLoS Comput. Biol.* **8**, e1002645 (2012).

118. JetBrains Academy. The Jaccard similarity index | Text similarity | NLP | Data science | Computer science. *Hyperskill* https://hyperskill.org/learn/step/16590.

119. Statstutor. Spearman's correlation. *Spearman's correlation* http://www.statstutor.ac.uk/resources/uploaded/spearmans.pdf.

120. Guven-Maiorov, E., Tsai, C.-J. & Nussinov, R. Pathogen mimicry of host protein-protein interfaces modulates immunity. *Semin. Cell Dev. Biol.* **58**, 136–145 (2016).

121. Zhang, S., Paul, S. & Kundu, P. NF-κB Regulation by Gut Microbiota Decides Homeostasis or Disease Outcome During Ageing. *Front. Cell Dev. Biol.* **10**, 874940 (2022).

122. Cottier, S. *et al.* The yeast three-hybrid system as an experimental platform to identify proteins interacting with small signaling molecules in plant cells: Potential and limitations. *Front. Plant Sci.* **2**, 101 (2011).

123. Vidal, M. & Legrain, P. Yeast forward and reverse 'n'-hybrid systems. *Nucleic Acids Res.* **27**, 919–929 (1999).

124. Reimand, J., Kull, M., Peterson, H., Hansen, J. & Vilo, J. g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res.* **35**, W193–W200 (2007). Webserver: https://biit.cs.ut.ee/gprofiler/gost.

125. Ohno, H. & Hase, K. Glycoprotein 2 (GP2): grabbing the FimH bacteria into M cells for mucosal immunity. *Gut Microbes* **1**, 407–410 (2010).

126. Marcobal, A., Southwick, A. M., Earle, K. A. & Sonnenburg, J. L. A refined palate: Bacterial consumption of host glycans in the gut. *Glycobiology* **23**, 1038–1046 (2013).

127. Frick, J. S. *et al.* Colitogenic and non-colitogenic commensal bacteria differentially trigger DC maturation and Th cell polarization: An important role for IL-6. *Eur. J. Immunol.* **36**, 1537–1547 (2006).

128. EMBL-EBI, accessed October 2023. QuickGO::Term GO:0031623. https://www.ebi.ac.uk/QuickGO/term/GO:0031623 (2023).

129. Feng, Z., Sun, R., Cong, Y. & Liu, Z. Critical roles of G protein-coupled receptors in regulating intestinal homeostasis and inflammatory bowel disease. *Mucosal Immunol.* **15**, 819–828 (2022).

130. Geisler, F. & Leube, R. E. Epithelial Intermediate Filaments: Guardians against Microbial Infection? *Cells* **5**, 29 (2016).

131. Polari, L. *et al.* Keratin intermediate filaments in the colon: guardians of epithelial homeostasis. *Int. J. Biochem. Cell Biol.* **129**, 105878 (2020).

132. La Ferla, K., Seegert, D. & Schreiber, S. Activation of NF-κB in intestinal epithelial cells by E. coli strains isolated from the colonic mucosa of IBD patients. *Int. J. Colorectal Dis.* **19**, 334–342 (2004).

133. Jang, H.-M., Lee, K.-E., Lee, H.-J. & Kim, D.-H. Immobilization stress-induced Escherichia coli causes anxiety by inducing NF-κB activation through gut microbiota disturbance. *Sci. Rep.* **8**, 13897 (2018).

134. Bambou, J.-C. *et al.* In Vitro and ex Vivo Activation of the TLR5 Signaling Pathway in Intestinal Epithelial Cells by a Commensal Escherichia coli Strain*. *J. Biol. Chem.* **279**, 42984–42992 (2004).

135. Bhat, M. I., Sowmya, K., Kapila, S. & Kapila, R. Escherichia coli K12: An evolving opportunistic commensal gut microbe distorts barrier integrity in human intestinal cells. *Microb. Pathog.* **133**, 103545 (2019).

136. Wang, X. *et al.* 4-Hydroxy-2-Nonenal Mediates Genotoxicity and Bystander Effects Caused by Enterococcus faecalis-Infected Macrophages. *Gastroenterology* **142**, 543-551.e7 (2012).

137. Xu, X., Lai, Y. & Hua, Z.-C. Apoptosis and apoptotic body: disease message and therapeutic target potentials. *Biosci. Rep.* **39**, BSR20180992 (2019).

138. Altonsy, M. O., Andrews, S. C. & Tuohy, K. M. Differential induction of apoptosis in human colonic carcinoma cells (Caco-2) by Atopobium, and commensal, probiotic and enteropathogenic bacteria: mediation by the mitochondrial pathway. *Int. J. Food Microbiol.* **137**, 190–203 (2010).

139. Mirzarazi, M. *et al.* The OmpA of commensal Escherichia coli of CRC patients affects apoptosis of the HCT116 colon cancer cell line. *BMC Microbiol.* **22**, 139 (2022).

140. Gene Ontology Consortium *et al.* The Gene Ontology knowledgebase in 2023. *Genetics* **224**, iyad031 (2023). GO database accessed *via* https://www.ebi.ac.uk/QuickGO/.

141. Feng, Y. *et al.* Timing of apoptosis onset depends on cell cycle progression in peripheral blood lymphocytes and lymphocytic leukemia cells. *Oncol. Rep.* **17**, 1437–1444 (2007).

142. Wang, L. *et al.* Midline2 is overexpressed and a prognostic indicator in human breast cancer and promotes breast cancer cell proliferation in vitro and in vivo. *Front. Med.* **10**, 41–51 (2016).

143. Jiang, C. & Lin, X. Regulation of NF-κB by the CARD proteins. *Immunol. Rev.* **246**, 141–153 (2012).

144. Xiao, C. *et al.* TRIM27 interacts with IκBα to promote the growth of human renal cancer cells through regulating the NF-κB pathway. *BMC Cancer* **21**, 841 (2021).

145. Kumar, A., Takada, Y., Boriek, A. M. & Aggarwal, B. B. Nuclear factor-kappaB: its role in health and disease. *J. Mol. Med. Berl. Ger.* **82**, 434–448 (2004).

146. Visekruna, A. *et al.* Proteasome-mediated degradation of IκBα and processing of p105 in Crohn disease and ulcerative colitis. *J. Clin. Invest.* **116**, 3195–3203 (2006).

147. Vester-Andersen, M. K. *et al.* Increased abundance of proteobacteria in aggressive Crohn's disease seven years after diagnosis. *Sci. Rep.* **9**, 13473 (2019).

148. InfluentialPoints. Kruskal-Wallis ANOVA: Use & misuse - non-parametric ANOVA, test of dominance, test of medians, distribution of observations. https://influentialpoints.com/Training/Kruskal-Wallis_ANOVA_use_and_misuse.htm.

149. Moore, C. Introduction to Western Blotting. (2009) https://www.yumpu.com/en/document/view/28808744/introduction-to-western-blotting

150. Shifera, A. S. & Hardin, J. A. PMA induces expression from the herpes simplex virus thymidine kinase promoter via the activation of JNK and ERK in the presence of adenoviral E1A proteins. *Arch. Biochem. Biophys.* **490**, 145–157 (2009).

151. Cho, W., Hagemann, T. L., Johnson, D. A., Johnson, J. A. & Messing, A. Dual transgenic reporter mice as a tool for monitoring expression of glial fibrillary acidic protein. *J. Neurochem.* **110**, 343–351 (2009).

152. Bui, T. M., Wiesolek, H. L. & Sumagin, R. ICAM-1: A master regulator of cellular responses in inflammation, injury resolution, and tumorigenesis. *J. Leukoc. Biol.* **108**, 787–799 (2020).

153. Karunanathie, H., Kee, P. S., Ng, S. F., Kennedy, M. A. & Chua, E. W. PCR enhancers: Types, mechanisms, and applications in long-range PCR. *Biochimie* **197**, 130–143 (2022).

154. Lamesch, P. *et al.* C. elegans ORFeome version 3.1: increasing the coverage of ORFeome resources with improved gene predictions. *Genome Res.* **14**, 2064–2069 (2004).

155. Maier, C. J. *et al.* Construction of a highly flexible and comprehensive gene collection representing the ORFeome of the human pathogen Chlamydia pneumoniae. *BMC Genomics* **13**, 632 (2012).

156. Rual, J.-F. *et al.* Human ORFeome Version 1.1: A Platform for Reverse Proteomics. *Genome Res.* **14**, 2128–2135 (2004).

157. Wei, C. *et al.* Closing in on the C. elegans ORFeome by cloning TWINSCAN predictions. *Genome Res.* **15**, 577–582 (2005).

158. Carter, M. *et al.* A Trypanosoma brucei ORFeome-Based Gain-of-Function Library Identifies Genes That Promote Survival during Melarsoprol Treatment. *mSphere* **5**, e00769-20 (2020).

159. Altmann, M. *et al.* Extensive signal integration by the phytohormone protein network. *Nature* **583**, 271–276 (2020).

160. Walch, P. *et al.* Global mapping of Salmonella enterica-host protein-protein interactions during infection. *Cell Host Microbe* **29**, 1316-1332.e12 (2021).

161. Sontag, R. L. *et al.* Identification of Novel Host Interactors of Effectors Secreted by Salmonella and Citrobacter. *mSystems* **1**, e00032-15 (2016).

162. Dyer, M. D. *et al.* The Human-Bacterial Pathogen Protein Interaction Networks of Bacillus anthracis, Francisella tularensis, and Yersinia pestis. *PLOS ONE* **5**, e12089 (2010).

163. Yang, H. *et al.* Insight into bacterial virulence mechanisms against host immune response via the Yersinia pestis-human protein-protein interaction network. *Infect. Immun.* **79**, 4413–4424 (2011).

164. Memišević, V. *et al.* Novel Burkholderia mallei Virulence Factors Linked to Specific Host-Pathogen Protein Interactions. *Mol. Cell. Proteomics* **12**, 3036–3051 (2013).

165. Sarmiento, K. N. & Castillo, J. A. Genome-scale Solanum spp.-Ralstonia solanacearum interactome reveals candidate determinants for host specificity and environmental adaptation. *Eur. J. Plant Pathol.* **162**, 855–868 (2022).

166. Blasche, S. *et al.* The EHEC-host interactome reveals novel targets for the translocated intimin receptor. *Sci. Rep.* **4**, 7531 (2014).

167. Lamaze, C., Fujimoto, L. M., Yin, H. L. & Schmid, S. L. The Actin Cytoskeleton Is Required for Receptor-mediated Endocytosis in Mammalian Cells*. *J. Biol. Chem.* **272**, 20332–20335 (1997).

168. Büttner, D. Behind the lines-actions of bacterial type III effector proteins in plant cells. *FEMS Microbiol. Rev.* **40**, 894–937 (2016).

169. Friebel, A. *et al.* SopE and SopE2 from Salmonella typhimurium activate different sets of RhoGTPases of the host cell. *J. Biol. Chem.* **276**, 34035–34040 (2001).

170. Trosky, J. E., Liverman, A. D. B. & Orth, K. Yersinia outer proteins: Yops. *Cell. Microbiol.* **10**, 557–565 (2008).

171. Tahoun, A. *et al.* Mitotic Arrest-Deficient 2 Like 2 (MAD2L2) Interacts with Escherichia coli Effector Protein EspF. *Life Basel Switz.* **11**, 971 (2021).

172. Dean, P. Functional domains and motifs of bacterial type III effector proteins and their roles in infection. *FEMS Microbiol. Rev.* **35**, 1100–1125 (2011).

173. Lechuga, S. & Ivanov, A. I. Actin cytoskeleton dynamics during mucosal inflammation: a view from broken epithelial barriers. *Curr. Opin. Physiol.* **19**, 10–16 (2021).

174. Coch, R. A. & Leube, R. E. Intermediate Filaments and Polarization in the Intestinal Epithelium. *Cells* **5**, 32 (2016).

175. Chelakkot, C., Ghim, J. & Ryu, S. H. Mechanisms regulating intestinal barrier integrity and its pathological implications. *Exp. Mol. Med.* **50**, 1–9 (2018).

176. Dalton, W. B. & Yang, V. W. Mitotic Origins of Chromosomal Instability in Colorectal Cancer. *Curr. Colorectal Cancer Rep.* **3**, 59–64 (2007).

177. Ho, M. *et al.* Update of the keratin gene family: evolution, tissue-specific expression patterns, and relevance to clinical disorders. *Hum. Genomics* **16**, 1 (2022).

178. Uhlén, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015). Human Protein Atlas website: https://www.proteinatlas.org/

179. Ritter, S. L. & Hall, R. A. Fine-tuning of GPCR activity by receptor-interacting proteins. *Nat. Rev. Mol. Cell Biol.* **10**, 819–830 (2009).

180. Dyer, J., Salmon, K. S. H., Zibrik, L. & Shirazi-Beechey, S. P. Expression of sweet taste receptors of the T1R family in the intestinal tract and enteroendocrine cells. *Biochem. Soc. Trans.* **33**, 302–305 (2005).

181. Votteler, J. & Sundquist, W. I. Virus Budding and the ESCRT Pathway. *Cell Host Microbe* **14**, 232–241 (2013).

182. Wu, J. *et al.* Gut microbiota dysbiosis associated with plasma levels of Interferon-γ and viral load in patients with acute hepatitis E infection. *J. Med. Virol.* **94**, 692–702 (2022).

183. Brockhausen, I., Schutzbach, J. & Kuhns, W. Glycoproteins and their relationship to human disease. *Acta Anat. (Basel)* **161**, 36–78 (1998).

184. Bucurica, S., Gaman, L., Jinga, M., Popa, A. A. & Ionita-Radu, F. Golgi Apparatus Target Proteins in Gastroenterological Cancers: A Comprehensive Review of GOLPH3 and GOLGA Proteins. *Cells* **12**, 1823 (2023).

185. Foley, S. E. *et al.* Gut microbiota regulation of P-glycoprotein in the intestinal epithelium in maintenance of homeostasis. *Microbiome* **9**, 183 (2021).

186. Wilson, A. *et al.* Crohn's Disease Is Associated with Decreased CYP3A4 and P-Glycoprotein Protein Expression. *Mol. Pharm.* **16**, 4059–4064 (2019).

187. Kim, Y.-W., Park, J., Lee, H.-J., Lee, S.-Y. & Kim, S.-J. TGF-β sensitivity is determined by N-linked glycosylation of the type II TGF-β receptor. *Biochem. J.* **445**, 403–411 (2012).

188. Panneerselvam, D. & Vaqar, S. Peyer Patches. in *StatPearls* (StatPearls Publishing, 2023). https://www.ncbi.nlm.nih.gov/books/NBK557457/

189. Shi, J.-H. & Sun, S.-C. Tumor Necrosis Factor Receptor-Associated Factor Regulation of Nuclear Factor κB and Mitogen-Activated Protein Kinase Pathways. *Front. Immunol.* **9**, 1849 (2018).

190. Stuart, L. M., Paquette, N. & Boyer, L. Effector-triggered versus pattern-triggered immunity: how animals sense pathogens. *Nat. Rev. Immunol.* **13**, 199–206 (2013).

191. Yang, S. *et al.* Salmonella Effector SpvB Inhibits NF-κB Activity via KEAP1-Mediated Downregulation of IKKβ. *Front. Cell. Infect. Microbiol.* **11**, 641412 (2021).

192. Zha, J. *et al.* The Ret Finger Protein Inhibits Signaling Mediated by the Noncanonical and Canonical IκB Kinase Family Members1. *J. Immunol.* **176**, 1072–1080 (2006).

193. Yu, H., Lin, L., Zhang, Z., Zhang, H. & Hu, H. Targeting NF-κB pathway for the therapy of diseases: mechanism and clinical study. *Signal Transduct. Target. Ther.* **5**, 1–23 (2020).

194. Mahapatro, M., Erkert, L. & Becker, C. Cytokine-Mediated Crosstalk between Immune Cells and Epithelial Cells in the Gut. *Cells* **10**, 111 (2021).

195. Spahn, T. W. & Kucharzik, T. Modulating the intestinal immune system: the role of lymphotoxin and GALT organs. *Gut* **53**, 456–465 (2004).

196. Solari, E., Marcozzi, C., Negrini, D. & Moriondo, A. Interplay between Gut Lymphatic Vessels and Microbiota. *Cells* **10**, 2584 (2021).
197. Di Blasi, R., Marbiah, M. M., Siciliano, V., Polizzi, K. & Ceroni, F. A call for caution in analysing mammalian co-transfection experiments and implications of resource competition in data misinterpretation. *Nat. Commun.* **12**, 2545 (2021).
198. Shalon, D. *et al.* Profiling the human intestinal environment under physiological conditions. *Nature* **617**, 581–591 (2023).
199. Bradley, P. H. & Pollard, K. S. Proteobacteria explain significant functional variability in the human gut microbiome. *Microbiome* **5**, 36 (2017).
200. Medzhitov, R. Recognition of microorganisms and activation of the immune response. *Nature* **449**, 819–826 (2007).
201. Liu, T., Zhang, L., Joo, D. & Sun, S.-C. NF-κB signaling in inflammation. *Signal Transduct. Target. Ther.* **2**, 1–9 (2017).
202. Ogura, Y. *et al.* Nod2, a Nod1/Apaf-1 family member that is restricted to monocytes and activates NF-kappaB. *J. Biol. Chem.* **276**, 4812–4818 (2001).
203. Wullaert, A., Bonnet, M. C. & Pasparakis, M. NF-κB in the regulation of epithelial homeostasis and inflammation. *Cell Res.* **21**, 146–158 (2011).
204. Han, Y. M. *et al.* NF-kappa B activation correlates with disease phenotype in Crohn's disease. *PLoS ONE* **12**, e0182071 (2017).
205. Vainer, B. Intercellular adhesion molecule-1 (ICAM-1) in ulcerative colitis: Presence, visualization, and significance. *Inflamm. Res.* **54**, 313–327 (2005).
206. Haque, S. Z. & Haque, M. The ecological community of commensal, symbiotic, and pathogenic gastrointestinal microorganisms – an appraisal. *Clin. Exp. Gastroenterol.* **10**, 91–103 (2017).
207. Gill, N., Wlodarska, M. & Finlay, B. B. Roadblocks in the gut: barriers to enteric infection. *Cell. Microbiol.* **13**, 660–669 (2011).
208. Martens, E. C., Neumann, M. & Desai, M. S. Interactions of commensal and pathogenic microorganisms with the intestinal mucosal barrier. *Nat. Rev. Microbiol.* **16**, 457–470 (2018).
209. Slack, E. & Diard, M. Resistance is futile? Mucosal immune mechanisms in the context of microbial ecology and evolution. *Mucosal Immunol.* **15**, 1188–1198 (2022).
210. Viennois, E. & Chassaing, B. First victim, later aggressor: How the intestinal microbiota drives the pro-inflammatory effects of dietary emulsifiers? *Gut Microbes* **9**, 1–4 (2018).
211. Gillois, K., Lévêque, M., Théodorou, V., Robert, H. & Mercier-Bonin, M. Mucus: An Underestimated Gut Target for Environmental Pollutants and Food Additives. *Microorganisms* **6**, 53 (2018).
212. Carvalho, F. A. *et al.* Transient inability to manage proteobacteria promotes chronic gut inflammation in TLR5-deficient mice. *Cell Host Microbe* **12**, 139–152 (2012).
213. Jakobsson, H. E. *et al.* The composition of the gut microbiota shapes the colon mucus barrier. *EMBO Rep.* **16**, 164–177 (2015).
214. Qi, L. Gene-Diet Interactions in Complex Disease: Current Findings and Relevance for Public Health. *Curr. Nutr. Rep.* **1**, 222–227 (2012).
215. Correa, V. R. *et al.* The bacterium Pantoea stewartii uses two different type III secretion systems to colonize its plant host and insect vector. *Appl. Environ. Microbiol.* **78**, 6327–6336 (2012).
216. Pallen, M. J., Beatson, S. A. & Bailey, C. M. Bioinformatics, genomics and evolution of non-flagellar type-III secretion systems: a Darwinian perspective. *FEMS Microbiol. Rev.* **29**, 201–229 (2005).
217. Gemler, B. T. *et al.* Function-based classification of hazardous biological sequences: Demonstration of a new paradigm for biohazard assessments. *Front. Bioeng. Biotechnol.* **10**, 979497 (2022).
218. LeBlanc, M.-A., Fink, M. R., Perkins, T. T. & Sousa, M. C. Type III secretion system effector proteins are mechanically labile. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2019566118 (2021).
219. Sory, M. P. & Cornelis, G. R. Translocation of a hybrid YopE-adenylate cyclase from Yersinia enterocolitica into HeLa cells. *Mol. Microbiol.* **14**, 583–594 (1994).
220. Göser, V. *et al.* Single molecule analyses reveal dynamics of Salmonella translocated effector proteins in host cell endomembranes. *Nat. Commun.* **14**, 1240 (2023).
221. Fujiwara, S. Humanized mice: A brief overview on their diverse applications in biomedical research. *J. Cell. Physiol.* **233**, 2889–2901 (2018).
222. Lv, C. *et al.* Research Progress on Small Molecular Inhibitors of the Type 3 Secretion System. *Molecules* **27**, 8348 (2022).
223. Suez, J., Zmora, N. & Elinav, E. Probiotics in the next-generation sequencing era. *Gut Microbes* **11**, 77–93 (2020).

224. Nieth, A., Verseux, C. & Römer, W. A Question of Attire: Dressing Up Bacteriophage Therapy for the Battle Against Antibiotic-Resistant Intracellular Bacteria. *Springer Sci. Rev.* **3**, 1–11 (2015).
225. Arabidopsis Interactome Mapping Consortium. Evidence for network evolution in an Arabidopsis interactome map. *Science* **333**, 601–607 (2011)
226. University College London. The effect of substrate concentration on enzyme activity. https://www.ucl.ac.uk/~ucbcdab/enzass/substrate.htm.
227. Padovani, F., Mairhörmann, B., Falter-Braun, P., Lengefeld, J. & Schmoller, K. M. Segmentation, tracking and cell cycle analysis of live-cell imaging data with Cell-ACDC. *BMC Biol.* **20**, 174 (2022).

A Appendix

Table S1 | 44 T3SS-positive strains. Abbreviations (abbr.) are only stated for the 18 selected strains.

GTDB taxonomy	strain name	isolation Source	strain Repository ID	genome_accession	WGS Accession	abbr.
Aeromonas dhakensis	Aeromonas dhakensis CIP 107500	faeces of child with diarrhoea	DSM 17689	GCF_000820305.1	CDBH01000000	
Aeromonas enteropelogenes	Aeromonas enteropelogenes CECT 4487	human faeces	DSM 6394	GCF_000819845.1	CDCG01000000	
Aeromonas enteropelogenes	Aeromonas enteropelogenes CECT 4255T	human stool	DSM 7312	GCF_000820205.1	CDDE01000000	
Aeromonas jandaei	Aeromonas jandaei CECT 4228	human stool of patient with diarrhea	DSM 7311	GCF_000819955.1	CDBV01000000	Aja
Aeromonas tecta	Aeromonas tecta CECT 7082	stool of child with diarrhoea	DSM 17300	GCF_000820185.1	CDCA01000000	
Cedecea davisae	Cedecea davisae Grimont et al. 1981	stool	DSM 4568	GCF_000412335.2	ATDT01000001	Cda
Citrobacter europaeus	Citrobacter europaeus Ribeiro et al. 2017	faeces from a human with diarrhoea	DSM 103031	GCF_900079995.1	FLYB01000000	
Citrobacter youngae	Citrobacter pasteurii Clermont et al. 2015	human diarrhoeal stool	DSM 28879	GCF_000826205.1	CDHL01000000	Cyo
Edwardsiella tarda	Edwardsiella tarda ATCC 23685	gastrointestinal_tract	ATCC 23685	GCF_000163955.1	ADGK00000000	Eta
Edwardsiella tarda	Edwardsiella tarda Ewing and McWhorter 1965	human faeces	DSM 30052	GCF_000264805.1	AFJG01000001	
Enterobacter roggenkampii	Enterobacter cloacae (Jordan 1890) Hormaeche and Edwards 1960	human faeces	DSM 16690	GCF_001729805.1	NZ_CP017184	
Escherichia albertii	Escherichia albertii 19982	stool from diarrhoeal child	DSM 17582	GCF_000759775.1	BBMY01000000	
Escherichia coli	Escherichia coli MS 57-2	gastrointestinal_tract	BEI HM-342	GCF_000164615.1	ADUG00000000	
Escherichia coli	Escherichia coli MS 110-3	gastrointestinal_tract	BEI HM-343	GCF_000164415.1	ADTW00000000	
Escherichia coli	Escherichia coli MS 115-1	gastrointestinal_tract	BEI HM-344	GCF_000164235.1	ADTL00000000	
Escherichia coli	Escherichia coli MS 16-3	gastrointestinal_tract	BEI HM-345	GCF_000164495.1	ADUA00000000	
Escherichia coli	Escherichia coli MS 21-1	gastrointestinal_tract	BEI HM-346	GCF_000164355.1	ADTR00000000	
Escherichia coli	Escherichia coli MS 69-1	gastrointestinal_tract	BEI HM-347	GCF_000164315.1	ADTP00000000	Ec6
Escherichia coli	Escherichia coli MS 200-1	gastrointestinal_tract	BEI HM-356	GCF_000164535.1	ADUC00000000	Ec2
Escherichia coli	Escherichia coli MS 196-1	gastrointestinal_tract	BEI HM-365	GCF_000164555.1	ADUD00000000	
Escherichia coli	Escherichia coli MS 198 -1	gastrointestinal_tract	BEI HM-366	GCF_000164195.1	ADTJ00000000	
Escherichia coli	Escherichia coli D9	gastrointestinal_tract	BEI HM-87	GCF_000158395.1	ACDL00000000	
Escherichia coli	Escherichia sp.	human gut biopsy	DSM 24827	GCF_000157115.2	ACAC01000001	
Escherichia fergusonii	Escherichia fergusonii Farmer et al. 1985	faeces of 1-year-old boy	DSM 13698	GCF_000026225.1	CU928144	Efe
Klebsiella pneumoniae	Klebsiella sp. MS 92-3	gastrointestinal_tract	BEI HM-354	GCF_000195655.1	AFBO00000000	Kpn
Klebsiella pneumoniae	Klebsiella pneumoniae subsp. pneumoniae WGLW3	gastrointestinal_tract	BEI HM-748	GCF_000300935.1	AMLN00000000	
Klebsiella pneumoniae	Klebsiella pneumoniae subsp. pneumoniae WGLW5	gastrointestinal_tract	BEI HM-749	GCF_000300955.1	AMLO00000000	
Morganella morganii	Morganella morganii subsp. morganii NBRC 3848	stool of a summer diarrhoea case	DSM 30164	GCF_001598895.1	BCZU01000000	Mmo
Pantoea septica	Pantoea septica	human stool	DSM 24604	GCF_002095575.1	MLJJ01000000	Pse
Phytobacter massiliensis	Enterobacter massiliensis Lagier et al. 2014	human feces of a healthy patient	DSM 26120	GCF_000321045.1	CAEO00000000	Pma

Species	Strain	Source	Collection numbers	GCF	WGS	Abbr
Providencia alcalifaciens	Providencia alcalifaciens DSM 30120	gastrointestinal_tract	NCTC 10286, DSM 30120, CIP 82.90, ATCC 9886	GCF_000173415.1	ABXW00000000	
Providencia rettgeri_D	Providencia rettgeri DSM 1131	gastrointestinal_tract	DSM 1131, NCTC 7481	GCF_000158055.1	ACCI00000000	Pre
Providencia rustigianii	Providencia rustigianii DSM 4541	gastrointestinal_tract	DSM 4541, ATCC 33673	GCF_000156395.1	ABXV00000000	
Providencia stuartii	Providencia stuartii ATCC 25827	gastrointestinal_tract	ATCC 25827	GCF_000154865.1	ABJD00000000	Pst
Pseudocitrobacter faecalis	Pseudocitrobacter faecalis Kämpfer et al. 2014	stool from a hospitalized patient	DSM 27453	GCF_003315335.1	QNRL00000000	Pfa
Pseudomonas aeruginosa	Pseudomonas sp. 2_1_26	gastrointestinal_tract	BEI HM-214	GCF_000233495.1	ACWU00000000	
Pseudomonas_B luteola	Pseudomonas sp. HPB0071	gastrointestinal_tract	BEI HM-860	GCF_000478505.2	AQFP00000000	
Pseudomonas_E massiliensis	Pseudomonas sp.	human stool	DSM 29075	GCF_000826105.1	CCYK00000000	Pem
Vibrio fluvialis	Vibrio fluvialis ATCC 33809	human faeces, diarrhoea	DSM 19283	GCF_001558415.2		
Vibrio furnissii	Vibrio furnissii NCTC 13120	faeces of an adult woman with gastroenteritis	DSM 19622	GCF_900460225.1	UHIT01000000	Vfu
Yersinia aleksiciae	Yersinia aleksiciae 159	human faeces	DSM 14987	GCF_001047675.1		
Yersinia bercovieri	Yersinia bercovieri Wauters et al. 1988	human stool	DSM 18528	GCF_000167975.1	AALC01000001	
Yersinia enterocolitica	Yersinia enterocolitica subsp. palearctica Y11	human stool	DSM 13030	GCF_000253175.1		Yen
Yokenella regensburgei	Yokenella regensburgei ATCC 43003	gastrointestinal_tract	ATCC 43003, JCM 3961	GCF_000239335.1	AGCL00000000	Yre

Table S2 | Bacterial effector primer sequences.

gene ID	forward primer sequence	reverse primer sequence
WP_000013921.1	TTGGCCAACTCATCCTCAC	TGAGCTGGCAGGTGGA
WP_000020896.1	ATGAGTATTGATCGCACTTCCC	ATTACTCTGTAAATAGCTCTGAGTTTC
WP_000057371.1	ATGGCGCAGATAACGACG	ACATGATTTCGCCTCCAGATA
WP_000067801.1	ATGAGTAATCTGTCACTTCAACCC	CGGATGAAACGAATTATTCTGTCC
WP_000077885.1	ATGTCGCAACATAACGAAAGA	CGCAGGAATTTTGTCAATCTTAGG
WP_000083190.1	ATGAGTCAGTCTGACACAACG	CAGCAAGCTCTTGACATAAGAG
WP_000083435.1	ATGAGCCAGTCACTGTTTAGC	GACGCCGTAACGTTCGC
WP_000097994.1	ATGTCTTCACCTTTGTATGTTACC	ATGTAACACTGTTACCGCGT
WP_000099375.1	ATGAGTTCGACTGATAACATAGTCT	GAACACTTTGTAGACAATATTTGAAATCG
WP_000148644.1	ATGACCCATGAATACCAGCG	ATTTATCAGTACCAGCAAGGGG
WP_000178797.1	ATGACGATGAGCTTTAACACCA	TGCAGACTCCTCTGAATACTG
WP_000189184.1	ATGACACCAACAATTGAATTACTTTG	ACGCGTCGCCCAAC
WP_000192007.1	ATGACTCAACATACTCATAATGTAACC	CTGTGGATAAGGCACCCAG
WP_000208170.1	ATGACTACTATTGTTGACAGCAATC	GGCGTTTAACGCCGC
WP_000255032.1	ATGGTCAGCTCAACGACTC	AAGGTGCAAAGACGCAAGA
WP_000375129.1	ATGGATCGTATTGTTAGTTCTTCAC	ATCGCGGCTGGCGA
WP_000377424.1	ATGGACAGCATTACAACCAGG	TAACGACGATCCTTGAAGGTT
WP_000433381.1	ATGGAGCCAATCAGCTTGA	TCCCGCTTTTACCGCAG
WP_000437922.1	ATGGAACAACGCCGTATTACA	ATATGTTTCACCCAAGCCGA
WP_000438625.1	ATGGAACAGGTTGAATCCGG	GCCGTTTGTCATATGTGGC
WP_000508975.1	GTGGGCGGAATAAGCCA	CCAGACCAGTCCTCCATCC
WP_000557378.1	ATGCACACTAACTGGCAAGT	TGGTGTTTCCTCACCTTGC
WP_000558718.1	GTGCACTGGCAAACTCAC	CAGGCAAACGCCTCCC
WP_000781397.1	TTGAAGATGGCATCTAATCAAACTA	AAGGAAATGACAGAATTTACCGAC
WP_000786561.1	ATGAAAAATTGTGTCATCGTCAGT	ATTCAATCGTTCAATCACCATCG
WP_000829745.1	ATGGCAGAAAATAACCCATCATC	TCCCCAGTCCAGCCCC
WP_000904613.1	ATGTTAATTTTCCAGATTGCCAATAAG	AAAAAGTGAACCAAGAAGACCG
WP_000937458.1	ATGTTGCCGCAAGAAATTGA	ATCTTGAGGTGTTGAAGCCG
WP_000995825.1	GTGAATTGCAGCTGCAACG	ATCGTAATAAGCTGGAATTGATTTTC
WP_000999547.1	ATGAACGATCCCATCTCAACA	TCGCGCCTTACTCCAGG
WP_001016304.1	ATGAACCACAATCAAAATTATGGAATG	GTGAGTTGCCTTTCTGATGG
WP_001023055.1	ATGAATATATTATTCACTGAATCGTCGC	TTCACAACTAGCAAAACCTGG
WP_001066218.1	ATGAACAACACCGCAGTGA	TACAGCTTTCTCCTTCTCAGC
WP_001067513.1	ATGAACCCCGCAGTGGA	GCGGAATTTACGCCGATAC
WP_001143213.1	ATGCCCACTTCTCATGAAAATG	CTGACTAAAGCGCATCTGC
WP_001147116.1	ATGCCTTACTCTTATCGCACTA	TTCCTGTGGATCAATGGGC
WP_001182890.1	ATGCAACCACTTACATACCAAC	TATTGCCCCCAACATGGTG
WP_001197909.1	ATGCAGGTGTTACCCCCG	CGCCTCCTCTTCATGACG
WP_001235473.1	ATGAGACTCGAAAGCGTAGC	ACGTGTGACCGCGTTC
WP_001237041.1	ATGCGGCTTATTAATTATTTTCCAAA	TTCGCGGCTGATAATGTCC
WP_001272443.1	ATGAGTGCGGGGAAATTGC	GGATTTTTTACGTGAGGCTTTTTA
WP_001278605.1	ATGAGCGACGACAATTCACA	TTCATCCAGCTTTGGCTGG
WP_001299868.1	TTGGGAAACGTTCTGACATGG	GTGATTTTTCTGGTAAAAATTATCCAGA
WP_001317460.1	ATGTCTGTCACAATTCAGGGG	GACGTAAATATCAATGGTGCCA
WP_001406816.1	ATGATACGCATTATCTCAAGAGC	GTCATGCGGAATCTGTCCT
WP_002431388.1	ATGATGTCACTGACACTGAATGT	CACTATGGCAAAACTCAGGTC
WP_002431733.1	ATGAGTAAAGTGAAAAGCATCACA	CAAGAAAGATTTGAACGGCAGA
WP_015953472.1	GTGCGTAGCCGTTTTTCG	ACTGTAGGCATGTAGCGTAC
WP_015953559.1	ATGGATTGCCGTCCGGA	GGGAGCTGTCAGGGCT
WP_024256417.1	ATGGAAGCCGGAAAGATGAC	GAAATAATATTCAACGGTTTCTTCTTCT
WP_032243086.1	ATGCGAAGTGAACAGATTTCTG	TCTCATGTCTTTTCGCTGCC
WP_071821796.1	TTGATCAATAACATTGAGTGTCGC	CAGGCGATGGAACAGATTAGC
WP_077626319.1	ATGTATCATTCCATTTATACTCGTCATA	ATTCTTTTTTTCTTGCCCTGGA
WP_077626322.1	TTGAAGTGTTCATGGCAGTCT	ATGATCTTGTTCAATAACAATACTTGTG
WP_077626326.1	GTGCCGCAACGAAACG	CGCCGTACATTTCACCAGA
WP_005156400.1	ATGGAGTGGCAGACTCATTTA	CATAATATTGCCGGAGCCTTC
WP_005156427.1	ATGGCGTGGAATCAGCC	TTCTCTCCCTACGCGAGT
WP_005156531.1	ATGAATATTGATTTGATGCCATATCAC	TGAATGGCTTTCCGAGGTAA
WP_005156566.1	ATGAGAACAATTAGTTATAGTGAGGC	CTCTTCAAGTTCCCTTTGAATTCC
WP_005156690.1	ATGAATATGACAGTTAAACATCAAATACA	AAACTTCGCTGCTGTGGC
WP_005156692.1	ATGTCGCATTCCATTGGTATTAA	TACCAATGAACCGGCGATT
WP_005156700.1	ATGACATTGAATACTAATGTGTTAAGTAG	GCAAAGATTGCGGGCAA
WP_005157177.1	ATGGTAGCCCAAATAAGCAGTA	GAAAGAGAACTGTTTCAGCTGA
WP_005157407.1	ATGATCAGCGTTTCTATCGATATC	TTTGTGCAAAATTCGAACACCA
WP_005157433.1	ATGAAAATCGTGAGTAACTTTATCGG	ACCCAGCGTTGGCATAC
WP_005157512.1	ATGAGCGACATCACAAAAGGA	TTTTGGTGTCAGGCTACGT
WP_005157598.1	ATGCAGTCTGTATCTTCTATTTTTGT	CCAGGGTTTCTCCACAACA
WP_005157674.1	ATGAGTGAGGCTACTGTCAATG	TCGTAACGCCTCCACAATC
WP_005157827.1	ATGACATCGCTAATATTTAAAAACGTAA	AAGTATGTCCATACTTTCTTCATATATGA
WP_005157881.1	ATGACAACATTTAACCCCAGAATG	TAAATGAGTGACCTCACTCATGG
WP_005158044.1	GTGAAACAAAACTCCGAAAAAAGT	AGTGTCAGCGCTATAGTCATT
WP_005158077.1	ATGCGAGTGAACAAACCTGT	ATTCACCGGACGATAAATTCTCA
WP_005158295.1	GTGCCCCCCGCAGCG	ATCTCTTATAGGTGGTGAACTGGG
WP_005158416.1	ATGAGCGACGACCACTCA	TTCTTCCAGTTTGGGTTGCG
WP_005158559.1	ATGAAATACCAATTTCCAGATAACTTCT	ATCAAAACCATTATTTGCCGCG
WP_005158823.1	ATGAACATCACTACATCGGCG	CAGCAGGCCGCCAC
WP_005159058.1	ATGGGTGAAAACCTCTGGATAT	ATCCGGGTACTTTATCTGATATCG
WP_005159145.1	ATGATAACTCCATCGACTTCCTC	TGTTGTCTTCCCCTGCTCC
WP_005159272.1	ATGGCACAGCTAAATAATTGGC	CCAACTCAACTTATCCACAAAATCA
WP_005159550.1	ATGAATACTGAAGCCTCCCAA	CTCTACACCGACAGTCAGG
WP_005159572.1	ATGTCACAATCCTCCGCG	GTTAGCGGCTTTACGCTG
WP_005160046.1	ATGCGTACTCAATCACTAAAACC	ATGATACAAAGAAGAATGAATTTTTAACG
WP_005160467.1	ATGAGTATTGATCGCACTCAGC	GCTGCTCTTCAGGTCATCG
WP_005160776.1	ATGAGTACGCTTCTATATATTCACGG	AGCGGTCGCTAGCCC

111

gene ID	forward primer sequence	reverse primer sequence
WP_005160863.1	ATGAAATCATTAACCAGCCGAC	ATTAATATTGGTTTGTTTCAGCAATGC
WP_005161939.1	ATGTCACTGCAAACTCAAGAC	TTTCCGCGTCTCCCCG
WP_005162155.1	ATGCATCGAGCTAAACCTATGC	CGCTTTGTCGGCCTCA
WP_005162175.1	GTGACTAAACCTGTATCTCAAGAAA	AACGTAGTTCTGATACCAGTCG
WP_005162179.1	ATGCAAGTGATGCCTCCAA	CGGCTGTTCCCGTCC
WP_005162291.1	ATGGCCGATATTAATACAGCACA	GAGATGCAAGGATGCAAGAC
WP_005162623.1	ATGTCGACTGATAAAACAGGGT	CTCAGTACGACTGGTGCG
WP_005162781.1	ATGCGCATCACACAGCT	CTGCATCGTGGATGAAAAATTC
WP_005162811.1	ATGACTGATTTTATCCCCACTGA	AGCCAGTTGCACATCCC
WP_005162998.1	ATGTCACAATCGCCAATTGAG	ATTTAAAGGTTGTATAGTTAATGCGTTAT
WP_005163324.1	ATGTTGAGCGGAAATGGCA	GCTCAGAGTAAATAATTGAGCGA
WP_005163332.1	ATGGAAAACATAAATCAATCTCAGCA	GCTTACGAATAAGTTAGGTTGCA
WP_005163729.1	ATGCAAAATTCAGCACAAAAAATATG	GTGTTGCGCCTGCGC
WP_005163816.1	ATGGCCGTCACTATCACATAT	AGCAGATAATAATAATTCATCATTATCGG
WP_005163828.1	ATGACAGTAAGTATTCAATATAAAAGCCA	GGCAGCCATCAGTATTTCGG
WP_005163940.1	ATGTCCAGAGCTAGCCGC	AGGCCCCAACCGGC
WP_005164084.1	TTGGACCAACAGGCAAGC	ATACACGAAACTTACCGGCG
WP_005164130.1	ATGACTCAGAATCTCAGCCAG	AGCACGCAGGATATAGTCAC
WP_005164132.1	ATGGTAAGTACTCAATTTATCGGTAG	ACCATGAATTTTTTGCCAAATTAACA
WP_005164223.1	ATGAGCGAGAAGTCAGAAAACC	TCCGCGCTTATTAGACCC
WP_005164331.1	ATGAATCCGTTGGTGTATTTCTC	GTTCTGTCGTTGCCAAAATTC
WP_005164342.1	ATGAGTGATCAAATCCAAAGTAGC	CTTAGTGTTACGAAATGGTTTGGT
WP_005164542.1	ATGAATAGGCAACAACGAGTAGA	CTCCTGTATCTGATGAACAACATC
WP_005164848.1	ATGTCGAATACTATCTCTACAGCA	AATATAAACGTCAATCTGGTTTTGAG
WP_005165009.1	ATGCAAAGCCAAACCACCG	TGGCTTCGGCTCCTCT
WP_005165338.1	ATGCGTATAAGCGAAATTGATTATG	CATTAAGCCGGGGTTAAACTC
WP_005165350.1	ATGTCTAACTGGAACGATATAGCA	AGCCAAGACATGGAAGGATATG
WP_005165744.1	ATGAAAGGGTATAGAACAAACAGC	TAGCGAAAATTTTTTAGTTGCGG
WP_005165873.1	ATGTTCCAAGACCAAGACACG	ATCCAGTAAAAGGTGAGGGTG
WP_005165949.1	ATGAGCGAAACCCATACTACC	GCCTTTCCAGGGATGAATAAC
WP_005165954.1	ATGGCATCTAATACTCAAGTTTTGC	CAAAAAGTGGCCAAACTTGC
WP_005166097.1	ATGTCTGAAGGGTTCAGCC	TTGGGATATGCGCCCAAA
WP_005166456.1	ATGCCAGTTGGACTGGC	GACCGTTAGTGCTGTATCACTA
WP_005177203.1	ATGTCTTTAACAAAAATATCGTCATCTT	CGCTTTTTCATGAGGATTACGG
WP_005177600.1	ATGCAAAATAACTCGTTATCTACCA	TGGGTTACCTTGCTGCAA
WP_005178231.1	ATGATTTCCGCTTCACAACG	TTCAAAGAGATTGTGATGCAGC
WP_005178976.1	GTGAATGGTGTAGTGAGTGGG	AAAAGCTGGCCCTGTTTCG
WP_005179029.1	ATGAGTTTCTCTATTGGACAAGC	GCCCAGTAATTTTTCGATGCG
WP_005179266.1	GTGAGTGACCAATTTGTGAGTC	AATATTGACCTGCTGCGCC
WP_005179775.1	ATGCACAGCATCCAAGGAG	GGGTGTTAAATATTGTGCAAAAGTA
WP_005180212.1	ATGCACAGCATCCAAGGAG	CCAAAATGTTAAATATTGTGCAAAAGTA
WP_013650396.1	ATGCTTAATCCCATGATAATTTTCTCA	CTTCACTCTTTTATCTTTTTTAATTTGGC
WP_014609009.1	ATGCAGAATACAGCACAAAATATATG	ATGCTGGTCATAAGACTGAGTG
WP_014609104.1	ATGACCGAGCCACATGATT	ATCTAATCGACGATGAGTTTTCC
WP_014609110.1	ATGAGTAGCGTAGATATTAATGTTCC	TACATCCAGCAGTAAGCGAG
WP_014609134.1	TTGAACACCAGTATCAATGCAC	CTTCTGCTTAGCGGTCTGG
WP_014609219.1	ATGAAAAATATTAATCCTAGTCAAACCG	ACGCCAGCTCTTGAAACG
WP_014609336.1	ATGCATATCGATAATGCTATTACAACC	CAATAATTTGTAGACGATGTTGAAGAT
WP_014609358.1	ATGGAAAACATACCTTCTAATATCGAC	CAGTGCGGCGATAGTCG
WP_016266096.1	ATGAAAATAACATCACCTAGCTATAAAGA	GTGGTAACTGAGCAAATGGTATC
WP_019083735.1	ATGAATTTCACTCTTCTTCTGCAC	TGGATGGAGAATGGTCGGT
WP_020282365.1	TTGATGATAATCCTACCTGATTGC	GAAATCAACACAAACAACAGCA
WP_020283316.1	ATGCGCATACCCATAATTTGG	TGGCGGCGAACTTCCTA
WP_020283496.1	ATGGTGTCTTCGATGTCTAATTCG	ACGCTCATGCCCCACA
WP_023160234.1	TTGATTGGTATCTGTATTTCGACTC	AAACTGGGGTATTCAAGCTAACAAA
WP_023160440.1	ATGCCGATTGCAGTGGC	TTTTGCCCCCAGTTTAAAATTAGC
WP_023160782.1	GTGAAGATCAACAAACCTGCC	TTCTTTACCCTTATCGCGGT
WP_023161026.1	GTGATTGGCTGCCGCC	GCTTAACCCGCGGATACA
WP_023161097.1	ATGAGCTTGCATCGCGC	TCTGACGCACATATAAGCCA
WP_071598577.1	ATGTGCGTTGGCTACATTCA	TGCCCTAAATAATTCGAGTTGC
WP_071598586.1	GTGTCTATCGGCGGTCG	TGGTATTTGAATAATAAATCTTGGTGGC
WP_080010975.1	GTGTTTTATCCAATGAAACACGG	AATAACAGCCTCAGGAACCAC
WP_080366037.1	ATGTGTGTTGGCTGCCT	TGGTGCAGCGCAGATAAA
WP_100206134.1	GTGTTCTCTGATAACACCACATC	GCTGGACTGACTCTCTATCG
WP_102047494.1	GTGGTTCGGGTGCCAA	TGCACCCGAATCACTGAC
WP_002229817.1	ATGTCGCAGATCGAAAATGC	AGCTGCATTAGTCCTAGCG
WP_010891203.1	ATGGACAGTATTCACGGACAC	AACCTCCTTGGAGTCAAATGT
WP_010891207.1	ATGACAATAAATATCAAGACAGACAGC	GACAACACCAAAAGCGGC
WP_010891236.1	ATGTATTCATTTGAACAAGCTATCAC	ACTAAATGACCGTGGTGGTG
WP_010891241.1	ATGTGTGTTCCATCGCCA	ATATCCGATACCATCCATAGCG
WP_014609444.1	ATGAAAATCATGGGAACTATGCC	CATCCATTCCCGCTCCAA
WP_014609447.1	ATGTTTATTAAAGATGCTTATAACATGCG	TCCCATAATACATTTTTGATCGCA
WP_014609473.1	ATGAACTTATCATTAAGCGATCTTCA	GCTATTTAATAATGGTCGCCCTT
WP_014609475.1	ATGACGGTTACCCTTAATAGAGG	TGTATCCATATCAATTTGATGGCTG
WP_016266437.1	ATGAGTGCGTTGATAACCCA	AACAGTATGGGGTCTGCCG
WP_071598607.1	ATGCACAGCATCCAAGGAG	GTGATTGTAGATATTTGGATAGCCC
WP_080098316.1	ATGCTCATCGACTTAATCGCT	AGCTGAACTGTGGTCCCA
WP_004235416.1	ATGTCTATCACAACCGGCTC	GAACATACGGCGGTTTAATCC
WP_004235425.1	ATGACAACGCACAGCACC	AACCCCGAAGCTTTCACC
WP_004235474.1	ATGATTCGCCACTGTACTGA	GAACGGACAGGGACTGC
WP_015422755.1	ATGATCCCACGATTAACGCA	GGATACCAGTGTCAGCAGC
WP_032098087.1	ATGACTGAAAATAATCAGAATTCTGTG	TTTGGTCAGAATAAGTTTTCCGT
WP_036417208.1	ATGACTCAGACTTCTGCCTTT	TACTTTCACAAATTCAATATGCTGC
WP_062771418.1	ATGTCGGGACCACCGA	TTTGCGCTTTTTCATAAATGCC
WP_062771467.1	ATGTCACATCCGGCATGG	GGATTTTTTCGGCAGCGG
WP_062771490.1	ATGTCCGGAAACAATCCTGA	TGATGAGGCCCCCGTG
WP_062771533.1	ATGCAGCCAACATCAGCA	AAGTTTCATCAGTTCCGGAAGG

112

gene ID	forward primer sequence	reverse primer sequence
WP_062771613.1	ATGACTAAAAAAACCGTTTCAGATTT	ATGCGCCATACGTGCA
WP_073970164.1	ATGATTTGGAACACGTTGTGG	TTCCTCTGCGTGTCTCCGG
WP_080654118.1	ATGCAGAAAACGGATTCTTTTTC	CAATACGGCTGAAAGTCCGG
WP_004234458.1	ATGTTATTTATGGCGATAAATACGACA	GGAAAATGCCAGGAACGGC
WP_004234829.1	ATGACGACGATAACCACAGC	TGTCCGCATATTATTAAGCAAGTTA
WP_004234853.1	ATGAGTCAGCAATCTGTACCTT	ACTCCGTACCGGTTTTATCAT
WP_004240526.1	ATGAATATCACTTCTCAGATTATGGC	CTGCCCCGCAATACTCA
WP_004241150.1	ATGCAGAATAATGAGGGAAACAC	CTTCAGCGGGCATAATATATTCGA
WP_004241218.1	ATGAAACTTGAAACGCTCTCTG	TTTCGCACTGTCCAGTGC
WP_046024762.1	ATGGGACCAATACAGGGACA	AATATCGTCATCAGGGCCG
WP_062771915.1	ATGCTTTATAACCCGTCACAGG	CTGTGCCCGGATCCGTA
WP_062771956.1	ATGCACCCGAACAAAGAACA	TTTTTTCATGACCCGCAGC
WP_073970171.1	ATGATCAATAACCCAGAAAAAATGC	TGGTTGGCGATCCAATGG
WP_004235744.1	TTGCCTGTAACTATCGGCTT	ACCGATCCCGTCCACC
WP_004236839.1	ATGGGGATTAATGCAGCAACC	ATTTGACGGATACACTTTAAAGATTTC
WP_015422568.1	ATGGATACAACACAACCGCA	ACCAAAAACAGCTGATACAGAT
WP_015422612.1	ATGACGCAACTCATCATGCC	CTGTCCGGCTTTTTTCAGC
WP_036416987.1	ATGACACAACAGCGGCA	GTTCTCATCCAGTGCTTTTGT
WP_073970177.1	ATGGATAATCAACACATGTACAAGA	CAGGGTATTAAACAGTTTTCTGGTA
WP_004237750.1	ATGGACAACACATCAGCGC	ACTGGCGTTATCGGTAATGA
WP_004238584.1	ATGTCCAATATACCTGCACAGG	GGTTGTCAGCCCGGC
WP_004240712.1	ATGCTGATTGGCGGCTG	TAATTTTCTGACTTCATCGATAGCA
WP_004241031.1	ATGAGTATTCATAGTGCCGATACA	TTTAAATGCCGGCAGCCA
WP_004904012.1	ATGAGTACTGACAAAACCCAGA	TACGAACAGTTTGAACACGATG
WP_036416809.1	ATGACTCAAACTTTAACGCAGC	CTCGTAATCACTCATCGGTGC
WP_062772817.1	ATGAAATTTAAGGGTACGCCTG	TTCTCGGCCAAGTGCTTT
WP_062772924.1	ATGGCTACTATCCCTACTCAAAAT	TACGCAAATGATCTCTATGATAAAAATAG
WP_004235960.1	ATGGGAAAATCACCGGCC	CTGACTGACCGGTGGTG
WP_004235986.1	ATGTCCATGCAATCTCAAGACA	CTCATGGTGAGTTTCCTCCTC
WP_004238850.1	ATGTTCACCTCATCGTTAACCA	ACGGATCCCGCCGAA
WP_004238854.1	ATGAGCGTTTCTCTTTCTAAAGG	GGACGCGTTGATACCGTA
WP_036413944.1	ATGAGTCATCAATCCGGCC	TGCTTTTTCCACCTCAACTAATAG
WP_073970193.1	ATGATCAACGGCCCGGTA	GTAATAACGATAGTCTGATGTCACA
WP_081113481.1	TTGGGGTTACAGACACAGTATG	GAGTGACTTTCTGCTGAGCA
WP_004242103.1	ATGGCAACTTCAACAGTCAGT	CTGACGGTGTTCCGGC
WP_004242218.1	GTGACAGTATCCAGTGAAAATGC	AGCAACATTTCCTCCCGG
WP_064483359.1	ATGACAGACAGGAACCCCATGC	GCCCGCACCGCCGG
WP_004236571.1	ATGAACTCTGCAATTTCTGACAA	TTTCTCCGGCGGGAAAAT
WP_032098021.1	GTGCCTGTTACTGTGCAAAC	GCCTGTGGCGGTGTT
WP_036413302.1	ATGATGAACCCTTTGCTTACCC	CCCGTTGATTCCGTAGCC
WP_062773389.1	ATGACATCGCCGATCACAA	GTAAACATCCCGCTGATAACG
WP_004236694.1	ATGTCAATGCAATCACAGGAC	TTTCTGCTCCTCCTGACCT
WP_004240204.1	ATGCCGATCAATGCCGT	TGCCTGAGCTGATGCG
WP_036417499.1	ATGAAAGATGTGGCGCGG	ACCGCAGATAACAGATCCC
WP_062773486.1	ATGAAAAACAACATGCCTGTCA	GGCATGATTGCCGAAAACC
WP_024475195.1	ATGAGAAGTGGCAAAAAAAGCA	TGACTGGAGGTCTGTGATAATTG
WP_062773522.1	ATGCCTGAACAAAAAAAACAAAGC	GTGTATATTTCTTTTTGTCAGAGTATCTT
WP_062773581.1	ATGACTATGACACATTCTGCGTA	TTCATCGTCATTCTCACTTTTATAAAAA
WP_004242398.1	ATGGATGGTATTCAGCAGGC	GGGGAGCCTCGCTTTG
WP_062773608.1	ATGAGCCAGCAGATAGCTG	CCCCATCGCTCCGCC
WP_024474672.1	ATGGATGGTATTCAGCAGGC	GGGGAGCCTCGCTTTG
WP_062773673.1	ATGCCAAATCCCGGGAC	GTGTCACTGCCCTGAGAAG
WP_040259375.1	ATGCCCACCAATGAAAACCA	TTTGCCGACCTGTTTTTCC
WP_040259600.1	ATGATCGACCTGTCCACTTG	GGAGCCGGTTACATGGC
WP_040259685.1	ATGACGTTGTCCAACGCC	TAAACGAACGCTGGCGA
WP_040259742.1	ATGCAGATTACCTCTTCGATGA	GGCATAGGTATCGATCATCGT
WP_052469135.1	ATGACCTCCACAAGCTCCG	AGCCTGTTCCAAAGCAGC
WP_084596144.1	ATGAACACTCCCTATTTGGTGG	GTCCAGATGCAGCACTTTTT
WP_088776180.1	ATGGCCAGCGCAGC	GCCCCGTGCGATGATT
WP_040259943.1	ATGCTCGACTGGAAAAAACG	TCCCTGGGACAACAGCT
WP_040259953.1	ATGACCCAAAGCTTCGTTACC	CTCCGGTGAATGAAAGGACA
WP_040259956.1	ATGACGATGATGACCGAGAC	GAGCAGATCGTACTGCTTGA
WP_040260490.1	ATGAAACAGTCAGGTCAACGC	AAACACCCAATGTGATTCGG
WP_040260539.1	ATGTCACTCCACGAGTTGAT	GAGATCTTCGGCCAGCAC
WP_084596153.1	ATGCCTTATACCGTGCCG	GGGGGCCTCCGTGG
WP_084596156.1	ATGTCGACAACCCCGAGT	CTCGGCTATCGAATTATTACGAAG
WP_084596157.1	TTGTGGGTATCGAACTCAATAAGC	GTAGACCACTACATGACCACTAC
WP_084596159.1	ATGATCGCCTCCGCAAG	GAAGCCTCGACCTTGGATT
WP_040260598.1	TTGACTCTCTCCCATACGTCA	GGGCTGGTTCAGGAACAG
WP_040260599.1	ATGTCGCAACCTTCCAGC	CAGTAGCGATAAAAACAGCGT
WP_040260708.1	ATGCGTCCGGAACTAGC	TGAGCGCTTGGCAGGTA
WP_040260715.1	ATGCAGTTTTCCCCGATTACA	CCCCAAACGAATGTTCAGG
WP_040260799.1	ATGATCAGTGGCGTCAGTAG	GGCGGCGGTGCTCA
WP_040260848.1	ATGCGCAACAATCAGCCC	GCGGTTGAAGCGTTCCA
WP_040260864.1	ATGACAGTCGAGACGAATGT	ATAGACCTGCCTATAGCCAGC
WP_040260877.1	ATGCACACCGACACCAC	CTGGCAGGCTTCGCAA
WP_040260926.1	ATGACAACACATATCGTCCACT	ACCGTGCACGCTGTAG
WP_040260933.1	GTGTATAACGTCGTGATAAGTGG	CCGCTTGCGCAGAATCA
WP_040260992.1	ATGCCTGCCCCCTGG	GCCGCGCTGGCTTC
WP_040260997.1	ATGCAAACTCTCTACACCGC	CACCGCGACGTGCA
WP_040261043.1	ATGCCTAACGCCAACACG	CCCCAGCGCCGAGC
WP_040261105.1	ATGTCTGACCTGTCCAACC	GACCCGGTTCTTCAATGTATAGA
WP_040261144.1	ATGTCCCAGCCCTCCC	TCGCAGTAGCTCCAGGG
WP_040261153.1	ATGTCCGATCTGATCACCCC	GGCCGTGACGCGCA
WP_040261387.1	GTGCCAGTCCTGCGC	GCGCGTGCCATTGGC
WP_084596184.1	TTGAGTTGCGTTAATCAACATTCT	GGGGATACGTGTAGGGTGGA
WP_088776197.1	ATGAGCCAGCTTTCGCA	TACGAGGCGGGCGC

gene ID	forward primer sequence	reverse primer sequence
WP_088776203.1	TTGCCAGTTCCTGAACCC	CGCTTTTTCCGCTGGC
WP_040261650.1	TTGAGCAATCTCGATCCGT	GTCGAACATGAAGAGCGC
WP_040261701.1	ATGAGTCAACCCACCCTTG	GTGCTGGCTGGGCG
WP_040261863.1	ATGGCGGTAGAACAGGC	TCGCTCTAACCCTTTGACAA
WP_040262295.1	ATGGACATGCCTTCTGGTG	TAAAGTCGTGGAAGAGCCG
WP_040262478.1	ATGGCTACAACCAGCGATAA	GACTGCGATATGCTCCTGC
WP_040262495.1	ATGGACAACAGCAACCCA	GAACGGCACTTTCATGCTTA
WP_040262503.1	ATGTCTGCGTCCACCCA	TGGGGCAGTCCTGAGC
WP_040262707.1	ATGATCGATCTTTCAACCTGGA	GCTGCTTTTGTGCTTGTGG
WP_040262743.1	ATGACGTCCAGTGATGCCT	GCCGCCCAGCTCGC
WP_040262756.1	ATGTTGTGCCGTGCCG	CGAGACGCTAGAAGCCC
WP_040262907.1	ATGAACCCACAGACCGAAG	GTCCTGATCATCCTCGACC
WP_040262982.1	ATGCATCTGCTGCGTACC	GCCCTCTTCTTCACTGTCG
WP_040263025.1	ATGACTTCCTCGATTACCGC	GTGATGGGTGCCGAGC
WP_040263106.1	ATGACTGAACCGACTCAAACC	TGGAGCGAGTGCCGC
WP_040263221.1	ATGGACTCAAGCAACCCTG	GAAGGGTGCGACGACC
WP_040263293.1	ATGATAGACCGAGCCCAAC	TTGTTCGCCAGGCTTCA
WP_040263346.1	ATGAACGCCTACGCCG	GTGGAAATCGCGGCTGC
WP_040263420.1	ATGACTGCCTTGCAAAACC	TCGGCTGATCCCCGC
WP_040263598.1	ATGTCCAACGCCTATCTCG	TCGTGCTCCCAGGATCC
WP_040263703.1	ATGAAACTGATCGGAGCACG	TGCCGCGTTTTCAGCC
WP_052469277.1	ATGGAAATGAACAACCGTACG	GGTCTTCAACCAATATAGCGC
WP_052469283.1	ATGATCCAGCCAATCCCG	CAAGGCTTGGGAACTCGG
WP_052469295.1	ATGGCTTCGACAACATCCA	ATCGTCGGAGCTGTTCATG
WP_052469314.1	ATGGAGCCAACCGCG	GAAAGCGGGGTGGCG
WP_052469327.1	ATGGACACTGTCACGCTC	CATGGGCAGCGTCTGAT
WP_052469364.1	ATGGACCCTACAAGTAGTCGC	GTCCGCCGGGCCAA
WP_052469369.1	ATGACCATCCCAAACCAGTC	TACCAGGCCTCCTTTGTTC
WP_052469370.1	ATGAATGTCCCGGACTACG	AGCCCCTGTAGACTTCTCC
WP_052469427.1	ATGAATATGATGGTAGTCCGGG	TATTGAGAGATTCCCGAGATTCA
WP_084596217.1	ATGTACAGCCTCACCACTCC	CCCAATGAACAACCGGTACG
WP_084596219.1	ATGGAGAGCGGCCCA	CAAGTCCTCCGCATCCTG
WP_084596252.1	ATGAACAGGAGGCTACCGA	CACCCCCAACAACGTCA
WP_084596260.1	ATGCAGGGTAGCCCCA	GAAGGGCATCGCTACCTT
WP_084596289.1	ATGGTGAAGCCCACCAC	GATCCCGGCCAGCCA
WP_084596294.1	ATGCCAGTTATGGCACCCG	CGGCGCCAGGCGCT
WP_084596314.1	ATGATTGACCTCAGCACCTG	CTTCTTCGATCCGTTGATGC
WP_088776216.1	ATGGGGCAGCTGTATGAAGAG	ATCTGCAAGCAGGGTGC
WP_042029433.1	ATGCCGAACTCGTCTCAC	GGGCTCCTCGGTCGC
WP_042029474.1	ATGAACATCAATGTCAGTCTGC	GATGCGCAGACTAAGTGACT
WP_005354370.1	ATGGCATCTAACAACACAGCT	GAGGAAACCGGCGAACT
WP_042029697.1	ATGGAGATAAATGTGGTCGCG	GACCTGGGTATCGACCCG
WP_042029790.1	ATGACAACACCAACCACTTCG	GATTGCTTGTTCGGTGTTGC
WP_042029810.1	ATGCCAGCCAAGCCG	GATGGCGCGGGCAA
WP_042029960.1	ATGAATGCCACCTCTTCCG	CGCATTGCTCGCTCCC
WP_042030090.1	ATGCAGATTCAGCAAACCCG	CGTCAGCAGAGCCACC
WP_042030140.1	ATGACGCACCAAAAGCGG	AAAATGGAGGCTGAAAATGATTTTC
WP_041208069.1	ATGTCTGCGACGCCG	GGCAGAGGCAAACAGGC
WP_042030289.1	ATGAGTGGCACTATTATCAAGTCG	CGCTTGCGGCGGGG
WP_042030321.1	ATGAGTTTATCTGTAGAGGCTACC	CTTCAACAGGCTGAAGATCAG
WP_042030361.1	ATGAGCAAATCCTGCTGCT	ACGCTTGCGCAGCA
WP_042030436.1	ATGAGAGTCAATCAACCGGT	GCGAACCGGCACCAC
WP_042030437.1	ATGGCCGATCTGCCCC	GGGATGGCTTGCGGG
WP_042030528.1	ATGGCTGTGAATACCCCTTC	GCCGGCAACCAACGG
WP_042030885.1	ATGTCAAATACTGTCACTGGCG	ACCGACCAGCTGATAGAGG
WP_042030958.1	ATGAGCCAGGAAATCAACGA	CTCGACACTCTCTTCCAACA
WP_042030965.1	ATGACCCATCAATCCCACTC	AGCCAGCACGCTGC
WP_052448106.1	GTGATTGGTGATGCAACGG	CCTCTTGAATACAATTGCGAAAGC
WP_041208848.1	ATGAACCCGATCAGCAATGA	TATGGCGGCCGGTCT
WP_042031153.1	ATGACTTACTTCGCATGGGC	CTCATGGCATGTCCCTCC
WP_042031194.1	ATGTCTGCGCCCCTTG	CTCATGATCAGGCGCTGC
WP_042031310.1	ATGTCTGCCTCCACCCT	GCTCATCAGCAGGGCC
WP_042031415.1	ATGAGTGGCGCATCATTTCG	TGCCAATGCCATCAGCC
WP_042031493.1	ATGAACATTAACGACAGCACC	AGCAACCCCGAAGGCT
WP_042031495.1	ATGAGCATTATCTCTGATTACACCC	AATGGCGGCAAGAATGTCG
WP_042031532.1	ATGCAAATCAACTCAGCCTCC	GGCAGTGATATCGAGCAGG
WP_042031545.1	ATGATATTGGATAGTGTCGTTAGCA	GAACAGCTTGGAAACAATATTCAG
WP_052448118.1	ATGAATACCACCTCAGCGAC	GAGCGTCAACAGTTGTGC
WP_082035465.1	ATGCAGATTCAGCAAGCCC	CTCCTTCATCATCCAGGAGC
WP_041209466.1	TTGACCCAGAGTGTGACAC	GGCAACCTTACCCATCAAGG
WP_041209524.1	ATGAACAAAGCGGTGGAGA	TTTGACCCGCGCCAG
WP_042031937.1	ATGACATACAGGGTCAGCAAC	CTTCATGGTTGGCATGGAGA
WP_042032056.1	ATGAACCAGACTCTGCACTC	GTGATTGAGATAGATATCCTGTTCG
WP_042032153.1	ATGAACAGTTCAACCAATATCACG	GGAAAGTTGCAGATTACTCTGC
WP_042032213.1	ATGAGTGTTTCTATTGCTGAAAAGA	CAGCAGATCGTAGATAAGGGC
WP_042032223.1	ATGTCTACTTCCGTATCTGATACC	CGCCTCCTCCCCTGC
WP_042032269.1	ATGCCCAGAGTGAAATCCAG	AATATCCGCTTCATTGTTTACCTC
WP_042032390.1	ATGAGTGAGCCAGTAAGTATTGC	GCCCTTGCGTGCCA
WP_042032417.1	ATGAGTCAGGCGAACAGC	GCCTTCTACCTTGCCCC
WP_042032439.1	ATGTCAGCCGCCATGTT	CAGCTGGCGCAGTGC
WP_042032510.1	ATGATCATCAACAATCGGGTAGG	TGCCTGCACGTTGACC
WP_005335772.1	ATGACAACTGAATCCCCAAGC	CTCGTTGTTGCCGCCG
WP_042032808.1	ATGAATCGTTCAGGTTCCGG	GTGAGCCGAGTCATTGGC
WP_042032817.1	ATGGATATCAAACCCTACAGCC	GGCGGTTTTGGCCTCA
WP_042032842.1	ATGGGCGCATCAGAACAG	CTGAGTGGCCAGACACTG
WP_042032909.1	ATGAGCTATAGTGCCGAAATCA	ACGTAGCTGGGCGGC
WP_042033067.1	ATGACCGCTATCGTGATCAG	GCACCGTTGCAGCAGC

114

gene ID	forward primer sequence	reverse primer sequence
WP_042033145.1	ATGAATTATTCACTGGTTACGGTC	TAATGTCACCACCTTGCGG
WP_019446232.1	ATGGACATATCCAGCTCTGC	GACGCGGATATCGATATTGC
WP_042033236.1	ATGATCAGCAGCCTGCTT	GGCGCTGGCATTGAGC
WP_042033246.1	ATGGCCAATCTGGATTTTTTTTCA	GAGGCAGAAACGACGAACC
WP_042033269.1	ATGACCGCACAACACCA	CAGCATCTGCTTGAGGCG
WP_082035530.1	ATGTTAATAACATCCACATCTTGTCA	GGGTACCGCAATCAACGG
WP_042033293.1	ATGAAGCATCAAGCCACGC	TTTTTCTACGCCGCTCAGC
WP_042033505.1	ATGAATCAGACCATCGATCTTATCC	CTTGAGGTTAAAGCCTCGGG
WP_042033539.1	ATGAGCCTGCAGCAGAG	CCAGCGTTCACCCAAGA
WP_071910665.1	TTGCTGCTCCCTATGTGG	GGAGGGGCGGTCACT
WP_005132825.1	ATGGTTGCCCCCGTTCC	TTTAATGCCGCCATACGCC
WP_001149870.1	ATGCAGTGTATTCCTGGCG	TTTGATTTCCTTTGTATTTTCATGTGG
WP_040229922.1	ATGAACAACATACTCACCCTGA	ATCAATGATACGCATAATTGCGT
WP_080721914.1	ATGACGAATCCGGCCCC	CAGCCAGCCGCGTT
WP_040230127.1	ATGATCCAGCAGGAAATGCC	ATGTTTTCCTAATGTTAATACTGAAGC
WP_005134131.1	ATGACAACGACAATTGCCAA	TGCGTCAGCGCTTAATGC
WP_040230211.1	ATGCCCACTTCAAGCGAA	CTTACCGAAGCGCATCTGC
WP_040230231.1	GTGGAAAAAAATAATATTACCCTCGAC	CTCCTGACGTGGGTCGC
WP_005131490.1	ATGATTAGCGCATTCGATATTTTTAAG	GGTACAAACCACTTTAATTGCCA
WP_005131699.1	ATGAGTACTACTGATAGCATTGTATCC	AAACAGTTTGTAGACAATGTTCAGG
WP_005129057.1	ATGAGTATTGATCGTACCTCACC	TTTACTCTGTAGGTAGCTCTGTG
WP_005129187.1	ATGTCTTCCATTACAACAACTGAT	CGATTCTTTCCAGTACTTCTGC
WP_005129207.1	ATGGTCAGCACACATACTCA	AAGGTGCAAAGACGCAAGA
WP_040230605.1	ATGGAATGCCGTCCGGA	AGGCGCGGTCAGCG
WP_040230709.1	ATGGCAGGCATTAACACGC	ACGAACTGCAGTCTGTTGG
WP_040230744.1	GTGGGCGCAATAAGTCAGA	CCAGAGAAGTCCCCCGT
WP_072041421.1	ATGAACACCCCATCGTCAA	GTTTCGGAAACCATGTGAAAATG
WP_082031712.1	ATGGGGGGAGGTGTATTCGA	GTTAATTAAACCTGCCGCTTTG
WP_005126657.1	ATGAATCATCATCCTGTAAAATCATCC	CTTTATTTTGATTTCAGGAAACTTGC
WP_005126712.1	ATGGCTATGCAATCACAAGATATTA	CTCATGACCGTTCTCCTCTT
WP_005126769.1	ATGAAAAGTAACATGCCCGTTAC	TGCCGACACGCCAAAG
WP_005126801.1	ATGTCTATGGAAACGCAGGA	CTCATGATGATCCTCCTCGTC
WP_040231102.1	GTGGCAACAATCAATAGCACC	CTCATGATTCACTCCTTTGGTATT
WP_040231280.1	ATGTCGTCAAGCGAAGAGG	GACTGATTGCCCGGAAGA
WP_040231301.1	ATGGTTAAAAGGCCGAGCC	CTTTTCACCAGTCTTCTGTAGATAA
WP_052463753.1	ATGCAATTTAATTCTGAATCAACTTCT	GCTGTTGAGCGTGAAGATC
WP_003844491.1	ATGTCAACGGCTAAACGCG	CATTTCTGGCTCAATACCTAATTTAC
WP_005121577.1	ATGTCGACACCAGAGACCA	TCCTTTACGTAACAACCGCA
WP_005133830.1	TTGGCTAATTCACCCGCA	TGGGCTGGCAGGTGG
WP_005133841.1	ATGTCTGACTGTCGTACTTACG	GAATGTTGCCTTTCTGATGGC
WP_040231828.1	ATGTACACTTCAGGCTATGCA	ATGATACAAATTCGAGTGAATTTTTAGC
WP_040231861.1	ATGAGCCAGCTAACAGAAAGC	GCAGCGAACATGACTGA
WP_040231990.1	ATGCAAAATACCAACATCATTACG	GCCCACGTATTCCGGTT
WP_005130943.1	ATGATTAGCGTATTCGATATTTTCAAAA	ATCGCAGGTAACGATCTTCA
WP_040232070.1	ATGCATATCATCCACAACAATAATGA	ATTTTTGGTCGCGCCGG
WP_040232101.1	ATGCGGTTGATTGATTACTTTCC	TGCGAAGGCGATAATATCCG
WP_040232276.1	GTGATCAACCCGAAAGTTGAT	CATCCCGCCAGTGTTACG
WP_052463763.1	ATGGCCATTAATTCATCAACCATA	CTTTTTCGGGGAAATTTTTCGG
WP_072041464.1	TTGGTTTCTGCCAGAGCTAA	TTCGCCCGGTGCTCG
WP_005120762.1	GTGAGTATGTCTATTAATCAGATATCACC	TGTTTTTATTAATGTTTCGATCGCG
WP_005122303.1	ATGGATAAGTTATCTTACGCTTCAG	TCCCGGATAAATACCACGGT
WP_040232310.1	ATGGTCGATCATTCATCTGCT	TTTGACGACGTTATAAATGACCC
WP_040232312.1	ATGTTCGCATTCAACAATGCA	CTCTTCCTTGCCATTAAAATCGA
WP_040232340.1	GTGATTAGTCTATTCGACATGTTTAAG	GTCACACTGGACTTTGATTGC
WP_040232351.1	ATGGGCGATATGACCATCAG	CCACTCCGGTTGGTACATAG
WP_040232374.1	ATGGTTGGATCGATTAATTCAGTG	GGCTTTCAGATTGCTACTGAC
WP_040232375.1	ATGTACGGAATAACAAATAAGCCTG	CGCACGGGTATGTGCC
WP_040232376.1	ATGTCAAACACAATTAAAGCATTCTC	TGCAGGTGGATTAATTTCAGGG
WP_072041467.1	GTGTTTACTCAGGACTTCTTTACTG	TGCATTACTAAGCATAAAAAATGATATCT
WP_072041472.1	ATGTCAGCATTATTGAGTTCTTTTCT	ACCATATGTCCTTCCGATATGAG
WP_005121418.1	ATGAGCGACGACAATTCACA	TTCGTCCAGTTTGGGCTG
WP_040232565.1	ATGACGCAGAAAATCGAACAG	TTCCTTCACCTGATGCACA
WP_040232739.1	ATGTCTATTCGGGAACTGATTGA	GGCTTTACGTTCCGGCG
WP_040232745.1	ATGAGGCTCTTTATGGGTGG	TTGGCTATCAGGCAAATTGAAA
WP_040232896.1	ATGAATAGCTCCTGGGTTAAAAAC	GTGTGTATTCAGCAACAGGG
WP_040232962.1	ATGCACTGGCAAACGCA	CCGTATTCCTCCCGTCGC
WP_040232968.1	ATGAGTAAAGTGAAAAGCATCACG	CAGAAAGGATTTAAACGGTAAATCC
WP_005121691.1	ATGAGCATGTCCGACGATAG	CAGCAGATCGAAACGGTCC
WP_005121817.1	ATGGCTTATCAAACTCAAGATATTATCC	TTTGCGCACCTCGTCTT
WP_040233132.1	ATGTCAGATACAATGACTGTTGTAG	GTTTAACAAACCAGCATCAATATCA
WP_000550475.1	ATGCATCTCTTATACGTTGATGAG	GTGAGCTACGTAAAGTCCGT
WP_005126378.1	ATGAGTCACTCATCTGCCCC	CTTTTTAGAATGACGTCCGTATACC
WP_071887407.1	GTGCCGCTATTCTCAGCG	ACGCCTAAAATTTATCATTACCCTAA
WP_005122932.1	ATGGAAAGCCCTGTTCCC	TAACTCATCAGGACCTCCCT
WP_040233356.1	ATGAATATAGTTACAAGCGATAATGAAAG	TTTCTTTAACCCACCTTTGCG
WP_040233393.1	ATGCAGCAAAACTTATCATTTTTTATCA	TTTATTGATGCCCGGGTACT
WP_040233404.1	GTGACCCCCATGCTTTTAAG	GTAAAAAATCCTCACCGCACA
WP_072041502.1	GTGGATTTGCCAACGTTAGG	GGTTTCCTGACAGAGCCG
WP_072041503.1	TTGTTGTTCCTTTCACGTATCG	CACTTTTGGAGAATTTGCAGGC
WP_040233558.1	ATGAATACAATCATCCCCCAACA	TTTATCATTTGTGGATTCCTCGTC
WP_040233602.1	ATGGCTGATGTATTAGAGTGGC	AAATGCCTCCTCAAACTCAATTG
WP_040233716.1	ATGTCACGTATGGATAACACTACA	TTTGGTCAGAATGAGTTTTCCG
WP_040233766.1	ATGAGCTCTTCGACATCCTTAA	TTTCTGGTAGTGATAAAGGATAAAGTAG
WP_040233871.1	ATGAGTAACAGTGATAATTACCATCAA	CAGATAGCGGGCTTCCAG
WP_005123605.1	ATGACTCCAACCATTGATTTACTT	GCGCGTCGCCCATC
WP_072041520.1	ATGATGAATTGTCCTATGTGCG	GTAAGCTAATTGTTCCTGCATCC
WP_005123150.1	ATGTCAGACAATACCTATCAGCC	AGCCTGGCGCTTATCTTC

115

gene ID	forward primer sequence	reverse primer sequence
WP_040233949.1	ATGAGCAATCAAAATAACCAAAATGC	TAACGCGCTCCGGGT
WP_040233982.1	ATGCTTCCCGAGCATCTTG	TTTCTCTTCCAGAGAATAAAAAATAGGA
WP_082031767.1	ATGCCTGGGACAGTTCGC	TACCAGTGGATCGACTGGC
WP_023184674.1	ATGATGAATTGCCCAAAGTGTG	AAAATTCATGTGCCCTTGACC
WP_004253374.1	ATGAGAACTGAACATACCAGATCT	ATGGACAGACAACTTCTGATAATTT
WP_004253432.1	ATGACAGATTTTATGAGTGCCAAC	CCAGTCTTCTTTAAAGCCAGTG
WP_004253606.1	ATGTCGACTTCACAAAATAACGC	CTTCGATACTCTTTTGGTTTTAACTTT
WP_004253720.1	ATGTGCGGAAATATCCTAGGT	CTGACCAAGTAGCGTTTTACC
WP_004253752.1	ATGCCAATTGTACCAACGTATAA	TAGTGGTATATCCCTAATTTTCTCTCG
WP_004253755.1	ATGAGCTACTTTGGATTAAACCC	TCCTCTAAGAAAACAACTAACAGC
WP_004253768.1	ATGGCTAATCTCAATATTTCTGTATCA	CAGCAACTTACTAACTGTTGAATCA
WP_004254299.1	ATGGAACACTCACAAAATAGCC	CTGGTTATAACAAATTGATTTCAAACG
WP_004254558.1	ATGAAAACATTATTACCAACATCTACCG	CTTATTCAACAATTCTCTACGGATAATC
WP_004254983.1	ATGAACGCCTTTAGCCTTACC	AAGGATTAATTCGCTACGTCCT
WP_004255360.1	ATGAGTAGGAAAATAGCACGCG	CAGTAAGCTCCACGCTTTTT
WP_004256127.1	GTGATCAGTCAAGAACTAACATGG	ACTCATCCATTTTAAGAACCAAGG
WP_004256190.1	ATGAGTGATTTTTCCCAGACCG	ATATTTGTCGATGACATGTGCC
WP_004256437.1	ATGGATGCAACAAAAGTTGGT	GAAGAATAACTTATATGTTGCCGC
WP_004256475.1	ATGCCAAATAAAAACTTAAGTACGAC	TGCAATATCACCATTAATCGTTGT
WP_004256486.1	ATGCGAGTAAATAGTCAGTCTCTA	ATAGTAGCCATTTCCTGTTTTATAGTC
WP_004256696.1	ATGATATTAAAAACAACGACTGGCT	TTGGGTAAGTAATTCGATTAACGC
WP_004256728.1	ATGGCTATTCAACCGACAGC	TAATGATGGATATTTGCTATCATGGT
WP_004256756.1	ATGGCATTAACTGATAAATCAAGAATTAA	ATTTGCGTTAGGTATCATAGGTATTAC
WP_004256890.1	ATGGCTCAATTATCAGGTAGTAATAGTA	GTATTTAAGCTTAAATCCTGGATTTGC
WP_004256913.1	ATGCCACAGACACTTATACAACA	TACTGATACGGCATTGGTACTC
WP_004257109.1	ATGACCACCAGCGCATC	AAAGTCCCAGTCATCATCTTCG
WP_004257114.1	ATGAGTAAGCAAAATCAAGAAACAAC	ATGGTTTTTTTTACAAAATGGGCG
WP_004257253.1	ATGATTTCAAATTTAACCCTAAATACGAA	GTTTCGCATTGATATTAATATCTGTGG
WP_004257256.1	ATGAGTTCTACATTTAATAAAAATCGGG	ATCCCGTAATGGGGCTAATATAG
WP_004257539.1	ATGTCACACTACCAGAATAATCCC	TTGTCGTGATAAACGACCTAGT
WP_004257971.1	ATGGAACCCATTCAAAATATTAATCCG	GTAACCATTGACGAGAAGAAATTCG
WP_004258503.1	ATGAGCATCGACATTACTAATTTTTTTA	TAGCTGTGGTGGGTTCGT
WP_004259339.1	ATGAGTACAATGAGAGGGCTTAG	TTTAGTGTTGGTTGGCTCAATATAG
WP_004259388.1	ATGGCAAACACCCCCTATATT	ACTTATATCACTCGCATCGTGG
WP_004259957.1	ATGGCAGCTCAAGAAAGTGC	TTTCAGATAATTAAAGATTGCACCGA
WP_036957743.1	ATGAATATTCCGTTGTACTATTTAAGACA	TTTGCACCAAACATAATTGATATAGC
WP_036957904.1	ATGAGCATCTATTCTTGGAACAAC	GTGGTAATGATTTGAATGAATTTTTAAGT
WP_036957912.1	ATGCAATCCCTGACATTACAAC	ATGACTTAGACGTGCCAACT
WP_036957920.1	ATGAAAACAAATAGTTGGCTATCGG	CTTATTTATAGGTGTTAACACTTTGCTT
WP_036957968.1	GTGAAAATTGTCCAATCTGGCG	AGCTATATCAACTGTAGTCGATGT
WP_036958038.1	ATGCACTCACATTCCCATAAAG	TTTTTTCACTTGAGGTAACTTGACT
WP_036958071.1	ATGAATGCATATTCTATAAGTTGGGG	ATTGAAAGAAAACAAAATAACTGCATTAG
WP_080544449.1	ATGAGGGGGTTCCTATATGCTAAAA	ACCGCGATGTGAATAATGTCT
WP_004260347.1	ATGAGTGAGCAAACCACAAC	TTTTTTGCCAATATTTGGCAGC
WP_004260627.1	ATGTCTAGCTCACTGAGGCG	ATATTTATAGTAAACATCGACATCAAGC
WP_004261076.1	ATGATAGCAAAGTTTTATAGTTACCAAC	CTTAGGTCGGATACATTTTGAGGTT
WP_004261181.1	ATGAAACGCATTCAATTCAGTACT	AAAACGGAATTCCCCAGAAATATAA
WP_004261326.1	ATGAATATCGCAGCATCTTCAAC	TTCGAATAAGTTACTGTGCAAGG
WP_036958274.1	ATGCTTTTTTGGTATGATGATATCGC	CTTTATTGGGAGGATGGGAGG
WP_036958341.1	ATGTTGAGCAACTCGTGGT	GGCTTTTTTCATACTCTGACTTGG
WP_004261604.1	ATGAACTTTAAAAGTATAGTATCAACCAC	TTTGGAATCAGCTTTCTGTTTCC
WP_004261608.1	ATGACAGTCAATAAGAAGCAAAGT	TTTTTTTACTAGTACGTTCTTTATTTGCC
WP_004261611.1	ATGGATAACCCTAAATTATTCATATCGTA	TGGTATAGTTGAAAGCTTCTGATATC
WP_004261691.1	ATGACGAATATCTGGTCGCA	GTCATTGCTATATTCCTTTTTGATTTTTT
WP_004261765.1	ATGACTATGAATAATGCTTCAACACA	ATGCTTAGCGTCTTTGTCATAC
WP_004261788.1	ATGCAAGCAATAAACTCACCAA	AGCTGCTATTCTTGACCGAC
WP_004262117.1	ATGCAAATAATCAATCCTCATGATTTT	TAATTTTTCATTCCACGGGTCTG
WP_004262300.1	ATGAAAATTAATACCAAGGTTATTCATGG	TCTCTCTTCTTCAGGGATACAAG
WP_004262383.1	GTGACAGGTAGTTGGCCAC	AAAGCGGGTTTCTAACCCA
WP_036958442.1	ATGAATATTTTGCCTTTTAACCCTTC	TTTAGTAGTTTGACTTTGTTTATAATGGC
WP_036958456.1	GTGGTCCTGTGGAAATCTAATCT	AATATCAGGGTATTGCAGCAATAAT
WP_036958458.1	ATGTCGATGTCACTGAGTCAA	CATATAATCTTCAATCGGCACACA
WP_036958519.1	ATGGTAAATAATATGACAGCCCCA	CTTAGTTAAGATCAACTTACCTGTACG
WP_004905318.1	ATGAGCATAAATACCAATTCTACGT	ATTGGAAAGACCACCAAAGCA
WP_004905395.1	ATGCGATTATTATCGATAACAGCA	TTGAATATCGGGCCCGGT
WP_004905473.1	ATGGCTAAAGAAAAAACTCAGCA	TTCTGTATCAGGTGAAAAATCGG
WP_036958541.1	ATGCACAGCGATTACCAATC	GATGGGTATCCGTTTCTGTTCT
WP_004262673.1	ATGAATAATGGGCTACTTGATACCA	AGCGTGGCTGTATTGTTCT
WP_004262890.1	ATGAAAACTGAAGGTCACAGTTC	ATTTACAAAAGTATTTTTAACACGCCT
WP_004262981.1	ATGATGTCAATGTCATCCATTCAC	TTCACCTTGAGACTGCTTCG
WP_004262987.1	ATGACACCCATTATGACACAAC	CCCTCGTATTGCAGATTGGA
WP_004263017.1	ATGTCAGCAATAAAAATAAGCTCAAA	ATCTTCTTCTTGCTGTTGGTCA
WP_004263051.1	ATGTTAGCTATTACACCACATTTACTA	CTGTGCTTTTAAGACGGCAT
WP_004263067.1	ATGCAGACTAATCTACAAAATAAGGT	GATTACTAGTTCATCATCCTCATCG
WP_004263203.1	ATGGACGAATTTAAACCAGATAACC	TTTTTGATCCTGTTTGACTTGTTTC
WP_004263536.1	ATGTCCACTCAACCACAAAATAA	ATCATTAGGTTTTAATAATGAAAGCGAC
WP_004264245.1	ATGAAAACGAAAAATAACAGCTTAATTG	GTTAGCCCCCTCATTAATAAAATCA
WP_004264313.1	ATGAGAACCAAGAACTCCTCAA	GCTAAAAAGTGTATAACCGAGTGC
WP_004264319.1	ATGAACATATCTTCACTTTTCTCACA	GCATTGCTGACTTAGCAGG
WP_004906048.1	ATGGGCAATACCTTACATACTGT	GCAATCTCCACTGGTTTTATAGG
WP_004906105.1	GTGAATAACTCAGAAACACAACCG	AGATGGCGGCGGC
WP_036958810.1	ATGGAATCGTCATCAACTCAGT	CAGCTTACCGCCGTAATAGTA
WP_004264499.1	ATGCATATCATAATTAACGACCAACC	GCCCCCAGCGATAGCT
WP_004264504.1	ATGAAAATGCCAAAATTGCCTAG	GTTGGCAGTATTTCGCTCAA
WP_004264615.1	ATGCCTAATTCATTCATTCGCG	GAATGCGTAATTGAACTTCACAC
WP_004264748.1	ATGTGCGAGTTACTTGGCA	CAATATGCGCTCACCAAGATAG
WP_080544493.1	ATGCAAGGAACAAACAATACACT	AAAACGGAGCTTAACACCTAGC

gene ID	forward primer sequence	reverse primer sequence
WP_004264896.1	ATGACTATCGAAAATAATACCGAACA	GGCTTCAGGTAACTTCAATTTTG
WP_004264902.1	ATGTCGCAGCAATCGAATAATAA	GGTATTTTTTACTGGGATTGTGTCC
WP_004912645.1	ATGTCTGATATGACGACAGATGC	GACTGTTTTTACCTTGATAGGTTCC
WP_004265131.1	ATGTTCTTCAATAAAGAAAATGCGG	ATTTAGTTTAACCCAGAAATACTGAAGA
WP_004265317.1	ATGACTTACAAGCCGTCAACC	ACGTTGTGCCGCAACC
WP_036959118.1	ATGACCCGAGTGACTTTATCAT	TTTCTTTAGCCACTCTATCAGTTTT
WP_004265549.1	ATGTTGCAAGAGGAAGGCA	CGGCTGGATATCACCGTT
WP_044172271.1	ATGCAGGTTAATTATCGCTATCAG	TGCTTTTACCTTCGCCGC
WP_044172624.1	ATGACATCAGTACAGGCAACC	GATGTAGATGTCAATCTGATGATCG
WP_044172978.1	ATGGAATGCCGTCCGGA	CGGCGCAGTCGCGT
WP_044173012.1	ATGGTCAGCAATAATACTCAGTCG	AAGGTGCAAAGACGCAAGA
WP_044173054.1	ATGCCTATCGCTACGCG	CGCCCAGTTTTTTCTCTTGC
WP_044173232.1	ATGAATCCGACCGAACTCG	GCGGATAGAGAATCCATTGGT
WP_044173357.1	ATGACGAATTCCCTGCATACC	AAGTTTATCCATCAGGTAACGATG
WP_044174146.1	ATGAATATCAGCACCGCCT	GTCACGAAGGGGTAAAAACTG
WP_044174298.1	ATGCCCCAGCGCACA	TTGTCTTTTTTCCAGCCGC
WP_044175142.1	ATGAAAGACAACCAGGAACAGA	AAGTTTTTTTCGCGCCAGA
WP_044175145.1	ATGGAACAACGCCGCT	GGGAAAAGTGACATCGCCC
WP_044175597.1	ATGAGCGCAAGCATAACCA	AGCGCTTAACAGCGTATTCA
WP_044176371.1	ATGGGACAGCTCGTTAGC	CGTGGTCTGGCTGCC
WP_044177343.1	ATGTCGAACAACGATGAACATC	GGCAGGAATACGCTCAA
WP_044177448.1	ATGCGATTTTTGATCCCGC	CGAGCCGGATTTCGGAT
WP_044177550.1	ATGTCTTTCACCATTAACGCC	GAAGCTGTAGCTGGCGC
WP_044177605.1	ATGAGCATTGATCGTACATCGC	TTTACTCTGTAAGTAACTCTGCACT
WP_044177823.1	ATGTACCACTCTTTAAGCCGT	ATGATACAAACTAGAGTGAATTTTTATGC
WP_044177883.1	ATGAGTGAGTTTTCCCAGACC	ATATTTGTCAATGACATGCGCC
WP_044178555.1	ATGAGCGACGACAATTCACA	CTCGTCCAGTTTTGGGGC
WP_052332698.1	ATGCTCTTTCCCGGCAC	GAACCCCTGAGCACTTGC
WP_071825796.1	GTGAGTAATAGCCCATATTCTCTAAG	ACCCTTTCCTCGACAATCAG
WP_071825830.1	ATGACTGCAATAGCCATGAAC	TTTGCCGAACCAATTATTACCTAC
WP_081653585.1	ATGATAAGCGATATCGATTGTACG	ACCCCAGAAATTCCAGATGAT
WP_081653590.1	ATGAAACCTTGTTTCAGTTCAAATT	CAAGGGGATAACCAGAAGCT
WP_081653604.1	ATGACTCAATCTCATATCCGGC	GAACTTCCGCTTTCTGAAACC
WP_044179714.1	ATGAGTGAAGACGTTACCCC	TACCTCCTGCTGCCAGG
WP_044179802.1	ATGTCTTCTCCTCCCTTCGT	ATGCAGTACCGATACCGC
WP_044179949.1	ATGTCAGAAAACAGCTATCAGC	GCCGTTGCGTTTATCTTCG
WP_044180054.1	ATGCGCTTAATTGATTATTTTCCTC	CGCACGTGAAATGATGGC
WP_044180111.1	ATGACACAGACGTTAAGCCA	CTGATAGTCGCTCATCGGC
WP_044180332.1	ATGCGTAATAGTCAGAACATCAC	CCCCATCAGCCACATAAAGA
WP_044180369.1	ATGAAATTTCATCAGGATGGTTGTA	GTCATTTGTTAAGTACCGTTCCG
WP_044180423.1	ATGTTTTTCCGCACTCAAAACG	TAGCTTATTAAGGTCACTGGCA
WP_044180429.1	ATGAATACGCAACGAAAATACGG	TGACGGCTCTCCGTACT
WP_044180441.1	ATGGCTGAATCAAACCCGT	TTTAATGAAAGGACGGGTAATCAA
WP_044180510.1	ATGCAAAGTAACATCTCCACCC	AAAACCGGCGGCGG
WP_044180562.1	ATGCAAAAAAGAGTCTCCACTTAC	CAGGGGTGATAATTCTTGGGA
WP_052332703.1	ATGAAAAATATTAAAATCACAGGCAACG	TCTTTCCCAGCGTATGTTTTG
WP_044181333.1	ATGCACTGGCAAACCCAC	GCCAACGCCGCCCG
WP_044182930.1	ATGAAATATGCATTTCCCGAGC	CTCTTCATCAAAGCCATTATTTTTGG
WP_044182945.1	ATGTCAACAGCGAAACGC	TGCTTTAGGCTCAATGCCC
WP_044183152.1	ATGAATACCAAAGAGCATGACAATAGC	TGCCTGGGCGCGCA
WP_044183301.1	ATGACTGAAACTTATACGACTTCC	CTGCGGGTACGGCAC
WP_044183362.1	ATGAGCCACACAGCCTCC	ATCAACCGCATCCAGAATAGC
WP_044183672.1	ATGCAATTCTCACCTGAAAGC	ACGGGAAAGGGCTTCCA
WP_071825907.1	GTGGCTGTAGCTTGCATAAAA	ATCTTTTAAAAATGTGAAGCATTTCATAG
WP_044184806.1	ATGGCAATCAGTGAAACCCC	TTCTTTATGTCGGTTATCGGGC
WP_071825927.1	ATGCCTGAGATTATGGAGACG	CGCCTCCGGCGAAC
WP_044185675.1	ATGACGGAACAACAACACTCC	TTCTCCCTGCGGCGC
WP_004864811.1	ATGGAATGCCGAATCGACT	TGATGAGGTCAGCTTTTCAAGA
WP_016517497.1	ATGGACGCCAGCCATTT	TGCGTTATCTCCGCGAAC
WP_016517519.1	ATGTCCGAACACAACTATCAGC	TAAACCTAAGGTGTCAAAATCAGAA
WP_016517591.1	ATGAGCGAAACCCCGGC	GCCTTTCCAGGGATGAATCA
WP_016517593.1	ATGGCATCTGATACCCAAACT	CAGGAAGTGAAGGAATTTCCCG
WP_016517623.1	ATGCGCTGGCAAACTCA	GGCCATGCCTCCCGT
WP_016517628.1	ATGAGTAAAATTGCTTCGATTACCC	CAGGAATGATTTAAACGGCAAATC
WP_016534698.1	ATGTCCGCCCATGTGC	TCCCAGCACGGTTACCG
WP_016535010.1	ATGCCAGGTATAAAGAACAACG	CTCTTTAGAAGTCGCGACGC
WP_016535037.1	ATGGAAAAGAAAAACCGCCC	AGCGGCGGTAAGGTTG
WP_016535212.1	ATGCCCGCTAATATTCAGGAA	GTTAATTATGCTCTGGTATTCATCGA
WP_016535238.1	ATGTCCATGCAATCTCAGGA	CTCATGGCTATCTTCCTCGTC
WP_016535301.1	ATGATCAACGCAATTAACCGC	TTGCTCCTGGCCGCTA
WP_016535304.1	ATGCAGGTTGGTTTAGCTACA	TCCCAATAAATTTTTAACCGTCGC
WP_016535485.1	ATGAACTCACTTATTAATCATTCACTTG	ACGGCTAAAAGCCGCTT
WP_016535497.1	ATGCAGATATCATCGGGACC	GGCGAATTTCGCCATATGG
WP_016535500.1	ATGCCTTCAGTCAACACCG	GAAGGAAATTTTGTTCTGCGC
WP_016535503.1	ATGGATAACACATCCCCGG	GTCGTTCGGCATAAAAGAGTG
WP_016535516.1	ATGATCCATCACGCGCTAA	TACCGCCTCCCGCC
WP_016535531.1	ATGGAGATAAAGCTGCACTCC	TTTACCAGCCGACGGC
WP_016535715.1	ATGTTAATGCTATTTTTCAGCGC	AAAAACGACGGTCGAGCC
WP_016535835.1	ATGAGCATTGATCGCACGT	TTTACTCGATAAATAGCTTTGGGC
WP_016536039.1	ATGAAATTCAGTGAACTCTGGTTAC	ATCCCTCAATGATGCCTGG
WP_016536076.1	ATGAGTTTTCCCATTGAGCAGG	TCCTAACAGGCTGTGGATACG
WP_016536247.1	ATGAAAGAGTTTGCATTTACTATTAAGAG	ACGTGCGTAATTAGCGGA
WP_016536337.1	ATGAGCAATTCTACAGACTTACAC	CTGTTGCAGGTCGAAACG
WP_016536352.1	ATGCCAAATATTCATATCAGCCC	TGCTTCCAGAACATCTTTCTCC
WP_016536353.1	ATGCCGAATGCTCATATCAGC	ATCAGGAGGATTGAATATTCTGGA
WP_016536389.1	ATGAGCGTTGAGTTTCTTCC	CTTAATGCCTGAAAAACCAAAACT
WP_016536444.1	ATGCCGATAGATAAAATAGCTGC	TCGGCCAGTGTATTTATCCAC

117

gene ID	forward primer sequence	reverse primer sequence
WP_016536503.1	ATGCTTTTACTGCAAAACTGGG	TCTGCTGGTAGATGTTTCCC
WP_016536512.1	ATGAGTATCTCCACCCCGA	TCGAGTCATCGCAAATGCA
WP_016536835.1	ATGACAACACAAACTGGCTC	CTCATTTTCGTTCTCCTCAATCA
WP_016536850.1	ATGATTATTTGCTGGAGCGC	GCTCTTATCTGGCCCTAAGG
WP_016536932.1	ATGAATAACTGGTTACAGCAGC	GGCAGGAAGAGCCAGC
WP_016537145.1	ATGCCGATTTCCCTACATCC	GAGTTTATCCATCAGGTAACGATG
WP_016537273.1	ATGTTCACTGCCCTGAAAAAG	GGACAGCGCGGCAA
WP_016537316.1	ATGACCTTTTCAACCAATACTTCTG	GATGGCGAGATCGGTCAG
WP_016537327.1	ATGGTCGATAATAATTCAGCTACG	CAGGTGCACAGACGCA
WP_016537334.1	ATGTCTTTCGGGAGAAAATTGAC	GATGAAAGCTTTATCCCACAATGA
WP_016537335.1	ATGGAAAATGAAAGCTTATCTGTTG	AAAGTGCTCGCTCTGGC
WP_016537607.1	ATGACCAATGGAATACAACAGATAACCG	GCCCGCGGCGGAGG
WP_016537639.1	ATGGCAAGTTCTCAAACCCT	GGATTTATGGCGGTTATCGGG
WP_016537678.1	ATGGGGGAGTCGCGAAGT	CGCGTTTTTCTCGACGA
WP_016537795.1	ATGAAATACCGCAGAGAGAGTAA	TGCCTCGCTTGTTATTTTGG
WP_016537847.1	ATGTCATGGTCACAGGCTAC	GCCAATCATTGCCTCCAGT
WP_016537909.1	ATGAGTATAGCTTTGAATTCTATTCAGG	GGAACGCAGTAAAGTCTCTACA
WP_016537915.1	ATGGGGCGGCGGAATTT	TTTTTTATTTGCTGGGCCGT
WP_016537917.1	ATGTCCTCCATTCCAGCAAA	GCCTTTAATATTATTGGTCATTGAGC
WP_016537918.1	ATGGATTTGAGTTCTGCAAAATCA	AGCCTTCATGTTTCGAGTAACA
WP_016537985.1	ATGAGTATGTTTTCTGCAGTG	TTTCTGATAATTAATCACCAGATGGC
WP_016538127.1	ATGAGTAATGCAGAAACAACCC	CAATATTTTGTATACGATGTTCAGAATCG
WP_016538147.1	ATGAGCGATACCCTGAACTTC	ATAGCCGTTAACCTGTAAGAATGC
WP_016538154.1	ATGACTGACAAAATCACTTCTGTT	CTCGTCTTCCAGATACGTATAACC
WP_016538267.1	ATGAAAGTTGGCGTTCCTCT	ACGGTGCCTTGCCAAC
WP_016538739.1	ATGCCACATATCAACTCAGCA	CTGATGAGTATTTACTGTTTCTCCG
WP_016538749.1	ATGACGCCATCTACTGAAAATATAAG	CTGCTTCAGTTTGGGATCGA
WP_016538985.1	ATGACGGAACAAAATAATTCCTCC	TTCCCCCTGGCTTACGG
WP_039898053.1	ATGCACTCACAGAACTTACAAAC	ATGATACAAACTAGAATGAATTTTTAGGC
WP_039898129.1	ATGCCTAACACCGTTCTCG	ATAGCCTAATGATTTCTGAATAGCAG
WP_039898184.1	ATGATCTTTCAAGTGAGTAATGGATC	AGACCACCTGGGCTTTAGA
WP_039898190.1	ATGCCAATTTCAGCATCGC	TGTTTCTTGAGCTTCTCCGG
WP_039898226.1	GTGAATACTACAGACACCTCAGC	AAACTCTCCCATGTCATAGAGTT
WP_039898229.1	ATGCCTTATTCAGCCTTGCT	GGTGATAACATCTGCAATAGAGG
WP_039898312.1	ATGAGAAACTTCATTAACACTTTTCGC	GGCGCAGCGGGCTT
WP_039898519.1	GTGAGCCACTCACATTCTCA	TGGCTTCTTGTCCGCC
WP_039898535.1	ATGACTACAATATCCAGCAGTGC	GATATAAATATCAATCTGGTGGTCGT
WP_039898704.1	ATGAATATAGATAGCTCTGCATTCAC	CCAGTCGGCATTTTCATCATC
WP_039898721.1	ATGCTGCCCAACCCCT	ATCCACCGGAAGAGCGT
WP_039898948.1	ATGACGACGCCGGATAATAG	TTTATGCGTCTTGCCCGG
WP_039898953.1	TTGGCAAACTCATCTTCTTCAC	TCCAGGATATCCATCAAAAGCC
WP_055696358.1	ATGCCGGTATCTGCATCG	AAAAGATCGGCACATATACAGATG
WP_055696359.1	ATGCGGGATGAAGTACTGAA	GTAAGGCAGCAGGCTGG
WP_055696403.1	ATGAAAAATCTGCGCGTCAC	ATCCACCACCTTACCAAGC
WP_055696404.1	ATGAGCTTCCTGCTGACG	ACAAGAAGGTCTGAGGTAGC
WP_055696405.1	ATGCATAATGGACACAATTCGC	CGCCCGAGGGTGGT
WP_083478381.1	TTGGCAACGGTTGTACCG	ATAGTGGAGTATTGCTATGCGC
WP_083478383.1	ATGATTTCAGCGAGGTTAGTACA	GGACAGCGCTCCCGG
WP_033747833.1	ATGACTCAGCGACGTGG	ACTGGCGCTTTTATCTGCG
WP_033789480.1	ATGAGCATCGACAGAACGC	TTTACCCTCTAAATACGCCTTAGT
WP_084881949.1	ATGAATCCATACCGTCAACATCC	CGCGCGCGCTTTGT
WP_084882006.1	ATGAATATCAACAGCACGGTAATTA	TTCCTGATGTGAAGGGAACAG
WP_033790137.1	ATGTCATCATCGGTTTCCTCT	CGCCGCGAACTGCT
WP_084882062.1	GTGAAGATCAGCGCAATCG	CCAGCGGTTGAGCGT
WP_084882072.1	ATGCAATCAATAAGTCTGCAAAAC	ACTTTCCAATGAAAAGCATGTTG
WP_084882135.1	ATGACCCAGACTTTGAGCC	CTGATAATCGCTCATCGGCA
WP_033753257.1	ATGAACATCACAGGAAACTTCATT	GCTCAGGGTCGGCATG
WP_033792699.1	ATGCAGCTCACATTTCTCGG	CAGCGTAAACACCGCCA
WP_033791249.1	ATGCAACAGGAAACGCCG	GGGCGGCGGCTCAGG
WP_084882856.1	ATGACGTATGGCGGTCAG	CGCTCTCAGGCCAACAT
WP_033792330.1	ATGACCTGGACAACCCATACCG	CCCGCCGCTGGCGC
WP_033792202.1	ATGACTGCCTCTTCTTCCAG	GAAATGGTCGGCGATCGG
WP_084883315.1	ATGTTTTCCACTATTTCATCAGGT	CTCCTGGCTTAAGGTCTGG
WP_084883443.1	ATGATGCATTGCCCGCT	GGCGCAAACCTCAGG
WP_033754389.1	ATGGCAGAACATCGTGGT	GGAAGAGCTGCGGCC
WP_084883617.1	ATGACGATTTCAGCCGGC	CAGCAGTTGAGCAAGCCG
WP_033790314.1	ATGACCGATGAACAAGGGC	AACGCGGGCAGGATG
WP_084883665.1	ATGTGGGCAAACACACTCA	TTTAGCTTTCGTGTCCATGAAA
WP_033791692.1	ATGCAATGCCGTAGTCAGTG	GGGCGCGGTGTCCG
WP_084883854.1	ATGCAGGTTACGCCTCC	CCGCTCCGCCTCCC
WP_084884045.1	TTGACCAGCTCCCAACTTT	GTTCTGGTTGAACCAGTTTAGC
WP_084884150.1	ATGTCACAGCAGAACGTCA	TTCCGGTTTGTACGCCAT
WP_084884154.1	ATGCACACAGAAAACGTCAAC	GAAGGGAATCTCGTTCAGAGT
WP_084884203.1	ATGCATAAGCAGACACCTCT	GTGAACTATATCAGGTACTTCAGGC
WP_084884216.1	ATGCAGAGCGTTTTTATCTATGT	TACCTGATGATCTGCTACATACTT
WP_084884247.1	ATGTCCATTATAACTACGCCTGC	GGCGGCCGCCTGAT
WP_033751802.1	ATGAGCGACGACCATTCTC	TTCGTCATCCAGTTGGGG
WP_084884338.1	ATGGCTAGTTCGACAATGAAAA	AGGCTGGAAGGTGCCG
WP_084884455.1	ATGACGTGGTTAACCAACGA	AAGAAGATGTTGCCACATTGC
WP_084884529.1	ATGAAACCCATTATTGGATTCAATCG	TAGGAAAATAATGGCATCATTAGCT
WP_084884582.1	ATGACGGAACAAAAAGCGAC	GCCTCCTTGCTGTTCAGC
WP_071997399.1	ATGGCCGAACATCGTGG	ATCAGACTTACGCCCGC
WP_084884647.1	ATGGAACAACGCCGCTAT	CGCCGCTTCGCCCA
WP_033793256.1	GTGACCATCCATATCGCCA	AAAACCTACGGATTTGAGGTAGT
WP_033793269.1	ATGCGCTCTCGATTAGCG	TTCATCGCCGGTACGC
WP_084884932.1	ATGACAACCCTCTCTGAGCA	TCGGCCGCCTCCGA
WP_084884937.1	ATGGAAAATAGCTTGTACAATTCATATG	GCTGATAAGGCTGCCTGC

gene ID	forward primer sequence	reverse primer sequence
WP_033753922.1	ATGTCCGAACTGATTGTGACG	GAGCGCCTTCACAGATATGG
WP_084885202.1	ATGAGTAGTAAAGAACAGAACACCC	GCTTTTCGCTTTCGCGA
WP_033793566.1	ATGTCGTCCAATAATCACAACAAAA	GGCGCCGCCTGGTG
WP_084885483.1	ATGACTTTTGACTCACCGGA	AGCCCGAATCGGTTGAG
WP_084885824.1	ATGACGACTAAGCATTTGCG	CGCATTCTTTAGCCTCCCA
WP_084885933.1	ATGAAAGATGTCGCCCGC	AGTCGAGGCGCGCA
WP_084886065.1	ATGAAACAACCTGCACACCG	GTCTTCCTTAGCGGATGGC
WP_033750506.1	ATGGCGAATGATTCTGGCA	AAGATCGTCGAGGAAGGTG
WP_084886210.1	ATGAAACTTAACCAGAACCCTACA	CATGTCCGCACGCGC
WP_084886311.1	ATGGCTGAACATCGTGGT	TTTTTCTTTACGAGCGCCG
WP_084886441.1	ATGTCACAGCAGAACGTCA	CGGGAATAAGCTGATCGGG
WP_084886471.1	ATGCACACACAAAACGTCAA	GAAGGGGATCTCGTTCAGAG
WP_048785467.1	ATGAAGATCATTTCGTTTCTGAATCC	CTCAAAAACTGACAATATTTCTTTAGTCA
WP_084886567.1	ATGTCATCGATATCCGCTTTATCT	GATATAGACGTTAAGCTGGTTGG
WP_108473781.1	ATGGAATGTCGCCCGGA	AGGTGCCGTTAGCGC
WP_108474137.1	ATGACCCCCGTTGTAAATCG	ACGGCTCTTACGGCGG
WP_108474230.1	ATGATGACTTACGATCGTAACCG	GTGCATGCCGAGACGA
WP_108474287.1	ATGGATATGCAATCTCAGGATATTATC	CTCATGGCTGTCATCCTCTT
WP_108474309.1	ATGACCTCGAGCGCTAATT	CGGACGTTTGCCGAACT
WP_108475277.1	ATGACATCACTTCCCTCGTC	TCCGTTTTCCTTACCCAGC
WP_108475341.1	TTGAACTATCAGATGATCACCACC	GCGCGGATACTCGGC
WP_113856782.1	ATGGAGCAACGCCGC	CGCAGGTTGCCGCT
WP_113857077.1	ATGTCAAACACGCCAATCG	ATTAAACGGTTGAATTGTCAAAGC
WP_113857093.1	ATGCCACTAAAAACATCACGTAT	GTCGGTGAAGCGGGC
WP_113857237.1	ATGCACGTTTTACCGCCG	CGCCTCCTCCTCATGACG
WP_113857245.1	ATGAAAATTGAATCTGTAAACGTCAC	GTCGTGGCTGACGGTATTA
WP_113857302.1	ATGGACATCGATAACATCATTATCTC	CCACATGTTTTGCCCGATT
WP_015965366.1	ATGAGCGACGACAATTCACA	GTCTTCCATTTTCGGCTGTG
WP_108474549.1	ATGTATACATCTGGTGTTGTAAATCG	ATGATACAAATTTGAGTGAATTTTTAGCC
WP_108474640.1	ATGACGCCGACCATTGATC	GCGGGTCGCCCAGC
WP_113844795.1	ATGAGGCATTTTACAAGTAATCAAATT	CCACCCATAGTTATTAATGGATGTTA
WP_113857422.1	ATGACATTATGGAAACAAACACAATAT	GTTGTCTTTTGCAACCTTTGTATAT
WP_113857452.1	ATGGATAATAATCGCAATTTACCGT	GAAATGCACGTTCGCGG
WP_113857471.1	ATGTCATCATTGCAAATTTCGC	TATTTTACCTACCTGGTAGCGAT
WP_113857569.1	ATGGACATCGATAACATCATTATCTC	CCACATGTTTTGCCCGATT
WP_008323346.1	ATGCAATATCCAACCGTATCTGTA	AATACCGGGGAGATACATCTGTA
WP_015963067.1	ATGCTTGATGTATCAGAGAGTGTA	GTGGCGCAAATTTTCATTCG
WP_015963250.1	ATGCGCAATACTCAAAACATCA	GCCCATCAACCATAGAAATAATAGA
WP_108475953.1	ATGTTATCCACAGAAAATTGGGAT	GTCATTACAGCGTAACGTGATG
WP_108476339.1	ATGGAAAGCACTCAAACCAGC	GCTGAACAGAGAGTAGAAGATAGC
WP_113857601.1	ATGATTTATAACACCGGAACCATC	ATCAGAAATCACTTCGCATTTCT
WP_113857629.1	ATGCACCAGGTTGAAGCT	GGAACGCGTTGACGTTC
WP_113857724.1	GTGAAGGAGGCTTCAGCG	TGTTGAAACATCTGTAGTTATGTTTCT
WP_108473431.1	ATGCGTTGGCAAGGTCG	CAGCATTTTGCCGAACGTA
WP_108473773.1	ATGGCGACAACATTTTCACC	TTTCAACTCTTCCAGTTTATTGATTG
WP_113857886.1	ATGTGTATCGGCAGCAAGC	CTGACCGAGCAGCGTT
WP_113858097.1	ATGTCCTGGCAAACGCAC	CACGCCCCCTGTCGC
WP_113858203.1	ATGCCAACATCTACCCCTTC	CTGACTAAAGCGCATCTGC
WP_015962672.1	ATGCAATTTACTCCTGATAGTGC	ACGAGAAATGGATTCAAGATCCA
WP_108476447.1	ATGAACGCTGCAACTCAAAA	GTGAGCGTGTATCTGGCG
WP_113858369.1	TTGGCTAATACAGCGAAAACG	TGCGTTGGCAGGCG
WP_113858376.1	ATGACCCAATCAACCCAAACC	CTGCGGATACGGTACCCA
WP_113858384.1	ATGATGCCAACTCCATCGC	TTTAATCACCGCATAAATCTCGC
WP_108475618.1	ATGATAAGTAATATCGATTGTAAGAGGTT	GCCCCAATACTTCCAGACA
WP_108475752.1	ATGAGCATTGAACGTACATCG	GTTACTCTGTAGATAATTCTGCGTT
WP_113858462.1	ATGAATAACGAATTTACTTTTACCATCAA	AAGCGCGTTGTGAGCG
WP_113858483.1	ATGAATGCACTGACCGCC	CTGATGCGTCTGGTAGTGG
WP_113858572.1	ATGGTAGCGTCCTTTATCGG	GTAATACGCTTGCAGCGTC
WP_113858620.1	ATGCGCGAATACATAACATCTG	TTCGGCGTCACGCCA
WP_113858661.1	ATGTCGGATCAACACAATAAATCAA	GGCCGGTATCGCCGC
WP_015965623.1	ATGTCTATTCGGGAACTGATCG	CGCTTTACGTTCCGGCG
WP_113858733.1	ATGAATAACACAGCAGAAGGTCG	TTTAGTCAGGATCAATTTGCCG
WP_113858742.1	ATGGGATTGATTTCTCAGGGG	GCCTACGATTAACTGATGGTTTG
WP_113858745.1	ATGGGCATGTTCCTGACAG	CTGCGCCGCGCAGA
WP_113858811.1	ATGGCAATCAGTGAAACCAGT	CTGTTTTTGTCGGTTATCAGGC
WP_113858817.1	ATGAAGATCACACTGCATTGG	AAACCGGTAGGGAAAAGCG
WP_108474922.1	ATGCCACGAAATTTTTTCAACATA	TGCAAGTCGACGACCTTT
WP_108475013.1	ATGAGCGACTCGACTTCTG	TTCCGCCACTACAACAACG
WP_113858928.1	ATGAGTACGACAGAAAGTATTAATGTAA	AAACAGTTTGTAGACGATGTTCA
WP_113858941.1	ATGTCTTCACAAACCTGGGT	ATGCAGAACCGTTACGGC
WP_113858958.1	ATGACCCCGCTTTCATCG	CGGCGTGACCAGCC
WP_113858981.1	ATGTCAGACAATACCTATCAGCC	ACCCTGACGTTTATCTTCCG
WP_000116680.1	ATGAGTGTTTCTCTTTCCAAAGG	GGACGCGTTAATGCCGTA
WP_000703842.1	ATGAAAATCAGCCAGCTGGA	CTCTTTTTCCTGCATAGCGG
WP_011152995.1	ATGAAGCTATTTTCAATTTCATCCCG	AGACCAGCGGTAATACTTAAAATATT
WP_113859044.1	ATGAGCCCGTCACTGTTTAG	GACGCCGTAACGTTCGC
WP_113859080.1	ATGAGGGAGAATATCATGTTGCC	TGCTCTGAGGCCATGCT
WP_000817037.1	ATGAAGCGTGCTCCTGTC	GACTCCTGGTTTCTCAAGCT
WP_001531161.1	ATGACGATGAACCTGCACA	CTCCACTTCCAGCCAGGC
WP_001743098.1	ATGACGATGAACCTGCACA	GCAGTCCATCAGCCAGC
WP_020319858.1	TTGCACAATACACAAACGCC	GAATTTTTTATTGAACATCTTAGGGTCG
WP_001516695.1	ATGGAAACCTACAATCATACATATCG	GTCAGGATAAACAACAATACCCAG
WP_001297012.1	TTGGTGCAGATCACTTCTGA	TGAGAGATACCAAATGTCATTTTCA
WP_000932975.1	ATGCTGAACAAAACCGAAAACA	TTCAGACACGGTAATGCCTAG
WP_000490639.1	ATGTTCTCTGAACGCTCAGTA	GGAACTTGTTCGCCACCTTC
WP_004723723.1	ATGTTGAAAAATATCAATCCAACGC	TTCACCACGAAACGCTTTG
WP_004724218.1	ATGGGTAAGTTGAGTATTCCAAGT	CATCCAACCAGATCCTTCGC

gene ID	forward primer sequence	reverse primer sequence
WP_004724221.1	ATGACTTATAAAAGGGATCGTCGC	TAATTCAAGAGTATCGTCACCTGA
WP_004724237.1	ATGACCACCGCGATTACC	GGCGATATCATCACTGCTATGG
WP_004724260.1	ATGAGTAGTATTCAACAAAATCTTGAAC	CGCGCGGGGGTAATC
WP_004724348.1	ATGACAAGCAATAAAAATAATCCTCAG	GTTAGCCTCCGTTGTCATCG
WP_004724472.1	ATGAATTTATCGCCAATAGCACA	GTTGTTATTGATTACCTGATAGCGC
WP_004724690.1	ATGACCAATAAAATTGTTCCATCCC	TGCGGCGTTTTGCATCC
WP_004724988.1	ATGAACAGCGTCATCGACA	TTTCTTCGCTAGGCCTTTGC
WP_004725080.1	ATGGTAATGAGCACAGTGGAA	CCGCAATAACGCCACCA
WP_004725381.1	GTGAACCCTCAAGCCGAT	TTTCTGCAAATCGTACTTTTTGAG
WP_004725399.1	ATGCCTACAGAAAAACCTACAATTA	CGCGCGATCCACGAT
WP_004725469.1	ATGTGGTCGCACCCTGC	GAACATCGCCAGTTGGTTG
WP_004725751.1	ATGTCGTATTTGCCTTTGGATC	ATCGTTGTGTTGGCCAATTT
WP_004725800.1	ATGTCTCACGAACAATCACAGA	CAGCGACAGATGACGTTGA
WP_004725807.1	ATGTGTAATGTTTTAGTGTCTTTTGC	ACTCAACATTTCCTCGGCC
WP_004725809.1	ATGAGTCACATCCACTCGC	GCTTTCAAGGCTGAGGTAAGT
WP_004725893.1	ATGAGTTCACATCATTCTTCACG	GATTCTGCCTTGTAGTGGGA
WP_004726052.1	TTGACTCAGTCACAGCCCC	ACAGGCATCAATCGCGT
WP_004726083.1	ATGAACAGTCCTATGTATTCACGC	ATCGTCGCGCATCACG
WP_004726095.1	ATGAAATCTTGGGTGCGTTC	GGCGGCGTGAGGTG
WP_004726164.1	GTGATCACCTTGGCGACT	ATTAATTTTAAAACTATCTCGCCACAC
WP_004726209.1	ATGAAATTCAGCGAATCTTGGC	GTCGCGCAGTGTGGC
WP_004726225.1	ATGACGCAGCACCACG	CTTCAGTTGAGCCAGCTCG
WP_004726235.1	ATGCTTCACGCAGATTTCAAG	ATGGAAAGCGGTCGGGA
WP_004726312.1	ATGAAAATTGATGCCAGCAACG	GTTTGAGTTCACTAACGCGT
WP_004726386.1	ATGGAATCTAATGTAATCAACCAAGC	CAGAACGCCCATAAATCGC
WP_004726428.1	ATGCTTACTACCAAAATTCAACAGT	GTATTCAATGGTGCGTTTAAACG
WP_004726603.1	ATGAACCCAGTAGGCGTAGG	AAGCAGTAACAAAATACTGCCG
WP_004726677.1	ATGAGTAACACAGGCACAAAGT	TTGACAGATCTTAGCGATGTCT
WP_004728288.1	ATGACAACACAAGCAACCCG	GCTTTCGCTTTTGTATGAATAGTT
WP_004728297.1	TTGTATTCTCTAGTTTCTATTATTACCCC	CTTACGAAAAACGGCTCGC
WP_004728470.1	ATGATCGTGTCACCGACAA	CAAGACCGCAATCACTTCTG
WP_004728834.1	ATGCAAATGAACATCACTGGTAA	CTCTTCCACTTCTTCTTCCAGC
WP_004729099.1	TTGAACGTGTCGTCCAGTT	ATATCCGCTGGAGGTGAAATC
WP_004729342.1	ATGACGATTAAAGATGTAACGGATG	CACTTCAATATCCCGTGAAACC
WP_004729371.1	ATGGCCCCCGTATCAACG	CAATTTTAGCCGCTTTAAGAGATAG
WP_004729387.1	ATGTCTTACACGCCAACGA	TTCTTCCGAGGCTGGTGT
WP_004729399.1	ATGAATACAGAACAAGACAACATCA	AGGTAAGGTGAATCGTGCTAC
WP_004729437.1	ATGATCATCACGGTCAGTAGC	TGCATTCAATTTGAAGAAGCTCA
WP_004729460.1	ATGAACCAACCCAACATAACTC	AAAATCCAGTTGCCAGCCC
WP_038150968.1	ATGTCAGTTACTGGTTCGGC	GGCGTGATCAATCGCGT
WP_038151086.1	ATGAATACAACAACTACGGTTTCG	AAGAGCGCCGTAGATTACAG
WP_038151258.1	GTGAGCACACACCATTCG	TTCGACATGAAAATGCGTTTTC
WP_038151273.1	ATGAACGCCAGCATAACACT	GCCCATTTGTTCGGCAG
WP_038151305.1	ATGGCTCCGAAAAATAGCACC	TTCATGATCGGCCCCCT
WP_038151352.1	ATGGTTTCAATTAACGGACTGC	GATATAAATATCGACCCCAAGCATC
WP_038152552.1	ATGTCGTCATCACACTCTTCA	CAGCTTTACTTGATATGGGTTGA
WP_038152592.1	GTGTCAAATCCTATTACTATTACAGAATC	GTAGTAAGAGTGCTCGCCTC
WP_038152705.1	ATGCCCTCGCAAAATATACAGC	GAGTGATTTACGATGACGGC
WP_038152727.1	ATGAGTAACCAAGCCGTAAACT	AAGCGTGTACGTGTACGG
WP_049781808.1	TTGATATGTATTATATTGCCGCTAATTTC	TTGAACGGTGGAGTGAATCT
WP_081454515.1	ATGCTCTTGGGCAGTATGAC	ATGACAGTTCAGTAAGCGAATC
WP_086027288.1	ATGCTGTCTCAGCCCCTA	GCCGTGGTTGATGGTTTC
WP_115333225.1	GTGCTCCATGCTTCCAATC	ACGAGTCAGCCAGGTGC
WP_115333235.1	GTGTGCCCTGCTCTCAC	ACCACTGAAATCGTATGGATTTAAT
WP_115333239.1	GTGCAAATGCAAAAGATGCAG	AACTTTGAACGCCTGCAGT
WP_004726765.1	ATGAACGAAGACATCTCGAGT	ATTCAGTGAAGGCAGTGGAT
WP_004726788.1	ATGAATCAGCTAGCCAGAAACG	ATGGGTGTTGCTGAACTGG
WP_004726842.1	ATGAGCATCAGCTTCTCAGAA	CTTCCTGAATTGAATTTTGCCG
WP_004726871.1	ATGACCACATTCAACACACG	TTTATTTAACGCTTGGTTGAGCC
WP_004726981.1	ATGACGCAGCATTCACAGC	ACGAACATTCGCGTGACC
WP_004726990.1	ATGAGTACCCAACCGACAC	GGCGGAAACGTATCGGC
WP_004727003.1	ATGCAAATTCAAACCGTTAACCA	GGTCAAAATGCCCAGCG
WP_004727165.1	ATGCTCGCAAGTACGGTT	ACGCCAAACGCCCC
WP_004727345.1	ATGAGCGAAACAAAACAATCTGC	CCTTTCTGGTGGCAACGC
WP_004727405.1	ATGGTGCATCAAGTAACTCGT	GTGGGCCTCGTTCCCA
WP_004727567.1	ATGCCAAGTACTTATACCTGGG	ACGCTTGGCGTGAATATTG
WP_004727615.1	ATGTTAGAGAACCCAGCAAACA	GTCTACGGAGCGCGAAT
WP_004727628.1	ATGTCTACAACCATCGCTGC	GCCTTTAATGCCGCTGC
WP_004727644.1	ATGTCTGATTTTTCATTTACCTCGC	AGAAAACTGGCTTTGTAACCAA
WP_004727658.1	ATGGTCAGCCTGCTCGC	CAGCCCAGACACTTGAATGT
WP_004727675.1	ATGAACTCTGCATACACCTACC	CTTAATCTGAGTCAACACATAGTCC
WP_004727709.1	ATGGCATCGACGTCAATTCA	CTGAGCTAGAAGTCGCGT
WP_004727750.1	ATGAATACAGGCTTTAAATTTACGATTAA	CAAAGGTGTGTTGGCAGAC
WP_004727751.1	ATGAAAAAACTATTACCAACTTCAACC	GGACTTCGCCAAAATCTCTTG
WP_004727769.1	ATGACTGAATTACTTCACAGCCTA	CTCTTCGTAATTCTCAATGCTCG
WP_004727788.1	ATGTCGAAAATCAATACACAGTCA	CTTTGATGCTAAGTGGGTGTTG
WP_004727792.1	ATGACGATCGGCAGCG	CGCCATGACCGACCAC
WP_004727926.1	ATGCCATTTTTACAAGCCACT	AATCGACCTGTTTTCACTCAATG
WP_004728048.1	ATGAGCCACGCGAAGC	CGAATGCTTGATGGCGTAG
WP_004728083.1	ATGTGGATTTGGCAGCAGG	TATTGGGTATGAGGTCGCTAATT
WP_004728100.1	ATGACCAATAATCACCACTCACT	TTCCACTTCGTTCAATTGCG
WP_004728158.1	ATGACAAGAAAGAGAGAGATTACATC	TTGCATTGAGTTCATCGCTTG
WP_004728174.1	ATGAAATCAAACACCATCACGA	TCCCGTCGTTTCGTTGAT
WP_004728215.1	ATGCCTATTGAACCTTCTCTTATCT	GTGGCTGAGAAGCTCTCG
WP_014257346.1	ATGCCAAACCATATTCCTGATAC	ACGCTGTTTGGCGTATTCC
WP_014257409.1	ATGCCTGTTTCAGCCGTA	GACGCTGATATCAAGGAGGC
WP_014257429.1	ATGAACTATTTCCGTTTAAATGCG	TAGTGTGACCTTAAACGTGGT

gene ID	forward primer sequence	reverse primer sequence
WP_014257562.1	ATGAGAAAGCCTGTAAGAAAATCG	TTTTGAGTTGTTAATGTAGCGCT
WP_014258130.1	ATGCAAACACCTGTTATTTCAGC	TGTGCGCTCCTCCAGC
WP_038151743.1	ATGACCATCGCGACAAATAGT	CCTCGCTTTCTCACGGT
WP_038151757.1	GTGAATGGGGTGGGGC	CGGTTCCACATGAGCTTGC
WP_038151811.1	ATGAAATTAAACCATCACACTTACCA	GTCAGGAACCACAATCACAC
WP_038152009.1	ATGAGTCATACAATTCCTGGACC	AACGCTTGGAAACTCGAAC
WP_038152060.1	ATGAAAATACTATTGATTAACAGTAGCCC	TTGGGTTTTGATGACGTATTCTT
WP_004729624.1	ATGGCCCCCGTATCAACG	CAATTTTAACCGCTTTAAGAGATAGTG
WP_038152841.1	ATGTACCAACTAAATCTAAATTCAGGT	GCTGATCAGATCGTAGGCTG
WP_081454518.1	ATGTCTGAAACATCAAACTCTGAG	TTCAAAGTTCTTCATGGCCTCA
WP_000288735.1	ATGTACACTTCAGGATATGC	ATGATACAAATTAGAATGAATTTTTAAC
WP_001276378.1	ATGTCAGCATCTATCAATACA	CTGACTGGTTATATTATTTCGAT
WP_005158896.1	ATGAAACAATCTAATGATAAAAAAAC	TTTTGTCTGCCAGTATTTAAG
WP_005159587.1	ATGAATATCAATAATCAGATTAACAAG	GTTATTCTCCCATTCGATTAAT
WP_005162234.1	ATGCTGGCTATATCTGGA	TATATGTGTATTTTGTAATGACTGT
WP_005162694.1	ATGAGTGAATTAATAACACAAATTAA	ATATAACGCCGGTGCC
WP_005163567.1	GTGAACAATATTAATCACAAAAATAT	AATAAATACGCAGCCCC
WP_005165716.1	ATGGCCCAGCAATGT	TTTTTTCGTATACTTTTCAATATGT
WP_005178982.1	ATGGATTATGGTTTAACATTATTTTATCT	GGGGCAGCAACGCG
WP_005179706.1	ATGGGGGGTAAATAATACAAATTAT	ATAGACATACCAACTGCGT
WP_005181011.1	ATGCACAGCATCCAAGG	AGATATTGTTAAATATTGTGCAAAA
WP_005176491.1	ATGAAAATCAATACTCTTCAATCGTTAAT	GCCGTCAGCCGCCG
WP_010891206.1	ATGTTTATAAACCCAAGAAATGT	CTCAAATACATCATCTTCAAGT
WP_010891237.1	ATGAAAATATCATCATTTATTTCTACA	CATCAATGACAGTAATTGCT
WP_004238406.1	ATGACAAACATAATAAAAAACAGT	TAGTGTGGGTAAAGCCA
WP_062771682.1	ATGTCACCATCGCCAA	TATTAGATTGTTGATAGTGAGTTC
WP_004234827.1	ATGACTGTAATTATTGATAAAACATC	AACAGCTAATGGCCCT
WP_004242347.1	ATGCAGAACATTAGTTCTTTT	AGCCTTTATATCTTTATTTTCAATAG
WP_004238267.1	ATGATAATGAGTAATAAAACAGATAAT	GTGTTCCGGTGGCC
WP_062773651.1	ATGAATTTTTTACCTACATCTGA	CAACCCATAACGCAATCT
WP_040229899.1	ATGAAACATACTGATAATAAGACC	TGCCCAGTTTTTCTGTTT
WP_072041455.1	ATGAAATGCCCCTCATGC	AATGTTTCTGAATATGAAGATGT
WP_004254561.1	ATGAACAAAAATTTTACATTTACG	AAGCGCGTAGTTAGCA
WP_004254943.1	ATGCCACAATCGCCA	ATTTAAATATTCGATGTTAAGTTTGT
WP_004255002.1	ATGAAAAAAATATTTTTTTCAATCTGT	TTTGGTTTTTTTCTCTATAATAGC
WP_004255132.1	ATGAAATCAAATTTTTATAAAGCTATT	CTCACTAAAACAAAACATCTCA
WP_004255386.1	ATGGCGACTTGGAACC	TAAGTATTTGTTTAATAATGCTTCA
WP_004255405.1	ATGAATACTATATCAACAATTAACG	AGCTGGCAATGAATCCA
WP_004257237.1	ATGAGTATTAAAAATTTCTTTTACACA	TTTAATCTTATTGCTCCTGCT
WP_004258015.1	ATGATTAATAGAATTTTAATGGCTATA	ATTTTTACTATATGATAAATCCAGTG
WP_004258336.1	GTGAATATTTTATTCATTTGCAGT	GTTATGGTTGGTAAGAATACTAG
WP_004258949.1	ATGAAAAAATTTAAATTAATAATAAGTGC	TCTCTTATCCTTTTCATTTGATAA
WP_036957864.1	ATGTCTAACATTAGATCTTTTTCA	TGATTTTATATCTTTATTTTCAATAGCG
WP_036957944.1	ATGAAAAATAAACAAACTAAATTATTCA	TATTCCAAAAAAACTAATAAGCAG
WP_036958051.1	ATGTCTACCCCTTCTCCG	CACATCATTTGTTTGACTAATTT
WP_004260254.1	ATGAAATCAAGCCGCAC	TTGTGCTAATTTTTCATCATTTAA
WP_004261459.1	ATGTTTTATACTGACAAGAATATAATT	TTTCTCTAAAATTAAGTTATAAACTGT
WP_004905275.1	GTGAAAATAAGAAATATAGATAAGTTAAA	GTGAATTAAAAGTTTAAATTTTATGGA
WP_036958542.1	ATGAGTAACACTACACACTC	TTGTTTTAAATCATCATAAAGCAT
WP_004262559.1	GTGACGTTAGATACTAAAGTAAA	AGCTTTTACGCTATGTTGT
WP_004262990.1	ATGGTAGATATTAAAGCTTTAGGA	CGCAAAGCGTGCGG
WP_004263723.1	ATGCTCCTGCATAATTTAAC	GTATTTATTAGCCATATAAGTAGCA
WP_004264507.1	ATGTCCAATAATACAGAACAAAA	GGCAATTTTGGCATTTTCA
WP_004264858.1	ATGACTATTTCCTATAATGATGTAA	CTCCTTATCCTCTACTATTTCA
WP_004264927.1	ATGGATACTATTATGAAAACTCTAG	AAGTTTAATGTTTTTAGATTTGAAAA
WP_036959022.1	ATGAACAGTATAAATAGAACCTTAG	CTCCAAGTTAATTAATTTAGTGATG
WP_004265252.1	ATGAATATTCTTATTATTTCAGGAAG	GTGAGCGATAAAAAAACGT
WP_004906207.1	ATGACTTCTTCAAAACCGC	AATAACATTATTTTTTTTCATTTCCA
WP_036959188.1	ATGAGAAAAGATACAAGTATTTATATAA	ATAATTGTTATTTTCATCAATTTGAA
WP_044173350.1	ATGGCTGATAGTAATGAAAATTT	ACCAACTATTTTCTCGATTTC
WP_044177694.1	ATGACTATTTTTTCAAATAATATAAAAGT	TCGTAGATTCTCCGTCG
WP_071825876.1	ATGTTATTATTTAGTAATTGTTCGAA	CCCGGCACTGTAAACC
WP_016536514.1	ATGGATATGAATATTTCCGATTG	CTCTTTATTTGTTTTAGATATAATACGC
WP_016536523.1	ATGATTAACAATGTAAACTATTCAAT	AATGATGTTGGAAGTTCT
WP_016537275.1	ATGCTAAACCTAACCCCT	ATAACAAATTCTTTTTCCTTTATTTTT
WP_084882068.1	ATGAATTCACTAATTGATATTTTTACC	CTTGTCGGTGCCGA
WP_108473469.1	ATGGCCACTATATCTTCAAC	GATATAGATATCAATATGTGACGT
WP_004728310.1	ATGAATATTCTATCTCTAGATTCAC	TTTCACTCTGGATTCAATTTC
WP_115333279.1	ATGATATTTATTGTTTCAAATCAAAG	TTCATAATCGATTGTGACAATG
WP_004918212.1	ATGGCTATTTATTCTTGGAAC	GTGATAATAGTTAGAGTGAATTTTT
WP_005283179.1	ATGGGAAACAAGAACGAA	GCGCAGATCGTAGCG
WP_004918835.1	ATGTCACACAATAGTTCGA	AAGTTGTTCAAAATAGCTAAAGG
WP_014656597.1	ATGTCTATTGATTTATCATCTCC	CTTATATTCCTTAATATTAATGTCGTA
WP_005289768.1	ATGTCTAGCGATTCTTCTC	GATAGCCCATCCCCCG
WP_038256778.1	ATGAGCACACCTGATTTT	CTGACGGTAAGCCTGT
WP_071777555.1	ATGAATGACGACACCAATA	AATAGTGTTGCCATAGATGA
WP_004915527.1	ATGACGGAACACAATAAATC	CCCCTTCTGCTCATTAGT
WP_005284531.1	ATGCAGGTATTACAGAATCA	CCCTCTCGCCTCATC
WP_006819533.1	ATGCGGCAATTGACAA	GGCATCCCGCCATT
WP_006818304.1	ATGCATCACACTATTAGCA	ACGGTTAACCACTTTGG
WP_005285281.1	ATGCCGACGTTTCTTC	GAAGGGATTCGAGTGGC
WP_004916773.1	ATGAACTCTAGTGGTGAAAA	TTTAACGAGCTCATCATCA
WP_005287789.1	ATGACTATCCCAGAACAAAA	TACTGATTTCTCCTTCTCAGT
WP_004925628.1	ATGCCACAGACAGTGA	TACTGATACGGCATTGGC
WP_004919137.1	ATGACTGATAACATCACTAAAC	TTTGGTTAAAATCAGTTTACCA
WP_004918477.1	ATGTCGTCAGGCGG	CTTTTTAATAACAATGAGCGTT
WP_004922704.1	ATGAAGAAAAATAAGAGCAAGA	TTTATTAACGTTATCGGTATCTG

121

gene ID	forward primer sequence	reverse primer sequence
WP_004923882.1	ATGACCATTGACTTAAGCA	AACGGACTTTGCGACA
WP_014657167.1	ATGTCAATGCAATCACAAG	TTTTTGATCCTCCCCTTCA
WP_004924254.1	ATGTTTTTTGATAAACAAAATGC	ATTACTTAATTTAGCCCAAAAGT
WP_004919855.1	ATGACAGAGACCAATAAAAAC	ATCACTAAGTAGCGCTCG
WP_038258270.1	ATGAGCATTGATCGTACAT	CTTACTCTGAATCGAACTCT
WP_004916299.1	ATGCTTATCGACGCTATC	AAGCTGGGTAATTAAGAGTG
WP_036941148.1	GTGGTCCTGTGGAAATC	AATATCAGGATATTGCAGTAATAG
WP_004918442.1	ATGTTTTATTTTTCAACTAGTCG	TGCGCCTTTAGCGT
WP_042116289.1	ATGAATATCTATACTATTAGTAGTAGC	ACATTGTGAAAATTCAATATTCAA
WP_006820832.1	ATGTCATCTAACATTCAAACC	TTCTGCTTTTTGCCCTT
WP_006818629.1	ATGAAAATGATCCCCAGAG	GTTCTGAATGATCTCAACCG
WP_004921809.1	ATGGCTTTAATATTCATGTTATATAAT	TTGACATATTGCGTCTACTT
WP_004915379.1	ATGAAAATACGCAATAATACTTAC	CCAACGTTGCGGAAA
WP_004919570.1	ATGCACTTAATATTACGTAATGA	CAGTGTTTTACTTGTTACGC
WP_071777518.1	ATGCCGCCTTCATCA	GCAAAGGTTTGTGGTCG
WP_004921749.1	ATGAACTTTCCGTGTACC	TTTCCCCCTTATTTTGGTG
WP_006817037.1	ATGTCCACTACCGTTGA	CTGACTAAAGCGCATCTG
WP_005294857.1	ATGTCTCAGGCAGCC	TTGTTGTTTAGCCTGTGC
WP_004919091.1	ATGAAATCGACTTTATCAACAT	TCTTTGCAACAATGCGG
WP_087943223.1	ATGGCAATCTGCATGAG	ATTACGCTTGCTGTCAG
WP_005283191.1	ATGAGTAACACTGACAGTG	CAGCAACTGATAGACAATGT
WP_005282599.1	ATGACAGAGACGGAACA	TCCGCGGCACTGTT
WP_004926047.1	ATGACAAAAAATTTAAATCATTATCAT	TGTTGGAATATAATTTAATGGGT
WP_004917314.1	GTGTGTTGCATAACAAAGT	AACTAAAGATGAAAACCCTTG
WP_005295005.1	ATGAAAGCGAATAACGGT	TTTAGCGCCGAGTACC
WP_004926051.1	ATGAATATACTAAAAACCAATGATAAT	AAATGAATGAATGCTTATTTCTAAA
WP_004921623.1	ATGACTGAACAACTGAACA	ACGCTGTTCCATGCG
WP_004924795.1	ATGTCCCAAAATAATATTATCCC	GGCAATTTTTTCATCGGAC
WP_006820008.1	ATGACAATGATTCAAAGTTATTTAA	TTTAAAATTATTGTTATTTATCAGCTC
WP_005294393.1	ATGAATCAATCTCAAGACGA	GGCGTGAGTTAATTCTCG
WP_006820308.1	ATGTCTAACTCATTAAGCGT	GCTGTTTTGTTTTATATGGGA
WP_040902689.1	ATGAGTAAAGTGAACACCAT	GAGGAAGGATTTAAACGGC
WP_052309238.1	ATGGGTACTGGTTCTATAAATAT	ACAGCGACCATGTGT
WP_004920058.1	ATGGAAAATTCATCAAATAGTCT	ACAGCTAACTTTCAACTTGT
WP_004922896.1	ATGATTGACAGCCAGAATT	CTTATCCATTTCCTTTAAACGA
WP_006820270.1	ATGCCTGGATCTATTAATGG	TTCATGGTTAGTATCCCATTG
WP_004927350.1	ATGAGTATCAATGTCAATTCTAC	GTTTGTAGATAGCCCACC
WP_014656289.1	GTGACAACGATAAATAATAAATACA	GCGTGTTTTAACATCGG
WP_004922411.1	ATGAGCACTGATAGAGAGT	GCTATTATTCTTGTGATAGTCG
WP_004917276.1	ATGACAAATCCAATTGAAACA	TACTCTTGAATCGATAATACGG
WP_004921504.1	ATGAACAGTATTAACAGGACA	CTCCATATTAATTAGTTTAGTCATG
WP_005284092.1	ATGGCTTCAGTCACTGA	CTTCACTTGCTGACCCG
WP_004921520.1	ATGTCTTCTATCAATCAGAGTAT	TTGCATCATTATCATCTGCA
WP_005297258.1	ATGACACAAGATACCCAAC	TTCCTTGACCTGGCG
WP_004927466.1	ATGATGAAATTTAAATTAAGCATATTC	CTGATAATAAATATCATTGTTCTTCT
WP_004927136.1	ATGATGAAATTTAAATTAAGCATATTC	CTGACAAATTGCGCTTAAC
WP_006819026.1	ATGAGCGACGACAATTC	ATCTTCCAGCTTCGGTT
WP_036940590.1	ATGATGTTCGCGACTTC	TTCTTGGTTATGCGATTTGG
WP_006817659.1	ATGATGCAAACAGCCTC	TGCCCCCTTGTAAAAAAG
WP_004920813.1	ATGTCTGAGAACACAAGC	TTTTACTTGTGCCGCAC
WP_004918964.1	ATGGAACCGATACAGAATTT	ATACCCGTTTACTTGGAGA
WP_004927154.1	GTGAACAGACTCTCTAAAAAC	ATATTCACACACTACCTCTTG
WP_004917489.1	ATGATAATTTATGAGGGCAAAG	ATCATTATTGCGTTTTAGCTT
WP_004915712.1	ATGTCATCTAATCAGTCACAA	GTTTATGTTCTGTGGTGGT
WP_004918358.1	ATGAGACATGGTGATATTTCA	GCCTACAATTTTTATTTTATTGCC
WP_006819540.1	ATGAGTGATATTATCTTAAGCATC	GCGATGCCGCATCT
WP_004917712.1	ATGAATAATGATAAACGAGCTAG	CGTTTCTACCCCTGTTAC
WP_004915569.1	ATGGGCAATATTATTCATTCTG	TGCACACTCCCCACT
WP_005296212.1	ATGGAAAATCAGCTTTTAAGG	ATGTTTACAGGTGGGACA
WP_040903015.1	TTGGCACATTCATCGAC	TGCGTTGGCAGGAG
WP_004924913.1	ATGTACCATTTTTCTGCAAC	TTCATCATTCAGTATTGAGAGT
WP_004917132.1	ATGTCAAACAATATAAAAATGCC	AGGTTTAGCGACAAAATAAG
WP_004923523.1	ATGAACCAATCTGAAGAACA	ATTAACAAACGTATTTTTTACCC
WP_006820870.1	ATGGAATGCCGCCC	CGAAGTACTCTCTTCCAGA
WP_006820296.1	ATGAATAGCATCAACAGTATTG	TAACGACGGATCGCTG
WP_071599635.1	ATGTTGTTAGTGAGTTGCA	TTTCATCACCTGTCGTTG
WP_004920682.1	ATGCAACCCGGGATC	TGACACCTGTTGTTTGC
WP_004919757.1	ATGAATATTCCGCTCTACTAC	CTTGCACCAGACATAGTT
WP_004925244.1	ATGAGCATTCTAACTGATTTATC	CAAGTACGATACAAACAGTAC
WP_038258506.1	ATGTATACTTCTGGATATGCAA	ATGATACAAGTTTGAGTGAATT
WP_004926361.1	ATGCCGTTTTCTATTCCAA	ACCTCTGGTTGCAACA
WP_014657126.1	ATGGACGAATTTAAACCAGA	TTTCTGATCCTGTTGAACTT
WP_006818175.1	ATGAGTAACACAGATACCATC	GAACAGTTTGTAGACAATATTCA
WP_006818482.1	ATGGAAAACACTCAAACCA	GCTGAACAGAGAGTAGAAG
WP_004924278.1	ATGAAAGAATCAATAACAATAGTAAC	TTTAATTAGTTTTAGGTAAATACCTAAA
WP_006820335.1	ATGAGTGGAAATCTATCAGTAG	GTACATCTTCTGTGCAATCA
WP_004922817.1	ATGGCAATATTGAAGCAATG	GCGCCCTTTAATGGT
WP_005289077.1	ATGAGCGACGACCATT	GATCTCGACGTCCAGT
WP_004925512.1	ATGAATATTTTATTTATTTGTAGTCGT	CGGGATAAACTCGCTGA
WP_006820847.1	ATGGTCAGCAATAACGAAA	AAGGTGCAAAGACGCA
WP_004924049.1	ATGTCAATTCAGGATACGAG	TGCTTGCTGGCTTAATG
WP_004922318.1	ATGGATATAAATATGAATAGAAATAAGC	TAATTTCACAGGGGTACATG
WP_004922399.1	ATGAGCGACCAAACAAC	TTTTTGTTCTAAATTTGGTAGTAG
WP_071777524.1	ATGGGTAGCGTCCTGTAC	GCTATAGGCATGCAGC
WP_040903561.1	ATGACTGTGAATACCTCCA	ACTGATTGCATCCAACAA
WP_004916435.1	ATGACTAGACCAGCGC	CAGCAATCCCCACGC
WP_071777494.1	ATGATGAATTGCCCAAAGT	AAAATTCATATGCCCCTTGAC

gene ID	forward primer sequence	reverse primer sequence
WP_004925180.1	ATGACGAGCATTCCCTT	GTGCAAAATAGTTATCGCA
WP_014657408.1	ATGAGAAATTGTTATATATCAATAAGC	AAGATAGGCAATGCTCAAC
WP_004924072.1	ATGAACGTAGAATCAATTACTTC	AGGTTCGCTGATTAATATTGA
WP_004924665.1	TTGGTCGAAGATCCATCA	TACGGTAAGGGGCGC
WP_004924497.1	ATGAATGTATCTAGCTCATTGT	GCCCTGCTCTTTCTCA
WP_014658194.1	ATGAATATTGAAACTTTCATTTCG	TTGGAATATCACTGATATCGC
WP_004924065.1	ATGAGGTATATCATGGATTTTAATC	ATTAACATTAATTAAATCATAAGTTAACT
WP_004917956.1	ATGAGTCATCAATCCGGT	TGCTTTTTCCACCTCAAC
WP_050812366.1	ATGAAAAATCAAACTACATTTTCC	CCCCGGTTGCACGG
WP_004922117.1	ATGCAACAGATTCAGAATTTC	TTCCATAGTTGGCATTGTG
WP_004917889.1	ATGGTTTATTTATATATTTCAGTTGC	ATGATTGTTCTGGGCATT
WP_035595388.1	ATGACGGAACCCACTAC	TTCCCCCTGTGTCGC
WP_038254392.1	ATGGCAATCAGTGAAACC	GGCTTTTTGCCGGTT
WP_004922067.1	ATGCAAGTGAATCCCCA	CAGGTTTTCATCCCATGG
WP_004917602.1	ATGACAACGCACATTTCA	AAAATCCCAGTCATCATCTT
WP_004918732.1	ATGTCAATCACACATAAACATC	GTAGGATGTCAACAGTAATACA
WP_038254823.1	ATGACCACCATTCTGACA	GATATAAACGTCAATCTGGTTT
WP_005296556.1	ATGGAGCCTTCGCAC	TAGCGCCTCCTCGG
WP_004917987.1	ATGAAAGTATCTAGTCAAAGCA	TTCATACTTTGATCTGCTAGC
WP_004922581.1	ATGTTCCCCATATTTAGGGA	TTTTAATTCTTGCTCTTGACTAT
WP_004921265.1	ATGAAATACGCATCCGTG	TACTTCCGTTCTCCCAAT
WP_006818941.1	ATGAGCATTGAGCTAATCG	ATTTGTCATTACCGGGGC
WP_006818167.1	ATGAGCAGCACCGAC	TTCCCGAACCAGAATCA
WP_040902722.1	ATGACTACACGCTATTTTTCT	AGGTACGATAACCTGCA
WP_042116632.1	ATGAATCACATTAATCATAGTGAC	TTTTGTTGATTTTGATTTTCCA
WP_042116129.1	ATGAGTATAACAATAACAGTATGTG	TAATTTCTCGGGATTTGTTCT
WP_004927264.1	ATGAAAAGACACCCTATTCG	TTTAGGGAAGCTTATACTGAC
WP_004917597.1	ATGAGCAATAACACGAATGA	ATGTGATTTTCGGTTAAATGG
WP_005294415.1	ATGCCTGATGCTACCG	TTTGCTTCCACATTTGCC
WP_004921716.1	ATGACGATGCCTCTGAG	CATATACTCTTCAATCGGCA
WP_071599648.1	ATGTCATCCAGTCTTAATGC	ATAATTGACGTTAACAGTGAC
WP_052038327.1	ATGAGCCAAAAATCATCAGT	TTTGTGTAACTGTTTAAACTCT
WP_006817733.1	ATGACAACATCCGTATTTTATATC	CATGGCGGCGCGATA
WP_035595442.1	TTGTCAATGCAATCTCAAGA	TTTTTGCTCATGGCTATCC
WP_004919494.1	ATGGACACCTTTAACCTCA	CAGCACTAACTCACTTCTT
WP_004918153.1	ATGAGTGATTTTTCCCAGAC	ATATTTGTCGATGACATGTGC
WP_006820125.1	ATGTCCTCTGAACTCTCTG	GTTGCTGACGGAAAGC
WP_005280619.1	ATGTTTAGCAGTCAAGAAGG	TTGCAGGTAATGCTCCA
WP_004919715.1	ATGCCATTTTGTATTAATCAACA	TTCAAACAAGCCTTTATACATT
WP_004917178.1	ATGATGGAGATATTAACGATGG	TACTAAGCCATGCTCAGC
WP_004915373.1	ATGTATAACGCAGGAGCA	GATGTCTTCCTCGGCT
WP_004918480.1	ATGTATAACGCAGGAGCA	TTCTCTTTTTATATTAATTTCACTTTTC
WP_040903573.1	ATGAAGGAAAATGCCACG	TGAGAATGGGGATGTCTC
WP_006817197.1	ATGCAGGAAAACATCACG	GTAGACAAAACTTACCGGC
WP_004922866.1	GTGAAAAACAACGAAAAAACC	TTTTTTACGATTGCTAAGTTCA
WP_005287183.1	ATGAGTAACCCTTCATCCC	CTGCGCCTCCATCA
WP_006818356.1	ATGAGCGAATTCCGGG	GTATCCGTTAACCTGTAAAAA
WP_006818222.1	ATGTCTTCCCAGCCCT	ATGCAGTACGGTAACCG
WP_014656941.1	ATGGATCAGCAATTATCAGC	ATTAATAAATTGGTAACCAGGTAT
WP_004921358.1	ATGTCCGATATGACAACCA	GACTGCTTTCACCTTAATAGG
WP_004925688.1	ATGAAGATTAAAAAATCCATTTTATAGC	AGGTATTTTTATTAATGCCGG
WP_004919332.1	ATGATTAAAATCTCTCGAGGC	TAGCAAGTTCCACGCT
WP_004919841.1	ATGGATAATTCAACTTATACCCC	TCGAAGATCACTTGCATC
WP_004925746.1	ATGCACTCACACAACCA	TTTTTTAATTTGTGGTAGTTTTACT
WP_006817680.1	ATGACTGATTGTCGTCGT	GAATGCTGCCTTTTTGATG
WP_005288481.1	ATGACGATTAGTATGAACACG	TGCTGTCTCCTCTGCA
WP_004916524.1	ATGGTGAGTAAACAGGAGG	TCCCCTCGCCGTCA
WP_004919165.1	ATGACATACAACGTAACAGC	GCCTAACAATTTTTTAATGCG
WP_005284548.1	ATGCCGATGGATAAGCC	CATGGCGTGTTTAATGC
WP_006820021.1	ATGACCTCATCTTTACCCG	TAAACTTTCATTTTTTGCTTCC
WP_004926354.1	ATGATGATACAGCCTATTAATCG	TTGGGCTTTTAGTAGAGTATC
WP_087943245.1	ATGCTTGATTCACTTATCACC	GCTAAAAAGCCCAAAGAGT
WP_042117401.1	ATGCAATTAACGATTTCTTATCG	TTCGCTTTTACCATTTGGA
WP_006819050.1	ATGACTGATTTAAAAATTGCCG	TAATTTTTCCCCTGCCAG
WP_004920683.1	ATGAAGCGTTTCTTTGCC	TTCGTTTAGAGCTACCGC
WP_004926210.1	ATGATATCCCCACTAACGC	ATAAAATTTGAACTCATCATCTTG
WP_005286533.1	GTGGTAGTAGGAAATTTATCCG	CAGCATATTCCAGCCG
WP_042117315.1	ATGCCAACAATCTCTCCC	TGACCCATTTCCTAATTTAAATC
WP_006818160.1	ATGCCTGAAACCTATATCCC	GGCCAGCCATGGATG
WP_006818522.1	ATGCGCGATAGCAGC	GCCCATCAGCCAGTAA
WP_005283164.1	ATGATGAAGGAGATCATTATAAGC	CTCTATCGGGATATAGTGCT
WP_006819702.1	ATGGTACACAGCATGGC	AAGTGCATCAACCACAG
WP_006818297.1	ATGACAGACAATACCTGGC	CCCCTGACGCTTGTC
WP_006821162.1	ATGAATACTGAAGCCAACCA	CTGTGCCGGTGCGG
WP_005293940.1	TTGCCCTATTTTTTGACGC	TCGGCTGGCTACAGC
WP_081874660.1	ATGAGTCATTCTGCCACG	CAGCAAGCTCTTGACG
WP_005290551.1	ATGCAACGCGCACC	TCCGGCGGAAACGC
WP_006821080.1	ATGCGTTACATCATCAATGC	TTGCGGTTTCCACGA
WP_071586119.1	TTGTTGGCTGCGTGC	CCCTTTCGCCCATAAAG
WP_040903175.1	ATGAGGATTTCGCTTGAGC	CTCGGTGGCCAGCG
WP_050812352.1	ATGGAGCAGCGCCG	CTGCCTCAGCAAGCG
WP_004915478.1	GTGAAAAACTCAGAAACGCAAC	GGACGGGCGACGGC
WP_040903289.1	ATGCCGCTTTCTTCGCA	TTTTGCCGCGCGCG
WP_005282518.1	ATGGCGTGGGCGCA	TGCCCATCCCTGTAGC
WP_040902687.1	ATGCACTGGCAAACCCATAC	CCGCCCGCCTCCCG
WP_087943279.1	ATGCACATGAGTCAGGAAACG	ACGCGGCAGGGCGG
WP_071777502.1	GTGGCACAACGAGATTATGTACG	ACCCCCGGCGCCGA

gene ID	forward primer sequence	reverse primer sequence
WP_005288047.1	ATGTACAATACCGATTTTATCAGCGC	GGCGGCGCGCTCGA
WP_001539170.1	ATGGTTACAGTTATCAGCAATTATTGT	TGAGTGTTTTTTTAAAGACAGCAGTA
WP_001490312.1	ATGTGGAGTTCGGCAATTAACA	GAACGTCCAGACTACACCCG
WP_001335433.1	TTGCCTTTTTCTGTTTCTATTGAATCA	TGCAGTTTGCCATCCATGA
WP_001335497.1	GTGACTATCTCCTCAATGATTCATGC	GTTTACGCTTACCCATTCATAAATACG
WP_011579078.1	ATGAATACCAACAAAACTCTGGGC	ACCCCGTTCCGTCATGG
WP_001332471.1	ATGAGCTCCAACCGC	AATTAATATGAAAAATAATGACAGGA
WP_000120394.1	ATGTCTTACATCAAACCGGATACT	GGAATCAGCCAGAACCATAGTG
WP_001335175.1	ATGAATTATCCCATGGACTCATTAACA	AGGGAAAATAAGCGTAATGTTCATC
WP_000557383.1	ATGCACACTAACTGGCAAGT	TGGTGTTTCCTCACCTTGC
WP_001304629.1	ATGACCATGCCGTCAGGA	CAGTATCAGCACCTGAATCTCG
WP_001332782.1	ATGTCACACTCGCCTCATACA	TGTGTGACTGCACTGTTTTGT
WP_000020887.1	ATGAGTATTGATCGCACTTCGC	ATTACTCTGCAAGTCTTGCTGC
WP_000334996.1	ATGTGTATTAGTAGCCCGG	TGATTTTATCAGTTTTGTGACT
WP_000220141.1	ATGACATGGAATCCGTTGGC	GTGTGAATATTCTTCTGTAATTTCAGCC
WP_000213694.1	ATGACAACCTCTTCGCATAATTCC	GTCAATGATGCGCATAATGGC
WP_000288707.1	ATGTACACTTCAGGCTATG	ATGATACAAATTAGAGTGAATTTTT
WP_032294934.1	ATGAGTAAAGTGAAAAGTATCACCCG	GTTTTTTTGATCGGCGGGC
WP_000611436.1	ATGATTAACGTACAAAACGTCAGTAA	TGAAGAGACTCCCATTGGGT
WP_001115606.1	ATGCCATTTTGCAGTTCA	ATCTATTTCAAAATAACGTTCAAT
WP_000582830.1	ATGATCATCGAAAAAGTCATGAACA	GTCCATTAATACTTTGGTACGAATAGAG
WP_001298103.1	ATGATGAATAATAAAGTCAGCTTCACTAA	CAGAGTTTCCTCAAAGAACGGA
WP_000057389.1	ATGGCGCAGATAACAACGA	ACATGATTTCGCCTCCAGATAC
WP_000106767.1	ATGTCAACTCCACTTCAAGGAAT	GATGGCGTGGTTTTGCTTC
WP_032294942.1	ATGTGCCATTTATACAATGGTTTTGA	CTGCCCGGTCAGGCA
WP_000083477.1	ATGAGCCAGTCACTGTTTAGC	GACGCCGTAACGTTCGC
WP_001335297.1	ATGAATACAATCGCCTCCGTTA	AATTTTGAGGGTGTTTAATAGCGC
WP_001298277.1	ATGTCATCAATATCACATGGCGC	CATTAGACCTAAAATTTTCCACCAGG
WP_001139105.1	ATGCCTTCTCATCCGTATGTC	ATGCAGCACCGTCACCG
WP_000111853.1	ATGTCAGTTAAAATAGTCATTAAACCGA	GGCTTCCTCCCCTTTGCA
WP_001143232.1	ATGCCCACTTCTCATGAAAATGC	CTGACTAAAGCGCATCTGCT
WP_113698432.1	TTGACTGCTATTTTTGCTATTAATAGCC	CTTCGCTCTCAGCTTATTGTAGT
WP_000004564.1	ATGTCTGAATCCCGCAGCA	ATTATGACAGTCGCTCTCATTTGG
WP_001298077.1	GTGCACTGGCAAACTCACA	CAAACACACTCCTCCCCGTC
WP_000191565.1	ATGACTCAATTTACGCAAAATACCG	TTGCGGGTAAGGCACCC
WP_001088080.1	ATGAATTCACAATTACTGGCTAACG	GCTATTCAGTAGCTCACGCG
WP_000077881.1	ATGTCGCAACATAACGAAAAGAA	CGCCGGGATTTTGTCAATCT
WP_000013970.1	TTGGCCAACTCATCCTCACG	TGAGCTGGCCGGTGG
WP_001142370.1	ATGCCAACAATAACCACTGCA	TGCTGATGCTGTCAAAGTTATTG
WP_001197833.1	ATGCAGGTGTTACCCCCG	CGCCTCCTCTTCATGACG
WP_000369522.1	ATGGATAACTTGCGCTTCTCTT	TAACGTGTTCTCCGGTTGC
WP_001445787.1	ATGCGAAGTGAACAGATTTCTGG	TCGACTCATGTCTTTTCGCTG
WP_071587545.1	ATGAAAACCATTTCACATAATTCCACC	CTGTTCATCAGGGACTATGGTG
WP_077473175.1	ATGGGAATTATTTCGATGATATAT	GTTTCTCACTGTCAGTATGC
WP_032145775.1	TTGTTACAGCCCAGTCCGC	GCAGACGCCACGTCC
WP_071528128.1	ATGGTCTGTTCATCAGCGG	ATTACTTAACGTATCGTGCCGT
WP_001024524.1	ATGAACATAAACGGAAAAAACTCTTCT	TGGCAGATGTTTATTCCCATCTAA
WP_032140136.1	TTGTGCACAGCTCCCCA	TGTGATATTTACAGTTCCATCGGG
WP_032156687.1	ATGAGGATATGCGATAAATTTCAGTC	GTGCTGTTCTGACGGGG
WP_001327852.1	ATGCAGGCTGGAATTAACGC	CTCTATTTCGCCGTTATTTATCCC
WP_001328716.1	TTGCGTTTTTGTGAAAATGATCAAA	TCTTATCATCAGTTTCTGACCGTG
WP_023148198.1	TTGCACAGCTCCTCCAAC	GCTGCCTGAAAATTTTTTGAGC
WP_071528137.1	ATGACGGATCCGTCAGTAAGT	TTTTAAAAGTTATGATGTTCACCGATAG
WP_001122065.1	ATGCCGATTCTGACGCATG	TTTCTCACTCATCCAATTAGGAGG
WP_001135351.1	ATGCCTCCAGTATCTTGCGG	GTTAGCCCAAGTCCATAACGA
WP_000020875.1	ATGAGTATTGATCGCACTTCGC	GTTACTCTGCAAGTCTTGCTGC
WP_000206655.1	ATGACAACCTTTTCGCATAGCT	GGCTCATTTCAGCGCGG
WP_000102382.1	ATGAGCACAGATACACTTGAAATATTC	GCCATTATCTTCTGAATTATCGGTT
WP_000804518.1	ATGAAGCCACGAAATATTAATAATAGCC	ATGTGCATGACCTGGATTCAT
WP_000868324.1	ATGAAAGTAAGAAACCCAGAACAGA	GTCATACCAACGGCTATTGTTCG
WP_000189224.1	GTGACACCGACCCTTTTAAGT	ATAGAAAATGCGTACCGCGC
WP_000873388.1	ATGAAATACTTTAGCGATTCTAA	ATCTGTCGCTTCTCGC
WP_000611426.1	ATGATTAACGTACAAAACGTCAGTAA	TGAAGAGGCTCCCATTGGG
WP_001267298.1	ATGAGATGGATTTCAAAAATAAA	TTGCATCGTTCCCTTTTG
WP_001445815.1	ATGACATTACCAACCACTATTTATTCAT	GTTTCGCAAGTAGATCCATTCAC
WP_001445845.1	ATGATCATCGAAAAAGTCATGAACA	GTCCATTAATACTTTGGTACGAATTGA
WP_000062538.1	ATGTCTATGCCATTAAGCA	TTTACCAATATAATTCCATTCACC
WP_001059674.1	ATGAACAACATATTAGGCTCATCAATTA	TAGATAATTTTCGATTCTTTCGATAGTGT
WP_000004905.1	ATGAGCGAAACTTTTTTCCATCTG	CGTCGGTGCCCCTTCC
WP_001067519.1	ATGAACCCCGCAGTGGA	GCGGAATTTACGTCGATACTCG
WP_000981716.1	ATGATGAATAATAAAGTCAGCTTCACTAA	CAGAGTTTCCTCAAAGAACGGC
WP_000014226.1	ATGTCAGGTATAATTTCAAGTG	TTTTACTAGTTCAATACCAAATAGA
WP_000241053.1	ATGGTGAAATCGACGTCATGTA	ACCGCGAAACATCGTGG
WP_000057374.1	ATGGCGCAGATAACGACG	ACATGATTTCGCCTCCAGATAC
WP_001328837.1	GTGTCAGATCAGATTATCGCCC	GGAATGTAGCGCTGGATGC
WP_072108184.1	ATGGTATATAAAACAAAGGCGACCG	GGCATGGCACCCCCC
WP_000258580.1	ATGGTTACGCCAGTAAGCATC	ATAACGTTCAGGGAGGTCTGG
WP_001445771.1	ATGGAACAACGCCACATCA	ATATGTCTCACCCAACGCG
WP_000183751.1	ATGACCAATAATATTAGCAATATAAAC	TGAGCGAATGCCTTGA
WP_000786551.1	ATGAAAAATTGTGTCATCGTCAGT	ATTCAACCGTTCAATCACCATCG
WP_000139103.1	ATGACGGAACAAGAAAAAACCTC	TTCTCCTTGCGGCGCG
WP_000020636.1	ATGTCAATTGACTTTACCCC	TGGGATTATATTCTGGTAATGA
WP_000671701.1	ATGAAATTCCCTTCAATATTTAATAA	AAAAGGCCATGAACTGG
WP_000355772.1	ATGGATATTGTCGAACGCTTTATC	TGACTCACGATTCCCTGTGG
WP_077626097.1	GTGATTATGACCAATATTAATACAGCTTG	TAACTCTGTAGAAAAGGATTGGCT
WP_000338247.1	ATGTGCTACAATGGTTTAAAT	ATCTGCATCTCGCATAATAT
WP_075208399.1	GTGAGGTATATTATGGCAAGCCC	TAAATCAGCAGTTGTTGCGGC

124

gene ID	forward primer sequence	reverse primer sequence
WP_000075087.1	ATGAGCCCGTCACTGTTTAG	GACGCCGTAACGTTCGC
WP_001093944.1	ATGAATACCTTATTTAATCAGCCTTTGA	ACCATGATGGTTCTCCTTTGC
WP_001327845.1	ATGGCCAGCATTAAAGACATCC	CATAAAACGCTTACGTTTCATTGTT
WP_000961342.1	ATGTTAGTTAGTAAAAGCAACG	TACAAGGGGACTATTCACA
WP_000155927.1	ATGACAATTTCAAATTCGAATTT	TTTTTTAGCATTCGCATCC
WP_077626056.1	GTGTCATATTGCGAGGTTACTATG	TTTGTAGGTTGCAATTGCAATATTG
WP_000155738.1	ATGACGATAAGCTTTAACACCATTC	TGCAGACTCCTCTGAATACTGA
WP_001143217.1	ATGCCCACTTCTCATGAAAATGC	CTGACTAAAGCGCATCTGCT
WP_000900534.1	ATGTTAGGTCATATCTCAAAGT	AATACCTTTATTCTTTATACGATTAGC
WP_001328185.1	ATGAAAATTGGAACTGTGGCAG	GAGCTGTAACATTGTGCGC
WP_001445816.1	GTGCACTGGCAAACTCACA	CACACACACTCCCCCCG
WP_000191595.1	ATGACTCAATTTACGCAAAATACCG	TTGCGGGTAAGGCACCC
WP_000097400.1	ATGTCTTCCATGACAACAACTGA	CTCCACTTCCTGCCAGTTTT
WP_000456096.1	GTGGAATATCTTTTCCCTTGGGT	ATGAGTGGGAAGAGTCAACTTATATT
WP_000226319.1	ATGGTTGCCAATATTAATCTCATTAAACA	GACTCCCATCCCTACGGTAT
WP_000407090.1	ATGGAAATAGTATCATTTAAAGTTTC	TGCAGCTTGATGATCTG
WP_000859964.1	ATGAAAACCGTTAGGGAGTCC	TGCGCGCTTCAGATAGC
WP_001088085.1	ATGAATTCACAATTACTGGCTAATGC	GCTATTCAGTAGCTCACGCG
WP_000632861.1	ATGATTACTCGTATTCCTCGTATTCC	GCGTACGGTATGGGGAGA
WP_000077829.1	ATGTCGCAACATAACGAAAAGAA	CGCCGGAATTTTGTCAATCTTA
WP_001445762.1	TTGCAGGGTAATAATACGATTGTCA	GAACTTATATTTCAGGCCCCACCA
WP_000013994.1	TTGGCCAACTCATCCTCACG	TGAGCTGGCCGGTGG
WP_001328160.1	ATGAAAATACCCACTACTACGGATATT	TGAGCCAGAATGTGTGCAA
WP_001327854.1	ATGGGGCATCAAACAACACAATG	AAAGCTGAAGCGATATCCTCC
WP_001142374.1	ATGCCAACAATAACCACTGCA	TGCTGATGCTGTCAAAGTTATTG
WP_077626078.1	ATGGAAGGACACCCGATGG	TTTCACCTGCATCCCGGT
WP_000168878.1	ATGGCTTCAACAAATGCGC	AAGTGATTTACGGCAGGCG
WP_077266176.1	GTGCCATCAGGAGGAAGAATG	TATCCGCCGCTGACGA
WP_000369489.1	ATGGATAACTTGCGCTTCTCTT	TAACGTGTTCTCCGGTTGC
WP_004150458.1	GTGATTTCCGCCGGATGT	GTCTCCTGAATCCAGCAATATGA
WP_004140501.1	TTGACCATTGCCCCGTCG	TACTTTACCGCTACCTTGCTG
WP_009484109.1	ATGATAAACCATGTTACCGGGAA	ACGCTTTTTAGCGAAAAGATGAC
WP_020316952.1	GTGATGTCGCGCCATGG	TAAATTTTTATTTATCGCCTTTTTGTCGA
WP_004225268.1	ATGTGCCATACGCCG	AATTATTACCAATTTAACCACATATG
WP_032409076.1	ATGCTGACGCCCGCC	TTTTCCGCTACAGAACGCC
WP_009486504.1	ATGCATATCATTTTCACGCTAATCA	AAAAAGATGCATTTCGCAGTGG
WP_049245346.1	ATGTATTACGCTTCAGAGCTGAC	AAAGGCAGACAATCTGCAGG
WP_071526683.1	ATGCGGCCTACGGGC	ATGGATGGATAATTTGTGTCCA
WP_009484876.1	ATGATCACTATCGAATGCACCG	CTCTGTAATGTCACTGTTTTTTACCT
WP_009486019.1	GTGCAGGTCATCAGCGG	GTTCCAGTGTCCGCGG
WP_004149975.1	ATGCCACGAAATTTTTTTAACATA	GGCGTGGAGCGCTG
WP_004152718.1	ATGAAAGTATACGGCCATCAACA	TTTCACACCCAGACGGCG
WP_023328080.1	ATGAAGAACGGTAGCCGC	AAAAGCAATGAAGTGCTTTTTACTCA
WP_009483878.1	ATGCAGCTGACACACGC	CCACCGGAAAACGCCG
WP_002891634.1	ATGAGTCATTCTACCGATCATAGC	GTGCATCATCATGTTGTAGTTCAG
WP_009484937.1	ATGCCTGAGCAAAAAATGGAAA	AACTTGGACTGTCTCTGGATTG
WP_003026803.1	GTGGGACGTATAACCACGC	TGGAAGCTTTAGAAACAGAGTCC
WP_004179215.1	ATGTTTACTTCAGCTCACGCA	ATGATACAAATTTGAGTGAATTTTTAGCC
WP_009484324.1	ATGCTCAATACTCGTGTGCA	GGCTCGCCGCCCAC
WP_002898814.1	ATGAACGACGCAATAAACCAC	TGCCAGCAGGCAGGC
WP_004177339.1	GTGACACCTACCCTGATTAGC	TTCCACCACGATCCGCA
WP_002916607.1	GTGATTAATTTAAATGGAAATTCATC	GTGGCGCAGATGTTC
WP_004145550.1	ATGGCTAACTGGCTACATCAAC	GACGCTTAGCGCTTGTTTTT
WP_004210116.1	TTGAATATTTTAATTCACAGCAAT	AAACAATTTTCGAAACTCTTTC
WP_004171426.1	ATGACAGCCTTCAAGAGTCAAT	CAGGGACGGCAACCGT
WP_002889316.1	ATGTTGTCTGCTAATCAAACTGTCA	GGCATGACTGCCACCCG
WP_032420351.1	ATGCCAGGTAGTATAGATAAAGCG	AGGCACATAATCGCGTGG
WP_002916742.1	ATGTCAGAGCAAAACTACCAGC	CGCCTGGCGCTTATCTTC
WP_004197606.1	ATGAGCGACGACAATTCACA	CTCTTCCAGCTTCGGCCTG
WP_004179102.1	ATGAGCCAACCGCTACCC	AATATGGAACCCTTCCTCTTTCAG
WP_004210091.1	ATGGACAGCCTCACCTCG	CCTCAAATTGTTAGATACAAGGGGT
WP_000155904.1	ATGACGATTTCCCTTTATACCTCC	GAGTTTATCCATCAGATAAGAATGCG
WP_004118237.1	ATGAATCCTTCGTTAACCGATTCC	CAGCGGTGAAATGTTTATCCAG
WP_002913732.1	ATGAATCAGTTAGACAGCATCAAGC	AAGTGACAGTTTGGCGGC
WP_004152062.1	ATGAAGCGTGCCCCTGT	GACTCCTGGTTTCTCAAGCTC
WP_002889847.1	ATGAACGCCCTGACCGC	TTGATGAGTCTGGTAATGATTCGT
WP_002909008.1	ATGAATAATAAATTTACCTATACGATT	GCAAAGCGCGAAATTAG
WP_009485462.1	ATGAATACTGAAGCCACTCAAGA	CTGCGCCGGTGATGATT
WP_019705807.1	ATGAAAACATGGCTTCCGACA	CTTTGCCGCCAGCTCG
WP_004147894.1	ATGGTAAGGTCCCGTACCG	ATAATAGGCTTTCAGCGTCCG
WP_004174727.1	ATGGCTGATCAGACCAATCCG	TCGACAGAAAGGGCGAGT
WP_002912648.1	ATGAAAATTGAATCTGTGAATGTCACC	GTCGTGGCTAACGGTATTGG
WP_009484190.1	ATGTCGACTCCAGACGCG	TCAGCAGCAGCGTCAGC
WP_032457106.1	ATGGCAAGTTCGGGCAC	CGCGCCGATGGCTG
WP_004145486.1	ATGAGTACGACTGAGAGCATTG	GAACAGTTTGTAAGCAATGTTGAGG
WP_009486529.1	ATGACGATCACAGGAAACTTTATTG	ACCTAACGTCGGCATGCT
WP_004175074.1	ATGGTCAGCAATAATAGCGAACG	AAGGTGCAATGACGCAAGA
WP_002916277.1	ATGCATTCCTCTGTTAATAAAAACGA	GGCGTTGCTGAGGGATTT
WP_004222074.1	ATGCACTGGCAAACGCA	CGCCTGAGGCGCCT
WP_009484993.1	ATGAAAAACATCAACCCAACGC	TGCGCGCCAGGCTT
WP_020317218.1	ATGACCAATCCATTATTGACGCC	GCCTTTGATCCCGTAATGCT
WP_009485840.1	ATGAGCACGTCTAACGACCC	CAGATCGAAGCGGTCGAG
WP_002901554.1	ATGTCTGATAAGCACCCGAAC	CGCCGGAATAGCGGC
WP_009484142.1	ATGCAGAACGGCGCAAT	ATCGACGTTCAGCGCG
WP_009484412.1	ATGGATAACTTGCGCTTCTCTT	TAACGTGTTCTCCGGTTGC

125

Table S3 | Cloned effectors of the strains and the metagenomes. *Effector IDs and abbreviations (abbr.) used in this study are stated. Quality control indicates whether effectors were identified by end-reads or full-length sequencing. "complete alignment" refers to the complete alignment with the predicted sequence. Deviations from that are stated.*

effector ID	abbr.	quality control	full-length sequencing results
WP_042030958.1	Aja_1	full-length	complete alignment
WP_042033505.1	Aja_10	end-reads	
WP_042030965.1	Aja_11	end-reads	
WP_042031153.1	Aja_12	end-reads	
WP_082035530.1	Aja_13	end-reads	
WP_042032390.1	Aja_14	end-reads	
WP_042031545.1	Aja_15	end-reads	
WP_042032213.1	Aja_16	end-reads	
WP_042033246.1	Aja_17	end-reads	
WP_042032056.1	Aja_2	full-length	complete alignment
WP_042032153.1	Aja_3	full-length	complete alignment
WP_005354370.1	Aja_4	end-reads	
WP_042030140.1	Aja_5	end-reads	
WP_042031495.1	Aja_6	end-reads	
WP_042031532.1	Aja_7	end-reads	
WP_042032269.1	Aja_8	end-reads	
WP_042033236.1	Aja_9	end-reads	
WP_004864811.1	Cda_1	full-length	complete alignment
WP_039898535.1	Cda_10	full-length	complete alignment
WP_055696404.1	Cda_11	full-length	complete alignment
WP_083478381.1	Cda_12	full-length	complete alignment
WP_016517519.1	Cda_13	end-reads	
WP_016517593.1	Cda_14	end-reads	
WP_016535238.1	Cda_15	end-reads	
WP_016535500.1	Cda_16	end-reads	
WP_016536247.1	Cda_17	end-reads	
WP_016536444.1	Cda_18	end-reads	
WP_016536850.1	Cda_19	end-reads	
WP_016517497.1	Cda_2	full-length	complete alignment
WP_016536932.1	Cda_20	end-reads	
WP_016537145.1	Cda_21	end-reads	
WP_016537795.1	Cda_22	end-reads	
WP_016537847.1	Cda_23	end-reads	
WP_016538749.1	Cda_24	end-reads	
WP_055696403.1	Cda_25	end-reads	
WP_016517628.1	Cda_26	end-reads	
WP_016535304.1	Cda_27	end-reads	
WP_016536503.1	Cda_28	end-reads	
WP_039898184.1	Cda_29	end-reads	
WP_016535503.1	Cda_3	full-length	complete alignment
WP_039898519.1	Cda_30	end-reads	
WP_039898721.1	Cda_31	end-reads	
WP_039898948.1	Cda_32	end-reads	
WP_016538147.1	Cda_33	end-reads	
WP_039898229.1	Cda_34	end-reads	
WP_039898704.1	Cda_35	end-reads	
WP_016537275.1	Cda_36	end-reads	
WP_016535835.1	Cda_4	full-length	complete alignment
WP_016536389.1	Cda_5	full-length	complete alignment
WP_016536523.1	Cda_6	full-length	complete alignment
WP_016537909.1	Cda_7	full-length	complete alignment
WP_016538154.1	Cda_8	full-length	complete alignment
WP_039898226.1	Cda_9	full-length	complete alignment
WP_005129057.1	Cyo_1	full-length	complete alignment
WP_072041464.1	Cyo_10	full-length	complete alignment
WP_072041472.1	Cyo_11	full-length	complete alignment
WP_080721914.1	Cyo_12	full-length	complete alignment
WP_082031767.1	Cyo_13	full-length	complete alignment
WP_072041503.1	Cyo_14	end-reads	
WP_072041520.1	Cyo_15	end-reads	
WP_001149870.1	Cyo_16	end-reads	
WP_003844491.1	Cyo_17	end-reads	
WP_005120762.1	Cyo_18	end-reads	
WP_005122932.1	Cyo_19	end-reads	
WP_005129187.1	Cyo_2	full-length	complete alignment
WP_005126657.1	Cyo_20	end-reads	
WP_005126712.1	Cyo_21	end-reads	
WP_005132825.1	Cyo_22	end-reads	
WP_023184674.1	Cyo_23	end-reads	
WP_040229922.1	Cyo_24	end-reads	

effector ID	abbr.	quality control	full-length sequencing results
WP_040231861.1	Cyo_26	end-reads	
WP_040232101.1	Cyo_27	end-reads	
WP_040232739.1	Cyo_28	end-reads	
WP_040233356.1	Cyo_29	end-reads	
WP_005129207.1	Cyo_3	full-length	complete alignment
WP_040233602.1	Cyo_30	end-reads	
WP_040233766.1	Cyo_31	end-reads	
WP_052463753.1	Cyo_33	end-reads	
WP_071887407.1	Cyo_34	end-reads	
WP_072041455.1	Cyo_35	end-reads	
WP_072041467.1	Cyo_36	end-reads	
WP_072041502.1	Cyo_37	end-reads	
WP_005121691.1	Cyo_39	end-reads	
WP_005131699.1	Cyo_4	full-length	complete alignment
WP_005126378.1	Cyo_40	end-reads	
WP_040230231.1	Cyo_41	end-reads	
WP_040230744.1	Cyo_42	end-reads	
WP_040231102.1	Cyo_43	end-reads	
WP_040232312.1	Cyo_44	end-reads	
WP_040232375.1	Cyo_45	end-reads	
WP_040232376.1	Cyo_46	end-reads	
WP_040232565.1	Cyo_47	end-reads	
WP_040232896.1	Cyo_48	end-reads	
WP_040233393.1	Cyo_49	end-reads	
WP_040229899.1	Cyo_5	full-length	complete alignment
WP_052463763.1	Cyo_50	end-reads	
WP_005131490.1	Cyo_51	end-reads	
WP_040230127.1	Cyo_6	full-length	complete alignment
WP_040232070.1	Cyo_7	full-length	complete alignment
WP_040232968.1	Cyo_8	full-length	complete alignment
WP_040233404.1	Cyo_9	full-length	complete alignment
WP_000004564.1	Ec2_1	full-length	complete alignment
WP_000120394.1	Ec2_10	end-reads	
WP_001335175.1	Ec2_11	end-reads	
WP_001304629.1	Ec2_12	end-reads	
WP_001332782.1	Ec2_13	end-reads	
WP_000020887.1	Ec2_14	end-reads	
WP_000334996.1	Ec2_15	end-reads	
WP_032294934.1	Ec2_16	end-reads	
WP_001115606.1	Ec2_17	end-reads	
WP_000582830.1	Ec2_18	end-reads	
WP_001298103.1	Ec2_19	end-reads	
WP_000083477.1	Ec2_2	full-length	complete alignment
WP_000057389.1	Ec2_20	end-reads	
WP_000106767.1	Ec2_21	end-reads	
WP_032294942.1	Ec2_22	end-reads	
WP_001335297.1	Ec2_23	end-reads	
WP_000111853.1	Ec2_24	end-reads	
WP_001298077.1	Ec2_25	end-reads	
WP_000191565.1	Ec2_26	end-reads	
WP_001335433.1	Ec2_27	end-reads	
WP_000557383.1	Ec2_28	end-reads	
WP_000213694.1	Ec2_29	end-reads	
WP_000220141.1	Ec2_3	full-length	complete alignment
WP_001139105.1	Ec2_30	end-reads	
WP_000611436.1	Ec2_4	full-length	complete alignment
WP_001298277.1	Ec2_5	full-length	complete alignment
WP_001539170.1	Ec2_6	end-reads	
WP_001335497.1	Ec2_7	end-reads	
WP_011579078.1	Ec2_8	end-reads	
WP_001332471.1	Ec2_9	end-reads	
WP_000075087.1	Ec6_1	full-length	complete alignment
WP_001059674.1	Ec6_10	full-length	complete alignment
WP_001093944.1	Ec6_11	full-length	complete alignment
WP_001267298.1	Ec6_12	full-length	complete alignment
WP_001327852.1	Ec6_13	full-length	complete alignment
WP_001328837.1	Ec6_14	full-length	complete alignment
WP_001445762.1	Ec6_15	full-length	complete alignment
WP_001445771.1	Ec6_16	full-length	complete alignment
WP_001445815.1	Ec6_17	full-length	complete alignment
WP_001445845.1	Ec6_18	full-length	complete alignment
WP_032140136.1	Ec6_19	full-length	complete alignment
WP_000097400.1	Ec6_2	full-length	complete alignment
WP_077626056.1	Ec6_20	full-length	complete alignment
WP_071587545.1	Ec6_21	end-reads	
WP_077473175.1	Ec6_22	end-reads	

effector ID	abbr.	quality control	full-length sequencing results
WP_032145775.1	Ec6_23	end-reads	
WP_071528128.1	Ec6_24	end-reads	
WP_001024524.1	Ec6_25	end-reads	
WP_032156687.1	Ec6_26	end-reads	
WP_001328716.1	Ec6_27	end-reads	
WP_023148198.1	Ec6_28	end-reads	
WP_071528137.1	Ec6_29	end-reads	
WP_000155738.1	Ec6_3	full-length	complete alignment
WP_001135351.1	Ec6_30	end-reads	
WP_000020875.1	Ec6_31	end-reads	
WP_000868324.1	Ec6_32	end-reads	
WP_000189224.1	Ec6_33	end-reads	
WP_000873388.1	Ec6_34	end-reads	
WP_000062538.1	Ec6_35	end-reads	
WP_001067519.1	Ec6_36	end-reads	
WP_000981716.1	Ec6_37	end-reads	
WP_000102382.1	Ec6_38	end-reads	
WP_000057374.1	Ec6_39	end-reads	
WP_000191595.1	Ec6_4	full-length	complete alignment
WP_072108184.1	Ec6_40	end-reads	
WP_000183751.1	Ec6_41	end-reads	
WP_000786551.1	Ec6_42	end-reads	
WP_000241053.1	Ec6_43	end-reads	
WP_000355772.1	Ec6_44	end-reads	
WP_000338247.1	Ec6_45	end-reads	
WP_075208399.1	Ec6_46	end-reads	
WP_001327845.1	Ec6_47	end-reads	
WP_000155927.1	Ec6_48	end-reads	
WP_001445816.1	Ec6_49	end-reads	
WP_000206655.1	Ec6_5	full-length	position 154: amino acid T instead of A
WP_000407090.1	Ec6_50	end-reads	
WP_000859964.1	Ec6_51	end-reads	
WP_000077829.1	Ec6_52	end-reads	
WP_000013994.1	Ec6_53	end-reads	
WP_001122065.1	Ec6_54	end-reads	
WP_000139103.1	Ec6_55	end-reads	
WP_000020636.1	Ec6_56	end-reads	
WP_000671701.1	Ec6_57	end-reads	
WP_077626097.1	Ec6_58	end-reads	
WP_000004905.1	Ec6_59	end-reads	
WP_000258580.1	Ec6_6	full-length	complete alignment
WP_000611426.1	Ec6_7	full-length	complete alignment
WP_000804518.1	Ec6_8	full-length	complete alignment
WP_000961342.1	Ec6_9	full-length	complete alignment
WP_000067801.1	Efe_1	full-length	complete alignment
WP_001182890.1	Efe_10	full-length	complete alignment
WP_001235473.1	Efe_11	full-length	complete alignment
WP_001237041.1	Efe_12	full-length	complete alignment
WP_001272443.1	Efe_13	full-length	complete alignment
WP_024256417.1	Efe_14	full-length	complete alignment
WP_000937458.1	Efe_15	end-reads	
WP_000020896.1	Efe_16	end-reads	
WP_000189184.1	Efe_17	end-reads	
WP_000781397.1	Efe_18	end-reads	
WP_000995825.1	Efe_19	end-reads	
WP_000083435.1	Efe_2	full-length	position 252: amino acid M instead of L
WP_001066218.1	Efe_20	end-reads	
WP_001067513.1	Efe_21	end-reads	
WP_001147116.1	Efe_22	end-reads	
WP_032243086.1	Efe_23	end-reads	
WP_071821796.1	Efe_24	end-reads	
WP_077626319.1	Efe_25	end-reads	
WP_077626322.1	Efe_26	end-reads	
WP_000099375.1	Efe_27	end-reads	
WP_000786561.1	Efe_28	end-reads	
WP_001016304.1	Efe_29	end-reads	
WP_000148644.1	Efe_3	full-length	complete alignment
WP_001023055.1	Efe_30	end-reads	
WP_001143213.1	Efe_31	end-reads	
WP_000208170.1	Efe_32	end-reads	
WP_000438625.1	Efe_33	end-reads	
WP_000077885.1	Efe_34	end-reads	
WP_000083190.1	Efe_35	end-reads	
WP_000557378.1	Efe_36	end-reads	
WP_000097994.1	Efe_37	end-reads	
WP_000178797.1	Efe_4	full-length	complete alignment

effector ID	abbr.	quality control	full-length sequencing results
WP_000255032.1	Efe_5	full-length	complete alignment
WP_000375129.1	Efe_6	full-length	complete alignment
WP_000508975.1	Efe_7	full-length	complete alignment
WP_000904613.1	Efe_8	full-length	complete alignment
WP_000999547.1	Efe_9	full-length	complete alignment
WP_005285281.1	Eta_1	full-length	complete alignment
WP_035595442.1	Eta_10	end-reads	
WP_005283164.1	Eta_11	end-reads	
WP_005284548.1	Eta_12	end-reads	
WP_005283191.1	Eta_13	end-reads	
WP_005286533.1	Eta_14	end-reads	
WP_005289768.1	Eta_15	end-reads	
WP_005288481.1	Eta_2	full-length	complete alignment
WP_005294393.1	Eta_3	full-length	complete alignment
WP_005280619.1	Eta_4	end-reads	
WP_005282518.1	Eta_5	end-reads	
WP_005287789.1	Eta_6	end-reads	
WP_005294415.1	Eta_7	end-reads	
WP_005294857.1	Eta_8	end-reads	
WP_005296212.1	Eta_9	end-reads	
WP_002891634.1	Kpn_1	full-length	complete alignment
WP_009484876.1	Kpn_10	full-length	complete alignment
WP_009486019.1	Kpn_11	full-length	complete alignment
WP_009486529.1	Kpn_12	full-length	complete alignment
WP_020317218.1	Kpn_13	full-length	complete alignment
WP_032420351.1	Kpn_14	full-length	complete alignment
WP_004150458.1	Kpn_15	end-reads	
WP_004140501.1	Kpn_16	end-reads	
WP_009484109.1	Kpn_17	end-reads	
WP_020316952.1	Kpn_18	end-reads	
WP_004225268.1	Kpn_19	end-reads	
WP_002916607.1	Kpn_2	full-length	complete alignment
WP_032409076.1	Kpn_20	end-reads	
WP_009486504.1	Kpn_21	end-reads	
WP_049245346.1	Kpn_22	end-reads	
WP_009483878.1	Kpn_23	end-reads	
WP_009484937.1	Kpn_24	end-reads	
WP_002898814.1	Kpn_25	end-reads	
WP_004210116.1	Kpn_26	end-reads	
WP_004171426.1	Kpn_27	end-reads	
WP_002916742.1	Kpn_29	end-reads	
WP_004118237.1	Kpn_3	full-length	complete alignment
WP_004179102.1	Kpn_30	end-reads	
WP_000155904.1	Kpn_31	end-reads	
WP_002913732.1	Kpn_32	end-reads	
WP_004152062.1	Kpn_33	end-reads	
WP_002889847.1	Kpn_34	end-reads	
WP_002909008.1	Kpn_35	end-reads	
WP_009485462.1	Kpn_36	end-reads	
WP_019705807.1	Kpn_37	end-reads	
WP_004147894.1	Kpn_38	end-reads	
WP_002912648.1	Kpn_39	end-reads	
WP_004145486.1	Kpn_4	full-length	complete alignment
WP_002916277.1	Kpn_40	end-reads	
WP_009484993.1	Kpn_41	end-reads	
WP_004149975.1	Kpn_5	full-length	complete alignment
WP_004152718.1	Kpn_6	full-length	complete alignment
WP_004177339.1	Kpn_7	full-length	complete alignment
WP_004197606.1	Kpn_8	full-length	complete alignment
WP_009484324.1	Kpn_9	full-length	complete alignment
WP_004234458.1	Mmo_1	full-length	complete alignment
WP_062771682.1	Mmo_10	full-length	complete alignment
WP_062772817.1	Mmo_11	full-length	complete alignment
WP_062773522.1	Mmo_12	full-length	complete alignment
WP_062773651.1	Mmo_13	full-length	complete alignment
WP_004234829.1	Mmo_14	end-reads	
WP_004235416.1	Mmo_15	end-reads	
WP_004235744.1	Mmo_16	end-reads	
WP_004235986.1	Mmo_17	end-reads	
WP_004236694.1	Mmo_18	end-reads	
WP_004236839.1	Mmo_19	end-reads	
WP_004235425.1	Mmo_2	full-length	complete alignment
WP_004238267.1	Mmo_20	end-reads	
WP_004238584.1	Mmo_21	end-reads	
WP_004238854.1	Mmo_22	end-reads	
WP_004241150.1	Mmo_23	end-reads	

effector ID	abbr.	quality control	full-length sequencing results
WP_004242218.1	Mmo_24	end-reads	
WP_004242347.1	Mmo_25	end-reads	
WP_004242398.1	Mmo_26	end-reads	
WP_024474672.1	Mmo_27	end-reads	
WP_024475195.1	Mmo_28	end-reads	
WP_032098087.1	Mmo_29	end-reads	
WP_004236571.1	Mmo_3	full-length	complete alignment
WP_062771418.1	Mmo_30	end-reads	
WP_073970193.1	Mmo_31	end-reads	
WP_080654118.1	Mmo_32	end-reads	
WP_004234853.1	Mmo_33	end-reads	
WP_004240526.1	Mmo_34	end-reads	
WP_062773486.1	Mmo_35	end-reads	
WP_062773581.1	Mmo_36	end-reads	
WP_062773673.1	Mmo_37	end-reads	
WP_073970171.1	Mmo_38	end-reads	
WP_073970177.1	Mmo_39	end-reads	
WP_004238406.1	Mmo_4	full-length	complete alignment
WP_081113481.1	Mmo_40	end-reads	
WP_004241031.1	Mmo_41	end-reads	
WP_004241218.1	Mmo_42	end-reads	
WP_004904012.1	Mmo_43	end-reads	
WP_015422612.1	Mmo_44	end-reads	
WP_036417208.1	Mmo_45	end-reads	
WP_036416809.1	Mmo_46	end-reads	
WP_004235474.1	Mmo_47	end-reads	
WP_004240712.1	Mmo_5	full-length	complete alignment
WP_032098021.1	Mmo_6	full-length	complete alignment
WP_036413302.1	Mmo_7	full-length	complete alignment
WP_036417499.1	Mmo_8	full-length	complete alignment
WP_046024762.1	Mmo_9	full-length	complete alignment
WP_040259375.1	Pem_1	full-length	complete alignment
WP_040259956.1	Pem_10	end-reads	
WP_040260490.1	Pem_11	end-reads	
WP_040261863.1	Pem_12	end-reads	
WP_040263293.1	Pem_13	end-reads	
WP_084596159.1	Pem_14	end-reads	
WP_084596289.1	Pem_15	end-reads	
WP_084596314.1	Pem_16	end-reads	
WP_088776203.1	Pem_17	end-reads	
WP_040260997.1	Pem_18	end-reads	
WP_040259943.1	Pem_19	end-reads	
WP_040260715.1	Pem_2	full-length	complete alignment
WP_052469295.1	Pem_20	end-reads	
WP_088776180.1	Pem_21	end-reads	
WP_052469277.1	Pem_22	end-reads	
WP_040261153.1	Pem_23	end-reads	
WP_040262756.1	Pem_24	end-reads	
WP_040262478.1	Pem_25	end-reads	
WP_052469135.1	Pem_26	end-reads	
WP_040263025.1	Pem_3	full-length	complete alignment
WP_040263420.1	Pem_4	full-length	complete alignment
WP_040263598.1	Pem_5	full-length	complete alignment
WP_084596144.1	Pem_6	full-length	complete alignment
WP_084596156.1	Pem_7	full-length	complete alignment
WP_084596184.1	Pem_8	full-length	complete alignment
WP_040259600.1	Pem_9	end-reads	
WP_000116680.1	Pfa_1	full-length	complete alignment
WP_108474309.1	Pfa_10	full-length	complete alignment
WP_113857302.1	Pfa_11	full-length	complete alignment
WP_113857471.1	Pfa_12	full-length	complete alignment
WP_113857569.1	Pfa_13	full-length	complete alignment
WP_113857629.1	Pfa_14	full-length	complete alignment
WP_113858483.1	Pfa_15	full-length	complete alignment
WP_113858620.1	Pfa_16	full-length	complete alignment
WP_113859044.1	Pfa_17	full-length	complete alignment
WP_113859080.1	Pfa_18	full-length	complete alignment
WP_000490639.1	Pfa_19	end-reads	
WP_000703842.1	Pfa_2	full-length	complete alignment
WP_000817037.1	Pfa_20	end-reads	
WP_000932975.1	Pfa_21	end-reads	
WP_015963067.1	Pfa_22	end-reads	
WP_020319858.1	Pfa_23	end-reads	
WP_108474137.1	Pfa_24	end-reads	
WP_108474230.1	Pfa_25	end-reads	
WP_108474549.1	Pfa_26	end-reads	

effector ID	abbr.	quality control	full-length sequencing results
WP_108474640.1	Pfa_27	end-reads	
WP_108474922.1	Pfa_28	end-reads	
WP_108475013.1	Pfa_29	end-reads	
WP_001516695.1	Pfa_3	full-length	complete alignment
WP_108475618.1	Pfa_30	end-reads	
WP_108475752.1	Pfa_31	end-reads	
WP_108475953.1	Pfa_32	end-reads	
WP_113857422.1	Pfa_33	end-reads	
WP_113857724.1	Pfa_34	end-reads	
WP_113858462.1	Pfa_35	end-reads	
WP_113858733.1	Pfa_36	end-reads	
WP_113858817.1	Pfa_37	end-reads	
WP_113858981.1	Pfa_38	end-reads	
WP_011152995.1	Pfa_39	end-reads	
WP_001531161.1	Pfa_4	full-length	complete alignment
WP_108476339.1	Pfa_40	end-reads	
WP_108476447.1	Pfa_41	end-reads	
WP_113858661.1	Pfa_42	end-reads	
WP_113858928.1	Pfa_43	end-reads	
WP_113857601.1	Pfa_44	end-reads	
WP_113858376.1	Pfa_45	end-reads	
WP_108475277.1	Pfa_46	end-reads	
WP_113844795.1	Pfa_47	end-reads	
WP_015962672.1	Pfa_5	full-length	complete alignment
WP_015963250.1	Pfa_6	full-length	complete alignment
WP_108473469.1	Pfa_7	full-length	complete alignment
WP_108473773.1	Pfa_8	full-length	complete alignment
WP_108473781.1	Pfa_9	full-length	complete alignment
WP_044172624.1	Pma_1	full-length	complete alignment
WP_044180332.1	Pma_10	full-length	complete alignment
WP_044180423.1	Pma_11	full-length	complete alignment
WP_044180429.1	Pma_12	full-length	complete alignment
WP_044180562.1	Pma_13	full-length	complete alignment
WP_044183152.1	Pma_14	full-length	complete alignment
WP_044183301.1	Pma_15	full-length	complete alignment
WP_044183672.1	Pma_16	full-length	complete alignment
WP_071825927.1	Pma_17	full-length	complete alignment
WP_081653590.1	Pma_18	full-length	complete alignment
WP_044173357.1	Pma_19	end-reads	
WP_044173012.1	Pma_2	full-length	complete alignment
WP_044177694.1	Pma_20	end-reads	
WP_044177823.1	Pma_21	end-reads	
WP_044179949.1	Pma_22	end-reads	
WP_044182945.1	Pma_23	end-reads	
WP_052332703.1	Pma_24	end-reads	
WP_071825876.1	Pma_26	end-reads	
WP_081653585.1	Pma_27	end-reads	
WP_081653604.1	Pma_28	end-reads	
WP_071825830.1	Pma_29	end-reads	
WP_044173054.1	Pma_3	full-length	complete alignment
WP_044182930.1	Pma_30	end-reads	
WP_044183362.1	Pma_31	end-reads	
WP_044184806.1	Pma_32	end-reads	
WP_052332698.1	Pma_33	end-reads	
WP_044177883.1	Pma_34	end-reads	
WP_044173232.1	Pma_35	end-reads	
WP_044174146.1	Pma_4	full-length	complete alignment
WP_044176371.1	Pma_5	full-length	complete alignment
WP_044177448.1	Pma_6	full-length	difficult to sequence, quality always bad. DNA Sequence 1-21 perfect, several gaps until 65, 65-118 perfect, additional base T, 119-241 perfect
WP_044177605.1	Pma_7	full-length	complete alignment
WP_044178555.1	Pma_8	full-length	complete alignment
WP_044180054.1	Pma_9	full-length	complete alignment
WP_004253606.1	Pre_1	full-length	complete alignment
WP_004258336.1	Pre_10	full-length	complete alignment
WP_004258949.1	Pre_11	full-length	complete alignment
WP_004261326.1	Pre_12	full-length	difficult to sequence: DNA Sequence 1-1317 perfect, gap at 1318 with bad sequencing quality, no data until 1351, perfect until 1376
WP_004261459.1	Pre_13	full-length	complete alignment
WP_004261604.1	Pre_14	full-length	complete alignment
WP_004261691.1	Pre_15	full-length	complete alignment
WP_004262673.1	Pre_16	full-length	complete alignment

effector ID	abbr.	quality control	full-length sequencing results
WP_004263067.1	Pre_17	full-length	complete alignment
WP_004264858.1	Pre_18	full-length	complete alignment
WP_004264902.1	Pre_19	full-length	complete alignment
WP_004253752.1	Pre_2	full-length	complete alignment
WP_004264927.1	Pre_20	full-length	complete alignment
WP_004905473.1	Pre_21	full-length	complete alignment
WP_004906048.1	Pre_22	full-length	complete alignment
WP_004912645.1	Pre_23	full-length	complete alignment
WP_036957920.1	Pre_24	full-length	complete alignment
WP_004262987.1	Pre_25	end-reads	
WP_004253374.1	Pre_26	end-reads	
WP_004253768.1	Pre_27	end-reads	
WP_004254561.1	Pre_28	end-reads	
WP_004254943.1	Pre_29	end-reads	
WP_004254983.1	Pre_3	full-length	complete alignment
WP_004255405.1	Pre_30	end-reads	
WP_004256486.1	Pre_31	end-reads	
WP_004256890.1	Pre_32	end-reads	
WP_004257237.1	Pre_33	end-reads	
WP_004257539.1	Pre_34	end-reads	
WP_004258015.1	Pre_35	end-reads	
WP_004258503.1	Pre_36	end-reads	
WP_004260627.1	Pre_37	end-reads	
WP_004261076.1	Pre_38	end-reads	
WP_004261765.1	Pre_39	end-reads	
WP_004255002.1	Pre_4	full-length	complete alignment
WP_004261788.1	Pre_40	end-reads	
WP_004262117.1	Pre_41	end-reads	
WP_004262890.1	Pre_42	end-reads	
WP_004263051.1	Pre_43	end-reads	
WP_004263203.1	Pre_44	end-reads	
WP_004264245.1	Pre_45	end-reads	
WP_004264319.1	Pre_46	end-reads	
WP_004264499.1	Pre_47	end-reads	
WP_004264896.1	Pre_48	end-reads	
WP_004265252.1	Pre_49	end-reads	
WP_004255132.1	Pre_5	full-length	complete alignment
WP_004905275.1	Pre_50	end-reads	
WP_036957743.1	Pre_51	end-reads	
WP_036957864.1	Pre_52	end-reads	
WP_036957904.1	Pre_53	end-reads	
WP_036958071.1	Pre_54	end-reads	
WP_080544449.1	Pre_55	end-reads	
WP_004265131.1	Pre_56	end-reads	
WP_004905318.1	Pre_57	end-reads	
WP_004905395.1	Pre_58	end-reads	
WP_004906105.1	Pre_59	end-reads	
WP_004256127.1	Pre_6	full-length	complete alignment
WP_004906207.1	Pre_60	end-reads	
WP_036957944.1	Pre_61	end-reads	
WP_036958341.1	Pre_62	end-reads	
WP_036958442.1	Pre_63	end-reads	
WP_036958542.1	Pre_64	end-reads	
WP_004254299.1	Pre_65	end-reads	
WP_004261181.1	Pre_66	end-reads	
WP_004261611.1	Pre_67	end-reads	
WP_004262300.1	Pre_68	end-reads	
WP_004262383.1	Pre_69	end-reads	
WP_004256756.1	Pre_7	full-length	complete alignment
WP_004262559.1	Pre_70	end-reads	
WP_004262981.1	Pre_71	end-reads	
WP_004262990.1	Pre_72	end-reads	
WP_004263536.1	Pre_73	end-reads	
WP_004264507.1	Pre_74	end-reads	
WP_004264615.1	Pre_75	end-reads	
WP_004256437.1	Pre_76	end-reads	
WP_004256728.1	Pre_77	end-reads	
WP_004257114.1	Pre_78	end-reads	
WP_004260254.1	Pre_79	end-reads	
WP_004257109.1	Pre_8	full-length	complete alignment
WP_004260347.1	Pre_80	end-reads	
WP_004256913.1	Pre_81	end-reads	
WP_004257253.1	Pre_82	end-reads	
WP_004265549.1	Pre_83	end-reads	
WP_004254558.1	Pre_84	end-reads	
WP_004256190.1	Pre_85	end-reads	

effector ID	abbr.	quality control	full-length sequencing results
WP_036958458.1	Pre_86	end-reads	
WP_004257971.1	Pre_9	full-length	complete alignment
WP_033751802.1	Pse_1	full-length	complete alignment
WP_033791692.1	Pse_2	full-length	complete alignment
WP_033792202.1	Pse_3	full-length	complete alignment
WP_033792699.1	Pse_4	full-length	complete alignment
WP_033747833.1	Pse_5	end-reads	
WP_033750506.1	Pse_6	end-reads	
WP_033789480.1	Pse_7	end-reads	
WP_033793269.1	Pse_8	end-reads	
WP_033753922.1	Pse_9	end-reads	
WP_004915569.1	Pst_1	full-length	complete alignment
WP_004922866.1	Pst_10	full-length	complete alignment
WP_004924913.1	Pst_11	full-length	complete alignment
WP_004925512.1	Pst_12	full-length	complete alignment
WP_004926210.1	Pst_13	full-length	complete alignment
WP_004926361.1	Pst_14	full-length	complete alignment
WP_004927264.1	Pst_15	full-length	complete alignment
WP_042116632.1	Pst_16	full-length	complete alignment
WP_042117401.1	Pst_17	full-length	complete alignment
WP_052309238.1	Pst_18	full-length	Position 100: amino acid H instead of R
WP_071599648.1	Pst_19	full-length	complete alignment
WP_004915712.1	Pst_2	full-length	complete alignment
WP_004923882.1	Pst_20	end-reads	
WP_004921749.1	Pst_21	end-reads	
WP_004917889.1	Pst_22	end-reads	
WP_004922318.1	Pst_23	end-reads	
WP_004916773.1	Pst_24	end-reads	
WP_004917276.1	Pst_25	end-reads	
WP_004917489.1	Pst_26	end-reads	
WP_004918835.1	Pst_27	end-reads	
WP_004919137.1	Pst_28	end-reads	
WP_004919494.1	Pst_29	end-reads	
WP_004917132.1	Pst_3	full-length	complete alignment
WP_004919570.1	Pst_30	end-reads	
WP_004919841.1	Pst_31	end-reads	
WP_004919855.1	Pst_32	end-reads	
WP_004921504.1	Pst_33	end-reads	
WP_004922704.1	Pst_34	end-reads	
WP_004923523.1	Pst_35	end-reads	
WP_004924665.1	Pst_36	end-reads	
WP_004925688.1	Pst_37	end-reads	
WP_004926354.1	Pst_38	end-reads	
WP_014657126.1	Pst_39	end-reads	
WP_004917987.1	Pst_4	full-length	complete alignment
WP_014657167.1	Pst_40	end-reads	
WP_014657408.1	Pst_41	end-reads	
WP_014658194.1	Pst_42	end-reads	
WP_042116129.1	Pst_43	end-reads	
WP_042116289.1	Pst_44	end-reads	
WP_052038327.1	Pst_45	end-reads	
WP_071599635.1	Pst_46	end-reads	
WP_004915478.1	Pst_47	end-reads	
WP_004915527.1	Pst_48	end-reads	
WP_004917597.1	Pst_49	end-reads	
WP_004919332.1	Pst_5	full-length	complete alignment
WP_004918153.1	Pst_50	end-reads	
WP_004918358.1	Pst_51	end-reads	
WP_004918442.1	Pst_52	end-reads	
WP_004918480.1	Pst_53	end-reads	
WP_004919091.1	Pst_54	end-reads	
WP_004919165.1	Pst_55	end-reads	
WP_004920058.1	Pst_56	end-reads	
WP_004920817.1	Pst_57	end-reads	
WP_004921265.1	Pst_58	end-reads	
WP_004921358.1	Pst_59	end-reads	
WP_004919757.1	Pst_6	full-length	complete alignment
WP_004921623.1	Pst_60	end-reads	
WP_004922117.1	Pst_61	end-reads	
WP_004922411.1	Pst_62	end-reads	
WP_004922896.1	Pst_63	end-reads	
WP_004924065.1	Pst_64	end-reads	
WP_004924072.1	Pst_65	end-reads	
WP_004924254.1	Pst_66	end-reads	
WP_004924497.1	Pst_67	end-reads	
WP_004924795.1	Pst_68	end-reads	

effector ID	abbr.	quality control	full-length sequencing results
WP_004925180.1	Pst_69	end-reads	
WP_004920813.1	Pst_7	full-length	complete alignment
WP_004925244.1	Pst_70	end-reads	
WP_004925746.1	Pst_71	end-reads	
WP_004926047.1	Pst_72	end-reads	
WP_004926051.1	Pst_73	end-reads	
WP_004927136.1	Pst_74	end-reads	
WP_004927154.1	Pst_75	end-reads	
WP_004927350.1	Pst_76	end-reads	
WP_004927466.1	Pst_77	end-reads	
WP_014656289.1	Pst_78	end-reads	
WP_036940590.1	Pst_79	end-reads	
WP_004921520.1	Pst_8	full-length	complete alignment
WP_042117315.1	Pst_80	end-reads	
WP_004915373.1	Pst_81	end-reads	
WP_004917178.1	Pst_82	end-reads	
WP_004917602.1	Pst_83	end-reads	
WP_004918964.1	Pst_84	end-reads	
WP_004925628.1	Pst_85	end-reads	
WP_004922581.1	Pst_9	full-length	complete alignment
WP_004726235.1	Vfu_1	full-length	complete alignment
WP_004729624.1	Vfu_10	full-length	complete alignment
WP_014257346.1	Vfu_11	full-length	complete alignment
WP_014257409.1	Vfu_12	full-length	complete alignment
WP_014257429.1	Vfu_13	full-length	complete alignment
WP_014258130.1	Vfu_14	full-length	complete alignment
WP_038151258.1	Vfu_15	full-length	complete alignment
WP_038151743.1	Vfu_16	full-length	complete alignment
WP_038151811.1	Vfu_17	full-length	complete alignment
WP_038152552.1	Vfu_18	full-length	complete alignment
WP_038152592.1	Vfu_19	full-length	complete alignment
WP_004726603.1	Vfu_2	full-length	complete alignment
WP_038152705.1	Vfu_20	full-length	complete alignment
WP_115333225.1	Vfu_21	full-length	complete alignment
WP_004724260.1	Vfu_22	end-reads	
WP_004726312.1	Vfu_23	end-reads	
WP_004726677.1	Vfu_24	end-reads	
WP_004726765.1	Vfu_25	end-reads	
WP_004727405.1	Vfu_26	end-reads	
WP_004727644.1	Vfu_27	end-reads	
WP_004727675.1	Vfu_28	end-reads	
WP_004727750.1	Vfu_29	end-reads	
WP_004726842.1	Vfu_3	full-length	complete alignment
WP_004728100.1	Vfu_30	end-reads	
WP_004728215.1	Vfu_31	end-reads	
WP_004728297.1	Vfu_32	end-reads	
WP_004728834.1	Vfu_33	end-reads	
WP_004729099.1	Vfu_34	end-reads	
WP_004729399.1	Vfu_35	end-reads	
WP_038151273.1	Vfu_36	end-reads	
WP_038151757.1	Vfu_37	end-reads	
WP_038152841.1	Vfu_38	end-reads	
WP_115333239.1	Vfu_39	end-reads	
WP_004726871.1	Vfu_4	full-length	complete alignment
WP_004727628.1	Vfu_40	end-reads	
WP_004728470.1	Vfu_41	end-reads	
WP_004729387.1	Vfu_42	end-reads	
WP_038152009.1	Vfu_43	end-reads	
WP_115333279.1	Vfu_44	end-reads	
WP_004726990.1	Vfu_45	end-reads	
WP_004727567.1	Vfu_46	end-reads	
WP_004727751.1	Vfu_47	end-reads	
WP_004728048.1	Vfu_48	end-reads	
WP_004728083.1	Vfu_49	end-reads	
WP_004727003.1	Vfu_5	full-length	difficult to sequence, quality always bad. 880-1747 perfect
WP_004728310.1	Vfu_50	end-reads	
WP_004729342.1	Vfu_51	end-reads	
WP_038150968.1	Vfu_52	end-reads	
WP_081454518.1	Vfu_53	end-reads	
WP_086027288.1	Vfu_54	end-reads	
WP_004725800.1	Vfu_55	end-reads	
WP_004725809.1	Vfu_56	end-reads	
WP_004725893.1	Vfu_57	end-reads	
WP_004726164.1	Vfu_58	end-reads	
WP_004726428.1	Vfu_59	end-reads	

effector ID	abbr.	quality control	full-length sequencing results
WP_004727345.1	Vfu_6	full-length	position 241: amino acid T instead of P
WP_004725751.1	Vfu_60	end-reads	
WP_038151352.1	Vfu_61	end-reads	
WP_004727658.1	Vfu_7	full-length	complete alignment
WP_004727926.1	Vfu_8	full-length	complete alignment
WP_004729371.1	Vfu_9	full-length	complete alignment
WP_005156690.1	Yen_1	full-length	complete alignment
WP_005163729.1	Yen_10	full-length	complete alignment
WP_005163816.1	Yen_11	full-length	position 202: amino acid D instead of E
WP_005164084.1	Yen_12	full-length	complete alignment
WP_005164331.1	Yen_13	full-length	complete alignment
WP_005164542.1	Yen_14	full-length	complete alignment
WP_005166097.1	Yen_15	full-length	complete alignment
WP_005179029.1	Yen_16	full-length	complete alignment
WP_010891207.1	Yen_17	full-length	complete alignment
WP_010891236.1	Yen_18	full-length	complete alignment
WP_014609009.1	Yen_19	full-length	complete alignment
WP_005157598.1	Yen_2	full-length	complete alignment
WP_014609475.1	Yen_20	full-length	complete alignment
WP_005164132.1	Yen_21	end-reads	
WP_005156531.1	Yen_22	end-reads	
WP_005160863.1	Yen_23	end-reads	
WP_005156566.1	Yen_24	end-reads	
WP_005156700.1	Yen_25	end-reads	
WP_005157827.1	Yen_26	end-reads	
WP_005158044.1	Yen_27	end-reads	
WP_005158295.1	Yen_28	end-reads	
WP_005158416.1	Yen_29	end-reads	
WP_005157674.1	Yen_3	full-length	complete alignment
WP_005160776.1	Yen_30	end-reads	
WP_005161939.1	Yen_31	end-reads	
WP_005162175.1	Yen_32	end-reads	
WP_005162234.1	Yen_33	end-reads	
WP_005163332.1	Yen_34	end-reads	
WP_005164848.1	Yen_35	end-reads	
WP_005165338.1	Yen_36	end-reads	
WP_005165350.1	Yen_37	end-reads	
WP_005179266.1	Yen_38	end-reads	
WP_010891241.1	Yen_39	end-reads	
WP_005159058.1	Yen_4	full-length	complete alignment
WP_014609447.1	Yen_40	end-reads	
WP_016266096.1	Yen_41	end-reads	
WP_020282365.1	Yen_42	end-reads	
WP_023160440.1	Yen_43	end-reads	
WP_071598586.1	Yen_44	end-reads	
WP_080366037.1	Yen_45	end-reads	
WP_002229817.1	Yen_46	end-reads	
WP_005160046.1	Yen_47	end-reads	
WP_005156692.1	Yen_48	end-reads	
WP_005157407.1	Yen_49	end-reads	
WP_005159145.1	Yen_5	full-length	complete alignment
WP_005157433.1	Yen_50	end-reads	
WP_005158077.1	Yen_51	end-reads	
WP_005158896.1	Yen_52	end-reads	
WP_005159587.1	Yen_53	end-reads	
WP_005164223.1	Yen_54	end-reads	
WP_005165873.1	Yen_55	end-reads	
WP_005179706.1	Yen_56	end-reads	
WP_014609219.1	Yen_57	end-reads	
WP_014609473.1	Yen_58	end-reads	
WP_016266437.1	Yen_59	end-reads	
WP_005159272.1	Yen_6	full-length	complete alignment
WP_005156400.1	Yen_60	end-reads	
WP_005157177.1	Yen_61	end-reads	
WP_005157512.1	Yen_62	end-reads	
WP_014609110.1	Yen_63	end-reads	
WP_014609336.1	Yen_64	end-reads	
WP_005163324.1	Yen_65	end-reads	
WP_005163567.1	Yen_66	end-reads	
WP_005166456.1	Yen_67	end-reads	
WP_010891206.1	Yen_68	end-reads	
WP_005162291.1	Yen_7	full-length	complete alignment
WP_005162694.1	Yen_8	full-length	complete alignment
WP_005162781.1	Yen_9	full-length	complete alignment
WP_006817197.1	Yre_1	full-length	Position 229: amino acid V instead of G
WP_040903175.1	Yre_10	full-length	position 298: amino acid T instead of A

effector ID	abbr.	quality control	full-length sequencing results
WP_040903573.1	Yre_11	full-length	complete alignment
WP_050812366.1	Yre_12	full-length	complete alignment
WP_071777502.1	Yre_13	full-length	complete alignment
WP_071777518.1	Yre_14	full-length	complete alignment
WP_006818167.1	Yre_15	end-reads	
WP_006818297.1	Yre_16	end-reads	
WP_006818304.1	Yre_17	end-reads	
WP_006818522.1	Yre_18	end-reads	
WP_006819026.1	Yre_19	end-reads	
WP_006818175.1	Yre_2	full-length	complete alignment
WP_006819540.1	Yre_20	end-reads	
WP_006820008.1	Yre_21	end-reads	
WP_006820125.1	Yre_22	end-reads	
WP_038256778.1	Yre_23	end-reads	
WP_038258270.1	Yre_24	end-reads	
WP_040903561.1	Yre_25	end-reads	
WP_071777494.1	Yre_26	end-reads	
WP_087943223.1	Yre_27	end-reads	
WP_006817680.1	Yre_28	end-reads	
WP_006818160.1	Yre_29	end-reads	
WP_006818941.1	Yre_3	full-length	complete alignment
WP_006818482.1	Yre_30	end-reads	
WP_040902689.1	Yre_31	end-reads	
WP_071777524.1	Yre_32	end-reads	
WP_006820021.1	Yre_33	end-reads	
WP_006820270.1	Yre_34	end-reads	
WP_006820308.1	Yre_35	end-reads	
WP_040902687.1	Yre_36	end-reads	
WP_081874660.1	Yre_37	end-reads	
WP_006818356.1	Yre_38	end-reads	
WP_006820296.1	Yre_39	end-reads	
WP_006819702.1	Yre_4	full-length	complete alignment
WP_006820870.1	Yre_40	end-reads	
WP_038258506.1	Yre_41	end-reads	
WP_006820335.1	Yre_5	full-length	complete alignment
WP_006820832.1	Yre_6	full-length	complete alignment
WP_006820847.1	Yre_7	full-length	complete alignment
WP_006821080.1	Yre_8	full-length	complete alignment
WP_038254823.1	Yre_9	full-length	complete alignment

Table S4 | Y2H interactions. *Interactions detected by the Y2H between effectors and human proteins in the main and/or repeat screen and/or the homology test. "1" indicates detected interaction between the human protein and the bacterial effector. abbr., abbreviation.*

human protein	effector	abbr.	MAIN	REPEAT	HOMOLOGY
SGTA	WP_042030958.1	Aja_1	1		
RHOXF2	WP_042032056.1	Aja_2	1		
PAX5	WP_042032153.1	Aja_3	1		
KRTAP10-9	WP_004864811.1	Cda_1			1
KRTAP1-3	WP_004864811.1	Cda_1			1
OTX1	WP_004864811.1	Cda_1			1
GLRX3	WP_004864811.1	Cda_1	1		1
KRT31	WP_004864811.1	Cda_1	1		1
INADL	WP_039898535.1	Cda_10	1	1	1
HSF2BP	WP_055696404.1	Cda_11	1		
PDE4DIP	WP_055696404.1	Cda_11	1		
PDCL2	WP_055696404.1	Cda_11	1		
NAV2	WP_055696404.1	Cda_11	1		
ZNF175	WP_055696404.1	Cda_11	1		
MID2	WP_055696404.1	Cda_11	1		
KRT27	WP_055696404.1	Cda_11	1		
PICK1	WP_055696404.1	Cda_11	1		
IKZF4	WP_055696404.1	Cda_11	1		
TCF4	WP_055696404.1	Cda_11	1		
KIAA1024	WP_055696404.1	Cda_11	1		
RAB11FIP3	WP_055696404.1	Cda_11	1		
SHANK2	WP_055696404.1	Cda_11	1		
HMBOX1	WP_055696404.1	Cda_11	1		
SPAG5	WP_055696404.1	Cda_11	1		
LZTS2	WP_055696404.1	Cda_11	1		
TFIP11	WP_055696404.1	Cda_11	1		
TNIP1	WP_055696404.1	Cda_11	1		
TRIP6	WP_055696404.1	Cda_11	1		
UBQLN2	WP_055696404.1	Cda_11	1		
UBXN7	WP_055696404.1	Cda_11	1		
ZBED4	WP_055696404.1	Cda_11	1		
SP4	WP_055696404.1	Cda_11	1		
CALCOCO1	WP_055696404.1	Cda_11	1		
GOLGA6A	WP_055696404.1	Cda_11	1		
GOLGA2	WP_055696404.1	Cda_11	1		
GIGYF1	WP_055696404.1	Cda_11	1		
ETV6	WP_055696404.1	Cda_11	1		
CEP250	WP_055696404.1	Cda_11	1		
CCDC136	WP_055696404.1	Cda_11	1		
MTUS2	WP_055696404.1	Cda_11	1		
GOLGA6L9	WP_055696404.1	Cda_11	1		
SH3RF1	WP_083478381.1	Cda_12	1		
CXorf41	WP_016517497.1	Cda_2	1		
VAC14	WP_016535503.1	Cda_3	1		
MAGEA6	WP_016535835.1	Cda_4			1
MFF	WP_016536389.1	Cda_5	1		
CEP76	WP_016536523.1	Cda_6	1		
CYB5B	WP_016536523.1	Cda_6	1		
HSF2BP	WP_016536523.1	Cda_6	1		
PRDM6	WP_016536523.1	Cda_6	1		
NRBP1	WP_016536523.1	Cda_6	1		
GOLGA2	WP_016536523.1	Cda_6	1		
PICK1	WP_016536523.1	Cda_6	1		
TRAF2	WP_016536523.1	Cda_6	1		
INADL	WP_016537909.1	Cda_7		1	1
ZC4H2	WP_016538154.1	Cda_8		1	
TRAF2	WP_016538154.1	Cda_8	1		
AGR2	WP_039898226.1	Cda_9	1		
MAGEA6	WP_005129057.1	Cyo_1	1		1
HSF2BP	WP_072041464.1	Cyo_10	1		

human protein	effector	abbr.	MAIN	REPEAT	HOMOLOGY
GOLGA6L9	WP_072041464.1	Cyo_10	1		
DYNLT1	WP_072041464.1	Cyo_10	1		
KRT75	WP_072041472.1	Cyo_11	1		
HACE1	WP_072041472.1	Cyo_11	1		
KRT76	WP_072041472.1	Cyo_11	1		
SSBP3	WP_072041472.1	Cyo_11	1		
COG6	WP_072041472.1	Cyo_11	1		
IKBKG	WP_080721914.1	Cyo_12	1		
TRAF2	WP_080721914.1	Cyo_12	1		
VAC14	WP_080721914.1	Cyo_12	1		
TH	WP_082031767.1	Cyo_13	1		1
MAL2	WP_005129187.1	Cyo_2			1
AGTRAP	WP_005129187.1	Cyo_2			1
ZBED4	WP_005129207.1	Cyo_3	1		1
TRAF2	WP_005129207.1	Cyo_3	1		1
EFHC2	WP_005129207.1	Cyo_3			1
MFF	WP_005131699.1	Cyo_4	1		1
MAL2	WP_040229899.1	Cyo_5			1
AGTRAP	WP_040229899.1	Cyo_5			1
ACOT8	WP_040230127.1	Cyo_6	1		1
NOTO	WP_040230127.1	Cyo_6	1		1
REL	WP_040230127.1	Cyo_6	1		1
ZMYND12	WP_040230127.1	Cyo_6	1		1
COL17A1	WP_040232070.1	Cyo_7	1		
VAC14	WP_040232968.1	Cyo_8	1		1
UBQLN1	WP_040233404.1	Cyo_9	1		1
REL	WP_000004564.1	Ec2_1	1		
SORBS3	WP_000004564.1	Ec2_1	1		
TCF4	WP_000004564.1	Ec2_1	1		
TNIP1	WP_000004564.1	Ec2_1	1		
TRIM27	WP_000004564.1	Ec2_1	1		
UBAP1	WP_000004564.1	Ec2_1	1		
PAX6	WP_000004564.1	Ec2_1	1		
ZBTB7B	WP_000004564.1	Ec2_1	1		
TRIM9	WP_000004564.1	Ec2_1	1		
ZBTB8A	WP_000004564.1	Ec2_1	1		
ZMYND12	WP_000004564.1	Ec2_1	1		
VPS52	WP_000004564.1	Ec2_1	1		
EHMT2	WP_000004564.1	Ec2_1	1		
VIM	WP_000004564.1	Ec2_1	1		
BIRC7	WP_000004564.1	Ec2_1	1		
CCDC102B	WP_000004564.1	Ec2_1	1		
CCDC125	WP_000004564.1	Ec2_1	1		
NECAB2	WP_000004564.1	Ec2_1	1		
CNTROB	WP_000004564.1	Ec2_1	1		
FAM188A	WP_000004564.1	Ec2_1	1		
HOMEZ	WP_000004564.1	Ec2_1	1		
HSPB8	WP_000004564.1	Ec2_1	1		
KCTD6	WP_000004564.1	Ec2_1	1		
KCTD9	WP_000004564.1	Ec2_1	1		
KRTAP10-5	WP_000004564.1	Ec2_1	1		
CNNM1	WP_000004564.1	Ec2_1	1		
LBX1	WP_000083477.1	Ec2_2			1
IFT88	WP_000220141.1	Ec2_3	1		
REL	WP_000220141.1	Ec2_3	1		
DPPA4	WP_000220141.1	Ec2_3	1	1	
PRPSAP2	WP_000220141.1	Ec2_3	1	1	
ZBTB39	WP_000220141.1	Ec2_3	1		
LZTFL1	WP_000611436.1	Ec2_4		1	1
MFF	WP_001298277.1	Ec2_5	1		
LBX1	WP_000075087.1	Ec6_1			1
EFHC2	WP_001059674.1	Ec6_10	1		
ZMYND12	WP_001059674.1	Ec6_10	1		
VIM	WP_001059674.1	Ec6_10	1		

human protein	effector	abbr.	MAIN	REPEAT	HOMOLOGY
TRIM50	WP_001059674.1	Ec6_10	1		
TRIM27	WP_001059674.1	Ec6_10	1		
TCF4	WP_001059674.1	Ec6_10	1		
KRT75	WP_001059674.1	Ec6_10	1		
KIFC3	WP_001059674.1	Ec6_10	1		
IFT88	WP_001059674.1	Ec6_10	1		
VPS52	WP_001059674.1	Ec6_10	1	1	
REL	WP_001059674.1	Ec6_10	1	1	
COG6	WP_001059674.1	Ec6_10	1		
KCTD6	WP_001059674.1	Ec6_10	1		
GOLGA2	WP_001059674.1	Ec6_10	1		
LBX1	WP_001093944.1	Ec6_11			1
GOLGA6L9	WP_001267298.1	Ec6_12	1		
RALYL	WP_001267298.1	Ec6_12	1		
PNMA2	WP_001267298.1	Ec6_12	1		
PNMA1	WP_001267298.1	Ec6_12	1		
NUP62	WP_001267298.1	Ec6_12	1		
NECAB1	WP_001267298.1	Ec6_12	1		
MTUS2	WP_001267298.1	Ec6_12	1		
MID1	WP_001267298.1	Ec6_12	1		
MCM6	WP_001267298.1	Ec6_12	1		
LBX1	WP_001267298.1	Ec6_12	1		
KCTD9	WP_001267298.1	Ec6_12	1		
KANK4	WP_001267298.1	Ec6_12	1		
IKBKG	WP_001267298.1	Ec6_12	1		
HMBOX1	WP_001267298.1	Ec6_12	1		
SP4	WP_001267298.1	Ec6_12	1		
GOLGA6A	WP_001267298.1	Ec6_12	1		
CEP70	WP_001267298.1	Ec6_12	1		
CCNDBP1	WP_001267298.1	Ec6_12	1		
CCDC102B	WP_001267298.1	Ec6_12	1		
C12orf68	WP_001267298.1	Ec6_12	1		
BANP	WP_001267298.1	Ec6_12	1		
APPL2	WP_001267298.1	Ec6_12	1		
ZNF143	WP_001267298.1	Ec6_12	1	1	
VPS52	WP_001267298.1	Ec6_12	1	1	
SCYL3	WP_001267298.1	Ec6_12	1	1	
REL	WP_001267298.1	Ec6_12		1	
HOMEZ	WP_001267298.1	Ec6_12	1		
UBAP1	WP_001267298.1	Ec6_12	1		
SERTAD1	WP_001267298.1	Ec6_12	1		
ZFP161	WP_001267298.1	Ec6_12	1		
ZBTB10	WP_001267298.1	Ec6_12	1		
ZBED1	WP_001267298.1	Ec6_12	1		
SEPTIN3	WP_001267298.1	Ec6_12	1		
UBQLN2	WP_001267298.1	Ec6_12	1		
TSN	WP_001267298.1	Ec6_12	1		
TRIM54	WP_001267298.1	Ec6_12	1		
TFIP11	WP_001267298.1	Ec6_12	1		
TAX1BP1	WP_001267298.1	Ec6_12	1		
TACC1	WP_001267298.1	Ec6_12	1		
VCP	WP_001267298.1	Ec6_12	1		
HSPB8	WP_001327852.1	Ec6_13	1		
PRDM13	WP_001328837.1	Ec6_14	1		
LATS2	WP_001328837.1	Ec6_14	1		
KRT76	WP_001328837.1	Ec6_14	1		
DMD	WP_001328837.1	Ec6_14	1		
KRT75	WP_001328837.1	Ec6_14	1		
SYCE2	WP_001445762.1	Ec6_15	1		
COG6	WP_001445771.1	Ec6_16			1
REL	WP_001445815.1	Ec6_17		1	1
DAZAP2	WP_001445815.1	Ec6_17			1
SH3RF1	WP_001445815.1	Ec6_17			1
TFIP11	WP_001445815.1	Ec6_17			1

human protein	effector	abbr.	MAIN	REPEAT	HOMOLOGY
TP53BP2	WP_001445815.1	Ec6_17			1
VAC14	WP_001445815.1	Ec6_17			1
CCNB1IP1	WP_001445845.1	Ec6_18	1		1
AGTRAP	WP_032140136.1	Ec6_19	1		
ARRDC3	WP_032140136.1	Ec6_19	1		
MAL2	WP_000097400.1	Ec6_2			1
AGTRAP	WP_000097400.1	Ec6_2			1
DYNLT1	WP_077626056.1	Ec6_20	1		1
CLTCL1	WP_000155738.1	Ec6_3	1		1
MFF	WP_000191595.1	Ec6_4	1		1
RHOXF2	WP_000206655.1	Ec6_5	1		
VAC14	WP_000206655.1	Ec6_5	1		
TCL1A	WP_000258580.1	Ec6_6			1
TRIM23	WP_000258580.1	Ec6_6	1		1
PAX5	WP_000258580.1	Ec6_6	1		1
EIF2B1	WP_000258580.1	Ec6_6	1		1
REL	WP_000258580.1	Ec6_6			1
PAX6	WP_000258580.1	Ec6_6			1
KIAA1328	WP_000258580.1	Ec6_6	1		1
LZTFL1	WP_000611426.1	Ec6_7		1	1
PRKAR1B	WP_000804518.1	Ec6_8	1	1	
REL	WP_000961342.1	Ec6_9			1
PROP1	WP_000961342.1	Ec6_9			1
GOLGA6L9	WP_000961342.1	Ec6_9			1
PAX6	WP_000961342.1	Ec6_9			1
PAX5	WP_000961342.1	Ec6_9			1
LHX2	WP_000961342.1	Ec6_9	1		1
VAC14	WP_000067801.1	Efe_1	1		
LRRC73	WP_001182890.1	Efe_10	1		
CLTCL1	WP_001235473.1	Efe_11	1		
KIFC3	WP_001235473.1	Efe_11	1		
KCTD6	WP_001235473.1	Efe_11	1		
IKZF3	WP_001235473.1	Efe_11	1		
IFT88	WP_001235473.1	Efe_11	1		
HOMEZ	WP_001235473.1	Efe_11	1		
COG6	WP_001235473.1	Efe_11	1		
KRT34	WP_001235473.1	Efe_11	1		
CCNDBP1	WP_001235473.1	Efe_11	1		
CARD10	WP_001235473.1	Efe_11	1		
C15orf55	WP_001235473.1	Efe_11	1		
BNC2	WP_001235473.1	Efe_11	1		
KRT31	WP_001235473.1	Efe_11	1		
GOLGA2	WP_001235473.1	Efe_11	1		
USHBP1	WP_001235473.1	Efe_11	1		
SSX2IP	WP_001235473.1	Efe_11	1		
TCF4	WP_001235473.1	Efe_11	1		
TFIP11	WP_001235473.1	Efe_11	1		
TRAF1	WP_001235473.1	Efe_11	1		
TRAF2	WP_001235473.1	Efe_11	1		
TRIM27	WP_001235473.1	Efe_11	1		
TRIM32	WP_001235473.1	Efe_11	1		
KRT27	WP_001235473.1	Efe_11	1		
UBQLN2	WP_001235473.1	Efe_11	1		
SERTAD1	WP_001235473.1	Efe_11	1		
VIM	WP_001235473.1	Efe_11	1		
VPS37B	WP_001235473.1	Efe_11	1		
VPS52	WP_001235473.1	Efe_11	1		
ZBED4	WP_001235473.1	Efe_11	1		
ZBTB10	WP_001235473.1	Efe_11	1		
ZBTB7B	WP_001235473.1	Efe_11	1		
ZRANB1	WP_001235473.1	Efe_11	1		
TRIM54	WP_001235473.1	Efe_11	1		
PNMA1	WP_001235473.1	Efe_11	1		
KRT75	WP_001235473.1	Efe_11	1		

human protein	effector	abbr.	MAIN	REPEAT	HOMOLOGY
LATS2	WP_001235473.1	Efe_11	1		
LDOC1	WP_001235473.1	Efe_11	1		
LZTS2	WP_001235473.1	Efe_11	1		
MID2	WP_001235473.1	Efe_11	1		
PAX5	WP_001235473.1	Efe_11	1		
PAX6	WP_001235473.1	Efe_11	1		
SPAG5	WP_001235473.1	Efe_11	1		
PICK1	WP_001235473.1	Efe_11	1		
SERTAD3	WP_001235473.1	Efe_11	1		
PNMA5	WP_001235473.1	Efe_11	1		
PPP1R13B	WP_001235473.1	Efe_11	1		
PRDM16	WP_001235473.1	Efe_11	1		
PROP1	WP_001235473.1	Efe_11	1		
RBAK	WP_001235473.1	Efe_11	1		
REL	WP_001235473.1	Efe_11	1		
RINT1	WP_001235473.1	Efe_11	1		
GOLGA6L9	WP_001235473.1	Efe_11	1		
PDE4DIP	WP_001235473.1	Efe_11	1		
HOOK2	WP_001235473.1	Efe_11	1		
ZBED1	WP_001237041.1	Efe_12	1		1
CEACAM6	WP_001237041.1	Efe_12	1		1
USHBP1	WP_001237041.1	Efe_12	1		1
REL	WP_001237041.1	Efe_12	1		1
KANK1	WP_001272443.1	Efe_13	1		
PPFIBP1	WP_001272443.1	Efe_13	1		
NUP62	WP_001272443.1	Efe_13	1		
NECAB2	WP_001272443.1	Efe_13	1		
TFIP11	WP_001272443.1	Efe_13	1		
TNIP1	WP_001272443.1	Efe_13	1		
NECAB1	WP_001272443.1	Efe_13	1		
TADA3	WP_001272443.1	Efe_13	1		
NAP1L2	WP_001272443.1	Efe_13	1		
KRT27	WP_001272443.1	Efe_13	1		
GOLGA6L9	WP_001272443.1	Efe_13	1		
GOLGA6A	WP_001272443.1	Efe_13	1		
CEP63	WP_001272443.1	Efe_13	1		
CENPK	WP_001272443.1	Efe_13	1		
MCC	WP_001272443.1	Efe_13	1		
INADL	WP_024256417.1	Efe_14	1		
MPDZ	WP_024256417.1	Efe_14	1		
MPP7	WP_024256417.1	Efe_14	1		
CASK	WP_024256417.1	Efe_14	1		
VAC14	WP_000083435.1	Efe_2	1		1
LBX1	WP_000083435.1	Efe_2			1
TRAF2	WP_000148644.1	Efe_3	1		
REL	WP_000148644.1	Efe_3	1		
CLTCL1	WP_000178797.1	Efe_4	1		1
TRIM27	WP_000255032.1	Efe_5			1
GIGYF1	WP_000255032.1	Efe_5	1		1
TRAF2	WP_000255032.1	Efe_5	1		1
ZBED4	WP_000255032.1	Efe_5	1		1
EFHC2	WP_000255032.1	Efe_5	1		1
JSRP1	WP_000375129.1	Efe_6	1		
ZBTB22	WP_000375129.1	Efe_6	1		
REEP6	WP_000508975.1	Efe_7	1		
UBQLN2	WP_000904613.1	Efe_8	1		
REL	WP_000999547.1	Efe_9	1		
ZBED1	WP_000999547.1	Efe_9	1		
COL17A1	WP_005285281.1	Eta_1	1		
CLTCL1	WP_005288481.1	Eta_2	1		1
UBE4A	WP_005294393.1	Eta_3	1		
EDA	WP_002891634.1	Kpn_1	1		
PIBF1	WP_009484876.1	Kpn_10	1		
TEKT2	WP_009484876.1	Kpn_10	1		

human protein	effector	abbr.	MAIN	REPEAT	HOMOLOGY
FOXJ2	WP_009484876.1	Kpn_10	1	1	
CALCOCO1	WP_009484876.1	Kpn_10	1		
CARD9	WP_009484876.1	Kpn_10	1		
MCC	WP_009484876.1	Kpn_10	1		
MSANTD4	WP_009484876.1	Kpn_10	1		
TAX1BP1	WP_009484876.1	Kpn_10	1		
TBC1D5	WP_009484876.1	Kpn_10	1		
TRIM54	WP_009484876.1	Kpn_10	1		
TSGA10	WP_009484876.1	Kpn_10	1		
MDFI	WP_009486019.1	Kpn_11	1		
KRTAP10-5	WP_009486529.1	Kpn_12	1		1
KRTAP4-5	WP_009486529.1	Kpn_12	1		1
KRTAP5-7	WP_009486529.1	Kpn_12	1		1
BEGAIN	WP_020317218.1	Kpn_13	1	1	1
KIFAP3	WP_032420351.1	Kpn_14	1		
CEP70	WP_032420351.1	Kpn_14	1		
LONRF1	WP_002916607.1	Kpn_2	1		1
TCF4	WP_004118237.1	Kpn_3	1		
SF3B4	WP_004118237.1	Kpn_3	1		
SPAG5	WP_004118237.1	Kpn_3	1		
SPERT	WP_004118237.1	Kpn_3	1		
TRIP6	WP_004118237.1	Kpn_3	1		
SPTA1	WP_004118237.1	Kpn_3	1		
L3MBTL3	WP_004118237.1	Kpn_3	1		
TRAF1	WP_004118237.1	Kpn_3	1		
TAX1BP1	WP_004118237.1	Kpn_3	1		
TRAF2	WP_004118237.1	Kpn_3	1		
TFIP11	WP_004118237.1	Kpn_3	1		
TOX3	WP_004118237.1	Kpn_3	1		
TRA2B	WP_004118237.1	Kpn_3	1		
SERTAD3	WP_004118237.1	Kpn_3	1		
LBX1	WP_004118237.1	Kpn_3	1		
MTUS2	WP_004118237.1	Kpn_3	1		
TACC1	WP_004118237.1	Kpn_3	1		
PICK1	WP_004118237.1	Kpn_3	1		
TSN	WP_004118237.1	Kpn_3	1		
TRAF3	WP_004118237.1	Kpn_3	1		
NECAB2	WP_004118237.1	Kpn_3	1		
N4BP2	WP_004118237.1	Kpn_3	1		
PAX5	WP_004118237.1	Kpn_3	1		
PAX6	WP_004118237.1	Kpn_3	1		
MID2	WP_004118237.1	Kpn_3	1		
PIBF1	WP_004118237.1	Kpn_3	1		
SERTAD1	WP_004118237.1	Kpn_3	1		
PNMA1	WP_004118237.1	Kpn_3	1		
PNMA2	WP_004118237.1	Kpn_3	1		
MRFAP1L1	WP_004118237.1	Kpn_3	1		
LCE2C	WP_004118237.1	Kpn_3	1		
PRDM6	WP_004118237.1	Kpn_3	1		
PSTPIP1	WP_004118237.1	Kpn_3	1		
RFX6	WP_004118237.1	Kpn_3	1		
TRIM27	WP_004118237.1	Kpn_3	1		
KRT27	WP_004118237.1	Kpn_3	1		
EFHC2	WP_004118237.1	Kpn_3	1		
COG6	WP_004118237.1	Kpn_3	1		
UBQLN2	WP_004118237.1	Kpn_3	1		
COL17A1	WP_004118237.1	Kpn_3	1		
DNM2	WP_004118237.1	Kpn_3	1		
KRTAP5-7	WP_004118237.1	Kpn_3	1		
KRTAP5-1	WP_004118237.1	Kpn_3	1		
KRT76	WP_004118237.1	Kpn_3	1		
VPS52	WP_004118237.1	Kpn_3	1	1	
KRT31	WP_004118237.1	Kpn_3	1		
AKAP9	WP_004118237.1	Kpn_3	1		

human protein	effector	abbr.	MAIN	REPEAT	HOMOLOGY
KLHL2	WP_004118237.1	Kpn_3	1		
KIFC3	WP_004118237.1	Kpn_3	1		
KCTD9	WP_004118237.1	Kpn_3	1		
KCTD6	WP_004118237.1	Kpn_3	1		
HIP1	WP_004118237.1	Kpn_3	1		
GOLGA6L9	WP_004118237.1	Kpn_3	1		
GOLGA2	WP_004118237.1	Kpn_3	1		
FSD2	WP_004118237.1	Kpn_3	1		
KRT75	WP_004118237.1	Kpn_3	1		
USHBP1	WP_004118237.1	Kpn_3	1		
ZBED1	WP_004118237.1	Kpn_3	1		
ZBTB10	WP_004118237.1	Kpn_3	1		
ZBTB7B	WP_004118237.1	Kpn_3	1		
ZC4H2	WP_004118237.1	Kpn_3	1		
ZRANB1	WP_004118237.1	Kpn_3	1	1	
VPS37B	WP_004118237.1	Kpn_3	1		
CDR2	WP_004118237.1	Kpn_3		1	
ASPG	WP_004118237.1	Kpn_3	1		
THAP1	WP_004118237.1	Kpn_3		1	
ARL6IP1	WP_004118237.1	Kpn_3	1		
REL	WP_004118237.1	Kpn_3	1	1	
CADPS	WP_004118237.1	Kpn_3	1		
CCDC102B	WP_004118237.1	Kpn_3	1		
CIT	WP_004118237.1	Kpn_3	1		
VIM	WP_004118237.1	Kpn_3	1		
MFF	WP_004145486.1	Kpn_4	1		1
MAGEA8	WP_004149975.1	Kpn_5	1		
HSF2BP	WP_004152718.1	Kpn_6	1		
UBQLN1	WP_004177339.1	Kpn_7	1		1
CIB1	WP_004197606.1	Kpn_8	1		1
BIRC7	WP_009484324.1	Kpn_9	1		
ZNF143	WP_009484324.1	Kpn_9	1		
CDR2	WP_009484324.1	Kpn_9	1	1	
BANP	WP_009484324.1	Kpn_9	1		
ANKRD28	WP_009484324.1	Kpn_9	1		
VPS52	WP_009484324.1	Kpn_9	1	1	
THAP1	WP_009484324.1	Kpn_9		1	
ZRANB1	WP_009484324.1	Kpn_9	1	1	
ZNF398	WP_009484324.1	Kpn_9	1		
GOLGA2	WP_009484324.1	Kpn_9	1		
MDFI	WP_009484324.1	Kpn_9	1		
LZTS2	WP_009484324.1	Kpn_9	1		
LHX3	WP_009484324.1	Kpn_9	1		
KRTAP10-9	WP_009484324.1	Kpn_9	1		
KRTAP10-5	WP_009484324.1	Kpn_9	1		
KRT75	WP_009484324.1	Kpn_9	1		
KRT31	WP_009484324.1	Kpn_9	1		
KRT27	WP_009484324.1	Kpn_9	1		
KIFC3	WP_009484324.1	Kpn_9	1		
KCTD6	WP_009484324.1	Kpn_9	1		
IKZF3	WP_009484324.1	Kpn_9	1		
HSF2BP	WP_009484324.1	Kpn_9	1		
MID2	WP_009484324.1	Kpn_9	1		
HMBOX1	WP_009484324.1	Kpn_9	1		
GOLGA6L9	WP_009484324.1	Kpn_9	1		
REL	WP_009484324.1	Kpn_9	1	1	
EVI5	WP_009484324.1	Kpn_9	1		
EHMT2	WP_009484324.1	Kpn_9	1		
COG6	WP_009484324.1	Kpn_9	1		
CEP70	WP_009484324.1	Kpn_9	1		
CEP250	WP_009484324.1	Kpn_9	1		
CDR2L	WP_009484324.1	Kpn_9	1		
CCNDBP1	WP_009484324.1	Kpn_9	1		
CCDC136	WP_009484324.1	Kpn_9	1		

human protein	effector	abbr.	MAIN	REPEAT	HOMOLOGY
CCDC125	WP_009484324.1	Kpn_9	1		
CCDC102B	WP_009484324.1	Kpn_9	1		
CARD10	WP_009484324.1	Kpn_9	1		
C15orf55	WP_009484324.1	Kpn_9	1		
HOMEZ	WP_009484324.1	Kpn_9	1		
TCF4	WP_009484324.1	Kpn_9	1		
FSD2	WP_009484324.1	Kpn_9	1		
MSANTD4	WP_009484324.1	Kpn_9	1		
PNMA2	WP_009484324.1	Kpn_9	1		
PICK1	WP_009484324.1	Kpn_9	1		
PNMA5	WP_009484324.1	Kpn_9	1		
PPP1R13B	WP_009484324.1	Kpn_9	1		
RAB3IP	WP_009484324.1	Kpn_9	1		
ROPN1	WP_009484324.1	Kpn_9	1		
SERTAD1	WP_009484324.1	Kpn_9	1		
SPAG5	WP_009484324.1	Kpn_9	1		
SSX2IP	WP_009484324.1	Kpn_9	1		
TFIP11	WP_009484324.1	Kpn_9	1		
TP53BP2	WP_009484324.1	Kpn_9	1		
PNMA1	WP_009484324.1	Kpn_9	1		
PDE4DIP	WP_009484324.1	Kpn_9	1		
MTUS2	WP_009484324.1	Kpn_9	1		
NECAB1	WP_009484324.1	Kpn_9	1		
SPTA1	WP_009484324.1	Kpn_9	1		
NOTO	WP_009484324.1	Kpn_9	1		
TRAF1	WP_009484324.1	Kpn_9	1		
PIBF1	WP_009484324.1	Kpn_9	1		
ZBTB8A	WP_009484324.1	Kpn_9	1		
TRIM27	WP_009484324.1	Kpn_9	1		
NECAB2	WP_009484324.1	Kpn_9	1		
TRAF2	WP_009484324.1	Kpn_9	1		
ZBTB7B	WP_009484324.1	Kpn_9	1		
USHBP1	WP_009484324.1	Kpn_9	1		
VIM	WP_009484324.1	Kpn_9	1		
ZBTB10	WP_009484324.1	Kpn_9	1		
ZNF639	dime_meta_effector_140	met_10	1		
KRTAP1-3	dime_meta_effector_140	met_10	1		
MID2	dime_meta_effector_140	met_10	1		
KRTAP5-8	dime_meta_effector_140	met_10	1		
KRTAP5-7	dime_meta_effector_140	met_10	1		
KRTAP5-2	dime_meta_effector_140	met_10	1		
KRTAP5-1	dime_meta_effector_140	met_10	1		
SFN	dime_meta_effector_140	met_10	1		
KRTAP10-6	dime_meta_effector_140	met_10	1		
KRTAP10-5	dime_meta_effector_140	met_10	1		
KCTD6	dime_meta_effector_140	met_10	1		
HOMEZ	dime_meta_effector_140	met_10	1		
GADD45G	dime_meta_effector_140	met_10	1		
ATP6V0D1	dime_meta_effector_140	met_10	1		
KRTAP1-1	dime_meta_effector_140	met_10	1		
PAX6	dime_meta_effector_141	met_11			1
LRRC6	dime_meta_effector_144	met_12	1		
PAX6	dime_meta_effector_145	met_13			1
PROP1	dime_meta_effector_145	met_13			1
GOLGA6L9	dime_meta_effector_145	met_13			1
REL	dime_meta_effector_145	met_13			1
LBX1	dime_meta_effector_147	met_14			1
NOTO	dime_meta_effector_147	met_14			1
RAB3IP	dime_meta_effector_147	met_14			1
REL	dime_meta_effector_147	met_14			1
SERTAD1	dime_meta_effector_147	met_14			1
TCF4	dime_meta_effector_147	met_14			1
WASF1	dime_meta_effector_150	met_15	1		
TRIM37	dime_meta_effector_150	met_15	1		

human protein	effector	abbr.	MAIN	REPEAT	HOMOLOGY
TFIP11	dime_meta_effector_150	met_15	1		
CCDC102B	dime_meta_effector_150	met_15	1		
PNMA1	dime_meta_effector_150	met_15	1		
ALAS1	dime_meta_effector_150	met_15	1		
LBX1	dime_meta_effector_150	met_15	1		
COG6	dime_meta_effector_150	met_15	1		
RBAK	dime_meta_effector_157	met_16			1
ZNF263	dime_meta_effector_157	met_16			1
ZNF143	dime_meta_effector_157	met_16			1
ZBTB34	dime_meta_effector_157	met_16			1
ZBTB26	dime_meta_effector_157	met_16			1
TRAF2	dime_meta_effector_157	met_16			1
TFIP11	dime_meta_effector_157	met_16			1
SP4	dime_meta_effector_157	met_16			1
SH3RF1	dime_meta_effector_157	met_16			1
SFN	dime_meta_effector_157	met_16			1
SERTAD2	dime_meta_effector_157	met_16			1
REL	dime_meta_effector_157	met_16			1
RAB3IP	dime_meta_effector_157	met_16			1
PAX6	dime_meta_effector_157	met_16			1
NIF3L1	dime_meta_effector_157	met_16			1
LBX1	dime_meta_effector_157	met_16			1
KIFC3	dime_meta_effector_157	met_16			1
ETV6	dime_meta_effector_157	met_16			1
EFHC2	dime_meta_effector_157	met_16			1
CCDC102B	dime_meta_effector_157	met_16			1
BHLHB9	dime_meta_effector_157	met_16			1
SERTAD1	dime_meta_effector_157	met_16			1
SERTAD1	dime_meta_effector_159	met_17			1
REL	dime_meta_effector_159	met_17			1
RAB3IP	dime_meta_effector_159	met_17	1		1
LBX1	dime_meta_effector_159	met_17			1
TCF4	dime_meta_effector_159	met_17			1
GOLGA6L9	dime_meta_effector_15	met_18			1
PROP1	dime_meta_effector_15	met_18			1
RNF41	dime_meta_effector_169	met_19	1		
GOLGA6L9	dime_meta_effector_170	met_20	1		
OPLAH	dime_meta_effector_170	met_20	1		
IKZF3	dime_meta_effector_175	met_21	1		
HSF2BP	dime_meta_effector_183	met_22	1		
RNF41	dime_meta_effector_183	met_22	1		
TEX11	dime_meta_effector_20	met_23			1
PM20D2	dime_meta_effector_20	met_23			1
TP53BP2	dime_meta_effector_24	met_24			1
SERTAD2	dime_meta_effector_24	met_24			1
SP4	dime_meta_effector_24	met_24			1
ZNF451	dime_meta_effector_24	met_24			1
SYCE1	dime_meta_effector_24	met_24			1
TCF4	dime_meta_effector_24	met_24			1
SERTAD1	dime_meta_effector_24	met_24			1
TFIP11	dime_meta_effector_24	met_24			1
TRAF2	dime_meta_effector_24	met_24			1
VAC14	dime_meta_effector_24	met_24			1
VPS52	dime_meta_effector_24	met_24			1
CCDC102B	dime_meta_effector_24	met_24	1		1
ZBTB10	dime_meta_effector_24	met_24			1
IKZF3	dime_meta_effector_24	met_24			1
RPS18	dime_meta_effector_24	met_24			1
ZBTB1	dime_meta_effector_24	met_24			1
KRT75	dime_meta_effector_24	met_24			1
KIFC3	dime_meta_effector_24	met_24			1
REL	dime_meta_effector_24	met_24			1
CEP250	dime_meta_effector_24	met_24			1
EFHC2	dime_meta_effector_24	met_24			1

human protein	effector	abbr.	MAIN	REPEAT	HOMOLOGY
KCTD13	dime_meta_effector_24	met_24			1
KRT76	dime_meta_effector_24	met_24			1
LBX1	dime_meta_effector_24	met_24			1
LRRC6	dime_meta_effector_24	met_24			1
MAP3K5	dime_meta_effector_24	met_24			1
NEDD4	dime_meta_effector_24	met_24			1
NOTO	dime_meta_effector_24	met_24			1
PAX5	dime_meta_effector_24	met_24			1
PAX6	dime_meta_effector_24	met_24			1
RBAK	dime_meta_effector_24	met_24			1
ETV6	dime_meta_effector_24	met_24			1
CEP76	dime_meta_effector_28	met_25	1		
LONRF1	dime_meta_effector_28	met_25	1		
USP54	dime_meta_effector_28	met_25	1		
REL	dime_meta_effector_28	met_25	1		
NEBL	dime_meta_effector_28	met_25	1		
NAB2	dime_meta_effector_28	met_25	1		
KRT76	dime_meta_effector_28	met_25	1		
KRT75	dime_meta_effector_28	met_25	1		
KIAA1377	dime_meta_effector_28	met_25	1		
KCTD9	dime_meta_effector_28	met_25	1		
EFHC2	dime_meta_effector_28	met_25	1		
DLG2	dime_meta_effector_32	met_26	1		
KEAP1	dime_meta_effector_32	met_26	1		
MAGI1	dime_meta_effector_32	met_26	1		
MAGI2	dime_meta_effector_32	met_26	1		
SEC24C	dime_meta_effector_32	met_26	1		
PDZK1	dime_meta_effector_32	met_26	1		
MPDZ	dime_meta_effector_32	met_26	1		
PAX5	dime_meta_effector_33	met_27	1		
MED1	dime_meta_effector_35	met_28			1
RAB3IL1	dime_meta_effector_35	met_28			1
RAB3IP	dime_meta_effector_35	met_28	1		1
PROP1	dime_meta_effector_37	met_29			1
CLTCL1	dime_meta_effector_37	met_29			1
PAX6	dime_meta_effector_37	met_29			1
REL	dime_meta_effector_37	met_29			1
TCL1A	dime_meta_effector_37	met_29			1
ZBTB10	dime_meta_effector_37	met_29			1
PAX5	dime_meta_effector_37	met_29			1
LZTS2	dime_meta_effector_114	met_3	1		
KIFC3	dime_meta_effector_114	met_3	1		
TRIP4	dime_meta_effector_114	met_3	1		
FHL3	dime_meta_effector_38	met_30	1		
SHANK2	dime_meta_effector_39	met_31	1		
PICK1	dime_meta_effector_41	met_32	1		
LRRC6	dime_meta_effector_51	met_33	1		
KIFC3	dime_meta_effector_56	met_34	1		
TOX3	dime_meta_effector_56	met_34	1		
GOLGA2	dime_meta_effector_56	met_34	1		
TRAF1	dime_meta_effector_57	met_35	1		
PER2	dime_meta_effector_65	met_36			1
MAL2	dime_meta_effector_65	met_36			1
AGTRAP	dime_meta_effector_65	met_36			1
VAC14	dime_meta_effector_66	met_37			1
MAGEA12	dime_meta_effector_68	met_38	1		
RAB3IL1	dime_meta_effector_68	met_38	1		
PROP1	dime_meta_effector_73	met_39			1
USHBP1	dime_meta_effector_118	met_4			1
EXOSC8	dime_meta_effector_118	met_4			1
PPP1R13B	dime_meta_effector_118	met_4			1
TAX1BP1	dime_meta_effector_118	met_4			1
SYCE1	dime_meta_effector_118	met_4			1
SERTAD2	dime_meta_effector_118	met_4			1

human protein	effector	abbr.	MAIN	REPEAT	HOMOLOGY
SEC23IP	dime_meta_effector_118	met_4			1
PROP1	dime_meta_effector_118	met_4			1
NOTO	dime_meta_effector_118	met_4			1
NEDD4	dime_meta_effector_118	met_4			1
MID2	dime_meta_effector_118	met_4			1
CCHCR1	dime_meta_effector_118	met_4			1
KRT75	dime_meta_effector_118	met_4			1
KIAA1715	dime_meta_effector_118	met_4	1		1
E2F4	dime_meta_effector_118	met_4			1
USP54	dime_meta_effector_118	met_4			1
MID1	dime_meta_effector_118	met_4			1
HOMEZ	dime_meta_effector_77	met_40	1		
SGTA	dime_meta_effector_7	met_41	1		
KCNIP1	dime_meta_effector_7	met_41	1		
COL17A1	dime_meta_effector_7	met_41	1		
SGTB	dime_meta_effector_7	met_41	1		
UBQLN2	dime_meta_effector_7	met_41	1		
RAB3IP	dime_meta_effector_86	met_42			1
ZBTB26	dime_meta_effector_86	met_42			1
VAC14	dime_meta_effector_89	met_43			1
REL	dime_meta_effector_93	met_44			1
DAZAP2	dime_meta_effector_93	met_44			1
VAC14	dime_meta_effector_93	met_44			1
TP53BP2	dime_meta_effector_93	met_44			1
TFIP11	dime_meta_effector_93	met_44			1
SH3RF1	dime_meta_effector_93	met_44			1
REL	dime_meta_effector_95	met_45			1
PKNOX2	dime_meta_effector_97	met_46	1		
DLG3	dime_meta_effector_97	met_46	1		
PKNOX1	dime_meta_effector_97	met_46	1		
DLG2	dime_meta_effector_97	met_46	1		
REL	dime_meta_effector_98	met_47	1		
RAB3IP	dime_meta_effector_9	met_48			1
ZMYND12	dime_meta_effector_9	met_48			1
SEC23IP	dime_meta_effector_119	met_5			1
EPN2	dime_meta_effector_129	met_6			1
AGR2	dime_meta_effector_129	met_6			1
TRIM27	dime_meta_effector_130	met_7			1
TRAF2	dime_meta_effector_130	met_7			1
MID1	dime_meta_effector_130	met_7			1
COG6	dime_meta_effector_130	met_7			1
C11orf74	dime_meta_effector_131	met_8	1		
BAAT	dime_meta_effector_131	met_8	1		
CUTC	dime_meta_effector_132	met_9	1		
UBQLN1	dime_meta_effector_132	met_9	1		
MFF	WP_004234458.1	Mmo_1	1		
KLHL12	WP_062771682.1	Mmo_10	1		1
REL	WP_062771682.1	Mmo_10	1		1
TRAF3	WP_062772817.1	Mmo_11	1		
CEP250	WP_062772817.1	Mmo_11	1		
REL	WP_062773522.1	Mmo_12	1		
ZFP161	WP_062773522.1	Mmo_12	1		
CCDC125	WP_062773522.1	Mmo_12	1		
ZBTB33	WP_062773522.1	Mmo_12	1		
ZNF326	WP_062773522.1	Mmo_12	1		
WAC	WP_062773522.1	Mmo_12	1		
VIM	WP_062773522.1	Mmo_12	1		
TBC1D5	WP_062773522.1	Mmo_12	1		
SFMBT1	WP_062773522.1	Mmo_12	1		
RHOXF2	WP_062773522.1	Mmo_12	1		
PAX6	WP_062773522.1	Mmo_12	1		
NOTO	WP_062773522.1	Mmo_12	1		
NFIX	WP_062773522.1	Mmo_12	1		
KRT27	WP_062773522.1	Mmo_12	1		

human protein	effector	abbr.	MAIN	REPEAT	HOMOLOGY
KIFC3	WP_062773522.1	Mmo_12	1		
KIAA1328	WP_062773522.1	Mmo_12	1		
KCTD10	WP_062773522.1	Mmo_12	1		
CRTC2	WP_062773522.1	Mmo_12	1		
AMOT	WP_062773522.1	Mmo_12	1		
ZNF446	WP_062773522.1	Mmo_12	1		
INTS10	WP_062773522.1	Mmo_12	1		
EVI5	WP_062773651.1	Mmo_13	1		
MAGI2	WP_004235425.1	Mmo_2	1		
STX2	WP_004236571.1	Mmo_3	1		
PNMA1	WP_004238406.1	Mmo_4	1		
NFE2L1	WP_004240712.1	Mmo_5	1		
VPS53	WP_004240712.1	Mmo_5	1		
VPS52	WP_004240712.1	Mmo_5	1		
TRIM38	WP_004240712.1	Mmo_5	1		
TPD52L1	WP_004240712.1	Mmo_5	1		
TEX9	WP_004240712.1	Mmo_5	1		
TBC1D25	WP_004240712.1	Mmo_5	1		
SMARCE1	WP_004240712.1	Mmo_5	1		
PML	WP_004240712.1	Mmo_5	1		
ARID3A	WP_004240712.1	Mmo_5	1		
PCYT1A	WP_004240712.1	Mmo_5	1		
NEBL	WP_004240712.1	Mmo_5	1		
KIAA1958	WP_004240712.1	Mmo_5	1		
HINT2	WP_004240712.1	Mmo_5	1		
EVI5	WP_004240712.1	Mmo_5	1		
CEP76	WP_004240712.1	Mmo_5	1		
CDR2	WP_004240712.1	Mmo_5	1		
C17orf59	WP_004240712.1	Mmo_5	1		
ATPIF1	WP_004240712.1	Mmo_5	1		
PKNOX2	WP_004240712.1	Mmo_5	1		
DYNLT1	WP_032098021.1	Mmo_6	1		
ZBED1	WP_036413302.1	Mmo_7	1		1
SEC16B	WP_036413302.1	Mmo_7	1		1
RAI1	WP_036413302.1	Mmo_7	1		1
BEGAIN	WP_036413302.1	Mmo_7	1	1	1
RHOXF2	WP_036417499.1	Mmo_8	1		
REL	WP_036417499.1	Mmo_8	1		
FES	WP_046024762.1	Mmo_9	1		
ZNF699	WP_040259375.1	Pem_1	1		
FAM9B	WP_040259375.1	Pem_1	1		
VPS52	WP_040260715.1	Pem_2	1		
TCF4	WP_040260715.1	Pem_2	1		
REL	WP_040260715.1	Pem_2	1		
NOTO	WP_040260715.1	Pem_2	1		
LBX1	WP_040260715.1	Pem_2	1		
GOLGA6L9	WP_040260715.1	Pem_2	1		
PAX5	WP_040260715.1	Pem_2	1		
UBQLN2	WP_040263025.1	Pem_3	1		
APPBP2	WP_040263420.1	Pem_4	1		
CREB3L1	WP_040263598.1	Pem_5	1		
CLEC17A	WP_040263598.1	Pem_5	1		
ZBED1	WP_084596144.1	Pem_6	1		
SPTA1	WP_084596144.1	Pem_6	1		
APPL2	WP_084596144.1	Pem_6	1		
EDA	WP_084596156.1	Pem_7	1		
KRT31	WP_084596184.1	Pem_8	1		
NEDD9	WP_084596184.1	Pem_8	1		
REL	WP_084596184.1	Pem_8	1		
TCF4	WP_084596184.1	Pem_8	1		
TRAF2	WP_084596184.1	Pem_8	1		
FXR2	WP_084596184.1	Pem_8	1		
AGR2	WP_000116680.1	Pfa_1	1		1
KRTAP10-5	WP_108474309.1	Pfa_10	1		

human protein	effector	abbr.	MAIN	REPEAT	HOMOLOGY
KRTAP5-7	WP_108474309.1	Pfa_10	1		
KRTAP10-6	WP_108474309.1	Pfa_10	1		
GOLGA6L9	WP_108474309.1	Pfa_10	1		
CCDC57	WP_108474309.1	Pfa_10	1		
NOTO	WP_113857302.1	Pfa_11	1		1
CLTCL1	WP_113857302.1	Pfa_11			1
PSMC6	WP_113857302.1	Pfa_11	1		1
MYOG	WP_113857302.1	Pfa_11			1
GOLGA2	WP_113857302.1	Pfa_11	1		1
PNMA1	WP_113857302.1	Pfa_11	1		1
REL	WP_113857302.1	Pfa_11	1		1
LBX1	WP_113857302.1	Pfa_11			1
KIFC3	WP_113857302.1	Pfa_11			1
GOLGA6L9	WP_113857302.1	Pfa_11	1		1
TCF4	WP_113857302.1	Pfa_11	1		1
PNMA1	WP_113857471.1	Pfa_12	1		
GOLGA2	WP_113857569.1	Pfa_13	1		1
CLTCL1	WP_113857569.1	Pfa_13	1		1
KIFC3	WP_113857569.1	Pfa_13	1		1
MYOG	WP_113857569.1	Pfa_13	1		1
NOTO	WP_113857569.1	Pfa_13	1		1
PNMA1	WP_113857569.1	Pfa_13	1		1
PROP1	WP_113857569.1	Pfa_13	1		1
REL	WP_113857569.1	Pfa_13	1		1
TCF4	WP_113857569.1	Pfa_13			1
GOLGA6L9	WP_113857569.1	Pfa_13			1
PSMC6	WP_113857569.1	Pfa_13			1
LBX1	WP_113857569.1	Pfa_13	1		1
REL	WP_113857629.1	Pfa_14	1		
TRIM27	WP_113857629.1	Pfa_14	1		
FAM131C	WP_113858483.1	Pfa_15	1		1
GOLGA6L9	WP_113858483.1	Pfa_15	1		1
REL	WP_113858483.1	Pfa_15	1		1
L3MBTL3	WP_113858483.1	Pfa_15	1		1
MRFAP1L1	WP_113858620.1	Pfa_16	1		1
LBX1	WP_113859044.1	Pfa_17	1		1
PAIP1	WP_113859080.1	Pfa_18	1		
ZRANB1	WP_113859080.1	Pfa_18	1		
SERTAD1	WP_000703842.1	Pfa_2	1		
PROP1	WP_001516695.1	Pfa_3			1
ZNF263	WP_001516695.1	Pfa_3			1
ZNF398	WP_001516695.1	Pfa_3			1
RBAK	WP_001516695.1	Pfa_3			1
TFIP11	WP_001531161.1	Pfa_4	1		
ZEB1	WP_001531161.1	Pfa_4	1		
NOTO	WP_001531161.1	Pfa_4	1		
LBX1	WP_001531161.1	Pfa_4	1		
REL	WP_001531161.1	Pfa_4	1		
CAP1	WP_015962672.1	Pfa_5	1		1
TRAF2	WP_015962672.1	Pfa_5	1		1
MFF	WP_015963250.1	Pfa_6	1		1
MARCO	WP_015963250.1	Pfa_6	1		1
INADL	WP_108473469.1	Pfa_7	1		1
ACCS	WP_108473773.1	Pfa_8	1		
ROPN1	WP_108473781.1	Pfa_9	1		1
TCF4	WP_108473781.1	Pfa_9			1
OTX1	WP_108473781.1	Pfa_9			1
INADL	WP_044172624.1	Pma_1	1	1	1
MFF	WP_044180332.1	Pma_10	1		1
MARCO	WP_044180332.1	Pma_10	1		1
LHX3	WP_044180423.1	Pma_11	1		
NECAB1	WP_044180423.1	Pma_11	1		
MAGEB4	WP_044180429.1	Pma_12	1		
PSTPIP1	WP_044180562.1	Pma_13	1		

human protein	effector	abbr.	MAIN	REPEAT	HOMOLOGY
TAF7L	WP_044183152.1	Pma_14	1		1
MFF	WP_044183301.1	Pma_15	1		1
KIFC3	WP_044183672.1	Pma_16	1		1
CAP1	WP_044183672.1	Pma_16			1
PPCDC	WP_071825927.1	Pma_17	1		
DDX17	WP_071825927.1	Pma_17	1		
ZBTB1	WP_071825927.1	Pma_17	1		
CBY1	WP_071825927.1	Pma_17	1		
BNC2	WP_081653590.1	Pma_18	1		
ZNF408	WP_081653590.1	Pma_18	1		
TRIM27	WP_044173012.1	Pma_2			1
ZBED4	WP_044173012.1	Pma_2	1		1
TRAF2	WP_044173012.1	Pma_2	1		1
EFHC2	WP_044173012.1	Pma_2	1		1
GIGYF1	WP_044173012.1	Pma_2	1		1
MAL2	WP_044173054.1	Pma_3			1
AGTRAP	WP_044173054.1	Pma_3			1
PER2	WP_044173054.1	Pma_3	1		1
KRT76	WP_044174146.1	Pma_4	1		
ALAS1	WP_044174146.1	Pma_4	1		
C17orf28	WP_044174146.1	Pma_4	1		
CCHCR1	WP_044174146.1	Pma_4	1		
DCP1B	WP_044174146.1	Pma_4	1		
REL	WP_044174146.1	Pma_4	1	1	
KRT75	WP_044174146.1	Pma_4	1		
PAX5	WP_044174146.1	Pma_4	1		
ROPN1	WP_044174146.1	Pma_4	1		
TCF4	WP_044174146.1	Pma_4	1		
GSTA4	WP_044174146.1	Pma_4	1		
CCNDBP1	WP_044176371.1	Pma_5	1		
FAM9B	WP_044176371.1	Pma_5	1		
UBQLN2	WP_044177448.1	Pma_6	1		
MAGEA6	WP_044177605.1	Pma_7			1
CIB1	WP_044178555.1	Pma_8	1		1
REL	WP_044180054.1	Pma_9	1	1	1
FSD2	WP_004253606.1	Pre_1	1		
GOLGA6L9	WP_004253606.1	Pre_1	1		
KRT10	WP_004253606.1	Pre_1	1		
KRT27	WP_004253606.1	Pre_1	1		
MTUS2	WP_004253606.1	Pre_1	1		
NUP62	WP_004253606.1	Pre_1	1		
SERTAD1	WP_004253606.1	Pre_1	1		
SPTA1	WP_004253606.1	Pre_1	1		
TFIP11	WP_004253606.1	Pre_1	1		
TRAF1	WP_004253606.1	Pre_1	1		
MTUS2	WP_004258336.1	Pre_10	1		1
VCP	WP_004258336.1	Pre_10	1		1
C12orf68	WP_004258949.1	Pre_11	1		1
UBQLN2	WP_004258949.1	Pre_11	1		1
CCDC102B	WP_004258949.1	Pre_11	1		1
XIRP1	WP_004261326.1	Pre_12	1		
ZBED1	WP_004261459.1	Pre_13	1		
CENPH	WP_004261604.1	Pre_14	1		
ZBED1	WP_004261691.1	Pre_15	1		1
TSN	WP_004261691.1	Pre_15	1		1
SERTAD1	WP_004261691.1	Pre_15	1		1
RPIA	WP_004261691.1	Pre_15	1		1
COG6	WP_004261691.1	Pre_15	1		1
LPIN2	WP_004261691.1	Pre_15	1		1
ATP5B	WP_004262673.1	Pre_16	1		
ZBTB10	WP_004262673.1	Pre_16	1		
REL	WP_004262673.1	Pre_16	1		
CCDC158	WP_004262673.1	Pre_16	1		
BNIP2	WP_004262673.1	Pre_16	1		

human protein	effector	abbr.	MAIN	REPEAT	HOMOLOGY
SDCBP	WP_004263067.1	Pre_17	1		
INADL	WP_004263067.1	Pre_17	1		
SPATA18	WP_004264858.1	Pre_18	1		
BNC2	WP_004264902.1	Pre_19	1		
PAX5	WP_004264902.1	Pre_19	1		
APBB1IP	WP_004253752.1	Pre_2	1		
RFX6	WP_004253752.1	Pre_2	1		
USH1C	WP_004253752.1	Pre_2	1		
UBQLN1	WP_004264927.1	Pre_20	1		1
AMOT	WP_004264927.1	Pre_20	1		1
SGTB	WP_004264927.1	Pre_20	1		1
UBQLN2	WP_004264927.1	Pre_20	1		1
SGTA	WP_004264927.1	Pre_20	1		1
XRCC4	WP_004905473.1	Pre_21	1		
GOLGA6L9	WP_004905473.1	Pre_21	1		
EMILIN1	WP_004905473.1	Pre_21	1		
DPPA4	WP_004905473.1	Pre_21	1		
DCTN2	WP_004905473.1	Pre_21	1		
TEX13A	WP_004905473.1	Pre_21	1		
CREBZF	WP_004905473.1	Pre_21	1		
PKNOX1	WP_004905473.1	Pre_21	1		
PPP1R13B	WP_004905473.1	Pre_21	1		
RALYL	WP_004905473.1	Pre_21	1		
REL	WP_004905473.1	Pre_21	1		
PAX6	WP_004905473.1	Pre_21	1		
SPAG5	WP_004905473.1	Pre_21	1		
USHBP1	WP_004905473.1	Pre_21	1		
TRIM3	WP_004905473.1	Pre_21	1		
TRIM9	WP_004905473.1	Pre_21	1		
PCYT1A	WP_004905473.1	Pre_21	1		
KIFC3	WP_004905473.1	Pre_21	1		
KRT27	WP_004905473.1	Pre_21	1		
NECAB2	WP_004905473.1	Pre_21	1		
RNF135	WP_004905473.1	Pre_21	1		
CENPK	WP_004905473.1	Pre_21	1		
CCHCR1	WP_004905473.1	Pre_21	1		
CAGE1	WP_004905473.1	Pre_21	1		
BCKDK	WP_004905473.1	Pre_21	1		
LZTFL1	WP_004905473.1	Pre_21	1		
CEP250	WP_004905473.1	Pre_21	1		
CEP44	WP_004906048.1	Pre_22	1		1
BACH2	WP_004906048.1	Pre_22	1		1
KRTAP10-6	WP_004906048.1	Pre_22			1
PRKAR1B	WP_004912645.1	Pre_23	1		1
BANP	WP_036957920.1	Pre_24	1		
CXorf41	WP_036957920.1	Pre_24	1		
PDCL2	WP_004254983.1	Pre_3	1		1
UBQLN2	WP_004255002.1	Pre_4			1
REL	WP_004255132.1	Pre_5	1		
CCR9	WP_004255132.1	Pre_5	1		
VAC14	WP_004256127.1	Pre_6	1		
GADD45G	WP_004256756.1	Pre_7	1		
PAX6	WP_004257109.1	Pre_8	1		
REL	WP_004257971.1	Pre_9	1		
KRT75	WP_004257971.1	Pre_9	1		
SERTAD1	WP_004257971.1	Pre_9	1		
SDPR	WP_004257971.1	Pre_9	1		
PPP1R13B	WP_004257971.1	Pre_9	1		
PAX6	WP_004257971.1	Pre_9	1		
LDB2	WP_004257971.1	Pre_9	1		
CASK	WP_004257971.1	Pre_9	1		
ZBTB10	WP_004257971.1	Pre_9	1		
CIB1	WP_033751802.1	Pse_1	1		1
KRT31	WP_033791692.1	Pse_2	1		1

human protein	effector	abbr.	MAIN	REPEAT	HOMOLOGY
NOTO	WP_033791692.1	Pse_2	1		1
USP54	WP_033791692.1	Pse_2	1		1
TCF4	WP_033791692.1	Pse_2	1		1
RIMBP3	WP_033791692.1	Pse_2	1		1
RAB3IP	WP_033791692.1	Pse_2	1		1
POF1B	WP_033791692.1	Pse_2	1		1
PAX5	WP_033791692.1	Pse_2	1		1
OTX1	WP_033791692.1	Pse_2	1		1
KRTAP10-9	WP_033791692.1	Pse_2	1		1
KRTAP10-5	WP_033791692.1	Pse_2	1		1
KRT38	WP_033791692.1	Pse_2	1		1
GLRX3	WP_033791692.1	Pse_2	1		1
APBB1IP	WP_033791692.1	Pse_2	1		1
REL	WP_033791692.1	Pse_2	1	1	1
KRTAP1-3	WP_033791692.1	Pse_2	1		1
KRT39	WP_033791692.1	Pse_2	1		1
KCNIP2	WP_033792202.1	Pse_3		1	
PAX5	WP_033792202.1	Pse_3	1		
KLHL2	WP_033792202.1	Pse_3	1		
CIB1	WP_033792202.1	Pse_3	1		
CASK	WP_033792202.1	Pse_3	1		
REL	WP_033792202.1	Pse_3		1	
PUF60	WP_033792699.1	Pse_4	1		
DISC1	WP_004915569.1	Pst_1	1		1
FOXP2	WP_004915569.1	Pst_1	1		1
CNTROB	WP_004915569.1	Pst_1	1		1
KRTAP10-6	WP_004915569.1	Pst_1	1		1
BACH2	WP_004915569.1	Pst_1	1		1
KLHL2	WP_004922866.1	Pst_10	1		
TRAF3	WP_004924913.1	Pst_11	1		1
REL	WP_004924913.1	Pst_11	1		1
REL	WP_004925512.1	Pst_12	1		1
GMPPA	WP_004926210.1	Pst_13	1		
ZBTB10	WP_004926210.1	Pst_13	1		
TCF4	WP_004926210.1	Pst_13	1		
REL	WP_004926210.1	Pst_13	1		
PROP1	WP_004926210.1	Pst_13	1		
NOTO	WP_004926210.1	Pst_13	1		
AGR2	WP_004926210.1	Pst_13	1		
PAX6	WP_004926210.1	Pst_13	1		
TRIM2	WP_004926361.1	Pst_14	1		
TRIM3	WP_004926361.1	Pst_14	1		
TRIP10	WP_004927264.1	Pst_15	1		
CREB3L1	WP_042116632.1	Pst_16	1		
MFF	WP_042116632.1	Pst_16	1		
SCYL3	WP_042117401.1	Pst_17	1		
C12orf68	WP_042117401.1	Pst_17	1		
CCDC102B	WP_042117401.1	Pst_17	1		
DNM2	WP_042117401.1	Pst_17	1		
FAM9A	WP_042117401.1	Pst_17	1		
FOXJ2	WP_042117401.1	Pst_17	1		
PNMA1	WP_042117401.1	Pst_17	1		
ZBTB10	WP_042117401.1	Pst_17	1		
RRM2B	WP_042117401.1	Pst_17	1		
ZBED1	WP_042117401.1	Pst_17	1		
SP4	WP_042117401.1	Pst_17	1		
SRSF11	WP_042117401.1	Pst_17	1		
TACC1	WP_042117401.1	Pst_17	1		
TAX1BP1	WP_042117401.1	Pst_17	1		
TRIM72	WP_042117401.1	Pst_17	1		
TSN	WP_042117401.1	Pst_17	1		
UBAP1	WP_042117401.1	Pst_17	1		
REL	WP_042117401.1	Pst_17	1		
ZNF326	WP_042117401.1	Pst_17	1		

human protein	effector	abbr.	MAIN	REPEAT	HOMOLOGY
UBQLN2	WP_042117401.1	Pst_17	1		
VCP	WP_042117401.1	Pst_17	1		
BANP	WP_042117401.1	Pst_17	1		
VPS37B	WP_042117401.1	Pst_17	1		
VAC14	WP_052309238.1	Pst_18	1		
DYNLT1	WP_071599648.1	Pst_19	1		
TOX3	WP_071599648.1	Pst_19	1		
REL	WP_004915712.1	Pst_2	1		
TRAF2	WP_004915712.1	Pst_2	1		
PROP1	WP_004915712.1	Pst_2	1		
PICK1	WP_004915712.1	Pst_2	1		
KEAP1	WP_004917132.1	Pst_3	1		
FSD2	WP_004917987.1	Pst_4	1		
NECAB2	WP_004917987.1	Pst_4	1		
BCL6	WP_004919332.1	Pst_5	1		
REL	WP_004919757.1	Pst_6	1		1
TRAF6	WP_004919757.1	Pst_6	1		1
TCF4	WP_004920813.1	Pst_7	1		
ZBTB9	WP_004920813.1	Pst_7	1		
KCTD9	WP_004920813.1	Pst_7	1		
HMG20A	WP_004920813.1	Pst_7	1		
ATN1	WP_004920813.1	Pst_7	1		
CCDC91	WP_004921520.1	Pst_8	1		
UBQLN1	WP_004922581.1	Pst_9	1		1
TAX1BP1	WP_004922581.1	Pst_9	1		1
UBQLN2	WP_004922581.1	Pst_9	1		1
REL	WP_004726235.1	Vfu_1	1		
ZRANB1	WP_004726235.1	Vfu_1	1		
RINT1	WP_004726235.1	Vfu_1	1		
CUL9	WP_004726235.1	Vfu_1	1		
CNOT7	WP_004726235.1	Vfu_1	1		
SSBP3	WP_004726235.1	Vfu_1	1		
LBX1	WP_004729624.1	Vfu_10			1
IKBKG	WP_014257346.1	Vfu_11	1		
PICK1	WP_014257409.1	Vfu_12	1		
REL	WP_014257409.1	Vfu_12	1		
REL	WP_014257429.1	Vfu_13	1		
C10orf96	WP_014258130.1	Vfu_14	1		1
NOTO	WP_014258130.1	Vfu_14	1		1
REL	WP_014258130.1	Vfu_14	1		1
GMCL1	WP_038151258.1	Vfu_15	1		
IP6K2	WP_038151258.1	Vfu_15	1		
LHX2	WP_038151258.1	Vfu_15	1		
N4BP1	WP_038151258.1	Vfu_15	1		
NOTO	WP_038151258.1	Vfu_15	1		
ZBTB39	WP_038151258.1	Vfu_15	1		
TRIM41	WP_038151258.1	Vfu_15	1		
ZBTB2	WP_038151258.1	Vfu_15	1		
PAX6	WP_038151258.1	Vfu_15	1		
REL	WP_038151258.1	Vfu_15	1		
TCF4	WP_038151258.1	Vfu_15	1		
MFF	WP_038151743.1	Vfu_16	1		
REL	WP_038151811.1	Vfu_17	1		1
SP5	WP_038151811.1	Vfu_17			1
KCTD9	WP_038151811.1	Vfu_17	1		1
ZNF699	WP_038151811.1	Vfu_17	1		1
RBAK	WP_038151811.1	Vfu_17	1		1
SP4	WP_038151811.1	Vfu_17	1		1
TSNAX	WP_038151811.1	Vfu_17	1		1
VPS52	WP_038151811.1	Vfu_17	1		1
ZFP161	WP_038151811.1	Vfu_17	1		1
ZNF250	WP_038151811.1	Vfu_17	1		1
ZNF263	WP_038151811.1	Vfu_17	1		1
ZNF398	WP_038151811.1	Vfu_17	1		1

human protein	effector	abbr.	MAIN	REPEAT	HOMOLOGY
ZNF473	WP_038151811.1	Vfu_17	1		1
MAT2A	WP_038151811.1	Vfu_17	1		1
KLHL2	WP_038152552.1	Vfu_18	1		
ZBTB10	WP_038152592.1	Vfu_19	1		
UBQLN2	WP_004726603.1	Vfu_2	1		
NAB2	WP_038152705.1	Vfu_20	1		
EFHC2	WP_038152705.1	Vfu_20	1		
FOXP2	WP_038152705.1	Vfu_20	1		
HOMEZ	WP_038152705.1	Vfu_20	1		
PAK7	WP_038152705.1	Vfu_20	1		
PAX6	WP_038152705.1	Vfu_20	1		
PRDM16	WP_038152705.1	Vfu_20	1		
TCF4	WP_038152705.1	Vfu_20	1		
ZBED1	WP_038152705.1	Vfu_20	1		
ZNF398	WP_038152705.1	Vfu_20	1		
KRT76	WP_038152705.1	Vfu_20	1		
KRT75	WP_038152705.1	Vfu_20	1		
COL17A1	WP_115333225.1	Vfu_21	1		
USHBP1	WP_115333225.1	Vfu_21	1		
TFIP11	WP_115333225.1	Vfu_21	1		
SPAG5	WP_115333225.1	Vfu_21	1		
PNMA1	WP_115333225.1	Vfu_21	1		
NECAB2	WP_115333225.1	Vfu_21	1		
N4BP1	WP_115333225.1	Vfu_21	1		
MMS22L	WP_115333225.1	Vfu_21	1		
MID2	WP_115333225.1	Vfu_21	1		
GOLGA6L9	WP_115333225.1	Vfu_21	1		
COG6	WP_115333225.1	Vfu_21	1		
C15orf55	WP_115333225.1	Vfu_21	1		
TRIM9	WP_115333225.1	Vfu_21	1		
KRT27	WP_115333225.1	Vfu_21	1		
ZBED4	WP_115333225.1	Vfu_21	1		
VAC14	WP_004726842.1	Vfu_3	1		
VPS37B	WP_004726871.1	Vfu_4	1		
TCF4	WP_004726871.1	Vfu_4	1		
TRIP6	WP_004726871.1	Vfu_4	1		
USP54	WP_004726871.1	Vfu_4	1		
S100A1	WP_004726871.1	Vfu_4	1		
RUNX1	WP_004726871.1	Vfu_4	1		
RFX6	WP_004726871.1	Vfu_4	1		
REL	WP_004726871.1	Vfu_4	1		
PSME3	WP_004726871.1	Vfu_4	1		
POU2F1	WP_004726871.1	Vfu_4	1		
BCAR3	WP_004726871.1	Vfu_4	1		
C22orf13	WP_004726871.1	Vfu_4	1		
N4BP2	WP_004726871.1	Vfu_4	1		
FOXH1	WP_004726871.1	Vfu_4	1		
KIAA1958	WP_004726871.1	Vfu_4	1		
KRT31	WP_004726871.1	Vfu_4	1		
LCN2	WP_004726871.1	Vfu_4	1		
KPRP	WP_004726871.1	Vfu_4	1		
KRT76	WP_004726871.1	Vfu_4	1		
DDIT4L	WP_004726871.1	Vfu_4	1		
TOLLIP	WP_004727003.1	Vfu_5	1		
CEP70	WP_004727345.1	Vfu_6	1		1
ANKRD28	WP_004727345.1	Vfu_6	1		1
KEAP1	WP_004727658.1	Vfu_7	1		
WAC	WP_004727926.1	Vfu_8	1		
MIF4GD	WP_004727926.1	Vfu_8		1	
LBX1	WP_004729371.1	Vfu_9	1		1
SPERT	WP_005156690.1	Yen_1	1		
REL	WP_005163729.1	Yen_10	1		1
CDR2L	WP_005163729.1	Yen_10	1		1
ZNF639	WP_005163729.1	Yen_10	1		1

human protein	effector	abbr.	MAIN	REPEAT	HOMOLOGY
KRT75	WP_005163729.1	Yen_10	1		1
ZBTB10	WP_005163816.1	Yen_11	1		
RREB1	WP_005163816.1	Yen_11	1		
REL	WP_005163816.1	Yen_11	1		
L3MBTL3	WP_005163816.1	Yen_11	1		
KIAA1958	WP_005163816.1	Yen_11	1		
GMCL1	WP_005163816.1	Yen_11	1		
GGA1	WP_005163816.1	Yen_11	1		
DZIP3	WP_005164084.1	Yen_12	1		1
REL	WP_005164084.1	Yen_12	1		1
CADPS	WP_005164084.1	Yen_12	1		1
REL	WP_005164331.1	Yen_13	1		
EVI5	WP_005164542.1	Yen_14	1		1
GOLGA6L9	WP_005164542.1	Yen_14	1		1
VPS52	WP_005164542.1	Yen_14	1		1
REL	WP_005164542.1	Yen_14	1		1
PNMA1	WP_005164542.1	Yen_14	1		1
L3MBTL3	WP_005164542.1	Yen_14	1		1
REL	WP_005166097.1	Yen_15	1		
RDM1	WP_005179029.1	Yen_16	1		1
UBQLN2	WP_010891207.1	Yen_17	1		
UBQLN1	WP_010891207.1	Yen_17	1		
KLHL3	WP_010891236.1	Yen_18	1		
PRPSAP2	WP_010891236.1	Yen_18	1		
IKZF3	WP_010891236.1	Yen_18	1		
HOMER3	WP_010891236.1	Yen_18	1		
GRIP1	WP_010891236.1	Yen_18	1		
CCHCR1	WP_010891236.1	Yen_18	1		
REL	WP_010891236.1	Yen_18	1		
RREB1	WP_010891236.1	Yen_18	1		
TNIP1	WP_010891236.1	Yen_18	1		
ZBTB10	WP_014609009.1	Yen_19	1		1
CDR2L	WP_014609009.1	Yen_19	1		1
TCF4	WP_014609009.1	Yen_19	1		1
RREB1	WP_014609009.1	Yen_19	1		1
PRPSAP2	WP_014609009.1	Yen_19	1		1
PAX6	WP_014609009.1	Yen_19	1		1
PAX5	WP_014609009.1	Yen_19	1		1
KRT75	WP_014609009.1	Yen_19	1		1
ZNF639	WP_014609009.1	Yen_19			1
USP54	WP_005157598.1	Yen_2	1		
TCF4	WP_005157598.1	Yen_2	1		
REL	WP_005157598.1	Yen_2	1		
NOL4	WP_005157598.1	Yen_2	1		
HOMEZ	WP_005157598.1	Yen_2	1		
PAX6	WP_014609475.1	Yen_20	1		
AGR2	WP_014609475.1	Yen_20	1		
CCDC57	WP_014609475.1	Yen_20	1		
ZBTB10	WP_014609475.1	Yen_20	1		
CTDSP2	WP_014609475.1	Yen_20	1		
KCTD6	WP_005157674.1	Yen_3	1		
REL	WP_005159058.1	Yen_4	1		1
MFF	WP_005159145.1	Yen_5	1		
STAU1	WP_005159272.1	Yen_6	1		1
PROP1	WP_005159272.1	Yen_6	1		1
NEDD4	WP_005159272.1	Yen_6	1		1
NAB2	WP_005159272.1	Yen_6	1		1
KRT76	WP_005159272.1	Yen_6	1		1
KRT75	WP_005159272.1	Yen_6	1		1
IKZF3	WP_005159272.1	Yen_6	1		1
EFHC2	WP_005159272.1	Yen_6	1		1
CNOT2	WP_005159272.1	Yen_6	1		1
CECR2	WP_005159272.1	Yen_6	1		1
ZMYND12	WP_005159272.1	Yen_6			1

human protein	effector	abbr.	MAIN	REPEAT	HOMOLOGY
NOTO	WP_005159272.1	Yen_6			1
ZBTB10	WP_005159272.1	Yen_6	1		1
UBAP1	WP_005159272.1	Yen_6	1		1
RNF20	WP_005162291.1	Yen_7	1		1
GIGYF1	WP_005162291.1	Yen_7	1		1
SFMBT1	WP_005162291.1	Yen_7	1		1
TRIM27	WP_005162291.1	Yen_7	1		1
CEP44	WP_005162694.1	Yen_8	1		
TCF4	WP_005162694.1	Yen_8	1		
LZTS2	WP_005162781.1	Yen_9	1		
ZNF398	WP_005162781.1	Yen_9	1		
KRT76	WP_005162781.1	Yen_9	1		
N4BP2	WP_005162781.1	Yen_9	1		
NEBL	WP_005162781.1	Yen_9	1		
PPP1R13B	WP_005162781.1	Yen_9	1		
PRPH	WP_005162781.1	Yen_9	1		
RBM42	WP_005162781.1	Yen_9	1		
TFIP11	WP_005162781.1	Yen_9	1		
ZC4H2	WP_005162781.1	Yen_9	1		
KRT75	WP_005162781.1	Yen_9	1		
REL	WP_006817197.1	Yre_1	1	1	1
EIF2B1	WP_006817197.1	Yre_1	1		1
ZBED1	WP_040903175.1	Yre_10	1		
HNRNPC	WP_040903573.1	Yre_11	1		
ZBTB48	WP_050812366.1	Yre_12	1		
REL	WP_050812366.1	Yre_12		1	
CBR1	WP_050812366.1	Yre_12	1		
EFHC2	WP_050812366.1	Yre_12	1		
KRT75	WP_050812366.1	Yre_12	1		
PAX5	WP_050812366.1	Yre_12	1		
PAX6	WP_050812366.1	Yre_12	1		
TCF4	WP_050812366.1	Yre_12	1		
PROP1	WP_050812366.1	Yre_12	1		
KCTD6	WP_071777502.1	Yre_13	1		
TRAF2	WP_071777518.1	Yre_14	1		
MFF	WP_006818175.1	Yre_2	1		1
VAC14	WP_006818941.1	Yre_3	1		1
PAX5	WP_006818941.1	Yre_3	1		1
REL	WP_006818941.1	Yre_3	1	1	1
NAGK	WP_006819702.1	Yre_4	1		
ZBTB10	WP_006820335.1	Yre_5	1		
MAGEA6	WP_006820832.1	Yre_6	1		1
MAL2	WP_006820832.1	Yre_6			1
AGTRAP	WP_006820832.1	Yre_6			1
ZBED4	WP_006820847.1	Yre_7	1		1
TRAF2	WP_006820847.1	Yre_7	1		1
TRIM27	WP_006820847.1	Yre_7			1
EFHC2	WP_006820847.1	Yre_7			1
MFF	WP_006821080.1	Yre_8	1		
PTPN21	WP_038254823.1	Yre_9	1		1
NOL4	WP_038254823.1	Yre_9	1		1
INADL	WP_038254823.1	Yre_9	1	1	1

Table S5 | Curated interactions of the bhLit_BM-v1 tested in the Y2H. *Some interactions were excluded from the Y2H experiment to determine the assay sensitivity due to incorrect sequences (indicated by "FALSE") of the human interactors (h_sequ) and autoactivation of the effectors (eff AA, autoactivation indicated by "1"). Pubmed IDs of publications describing the interaction are listed. One yeast culture (underlined) did not grow. Seven interactions were detected in the Y2H (bold). No autoactivation of the human proteins was observed. effector, effector gene name.*

effector	effector ID	human symbol	h_sequ	eff AA	reference Pubmed ID
NleA	Q8XAJ5	DSCR4	FALSE		25519916, 25519916, 27018634
NleA	Q8XAJ5	SEC24A	FALSE		18005731, 22432415
NleA	Q8XAJ5	SEC24B	FALSE		22432415, 27018634
NleF	Q8XAL7	CASP4	FALSE		23516580
NleB	B7UI21	TRADD	FALSE		23955153, 24025841
EspI/NleA	B7UR60	SNTA1	FALSE		17979986
EspZ	Q7DB68	SLC7A5	FALSE		20374249
Tir	Q7DB77	HPCAL4	FALSE		25519916, 25519916
Tir	Q7DB77	PDE6D	FALSE		25519916, 25519916
EspW	Q8X9A5	KIF15	FALSE		28630074, 28630074
EspY1	Q8XA11	PIH1D1	FALSE		25519916, 25519916
EspB	Q8XC86	RBCK1	FALSE		25519916, 25519916
PipB2	A0A0F6B5H5	KLC1	TRUE	1	16938850, 16938850
YopM	A1JU68	RPS6KA1	TRUE	1	12626518, 20957203
YopM	A1JU68	PKN2	TRUE	1	12626518, 20957203
Map	P0AE20	RHPN1	TRUE	1	25519916, 25519916, 27018634
NleA	Q8XAJ5	SEC24D	TRUE	1	22432415
NleA	Q8XAJ5	FRMD3	TRUE	1	25519916, 25519916, 27018634
NleA	Q8XAJ5	PENK	TRUE	1	25519916, 25519916, 27018634
NleA	Q8XAJ5	PTP4A1	TRUE	1	25519916, 25519916, 27018634
NleF	Q8XAL7	TRNT1	TRUE	1	25519916, 25519916
NleF	Q8XAL7	CASP8	TRUE	1	23516580
NleF	Q8XAL7	CASP9	TRUE	1	23516580, 25519916, 25519916
NleF	Q8XAL7	DHFR	TRUE	1	25519916, 25519916
NleF	Q8XAL7	ERI3	TRUE	1	25519916, 25519916
NleF	Q8XAL7	HMGN2	TRUE	1	25519916, 25519916
NleF	Q8XAL7	LMO4	TRUE	1	25519916, 25519916
SseJ	<u>A0A0F6B1Q8</u>	RHOA	TRUE		18996344, 19887681, 22740689
NleD	A0A0H3JGR6	METTL2A	TRUE		25519916, 25519916
YopJ	A0A0N9NCU6	MAP2K2	TRUE		10489373
NleB	B7UI21	FADD	TRUE		23955153, 24025841, 24025841
NleB	**B7UI21**	**RIPK1**	TRUE		23955153, 24025841, 24025841
EspF	**B7UM88**	**SNX9**	TRUE		16585770, 16585770, 17893247, 17893247
Tir	B7UM99	STK16	TRUE		25519916
Tir	**B7UM99**	**HPCAL1**	TRUE		25519916
Tir	B7UM99	KRT18	TRUE		14710194, 14710194, 16367866
Tir	**B7UM99**	**NCALD**	TRUE		25519916
EspI/NleA	B7UR60	SLC9A3R2	TRUE		17979986, 20618342
Sirp	D0ZRB2	DNAJB11	TRUE		20335166, 20335166
Sirp	D0ZRB2	TXN	TRUE		19690162, 19690162, 19690162
SopE	O52623	CDC42	TRUE		9630225, 12093730, 12093730
IpaH4.5	P18009	RELA	TRUE		23083102
IpaB	P18011	MAD2L2	TRUE		17719540, 17719540
YopK	Q56935	RACK1	TRUE		21347310
BopE	Q63K41	RAC1	TRUE		12897019
BopE	Q63K41	CDC42	TRUE		12897019
EspG	**Q7DB50**	**NMI**	TRUE		25519916, 25519916
Tir	Q7DB77	BAIAP2L1	TRUE		25519916, 25519916, 19286134, 19366662, 22921828
Tir	Q7DB77	ARRB1	TRUE		25519916, 25519916
IpaH9.8	Q8VSC3	IKBKG	TRUE		20010814
EspY1	Q8XA11	ZNHIT1	TRUE		25519916, 25519916
EspY1	Q8XA11	CDKN2AIPNL	TRUE		25519916, 25519916
EspY1	Q8XA11	CLK1	TRUE		25519916, 25519916
EspY1	Q8XA11	DNAJC14	TRUE		25519916, 25519916
EspY1	Q8XA11	PSMC1	TRUE		25519916, 25519916
EspJ	Q8XB62	RIC8A	TRUE		25519916, 25519916, 27018634
EspJ	Q8XB62	CENPH	TRUE		25519916, 25519916, 27018634
EspJ	Q8XB62	MRFAP1L1	TRUE		25519916, 25519916
NleB1	Q8XBX8	DRG2	TRUE		25519916, 25519916, 27018634
NleB1	Q8XBX8	FADD	TRUE		24025841, 27018634
NleB1	Q8XBX8	POLR2E	TRUE		25519916, 25519916, 27018634
NleB1	**Q8XBX8**	**RIPK1**	TRUE		24025841, 27018634
SopA	Q8ZNR3	RNF5	TRUE		16176924, 16176924
OspG	**Q99PZ6**	**UBE2L3**	TRUE		16162672, 16162672, 24856362, 24856362
OspG	*Q99PZ6*	UBE2D1	TRUE		16162672, 16162672
OspG	Q99PZ6	UBE2D2	TRUE		16162672, 16162672
OspG	Q99PZ6	UBE2D3	TRUE		16162672, 24446487

Table S6 | Protein pairs of the bhRRS-v1 tested in the Y2H. *Some interactors were excluded from the Y2H experiment due to incorrect sequences (indicated by "FALSE") of the human proteins (h-sequ) and due to autoactivation of either the effector (effector AA) or the human interactor (human AA). Autoactivation is indicated by "1". No interactions were detected in the Y2H.*

effector ID	human symbol	h_sequ	effector AA	human AA
dime_meta_effector_102	KCNH8	FALSE		
dime_meta_effector_122	FLJ38668	FALSE		
dime_meta_effector_131	MPND	FALSE		
A0A6M7GVE3	GADD45G	FALSE		
dime_meta_effector_59	TTLL13	FALSE		
dime_meta_effector_77	EFHD2	FALSE		
WP_000097400.1	SPANXC	FALSE		
WP_000155904.1	CD19	FALSE		
WP_000407090.1	CORO1B	FALSE		
WP_001237041.1	C2orf61	FALSE		
WP_002891634.1	TRADD	FALSE		
WP_004256728.1	HOXC12	FALSE		
WP_004261459.1	LOC144742	FALSE		
WP_004921623.1	PAPPA2	FALSE		
WP_005164223.1	TBCCD1	FALSE		
WP_005285281.1	ATP8A1	FALSE		
WP_016537795.1	ECHDC3	FALSE		
WP_032156687.1	RBMY2FP	FALSE		
WP_036957743.1	PDE6D	FALSE		
WP_042032153.1	RHOXF1	FALSE		
WP_044180429.1	CCR10	FALSE	1	
WP_071777518.1	ACOT12	FALSE		
WP_077626056.1	TRDV2	FALSE		
WP_077626319.1	DEFB115	FALSE		
WP_113858462.1	COLEC12	FALSE		
WP_113858733.1	FUT8	FALSE		
dime_meta_effector_111	OMA1	TRUE		
dime_meta_effector_113	TMEM230	TRUE		
dime_meta_effector_125	EXOSC5	TRUE		
dime_meta_effector_154	ECM2	TRUE		
dime_meta_effector_178	BAZ2B	TRUE		
dime_meta_effector_185	TPPP2	TRUE		
A0A0F6B5H5	PTPRM	TRUE		1
B7UM88	TGOLN2	TRUE		
dime_meta_effector_2	SPIN2B	TRUE		
Q7DB77	TNFRSF25	TRUE		
Q8ZNR3	TRIP11	TRUE		
dime_meta_effector_11	AHDC1	TRUE		
dime_meta_effector_29	AKAP11	TRUE		
dime_meta_effector_35	PIPOX	TRUE		
dime_meta_effector_45	FBXL7	TRUE		
dime_meta_effector_56	HELLS	TRUE		
dime_meta_effector_65	COQ6	TRUE		
dime_meta_effector_82	ADAMTSL5	TRUE		
dime_meta_effector_97	TMEM79	TRUE		
WP_000490639.1	NAALADL1	TRUE		
WP_000611436.1	ZNF140	TRUE		
WP_000804518.1	LRRC28	TRUE		
WP_001149870.1	PRPF38A	TRUE		
WP_004171426.1	PATZ1	TRUE		
WP_004234458.1	SSTR3	TRUE	1	
WP_004241218.1	RBM33	TRUE		
WP_004253752.1	IPO13	TRUE		
WP_004255132.1	TMEM60	TRUE		
WP_004257971.1	ORC6	TRUE		1
WP_004258949.1	KIAA0930	TRUE		
WP_004261076.1	RPS15	TRUE		
WP_004262559.1	DOCK3	TRUE		
WP_004262987.1	SCP2	TRUE		
WP_004264896.1	PRNP	TRUE		
WP_004726842.1	ARL3	TRUE		

effector ID	human symbol	h_sequ	effector AA	human AA
WP_004728215.1	SLC35G1	TRUE		
WP_004728297.1	RDM1	TRUE		
WP_004905473.1	MED24	TRUE		
WP_004906048.1	PCDHGB5	TRUE		
WP_004915373.1	DCAF16	TRUE		
WP_004917132.1	LPAR1	TRUE		
WP_004918442.1	ACER2	TRUE		
WP_004920058.1	YPEL5	TRUE		
WP_005126657.1	HDAC10	TRUE		
WP_005158077.1	SYNGR2	TRUE		
WP_005159587.1	SLC39A11	TRUE		
WP_005162291.1	FAM72B	TRUE		
WP_006818304.1	PLK2	TRUE		
WP_009485462.1	ACAD8	TRUE		
WP_011579078.1	C7orf60	TRUE		
WP_014657167.1	FASTKD2	TRUE		
WP_015422612.1	CSN3	TRUE		
WP_015963250.1	FAHD2A	TRUE		
WP_016536850.1	DPP9	TRUE		
WP_020317218.1	GADD45GIP1	TRUE		
WP_032145775.1	C17orf64	TRUE		
WP_036958071.1	RPS26	TRUE		
WP_038150968.1	SETD9	TRUE		
WP_039898226.1	KRT33B	TRUE		
WP_039898721.1	ELL3	TRUE		
WP_040232565.1	KCNJ6	TRUE		
WP_040233393.1	GSTO2	TRUE		
WP_040902689.1	OLFML1	TRUE		
WP_042117315.1	EPS15L1	TRUE		
WP_044177605.1	NEK5	TRUE		
WP_044180423.1	USP1	TRUE		
WP_044182945.1	DNM3	TRUE		
WP_052463753.1	DGAT2	TRUE		
WP_055696404.1	C4orf27	TRUE		
WP_062773486.1	KAZN	TRUE		
WP_071528137.1	L2HGDH	TRUE		
WP_071777494.1	SYCP3	TRUE		
WP_071777502.1	SEPP1	TRUE		
WP_071777524.1	COA5	TRUE		

Table S7 | Protein pairs of the hsPRS-v2 and hsRRS-v2 tested in the Y2H. *Configuration (config.) 1: interactor 1 as DB-X, interactor 2 as AD-Y. Configuration 2: interactor 1 as AD-Y, interactor 2 as DB-X. If an interaction was detected between two proteins, one of the last two columns is marked with "1" respective to the configuration.*

interator 1	interator 2	dataset	interaction in config. 1	interaction in config. 2
NCK1	LCP2	hsPRS-v2		
STAC3	SHMT2	hsRRS-v2		
ARMC1	EMD	hsRRS-v2		
ZC3HC1	PDHB	hsRRS-v2		
NQO2	FKBP3	hsRRS-v2		
BOC	RAB3B	hsRRS-v2		
GRAP2	LCP2	hsPRS-v2		1
LSM2	LSM3	hsPRS-v2	1	1
MCCC1	GALK1	hsRRS-v2		
C22orf29	GPD2	hsRRS-v2		
SLC22A15	PSMD5	hsRRS-v2		
EIF3E	IFIT1	hsPRS-v2		
ZNF350	PPP6C	hsRRS-v2		
NPC2	CD81	hsRRS-v2		
RAD23A	PSMD4	hsPRS-v2		
BAT2L1	BTC	hsRRS-v2		
MCM3	MCM2	hsPRS-v2		
BCL2L1	BAK1	hsPRS-v2	1	1
MCM10	ORC2L	hsPRS-v2		
CKS1B	CDK2	hsPRS-v2		1
WDR41	CD151	hsRRS-v2		
BCL2L1	BAD	hsPRS-v2		
TP53	HIF1A	hsPRS-v2		
NTF4	BDNF	hsPRS-v2		
REM2	PDE9A	hsRRS-v2		
ATAD2	CANX	hsRRS-v2		
MAD1L1	MAD2L1	hsPRS-v2	1	1
BATF	JUNB	hsPRS-v2		1
ZCCHC9	GCDH	hsRRS-v2		
ARL6IP6	GRIK2	hsRRS-v2		
ABCF3	RGR	hsRRS-v2		
FOS	DDIT3	hsPRS-v2		
UBE2G2	LAMP2	hsRRS-v2		
ZNF688	OSM	hsRRS-v2		
GMPPA	MNAT1	hsRRS-v2		
STX5	FABP7	hsRRS-v2		
NFE2L1	MAFG	hsPRS-v2		
DCTN6	SERPINB3	hsRRS-v2		
SALL2	HCLS1	hsRRS-v2		
HSP90AA1	NR3C1	hsPRS-v2		
ZBTB25	FIGF	hsRRS-v2		
TMEM22	ITPK1	hsRRS-v2		
STK25	PROS1	hsRRS-v2		
ZBTB16	HDAC1	hsPRS-v2		
DDIT3	ATF3	hsPRS-v2		1
NDFIP1	PDGFRA	hsRRS-v2		
PSEN2	HLA-DMB	hsRRS-v2		
TM4SF4	SLC6A1	hsRRS-v2		
PCNA	FEN1	hsPRS-v2		
IGFBP4	IGF2	hsPRS-v2		
BEST1	NONO	hsRRS-v2		
PEX11B	PEX19	hsPRS-v2		
NCBP2	NCBP1	hsPRS-v2		
FANCG	FANCA	hsPRS-v2		
SKP2	SKP1	hsPRS-v2		
GRB2	CBLB	hsPRS-v2	1	1
LMBR1L	ETF1	hsRRS-v2		
GCG	FABP4	hsRRS-v2		
MUC7	APOD	hsRRS-v2		
RAN	RCC1	hsPRS-v2		
HGS	NF2	hsPRS-v2		
GINS3	ASS1	hsRRS-v2		
DBN1	ARSA	hsRRS-v2		
ZNF213	SLC25A6	hsRRS-v2		
FOS	CEBPG	hsPRS-v2		
PThsPRS-v2	CA2	hsRRS-v2		
KLHL6	RCC1	hsRRS-v2		
CASP3	XIAP	hsPRS-v2		
PPIL3	CENPA	hsRRS-v2		
TIRAP	NFIB	hsRRS-v2		
CXCL11	RXRB	hsRRS-v2		

interator 1	interator 2	dataset	interaction in config. 1	interaction in config. 2
UGGT2	LUM	hsRRS-v2		
PCNA	GADD45A	hsPRS-v2		
LMNB1	LMNA	hsPRS-v2	1	1
VAV1	GRB2	hsPRS-v2		
SLC39A14	COPB1	hsRRS-v2		
FANCC	FANCA	hsPRS-v2		
UBLCP1	INPP1	hsRRS-v2		
WDR62	ITPA	hsRRS-v2		
RAB3IP	DLX4	hsRRS-v2		
NUP62CL	RHOC	hsRRS-v2		
ARHGEF15	CANX	hsRRS-v2		
C19orf40	DUT	hsRRS-v2		
NOD1	RIPK2	hsPRS-v2		
PEX19	PEX14	hsPRS-v2		
ARFIP2	ARF1	hsPRS-v2		
RB1	LMNA	hsPRS-v2		
PEX3	PEX19	hsPRS-v2	1	
DNAJA1	NAT2	hsRRS-v2		
TSTD2	DEFA3	hsRRS-v2		
CRIPT	PSMD12	hsRRS-v2		
PHF21B	SCARB1	hsRRS-v2		
CASP9	XIAP	hsPRS-v2	1	
ARHGAP1	RHOA	hsPRS-v2		
MRPS25	MOBP	hsRRS-v2		
SYCE1	PSMD5	hsRRS-v2		
L3MBTL2	CLPTM1	hsRRS-v2		
CASP7	XIAP	hsPRS-v2		
LSM3	FAS	hsRRS-v2		
RIC3	PMCH	hsRRS-v2		
PEX16	PEX19	hsPRS-v2		
UBE2I	TP53	hsPRS-v2		
SNX21	CD34	hsRRS-v2		
PDPK1	AKT1	hsPRS-v2		
NUDT4	NDP	hsRRS-v2		
NPDC1	HIST1H1C	hsRRS-v2		
SEMA4G	RFX3	hsRRS-v2		
BTRC	SKP1	hsPRS-v2		
CSGALNACT2	NKX2-5	hsRRS-v2		
ARFIP2	RAC1	hsPRS-v2		
C10orf119	BMP5	hsRRS-v2		
PGAP2	CNN1	hsRRS-v2		
NRG1	ERBB3	hsPRS-v2		
HPCAL4	COPB1	hsRRS-v2		
LAT	GRB2	hsPRS-v2		
PLXNA4	MCM2	hsRRS-v2		
HNRPLL	GPR18	hsRRS-v2		
CRADD	CASP2	hsPRS-v2	1	
CTCF	MAOB	hsRRS-v2		
CWF19L1	ACVR1	hsRRS-v2		
MIIP	NUDT2	hsRRS-v2		
CGB5	CGA	hsPRS-v2		
GNB2L1	PDE4D	hsPRS-v2		
SF3A1	RBM3	hsRRS-v2		
DCP1A	SMAD4	hsPRS-v2		1
PPP3R1	PPP3CA	hsPRS-v2		
PVRL2	GP1BA	hsRRS-v2		
SMAD4	SMAD1	hsPRS-v2		1
VILL	PBX2	hsRRS-v2		
GTF2F2	GTF2F1	hsPRS-v2		
MCM5	MCM2	hsPRS-v2		
DRAP1	DR1	hsPRS-v2	1	1
CLEC2D	ATP5O	hsRRS-v2		
KIAA0907	BYSL	hsRRS-v2		
HBB	HBA2	hsPRS-v2		
ORC4L	ORC2L	hsPRS-v2		
HBZ	CKB	hsRRS-v2		
C3orf38	ERBB3	hsRRS-v2		

Table S8 | Subset of HuMMI tested in yN2H. *Configuration indicated by N1-X and N2-Y. Interaction indicated by "1" in last column if log₂ NLR ≥ 0.*

N1-X	N2-Y	interaction according to NLR (log$_2$) ≥ 0
ATP5B	WP_004262673.1	
BANP	WP_042117401.1	
C12orf68	WP_042117401.1	
C15orf55	WP_115333225.1	
CAGE1	WP_004905473.1	
CBY1	WP_071825927.1	
CCDC102B	dime_meta_effector_150	
CCDC57	WP_108474309.1	
CCDC91	WP_004921520.1	
CENPK	WP_004905473.1	
CENPK	WP_001272443.1	
CEP250	WP_062772817.1	
CEP76	WP_016536523.1	
CIB1	WP_033792202.1	
CIB1	WP_033751802.1	
CLTCL1	WP_005288481.1	
CLTCL1	WP_001235473.1	
CNOT7	WP_004726235.1	
COG6	WP_001235473.1	
COG6	WP_004261691.1	
CREBZF	WP_004905473.1	
CUL9	WP_004726235.1	
DCTN2	WP_004905473.1	
DPPA4	WP_004905473.1	
DYNLT1	WP_072041464.1	
DYNLT1	WP_077626056.1	1
EFHC2	WP_044173012.1	
EIF2B1	WP_000258580.1	1
FAM9A	WP_042117401.1	1
FOXJ2	WP_042117401.1	
FSD2	WP_004917987.1	
FSD2	WP_004253606.1	
GADD45G	dime_meta_effector_140	
GMCL1	WP_005163816.1	
GOLGA2	WP_001235473.1	
GOLGA2	WP_113857302.1	
GOLGA6L9	WP_115333225.1	
GRIP1	WP_010891236.1	
HMG20A	WP_004920813.1	
HSF2BP	WP_004152718.1	
HSF2BP	WP_009484324.1	
KCTD6	WP_071777502.1	
KCTD9	dime_meta_effector_28	
KIAA1328	WP_000258580.1	
KRT27	WP_009484324.1	
KRT34	WP_001235473.1	
KRT75	WP_038152705.1	
KRT75	WP_005162781.1	
KRT75	dime_meta_effector_28	
KRTAP1-1	dime_meta_effector_140	
KRTAP5-1	dime_meta_effector_140	
L3MBTL3	WP_113858483.1	
LPIN2	WP_004261691.1	
MDFI	WP_009486019.1	
MFF	WP_038151743.1	1
MPP7	WP_024256417.1	1
MSANTD4	WP_009484876.1	
MTUS2	WP_001267298.1	
N4BP2	WP_004726871.1	
NAP1L2	WP_001272443.1	1
NECAB2	WP_009484324.1	
NOTO	WP_040230127.1	
NOTO	WP_040260715.1	
NUP62	WP_001272443.1	
PAIP1	WP_113859080.1	
PAX5	WP_050812366.1	
PDE4DIP	WP_055696404.1	
PDE4DIP	WP_009484324.1	
PICK1	WP_055696404.1	
PICK1	WP_004118237.1	
PICK1	WP_004915712.1	
PKNOX1	dime_meta_effector_97	

N1-X	N2-Y	interaction according to NLR (log$_2$) ≥ 0
PNMA2	WP_004118237.1	
PPP1R13B	WP_005162781.1	
PPP1R13B	WP_004905473.1	1
PROP1	WP_050812366.1	
PROP1	WP_004926210.1	
PROP1	WP_005159272.1	
PROP1	WP_001235473.1	
PRPSAP2	WP_000220141.1	
PSTPIP1	WP_044180562.1	1
PUF60	WP_033792699.1	1
REEP6	WP_000508975.1	
REL	WP_038151258.1	
REL	WP_004262673.1	
REL	WP_040230127.1	
REL	WP_014257429.1	
REL	WP_113857302.1	
REL	WP_036417499.1	
REL	WP_005164331.1	
REL	WP_000220141.1	
REL	WP_113857569.1	
RNF20	WP_005162291.1	
S100A1	WP_004726871.1	
SDCBP	WP_004263067.1	
SGTA	WP_042030958.1	
SHANK2	dime_meta_effector_39	
SSBP3	WP_072041472.1	
SSX2IP	WP_009484324.1	
TACC1	WP_042117401.1	
TADA3	WP_001272443.1	
TAF7L	WP_044183152.1	
TCF4	WP_004118237.1	
TCF4	WP_005157598.1	
TCF4	WP_040260715.1	
TCF4	WP_038152705.1	
TCF4	WP_000004564.1	
TCF4	WP_009484324.1	
TFIP11	WP_004118237.1	
TFIP11	WP_001531161.1	
TRAF1	WP_009484324.1	
TRAF2	WP_004118237.1	
TRIM27	WP_005162291.1	
TRIM27	WP_004118237.1	
TRIM50	WP_001059674.1	
TSN	WP_004261691.1	
TSN	WP_042117401.1	
TSN	WP_004118237.1	
TSNAX	WP_038151811.1	
UBQLN2	WP_000904613.1	
UBQLN2	WP_040263025.1	
UBQLN2	WP_044177448.1	
UBQLN2	WP_001267298.1	1
USP54	dime_meta_effector_28	
USP54	WP_033791692.1	
VAC14	WP_000067801.1	1
VAC14	WP_006818941.1	1
VPS52	WP_000004564.1	
VPS52	WP_005164542.1	
ZBED1	WP_038152705.1	
ZBED1	WP_084596144.1	
ZBED4	WP_000255032.1	
ZBTB8A	WP_009484324.1	
ZFP161	WP_062773522.1	
ZFP161	WP_038151811.1	
ZMYND12	WP_001059674.1	
ZNF326	WP_042117401.1	
ZNF398	WP_005162781.1	
ZNF639	dime_meta_effector_140	
ZNF699	WP_040259375.1	
dime_meta_effector_132	CUTC	
dime_meta_effector_140	GADD45G	
dime_meta_effector_140	ZNF639	
dime_meta_effector_140	KRTAP1-1	
dime_meta_effector_140	KRTAP5-1	
dime_meta_effector_140	KRTAP10-6	1
dime_meta_effector_32	MAGI1	1
dime_meta_effector_39	SHANK2	

N1-X	N2-Y	interaction according to NLR (log$_2$) $\geq$ 0
dime_meta_effector_41	PICK1	
dime_meta_effector_97	PKNOX1	
WP_000004564.1	TCF4	
WP_000004564.1	VPS52	
WP_000220141.1	PRPSAP2	
WP_000220141.1	REL	
WP_000255032.1	ZBED4	1
WP_000258580.1	KIAA1328	
WP_000258580.1	EIF2B1	1
WP_000508975.1	REEP6	
WP_000904613.1	UBQLN2	
WP_001059674.1	TRIM50	
WP_001059674.1	ZMYND12	
WP_001235473.1	KRT34	
WP_001235473.1	KRT31	
WP_001235473.1	PROP1	
WP_001235473.1	LZTS2	
WP_001235473.1	GOLGA2	
WP_001235473.1	CLTCL1	1
WP_001267298.1	MTUS2	
WP_001267298.1	TAX1BP1	
WP_001267298.1	SERTAD1	
WP_001267298.1	APPL2	
WP_001267298.1	IKBKG	1
WP_001267298.1	UBQLN2	1
WP_001272443.1	NUP62	
WP_001272443.1	CENPK	
WP_001272443.1	TADA3	1
WP_001272443.1	NAP1L2	1
WP_001531161.1	TFIP11	
WP_004118237.1	TAX1BP1	
WP_004118237.1	TSN	
WP_004118237.1	PNMA2	
WP_004118237.1	TCF4	
WP_004118237.1	SERTAD1	
WP_004118237.1	HIP1	
WP_004118237.1	TRAF2	
WP_004118237.1	PICK1	
WP_004118237.1	TRIM27	1
WP_004118237.1	TFIP11	1
WP_004152718.1	HSF2BP	
WP_004240712.1	EVI5	
WP_004262673.1	REL	
WP_004262673.1	ATP5B	
WP_004726235.1	CUL9	1
WP_004726871.1	S100A1	
WP_004726871.1	N4BP2	
WP_004905473.1	RNF135	
WP_004905473.1	CENPK	
WP_004905473.1	CAGE1	
WP_004905473.1	DPPA4	
WP_004905473.1	CCHCR1	
WP_004905473.1	CREBZF	1
WP_004905473.1	PPP1R13B	1
WP_004905473.1	DCTN2	1
WP_004915712.1	PICK1	
WP_004917132.1	KEAP1	
WP_004917987.1	FSD2	
WP_004919332.1	BCL6	
WP_004920813.1	HMG20A	
WP_004921520.1	CCDC91	1
WP_004926210.1	PROP1	
WP_005129057.1	MAGEA6	1
WP_005157598.1	TCF4	
WP_005159272.1	PROP1	
WP_005162291.1	TRIM27	
WP_005162694.1	CEP44	
WP_005162781.1	ZNF398	
WP_005162781.1	PPP1R13B	1
WP_005162781.1	KRT75	1
WP_005163729.1	ZNF639	
WP_005163729.1	CDR2L	
WP_005163816.1	GMCL1	
WP_005164331.1	REL	
WP_005164542.1	VPS52	
WP_005288481.1	CLTCL1	

164

N1-X	N2-Y	interaction according to NLR (log$_2$) ≥ 0
WP_009484324.1	RAB3IP	
WP_009484324.1	KRT27	
WP_009484324.1	ZBTB7B	
WP_009484324.1	TRAF1	
WP_009484324.1	ZBTB8A	
WP_009484324.1	NECAB2	
WP_009484324.1	SSX2IP	
WP_009484324.1	PDE4DIP	
WP_009484324.1	TCF4	1
WP_009484324.1	HSF2BP	1
WP_009484876.1	MSANTD4	1
WP_009486019.1	MDFI	
WP_009486529.1	KRTAP10-5	
WP_014257429.1	REL	
WP_016536523.1	CEP76	
WP_024256417.1	MPP7	1
WP_033751802.1	CIB1	
WP_033791692.1	RAB3IP	
WP_033791692.1	USP54	
WP_033792202.1	CIB1	1
WP_033792699.1	PUF60	
WP_036417499.1	REL	
WP_038151258.1	REL	
WP_038151743.1	MFF	
WP_038151811.1	ZFP161	
WP_038151811.1	TSNAX	
WP_038152705.1	ZBED1	
WP_038152705.1	KRT75	
WP_038152705.1	TCF4	
WP_040230127.1	REL	
WP_040230127.1	NOTO	
WP_040259375.1	ZNF699	
WP_040260715.1	TCF4	
WP_040260715.1	NOTO	1
WP_040263025.1	UBQLN2	
WP_040263598.1	CREB3L1	
WP_042030958.1	SGTA	
WP_042117401.1	BANP	
WP_042117401.1	ZNF326	
WP_042117401.1	SRSF11	
WP_042117401.1	TACC1	
WP_042117401.1	TSN	
WP_042117401.1	FAM9A	
WP_042117401.1	FOXJ2	1
WP_042117401.1	C12orf68	1
WP_044177448.1	UBQLN2	
WP_044180562.1	PSTPIP1	
WP_044183152.1	TAF7L	
WP_050812366.1	PROP1	
WP_050812366.1	PAX5	
WP_055696404.1	PICK1	
WP_055696404.1	PDE4DIP	
WP_062772817.1	CEP250	1
WP_062773522.1	KIFC3	
WP_062773522.1	ZFP161	
WP_062773651.1	EVI5	
WP_071777502.1	KCTD6	1
WP_071825927.1	CBY1	
WP_071825927.1	ZBTB1	
WP_072041464.1	DYNLT1	1
WP_072041472.1	SSBP3	
WP_077626056.1	DYNLT1	1
WP_080721914.1	IKBKG	
WP_084596144.1	ZBED1	
WP_108474309.1	CCDC57	
WP_108474309.1	KRTAP10-5	1
WP_113857569.1	REL	
WP_113858483.1	L3MBTL3	
WP_113859080.1	PAIP1	
WP_115333225.1	C15orf55	
WP_115333225.1	GOLGA6L9	

Table S9 | Reference sets tested in yN2H. *Configuration is indicated by N1-X and N2-Y. Interaction indicated by "1" if log$_2$ NLR ≥ 0.*

N1-X	N2-Y	dataset	interaction according to NLR (log$_2$) ≥ 0
TRNT1	Q8XAL7	bhLit-BM-v1	
CASP9	Q8XAL7	bhLit-BM-v1	
CLK1	Q8XA11	bhLit-BM-v1	
RHPN1	P0AE20	bhLit-BM-v1	
DNAJB11	D0ZRB2	bhLit-BM-v1	
DHFR	Q8XAL7	bhLit-BM-v1	
UBE2D1	Q99PZ6	bhLit-BM-v1	
POLR2E	Q8XBX8	bhLit-BM-v1	
PDE6D	Q7DB77	bhLit-BM-v1	
RELA	P18009	bhLit-BM-v1	
ZNHIT1	Q8XA11	bhLit-BM-v1	
DRG2	Q8XBX8	bhLit-BM-v1	
HPCAL4	Q7DB77	bhLit-BM-v1	
D0ZRB2	DNAJB11	bhLit-BM-v1	
ERI3	Q8XAL7	bhLit-BM-v1	
PENK	Q8XAJ5	bhLit-BM-v1	
PTP4A1	Q8XAJ5	bhLit-BM-v1	
MAD2L2	P18011	bhLit-BM-v1	
Q8XA11	ZNHIT1	bhLit-BM-v1	
Q8XAJ5	PENK	bhLit-BM-v1	
RIC8A	Q8XB62	bhLit-BM-v1	
KLC1	A0A0F6B5H5	bhLit-BM-v1	
PSMC1	Q8XA11	bhLit-BM-v1	
Q8XAL7	ERI3	bhLit-BM-v1	
LMO4	Q8XAL7	bhLit-BM-v1	
CDC42	Q63K41	bhLit-BM-v1	
MRFAP1L1	Q8XB62	bhLit-BM-v1	
P18009	RELA	bhLit-BM-v1	
FRMD3	Q8XAJ5	bhLit-BM-v1	
RAC1	Q63K41	bhLit-BM-v1	
Q7DB77	HPCAL4	bhLit-BM-v1	
Q56935	RACK1	bhLit-BM-v1	
Q8XAL7	CASP8	bhLit-BM-v1	
CDC42	O52623	bhLit-BM-v1	
Q8XAJ5	PTP4A1	bhLit-BM-v1	
P0AE20	RHPN1	bhLit-BM-v1	
Q63K41	CDC42	bhLit-BM-v1	
CASP8	Q8XAL7	bhLit-BM-v1	
Q8XBX8	DRG2	bhLit-BM-v1	
Q8XAL7	CASP9	bhLit-BM-v1	
Q8XB62	RIC8A	bhLit-BM-v1	
ARRB1	Q7DB77	bhLit-BM-v1	
Q8XB62	MRFAP1L1	bhLit-BM-v1	
DSCR4	Q8XAJ5	bhLit-BM-v1	
RHOA	A0A0F6B1Q8	bhLit-BM-v1	
UBE2D3	Q99PZ6	bhLit-BM-v1	
Q8XBX8	POLR2E	bhLit-BM-v1	
HMGN2	Q8XAL7	bhLit-BM-v1	
BAIAP2L1	Q7DB77	bhLit-BM-v1	
CASP4	Q8XAL7	bhLit-BM-v1	
Q8XAL7	TRNT1	bhLit-BM-v1	
Q7DB77	PDE6D	bhLit-BM-v1	
Q7DB77	ARRB1	bhLit-BM-v1	
UBE2D2	Q99PZ6	bhLit-BM-v1	
Q7DB77	BAIAP2L1	bhLit-BM-v1	
P18011	MAD2L2	bhLit-BM-v1	
RNF5	Q8ZNR3	bhLit-BM-v1	
Q99PZ6	UBE2D3	bhLit-BM-v1	
Q63K41	RAC1	bhLit-BM-v1	
STK16	B7UM99	bhLit-BM-v1	
Q8XA11	CDKN2AIPNL	bhLit-BM-v1	
Q8XAL7	LMO4	bhLit-BM-v1	
SEC24D	Q8XAJ5	bhLit-BM-v1	
B7UM99	STK16	bhLit-BM-v1	
Q8XAL7	CASP4	bhLit-BM-v1	
Q8XBX8	RIPK1	bhLit-BM-v1	
B7UR60	SNTA1	bhLit-BM-v1	
Q8XAL7	HMGN2	bhLit-BM-v1	
B7UM99	HPCAL1	bhLit-BM-v1	
B7UI21	FADD	bhLit-BM-v1	
Q99PZ6	UBE2D1	bhLit-BM-v1	
Q8XBX8	FADD	bhLit-BM-v1	

N1-X	N2-Y	dataset	interaction according to NLR (log$_2$) ≥ 0
NMI	Q7DB50	bhLit-BM-v1	
B7UM99	KRT18	bhLit-BM-v1	
A0A0N9NCU6	MAP2K2	bhLit-BM-v1	
B7UI21	RIPK1	bhLit-BM-v1	
B7UR60	SLC9A3R2	bhLit-BM-v1	
TXN	D0ZRB2	bhLit-BM-v1	1
Q8VSC3	IKBKG	bhLit-BM-v1	1
Q8XB62	CENPH	bhLit-BM-v1	1
Q8XAJ5	FRMD3	bhLit-BM-v1	1
Q8XA11	PSMC1	bhLit-BM-v1	1
Q7DB50	NMI	bhLit-BM-v1	1
HPCAL1	B7UM99	bhLit-BM-v1	1
Q8ZNR3	RNF5	bhLit-BM-v1	1
Q99PZ6	UBE2D2	bhLit-BM-v1	1
Q8XAL7	DHFR	bhLit-BM-v1	1
D0ZRB2	TXN	bhLit-BM-v1	1
Q7DB68	SLC7A5	bhLit-BM-v1	1
A0A0F6B1Q8	RHOA	bhLit-BM-v1	1
Q8XAJ5	SEC24D	bhLit-BM-v1	1
NCALD	B7UM99	bhLit-BM-v1	1
A0A0F6B5H5	KLC1	bhLit-BM-v1	1
B7UM99	NCALD	bhLit-BM-v1	1
met_156	BAZ2B	bhRRS-v1	
SEPP1	Yre_13	bhRRS-v1	
CSN3	Mmo_44	bhRRS-v1	
MED24	Pre_21	bhRRS-v1	
LOC144742	Pre_13	bhRRS-v1	
TBCCD1	Yen_54	bhRRS-v1	
ORC6	Pre_9	bhRRS-v1	
Efe_12	C2orf61	bhRRS-v1	
Ec2_4	ZNF140	bhRRS-v1	
DCAF16	Pst_81	bhRRS-v1	
Pma_7	NEK5	bhRRS-v1	
NEK5	Pma_7	bhRRS-v1	
PAPPA2	Pst_60	bhRRS-v1	
TGOLN2	B7UM88	bhRRS-v1	
SCP2	Pre_25	bhRRS-v1	
PATZ1	Kpn_27	bhRRS-v1	
C2orf61	Efe_12	bhRRS-v1	
Pre_2	IPO13	bhRRS-v1	
Yre_13	SEPP1	bhRRS-v1	
met_36	COQ6	bhRRS-v1	
Pst_52	ACER2	bhRRS-v1	
TRDV2	Ec6_20	bhRRS-v1	
Pre_22	PCDHGB5	bhRRS-v1	
PCDHGB5	Pre_22	bhRRS-v1	
Pst_60	PAPPA2	bhRRS-v1	
Pst_40	FASTKD2	bhRRS-v1	
SYCP3	Yre_26	bhRRS-v1	
Pre_21	MED24	bhRRS-v1	
PTPRM	A0A0F6B5H5	bhRRS-v1	
met_50	SPIN2B	bhRRS-v1	
ARL3	Vfu_3	bhRRS-v1	
SLC35G1	Vfu_31	bhRRS-v1	
COA5	Yre_32	bhRRS-v1	
SETD9	Vfu_52	bhRRS-v1	
C17orf64	Ec6_23	bhRRS-v1	
SPANXC	Ec6_2	bhRRS-v1	
FAM72B	Yen_7	bhRRS-v1	
met_162	TPPP2	bhRRS-v1	
Yen_54	TBCCD1	bhRRS-v1	
Vfu_52	SETD9	bhRRS-v1	
Yen_7	FAM72B	bhRRS-v1	
Yre_31	OLFML1	bhRRS-v1	
LPAR1	Pst_3	bhRRS-v1	
met_117	OMA1	bhRRS-v1	
met_54	AHDC1	bhRRS-v1	
CD19	Kpn_31	bhRRS-v1	
Pre_13	LOC144742	bhRRS-v1	
Pre_5	TMEM60	bhRRS-v1	
Pst_56	YPEL5	bhRRS-v1	
Cda_9	KRT33B	bhRRS-v1	
Vfu_3	ARL3	bhRRS-v1	
CORO1B	Ec6_50	bhRRS-v1	
KCNJ6	Cyo_47	bhRRS-v1	
met_71	FBXL7	bhRRS-v1	

N1-X	N2-Y	dataset	interaction according to NLR (log$_2$) ≥ 0
KAZN	Mmo_35	bhRRS-v1	
HELLS	met_34	bhRRS-v1	
EPS15L1	Pst_80	bhRRS-v1	
PDE6D	Pre_51	bhRRS-v1	
FAHD2A	Pfa_6	bhRRS-v1	
EXOSC5	met_126	bhRRS-v1	
BAZ2B	met_156	bhRRS-v1	
A0A6M7GVE3	GADD45G	bhRRS-v1	
Yre_32	COA5	bhRRS-v1	
Cyo_49	GSTO2	bhRRS-v1	
B7UM88	TGOLN2	bhRRS-v1	
GSTO2	Cyo_49	bhRRS-v1	
USP1	Pma_11	bhRRS-v1	
Ec6_29	L2HGDH	bhRRS-v1	
Kpn_36	ACAD8	bhRRS-v1	
DOCK3	Pre_70	bhRRS-v1	
AHDC1	met_54	bhRRS-v1	
TTLL13	met_81	bhRRS-v1	
Ec6_23	C17orf64	bhRRS-v1	
Mmo_42	RBM33	bhRRS-v1	
Cda_19	DPP9	bhRRS-v1	
Pfa_6	FAHD2A	bhRRS-v1	
Pre_25	SCP2	bhRRS-v1	
Pre_70	DOCK3	bhRRS-v1	
Cyo_16	PRPF38A	bhRRS-v1	
RBMY2FP	Ec6_26	bhRRS-v1	
RPS26	Pre_54	bhRRS-v1	
RBM33	Mmo_42	bhRRS-v1	
Pma_11	USP1	bhRRS-v1	
NAALADL1	Pfa_19	bhRRS-v1	
Ec6_2	SPANXC	bhRRS-v1	
Vfu_31	SLC35G1	bhRRS-v1	
COLEC12	Pfa_35	bhRRS-v1	
Pre_9	ORC6	bhRRS-v1	
FLJ38668	met_124	bhRRS-v1	
met_110	KCNH8	bhRRS-v1	
Vfu_32	RDM1	bhRRS-v1	
Pfa_35	COLEC12	bhRRS-v1	
Kpn_31	CD19	bhRRS-v1	
GADD45G	A0A6M7GVE3	bhRRS-v1	
Kpn_27	PATZ1	bhRRS-v1	
ELL3	Cda_31	bhRRS-v1	
Pst_3	LPAR1	bhRRS-v1	
Pre_51	PDE6D	bhRRS-v1	
Efe_25	DEFB115	bhRRS-v1	
Mmo_35	KAZN	bhRRS-v1	
Cda_11	C4orf27	bhRRS-v1	
Mmo_1	SSTR3	bhRRS-v1	
SSTR3	Mmo_1	bhRRS-v1	
PRPF38A	Cyo_16	bhRRS-v1	
ACOT12	Yre_14	bhRRS-v1	
met_124	FLJ38668	bhRRS-v1	
met_66	AKAP11	bhRRS-v1	
met_34	HELLS	bhRRS-v1	
KCNH8	met_110	bhRRS-v1	
Pre_48	PRNP	bhRRS-v1	
Pst_81	DCAF16	bhRRS-v1	
Ec6_26	RBMY2FP	bhRRS-v1	
Pre_11	KIAA0930	bhRRS-v1	
C4orf27	Cda_11	bhRRS-v1	
Mmo_44	CSN3	bhRRS-v1	
Pre_38	RPS15	bhRRS-v1	
Cda_22	ECHDC3	bhRRS-v1	
DEFB115	Efe_25	bhRRS-v1	
RDM1	Vfu_32	bhRRS-v1	
FASTKD2	Pst_40	bhRRS-v1	
met_81	TTLL13	bhRRS-v1	
YPEL5	Pst_56	bhRRS-v1	
Cyo_33	DGAT2	bhRRS-v1	
C7orf60	Ec2_8	bhRRS-v1	
SYNGR2	Yen_51	bhRRS-v1	
AKAP11	met_66	bhRRS-v1	
Yre_26	SYCP3	bhRRS-v1	
RPS15	Pre_38	bhRRS-v1	
Yre_14	ACOT12	bhRRS-v1	
Pma_23	DNM3	bhRRS-v1	

N1-X	N2-Y	dataset	interaction according to NLR (log$_2$) ≥ 0
met_46	TMEM79	bhRRS-v1	
Pre_54	RPS26	bhRRS-v1	
met_126	EXOSC5	bhRRS-v1	
Ec2_8	C7orf60	bhRRS-v1	
Pfa_19	NAALADL1	bhRRS-v1	
Cyo_20	HDAC10	bhRRS-v1	
met_2	PIPOX	bhRRS-v1	
Yre_17	PLK2	bhRRS-v1	
Cda_31	ELL3	bhRRS-v1	1
met_119	TMEM230	bhRRS-v1	1
Yen_51	SYNGR2	bhRRS-v1	1
Cyo_47	KCNJ6	bhRRS-v1	1
PLK2	Yre_17	bhRRS-v1	1
DGAT2	Cyo_33	bhRRS-v1	1
FEN1	PCNA	hsPRS-v2	
ORC2L	MCM10	hsPRS-v2	
LMNA	RB1	hsPRS-v2	
RAC1	ARFIP2	hsPRS-v2	
PDE4D	RACK1	hsPRS-v2	
XIAP	CASP7	hsPRS-v2	
SMAD1	SMAD4	hsPRS-v2	
ERBB3	NRG1	hsPRS-v2	
XIAP	CASP3	hsPRS-v2	
XIAP	CASP9	hsPRS-v2	
GRB2	LAT	hsPRS-v2	
BDNF	NTF4	hsPRS-v2	
TP53	UBE2I	hsPRS-v2	
ARF1	ARFIP2	hsPRS-v2	
CEBPG	FOS	hsPRS-v2	
DDIT3	FOS	hsPRS-v2	
MCM2	MCM3	hsPRS-v2	
CGA	CGB5	hsPRS-v2	
PPP3CA	PPP3R1	hsPRS-v2	
MCM2	MCM5	hsPRS-v2	
NF2	HGS	hsPRS-v2	
IFIT1	EIF3E	hsPRS-v2	
PEX19	PEX11B	hsPRS-v2	
ATF3	DDIT3	hsPRS-v2	
CASP2	CRADD	hsPRS-v2	
CBLB	GRB2	hsPRS-v2	
GADD45A	PCNA	hsPRS-v2	
LCP2	NCK1	hsPRS-v2	
HBA2	HBB	hsPRS-v2	
JUNB	BATF	hsPRS-v2	
GTF2F1	GTF2F2	hsPRS-v2	
AKT1	PDPK1	hsPRS-v2	
PEX14	PEX19	hsPRS-v2	
SMAD4	DCP1A	hsPRS-v2	1
SKP1	SKP2	hsPRS-v2	1
RIPK2	NOD1	hsPRS-v2	1
LCP2	GRAP2	hsPRS-v2	1
SKP1	BTRC	hsPRS-v2	1
CDK2	CKS1B	hsPRS-v2	1
GRB2	VAV1	hsPRS-v2	1
BAK1	BCL2L1	hsPRS-v2	1
PEX19	PEX3	hsPRS-v2	1
BAD	BCL2L1	hsPRS-v2	1
DR1	DRAP1	hsPRS-v2	1
OSM	ZNF688	hsRRS-v2	
RGR	ABCF3	hsRRS-v2	
PSMD5	SLC22A15	hsRRS-v2	
ARSA	DBN1	hsRRS-v2	
PPP6C	ZNF350	hsRRS-v2	
NUDT2	MIIP	hsRRS-v2	
CLPTM1	L3MBTL2	hsRRS-v2	
MCM2	PLXNA4	hsRRS-v2	
PROS1	STK25	hsRRS-v2	
FIGF	ZBTB25	hsRRS-v2	
FAS	LSM3	hsRRS-v2	
PMCH	RIC3	hsRRS-v2	
ERBB3	C3orf38	hsRRS-v2	
ITPA	WDR62	hsRRS-v2	
FABP4	GCG	hsRRS-v2	
CD81	NPC2	hsRRS-v2	
PSMD5	SYCE1	hsRRS-v2	
NFIB	TIRAP	hsRRS-v2	

N1-X	N2-Y	dataset	interaction according to NLR (log$_2$) ≥ 0
DUT	C19orf40	hsRRS-v2	
INPP1	UBLCP1	hsRRS-v2	
GRIK2	ARL6IP6	hsRRS-v2	
SCARB1	PHF21B	hsRRS-v2	
DLX4	RAB3IP	hsRRS-v2	
CNN1	PGAP2	hsRRS-v2	
BTC	BAT2L1	hsRRS-v2	
BYSL	KIAA0907	hsRRS-v2	
PSMD12	CRIPT	hsRRS-v2	
DEFA3	TSTD2	hsRRS-v2	
PDHB	ZC3HC1	hsRRS-v2	
BMP5	C10orf119	hsRRS-v2	
RFX3	SEMA4G	hsRRS-v2	
COPB1	HPCAL4	hsRRS-v2	
MOBP	MRPS25	hsRRS-v2	
RHOC	NUP62CL	hsRRS-v2	
NDP	NUDT4	hsRRS-v2	
GPR18	HNRPLL	hsRRS-v2	
RXRB	CXCL11	hsRRS-v2	
ACVR1	CWF19L1	hsRRS-v2	
GPD2	C22orf29	hsRRS-v2	
COPB1	SLC39A14	hsRRS-v2	
MAOB	CTCF	hsRRS-v2	
EMD	ARMC1	hsRRS-v2	
CD34	SNX21	hsRRS-v2	
LUM	UGGT2	hsRRS-v2	
PDE9A	REM2	hsRRS-v2	
CANX	ARHGEF15	hsRRS-v2	
NONO	BEST1	hsRRS-v2	
HCLS1	SALL2	hsRRS-v2	
CANX	ATAD2	hsRRS-v2	
ATP5O	CLEC2D	hsRRS-v2	1
ASS1	GINS3	hsRRS-v2	1

Table S10 | Statistical analysis of the data obtained by the yN2H. *Dataset 1 and dataset 2 were tested together using the Fisher's exact test. P-values < 0.05 indicate that the two datasets are statistically different (bold).*

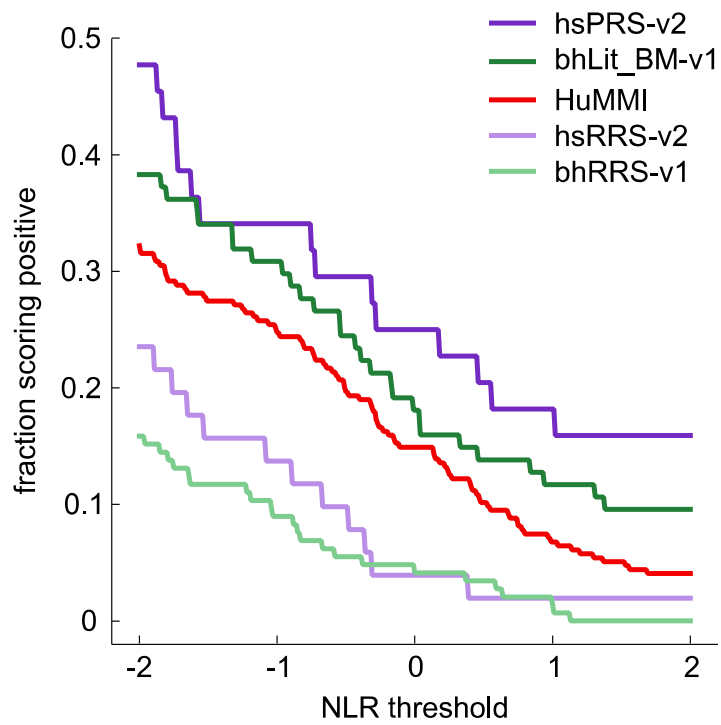| dataset 1	dataset 2	p-value
HuMMI	bhLit-BM-v1	0.5148
HuMMI	bhRRS-v1	**0.000644**
bhLit-BM-v1	bhRRS-v1	**0.000543**
HuMMI	hsPRS-v2	0.1218
bhLit-BM-v1	hsPRS-v2	0.36940
hsPRS-v2	hsRRS-v2	**0.005263**
HuMMI	hsRRS-v2	**0.04147**

Figure S1 | yN2H detection rates. *Fraction scoring positive of the four reference sets and the subset of HuMMI depending on the normalized luminescence ratio (NLR) threshold.*

Table S11 | Clustering of effectors according to sequence similarity. *"similarity (%)" shows the percentage of sequence similarity that the respective effector displays to the most dissimilar effector in the cluster.*

effector ID	cluster ID	similarity (%)
WP_042031545.1	0	58.78
WP_042032213.1	0	56.88
WP_005131699.1	0	57.40
WP_000099375.1	0	58.20
WP_005283191.1	0	57.85
WP_004145486.1	0	56.88
WP_004904012.1	0	58.66
WP_108476339.1	0	68.69
WP_113858928.1	0	57.85
WP_004256437.1	0	65.69
WP_014609336.1	0	58.31
WP_006818175.1	0	58.84
WP_006818482.1	0	67.27
WP_016535835.1	1	77.00
WP_005129057.1	1	75.00
WP_000020887.1	1	68.04
WP_000020875.1	1	67.01
WP_000020896.1	1	70.71
WP_044177605.1	1	76.00
WP_108475752.1	1	67.01
WP_033789480.1	1	70.71
WP_038258270.1	1	75.00
WP_005129187.1	2	81.16
WP_040229899.1	2	80.60
WP_000097400.1	2	80.46
WP_044173054.1	2	81.37
WP_005158896.1	2	80.46
WP_006820832.1	2	81.13
dime_meta_effector_65	2	81.34
WP_005129207.1	3	47.97
WP_000255032.1	3	45.12
WP_000937458.1	3	47.56
WP_044173012.1	3	47.56
WP_052332698.1	3	45.12
WP_005162291.1	3	49.11
WP_006820847.1	3	47.97
WP_016535238.1	4	86.64
WP_005126712.1	4	86.25
WP_035595442.1	4	84.77
WP_004235986.1	4	87.29
WP_004236694.1	4	84.77
WP_014657167.1	4	84.80
WP_005161939.1	4	84.49
WP_075208399.1	5	76.66
dime_meta_effector_105	5	76.66
dime_meta_effector_11	5	76.66
dime_meta_effector_146	5	77.57
dime_meta_effector_20	5	76.89
dime_meta_effector_3	5	78.03
WP_016536247.1	6	83.96
WP_002909008.1	6	81.23
WP_113858462.1	6	80.00
WP_004254561.1	6	84.00
WP_004727750.1	6	80.00
dime_meta_effector_155	6	83.69
WP_016517519.1	7	77.34
WP_044179949.1	7	79.14
WP_002916742.1	7	78.06
WP_113858981.1	7	79.50
WP_004919841.1	7	77.34
WP_006818297.1	7	78.78
WP_009486529.1	8	50.98
WP_004905318.1	8	50.11

effector ID	cluster ID	similarity (%)
WP_004922117.1	8	52.17
WP_004927350.1	8	50.76
WP_005157433.1	8	50.33
WP_077626056.1	9	64.19
dime_meta_effector_129	9	63.98
dime_meta_effector_17	9	68.99
dime_meta_effector_4	9	62.80
dime_meta_effector_79	9	62.80
WP_000961342.1	10	78.10
dime_meta_effector_141	10	76.89
dime_meta_effector_145	10	78.10
dime_meta_effector_15	10	76.89
dime_meta_effector_64	10	79.35
WP_000083477.1	11	63.60
WP_000075087.1	11	63.95
WP_001093944.1	11	63.60
WP_000083435.1	11	63.60
WP_113859044.1	11	63.95
dime_meta_effector_134	12	54.76
dime_meta_effector_29	12	58.45
dime_meta_effector_67	12	56.08
dime_meta_effector_81	12	63.66
dime_meta_effector_95	12	54.76
WP_001016304.1	13	60.43
WP_004241218.1	13	51.67
WP_004262300.1	13	51.67
WP_006817680.1	13	60.66
dime_meta_effector_86	13	53.20
WP_044178555.1	14	93.49
WP_004197606.1	14	93.15
WP_033751802.1	14	92.81
WP_005158416.1	14	94.52
WP_006819026.1	14	92.81
WP_039898535.1	15	65.29
WP_044172624.1	15	67.42
WP_108473469.1	15	64.44
WP_005164848.1	15	60.87
WP_038254823.1	15	59.55
WP_000191565.1	16	86.76
WP_000191595.1	16	87.12
WP_044183301.1	16	86.58
WP_113858376.1	16	86.76
WP_004261691.1	17	47.24
WP_004919165.1	17	49.74
WP_004924913.1	17	47.24
WP_005179029.1	17	51.94
WP_042033505.1	18	73.11
WP_005123605.1	18	73.11
WP_000189184.1	18	73.95
WP_108474640.1	18	75.21
WP_004864811.1	19	71.08
WP_040230605.1	19	71.08
WP_108473781.1	19	73.49
WP_033791692.1	19	72.29
WP_004922318.1	20	53.86
WP_004927136.1	20	53.97
WP_004927466.1	20	53.86
WP_005160863.1	21	50.31
dime_meta_effector_181	21	51.92
dime_meta_effector_182	21	51.91
WP_000077829.1	22	90.15
WP_000077885.1	22	89.88
WP_113858661.1	22	89.88
WP_001298077.1	23	88.85
WP_001445816.1	23	89.03
WP_040902687.1	23	88.85

effector ID	cluster ID	similarity (%)
WP_062773486.1	24	58.01
WP_004925180.1	24	58.01
WP_005158077.1	24	62.23
WP_000155738.1	25	92.57
WP_000178797.1	25	91.97
WP_005288481.1	25	91.97
dime_meta_effector_62	26	73.64
dime_meta_effector_84	26	78.10
dime_meta_effector_90	26	72.06
dime_meta_effector_27	27	93.06
dime_meta_effector_49	27	93.06
dime_meta_effector_61	27	93.54
WP_004912645.1	28	91.91
WP_033753922.1	28	91.91
WP_004921358.1	28	92.95
WP_016517628.1	29	95.76
WP_040232968.1	29	96.61
WP_040902689.1	29	95.76
WP_009485462.1	30	54.14
WP_004264896.1	30	50.00
WP_004729399.1	30	49.23
WP_016537145.1	31	93.02
WP_044173357.1	31	94.90
WP_000155904.1	31	93.02
WP_004238584.1	32	76.14
WP_004261765.1	32	74.91
WP_006818304.1	32	74.91
WP_001067519.1	33	89.70
WP_001067513.1	33	89.70
WP_108474137.1	33	89.70
WP_005354370.1	34	90.75
WP_016517593.1	34	94.44
WP_000781397.1	34	90.75
WP_002889316.1	35	78.19
WP_004242218.1	35	80.93
WP_004916773.1	35	78.19
WP_004921504.1	36	53.10
WP_004727345.1	36	52.54
dime_meta_effector_85	36	52.54
WP_000611436.1	37	55.20
WP_000611426.1	37	55.20
WP_014258130.1	37	55.20
WP_001516695.1	38	80.28
WP_038151811.1	38	80.28
dime_meta_effector_73	38	80.28
WP_040233404.1	39	80.83
WP_000189224.1	39	78.76
WP_004177339.1	39	78.76
WP_044177823.1	40	63.47
WP_036957904.1	40	61.08
WP_005160046.1	40	66.06
WP_040232101.1	41	86.39
WP_001237041.1	41	83.56
WP_044180054.1	41	83.56
WP_044180332.1	42	81.48
WP_015963250.1	42	81.48
WP_006818522.1	42	81.48
WP_023184674.1	43	73.61
WP_072041520.1	43	73.61
WP_071777494.1	43	75.00
WP_036416809.1	44	85.71
dime_meta_effector_178	44	85.71
WP_000083190.1	45	90.01
WP_081874660.1	45	90.01
dime_meta_effector_168	46	98.43
dime_meta_effector_184	46	98.43

effector ID	cluster ID	similarity (%)
WP_020317218.1	47	87.83
WP_036413302.1	47	87.83
WP_000859964.1	48	76.46
WP_005165873.1	48	76.46
WP_004264507.1	49	96.31
WP_004924795.1	49	96.31
dime_meta_effector_163	50	97.97
dime_meta_effector_165	50	97.97
dime_meta_effector_167	51	85.01
dime_meta_effector_173	51	85.33
WP_004729371.1	52	100.00
WP_004729624.1	52	100.00
dime_meta_effector_118	53	77.07
dime_meta_effector_119	53	77.07
WP_039898721.1	54	81.39
dime_meta_effector_138	54	81.39
WP_009484993.1	55	91.79
WP_014609219.1	55	91.79
dime_meta_effector_13	56	92.86
dime_meta_effector_76	56	92.86
dime_meta_effector_25	57	94.30
dime_meta_effector_43	57	94.30
dime_meta_effector_127	58	84.63
dime_meta_effector_82	58	84.63
WP_040232565.1	59	88.89
WP_005164542.1	59	88.89
WP_001335297.1	60	70.93
WP_000208170.1	60	70.93
WP_000155927.1	61	85.15
dime_meta_effector_139	61	85.15
WP_040230127.1	62	69.63
WP_005159272.1	62	69.63
dime_meta_effector_21	63	83.86
dime_meta_effector_60	63	83.86
WP_004924497.1	64	83.71
dime_meta_effector_143	64	83.71
WP_004915373.1	65	99.55
WP_004918480.1	65	99.55
WP_044177883.1	66	90.02
WP_004918153.1	66	90.02
WP_004254299.1	67	92.14
WP_004920058.1	67	92.14
WP_000106767.1	68	88.43
WP_004906207.1	68	88.43
dime_meta_effector_35	69	87.53
dime_meta_effector_9	69	87.53
WP_000139103.1	70	73.45
WP_004915527.1	70	72.78
WP_044184806.1	71	85.50
WP_004260347.1	71	85.50
WP_000786551.1	72	98.22
WP_000786561.1	72	98.22
WP_073970177.1	73	94.33
dime_meta_effector_72	73	94.33
WP_004257114.1	74	89.06
WP_004917597.1	74	89.06
WP_000183751.1	75	63.25
dime_meta_effector_123	75	63.11
dime_meta_effector_157	76	92.88
dime_meta_effector_24	76	92.88
WP_004906048.1	77	89.66
WP_004915569.1	77	89.66
WP_001445771.1	78	66.67
dime_meta_effector_130	78	66.67
WP_004728083.1	79	69.17
WP_005159058.1	79	69.17

effector ID	cluster ID	similarity (%)
WP_000258580.1	80	91.11
dime_meta_effector_37	80	91.11
WP_004147894.1	81	86.44
WP_071777524.1	81	86.44
WP_004256728.1	82	56.60
dime_meta_effector_133	82	56.60
WP_004906105.1	83	95.45
WP_004915478.1	83	95.45
WP_040232376.1	84	71.14
dime_meta_effector_107	84	71.14
WP_019705807.1	85	91.25
WP_004727751.1	85	91.25
dime_meta_effector_66	86	60.64
dime_meta_effector_89	86	60.64
WP_000057389.1	87	99.71
WP_000057374.1	87	99.71
WP_004265131.1	88	90.48
WP_004924254.1	88	90.48
WP_040231861.1	89	77.46
WP_000241053.1	89	77.46
WP_002889847.1	90	91.69
WP_113858483.1	90	91.69
WP_004152062.1	91	100.00
WP_000817037.1	91	100.00
WP_001298103.1	92	99.67
WP_000981716.1	92	99.67
WP_004923882.1	93	62.29
WP_006818941.1	93	62.29
WP_004263203.1	94	77.38
WP_014657126.1	94	78.03
WP_004254983.1	95	92.20
WP_004919494.1	95	92.20
WP_000582830.1	96	97.81
WP_001445845.1	96	97.81
WP_044183152.1	97	53.31
dime_meta_effector_148	97	53.31
WP_108475013.1	98	91.39
WP_006818167.1	98	91.39
WP_005164084.1	99	87.92
WP_006817197.1	99	87.92
WP_001445815.1	100	98.11
dime_meta_effector_93	100	98.11
WP_036957743.1	101	92.22
WP_004919757.1	101	92.22
WP_001066218.1	102	90.16
WP_005287789.1	102	90.16
WP_004255002.1	103	54.26
WP_004258949.1	103	54.26
WP_062771682.1	104	82.98
WP_004254943.1	104	82.98
WP_004264927.1	105	54.63
WP_004922581.1	105	54.63
WP_002916607.1	106	88.48
WP_015963067.1	106	88.48
WP_000873388.1	107	90.15
WP_113858620.1	107	90.15
WP_004238854.1	108	93.75
WP_000116680.1	108	93.75
dime_meta_effector_147	109	93.16
dime_meta_effector_159	109	93.16
WP_000868324.1	110	93.12
dime_meta_effector_16	110	93.12
WP_004242398.1	111	99.47
WP_024474672.1	111	99.47
WP_004242347.1	112	92.76
WP_036957864.1	112	92.76

| effector ID	cluster ID	similarity (%)
WP_042031532.1	113	84.17
dime_meta_effector_99	113	84.17
dime_meta_effector_42	114	83.46
dime_meta_effector_75	114	83.46
WP_082031767.1	115	92.45
WP_001147116.1	115	92.24
WP_004258336.1	116	88.68
WP_004925512.1	116	88.68
WP_016537909.1	117	55.24
WP_005120762.1	117	55.24
WP_005163729.1	118	93.33
WP_014609009.1	118	93.33
WP_113857302.1	119	100.00
WP_113857569.1	119	100.00
WP_044183672.1	120	94.67
WP_015962672.1	120	94.67
WP_003844491.1	121	91.94
WP_044182945.1	121	91.94
WP_081653585.1	122	90.57
WP_108475618.1	122	90.57
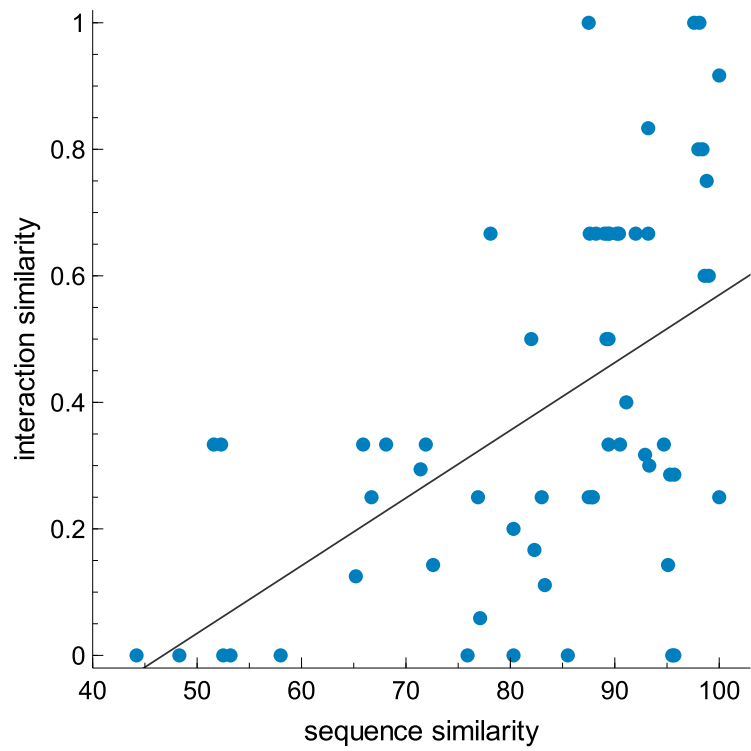
Figure S2 | Sequence similarity versus interaction similarity of homologous effectors. *Each dot represents the comparison between two homologous effectors having a union of at least three human interactors. The Spearman correlation coefficient ρ shows a moderate positive correlation (0.536).*

Table S12 | Human proteins subject to convergence. *Human proteins targeted by ≥ 4 strains. Column "in HuRI" indicates whether the human protein is part of HuRI. Proteins part of HuRI (TRUE) are included in the functional enrichment analysis as HuRI is used as background.*

human protein	in HuRI
AGR2	TRUE
BANP	TRUE
CCDC102B	TRUE
CCNDBP1	TRUE
CEP250	FALSE
CLTCL1	FALSE
COG6	TRUE
COL17A1	TRUE
DYNLT1	TRUE
EFHC2	TRUE
GIGYF1	TRUE
GOLGA2	TRUE
GOLGA6L9	TRUE
HOMEZ	TRUE
HSF2BP	TRUE
IKZF3	TRUE
PATJ	TRUE
KCTD6	TRUE
KCTD9	TRUE
KIFC3	TRUE
KLHL2	TRUE
KRT27	TRUE
KRT31	TRUE
KRT75	TRUE
KRT76	TRUE
KRTAP10-5	TRUE
LBX1	TRUE
LZTS2	TRUE
MFF	TRUE
MID2	TRUE
MTUS2	TRUE
NECAB1	TRUE
NECAB2	TRUE
NOTO	TRUE
PAX5	TRUE
PAX6	TRUE
PICK1	TRUE
PNMA1	TRUE
PPP1R13B	TRUE
PROP1	TRUE
REL	TRUE
SERTAD1	TRUE
SP4	TRUE
SPAG5	TRUE
TCF4	TRUE
TFIP11	TRUE
TNIP1	TRUE
TRAF1	TRUE
TRAF2	TRUE
TRIM27	TRUE
TSN	TRUE
UBAP1	FALSE
UBQLN1	TRUE
UBQLN2	TRUE
USHBP1	TRUE
USP54	TRUE
VAC14	TRUE
VIM	TRUE
VPS37B	TRUE
VPS52	TRUE
ZBED1	TRUE
ZBED4	FALSE
ZBTB10	TRUE
ZRANB1	TRUE

Table S13 | Functional enrichment analysis. *Terms detected in GO:Biological Process database. Term size refers to the number of genes that are annotated with the term in HuRI. Intersection size describes the number of genes that are annotated with the term in HuMMI_main and "intersecting effector targets" lists the respective genes. Odds ratio was calculated as described in Chapter 4.21. FDR indicates adjusted p-values.*

term name	term ID	term size	Intersection size	intersecting effector targets	odds ratio	FDR
intermediate filament-based process	GO:0045103	58	5	KRT31,KRT75,KRT76,KRT27,VIM	12.71	0.01492
intermediate filament cytoskeleton organization	GO:0045104	58	5	KRT31,KRT75,KRT76,KRT27,VIM	12.71	0.01492
intermediate filament organization	GO:0045109	54	5	KRT31,KRT75,KRT76,KRT27,VIM	13.66	0.01492
regulation of glycoprotein biosynthetic process	GO:0010559	15	3	GOLGA2,NECAB1,NECAB2	28.46	0.03238
regulation of glycoprotein metabolic process	GO:1903018	18	3	GOLGA2,NECAB1,NECAB2	23.71	0.03433
regulation of I-kappaB kinase/NF-kappaB signaling	GO:0043122	124	6	TRAF2,TRIM27,REL,MID2,TNIP1,TRAF1	7.22	0.03433
cytoskeleton organization	GO:0007010	598	13	EFHC2,TRIM27,DYNLT1,KRT31,KRT75,KRT76, CCDC102B,GOLGA2,ZRANB1,KRT27,PICK1,SPAG5,VIM	3.56	0.03433
I-kappaB kinase/NF-kappaB signaling	GO:0007249	132	6	TRAF2,TRIM27,REL,MID2,TNIP1,TRAF1	6.78	0.03933
regulation of viral process	GO:0050792	87	5	TRIM27,DYNLT1,MID2,TNIP1,VPS37B	8.44	0.03933
regulation of receptor internalization	GO:0002090	22	3	UBQLN2,NECAB2,PICK1	19.39	0.04527

Table S14 | Effectors targeting human proteins involved in host cell apoptosis. *Effectors targeting human proteins involved in apoptosis (targets) that were tested for their impact on cell viability and DNA fragmentation. abbr., abbreviation.*

effector ID	abbr.	targets
WP_004915712.1	Pst_2	TRAF2
WP_004919332.1	Pst_5	BCL6
WP_004919757.1	Pst_6	TRAF6
WP_004922581.1	Pst_9	TAX1BP1, UBQLN1
WP_042117401.1	Pst_17	DNM2, PNMA1, TAX1BP1, VCP
WP_071599648.1	Pst_19	TOX3
WP_004920813.1	Pst_7	ATN1
WP_004924913.1	Pst_11	TRAF3
WP_004926361.1	Pst_14	TRIM2
WP_042116632.1	Pst_16	CREB3L1
WP_040233404.1	Cyo_9	UBQLN1
WP_080721914.1	Cyo_12	TRAF2
WP_005129207.1	Cyo_3	TRAF2
WP_004258336.1	Pre_10	VCP
WP_004264927.1	Pre_20	UBQLN1
WP_001267298.1	Ec6_12	PNMA1
WP_009484876.1	Kpn_10	TAX1BP1
WP_009484324.1	Kpn_9	TRAF2
WP_004177339.1	Kpn_7	UBQLN1
WP_004118237.1	Kpn_3	TRAF2
WP_010891207.1	Yen_17	UBQLN1
WP_005159145.1	Yen_5	MFF
WP_005164542.1	Yen_14	PNMA1
WP_000148644.1	Efe_3	TRAF2
WP_001235473.1	Efe_11	TRAF2, PNMA1
WP_040263598.1	Pem_5	CREB3L1
WP_084596184.1	Pem_8	TRAF2
WP_038151743.1	Vfu_16	MFF
WP_115333225.1	Vfu_21	PNMA1
WP_001298277.1	Ec2_5	MFF
WP_000191595.1	Ec6_4	MFF
WP_004145486.1	Kpn_4	MFF
WP_044180332.1	Pma_10	MFF
WP_044173012.1	Pma_2	TRAF2
WP_044183301.1	Pma_15	MFF
WP_006821080.1	Yre_8	MFF
WP_071777518.1	Yre_14	TRAF2
WP_006818175.1	Yre_2	MFF
WP_004238406.1	Mmo_4	PNMA1
WP_062772817.1	Mmo_11	TRAF3
WP_004234458.1	Mmo_1	MFF
WP_015962672.1	Pfa_5	TRAF2
WP_015963250.1	Pfa_6	MFF
WP_113857302.1	Pfa_11	PNMA1
WP_113857569.1	Pfa_13	PNMA1
WP_005131699.1	Cyo_4	MFF
WP_000255032.1	Efe_5	TRAF2
WP_113857471.1	Pfa_12	PNMA1
WP_006820847.1	Yre_7	TRAF2
WP_016536389.1	Cda_5	MFF
WP_016536523.1	Cda_6	TRAF2
WP_016538154.1	Cda_8	TRAF2
dime_meta_effector_118	met_4	TAX1BP1
dime_meta_effector_132	met_9	UBQLN1
dime_meta_effector_24	met_24	TRAF2
dime_meta_effector_56	met_33	TOX3
dime_meta_effector_130	met_7	TRAF2
dime_meta_effector_150	met_15	PNMA1
dime_meta_effector_157	met_16	TRAF2

Table S15 | Effectors analyzed for their impact on NF-κB activation. *F/R values normalized to the F/R values of the respective control (A20 for TNF-treated cells and IKKβ for untreated cells) were subjected to a statistical analysis to identify significant differences compared to the normalized F/R values of the empty vector (pMH-FLAG-HA). The Kruskal-Wallis test was performed with Dunn's correction and adjusted by FDR-correction. Bold values indicate significant p-values. abbr., abbreviation; p unadj, unadjusted p-value; p adj, adjusted p-value.*

effector ID	abbr.	effector targets	untreated cells			treated cells		
			Z-score	p unadj	p adj	Z-score	p unadj	p adj
WP_080721914.1	Cyo_12	TRAF2, IKBKG	0.287277	0.7739	0.856906	-3.51362	0.000442	**0.000442**
WP_005129207.1	Cyo_3	TRAF2	-1.02757	0.304154	0.47448	-1.27065	0.203854	0.203854
WP_040230127.1	Cyo_6	REL	-1.5303	0.125942	0.24975	-1.18041	0.237836	0.237836
WP_001059674.1	Ec6_10	REL, TRIM27	1.653683	0.098192	0.210155	1.642634	0.100459	0.100459
WP_001267298.1	Ec6_12	REL, IKBKG	2.01462	0.043944	0.115974	-0.56166	0.574346	0.574346
WP_001445815.1	Ec6_17	REL	-1.90413	0.056893	0.137721	-1.81205	0.069978	0.069978
WP_000258580.1	Ec6_6	REL	-1.02757	0.304154	0.476598	0.005525	0.995592	0.995592
WP_000961342.1	Ec6_9	REL	-0.78264	0.433836	0.599513	-0.57455	0.565593	0.565593
WP_001235473.1	Efe_11	REL, MID2, ZRANB1, CARD10, TRAF1, TRAF2, TRIM27	0.57087	0.568087	0.712138	-1.15463	0.248241	0.248241
WP_001237041.1	Efe_12	REL	2.916964	0.003535	**0.019385**	-1.29643	0.194828	0.194828
WP_000148644.1	Efe_3	REL, TRAF2	1.511886	0.130563	0.254598	0.456696	0.647889	0.647889
WP_000255032.1	Efe_5	TRAF2	-1.14358	0.252797	0.412706	-1.83783	0.066087	0.066087
WP_000999547.1	Efe_9	REL	-1.62054	0.105117	0.222266	-2.12143	0.033886	0.033886
WP_009484876.1	Kpn_10	CARD9	2.517355	0.011824	**0.043231**	0.353571	0.72366	0.72366
WP_004118237.1	Kpn_3	REL, TRAF3, TRAF2, TRAF1, ZRANB1, TRIM27, MID2	2.697823	0.006979	**0.03101**	1.294587	0.195463	0.195463
WP_009484324.1	Kpn_9	REL, MID2, ZRANB1, CARD10, TRAF1, TRAF2, TRIM27	3.445479	0.00057	**0.00667**	2.080915	0.037442	0.037442
dime_meta_effector_157	met_16	REL, TRAF2	0.77712	0.437088	0.601638	1.500837	0.133398	0.133398
dime_meta_effector_24	met_24	REL, TRAF2	0.97048	0.331807	0.493493	-1.3351	0.181844	0.181844
dime_meta_effector_57	met_34	TRAF1	-2.23929	0.025137	0.077397	-3.16557	0.001548	**0.001548**
dime_meta_effector_130	met_7	TRAF2, TRIM27	3.239229	0.001199	**0.009783**	1.655524	0.097818	0.097818
WP_004924913.1	Pst_11	REL, TRAF3	-1.71077	0.087124	0.193547	-3.16557	0.001548	**0.001548**
WP_004925512.1	Pst_12	REL	-0.92444	0.355256	0.517407	1.075446	0.282175	0.282175
WP_004926210.1	Pst_13	REL	0.222824	0.823673	0.892312	-0.71635	0.473775	0.473775
WP_042117401.1	Pst_17	REL	1.743917	0.081174	0.18264	-0.97416	0.329976	0.329976
WP_004915712.1	Pst_2	REL, TRAF2	2.41423	0.015769	0.053219	-1.25776	0.20848	0.20848
WP_004919757.1	Pst_6	REL, TRAF6	2.246651	0.024662	0.077986	-1.07729	0.281352	0.281352

Table S16 | Statistical analysis of the effect on NF-κB activation by met_7. *F/R values of dime_meta_effector_130 (met_7) normalized to the F/R values of IKKβ were subjected to a statistical analysis to identify significant differences compared to the normalized F/R values of the empty vector (pMH-FLAG-HA). The Kruskal-Wallis test was performed with Dunn's correction and adjusted by FDR-correction. Bold values indicate significant p-values. p unadj, unadjusted p-value; p adj, adjusted p-value.*

effector ID, concentration	Z-score	p unadj	p adj
met_7, 2 µg	0.57522374	0.56513999	0.56513999
met_7, 4 µg	1.53392998	0.12504688	0.15005626
met_7, 6 µg	2.49263621	0.01267987	**0.03803961**

B Curriculum vitae

VERONIKA YOUNG

Academic experiences

Helmholtz Munich 07/2019 - 12/2023	**PhD in Biology** Network biology, Prof. Dr. Pascal Falter-Braun "A gut commensal microbiome-host protein network map reveals bacterial modulation of human immune signalling" *experiences* - cloned bacterial ORFs (PCR, Gateway™, digestion, ligation, transformation), conducted a large-scale Y2H and cell culture experiments (transfection, reporter assay, TUNEL, cell viability assay), prepared a library for NGS sequencing - taught a yearly 2-week practical course for master students, supervised a master student and her laboratory work for 3 months
TU München 10/2013 - 09/2016	**Master Nutrition and Biomedicine** *Master's thesis* Nutritional Medicine, Prof. Dr. H. Hauner „Energy and nutrient intake of pregnant and lactating women within the INFAT study"
LMU München 10/2010 - 09/2013	**Bachelor Biology** *Bachelor's thesis* Evolutionary Biology, Prof. Dr. W. Stephan „Correlation between fitness and cold adaption in *D. melanogaster*"

Relevant publications

A gut meta-interactome map reveals modulation of human immunity by microbiome effectors,
Young *V,* Dohai B, Hitch TCA, Hyden P, Weller B, van Heusden NS, Saha D., Fernandez Macgregor J, Maseko SB, Lin CW, Boujeant M, Choteau S.A, Ober F, Schwehn P, Rothballer ST, Altmann M, Altmann S, Strobel A, Rothballer M, Tofaute MJ, Heinig M, Clavel T, Twizere JC, Vincentelli R, Boes M, Krappmann D, Falter C, Rattei T, Brun C, Zanzoni A, Falter-Braun P. *bioRxiv* 2023.09.25.559292, 2023

A proteome-scale map of the SARS-CoV-2–human contactome,
Kim DK, Weller B, Lin CW, Sheykhkarimli D, Knapp JJ, Dugied G, Zanzoni A, Pons C, Tofaute MJ, Maseko SB, Spirohn K, Laval F, Lambourne L, Kishore N, Rayhan A, Sauer M, Young V, Halder H, la Rosa NM, Pogoutse O, Strobel A, Schwehn P, Li R, Rothballer ST, Altmann M, Cassonnet P, Coté AG, Vergara LE, Hazelwood I, Liu BB, Nguyen M, Pandiarajan R, Dohai B, Coloma PAR, Poirson J, Giuliana P, Willems L, Taipale M, Jacob Y, Hao T, Hill DE, Brun C, Twizere JC, Krappmann D, Heinig M, Falter C, Aloy P, Demeret C, Vidal M, Calderwood MA, Roth FP, Falter-Braun P.
Nat. Biotechnol. 41, 140–149 (2023)

Extensive signal integration by the phytohormone protein network,
Altmann M, Altmann S, Rodriguez PA, Weller B, Elorduy Vergara L, Palme J, Marín-de la Rosa N, Sauer M, Wenig M, Villaécija-Aguilar JA, Sales J, Lin CW, Pandiarajan R, Young V, Strobel A, Gross L, Carbonnel S, Kugler KG, Garcia-Molina A, Bassel GW, Falter C, Mayer KFX, Gutjahr C, Vlot AC, Grill E, Falter-Braun P. *Nature* 583, 271-276 (2020)

Full-time experiences

Impact Hub Munich　　　**Community Manager**
01/2017 - 06/2019　　　　*responsibilities*
　　　　　　　　　　　　　onboarded new participants, hosted weekly community events,
　　　　　　　　　　　　　contract management and member survey evaluation
　　　　　　　　　　　　　successes
　　　　　　　　　　　　　implementation of new intranet, smooth contract amendment

Important Skills

Languages
German, native
English, fluent
French, basic skills

Project management skills
classic project management (AZAV certified course "Introduction to classic project management")
agile project management (AZAV certified course "Introduction to agile project management")

IT skills
very good knowledge of Microsoft Office and Inkscape (Photoshop)
basic knowledge of statistical software (R and SPSS)

Laboratory skills
cell culture: transfection, reporter / cell viability / apoptosis assays, fluorescence imaging
molecular biology: PCR, Gateway cloning, enzymatic digestion, transformation, DNA purification &
extraction, Western Blot
sequencing: NGS library preparation, Sanger sequencing
microbiology: bacterial culture (S1 / S2), yeast culture
laboratory devices: FACS (Bio-Rad), ultracentrifuge, plate reader, liquid handling robots (TECAN /
INTEGRA), DNA extraction robot (QIAGEN)

C Acknowledgements