

# Functional metagenomic discovery of bacterial effectors in the human microbiome and isolation of commendamide, a GPCR G2A/132 agonist

Louis J. Cohen<sup>a,b</sup>, Hahk-Soo Kang<sup>a</sup>, John Chu<sup>a</sup>, Yun-Han Huang<sup>a</sup>, Emma A. Gordon<sup>a</sup>, Boojala Vijay B. Reddy<sup>a</sup>, Melinda A. Ternei<sup>a</sup>, Jeffrey W. Craig<sup>a,1</sup>, and Sean F. Brady<sup>a,2</sup>

<sup>a</sup>Laboratory of Genetically Encoded Small Molecules, Howard Hughes Medical Institute, The Rockefeller University, New York, NY 10065; and <sup>b</sup>Division of Gastroenterology, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029

Edited by Jerrold Meinwald, Cornell University, Ithaca, NY, and approved July 15, 2015 (received for review May 4, 2015)

The trillions of bacteria that make up the human microbiome are believed to encode functions that are important to human health; however, little is known about the specific effectors that commensal bacteria use to interact with the human host. Functional metagenomics provides a systematic means of surveying commensal DNA for genes that encode effector functions. Here, we examine 3,000 Mb of metagenomic DNA cloned from three phenotypically distinct patients for effectors that activate NF- $\kappa$ B, a transcription factor known to play a central role in mediating responses to environmental stimuli. This screen led to the identification of 26 unique commensal bacteria effector genes (Cbegs) that are predicted to encode proteins with diverse catabolic, anabolic, and ligand-binding functions and most frequently interact with either glycans or lipids. Detailed analysis of one effector gene family (Cbeg12) recovered from all three patient libraries found that it encodes for the production of *N*-acyl-3-hydroxypalmitoyl-glycine (commendamide). This metabolite was also found in culture broth from the commensal bacterium *Bacteroides vulgatus*, which harbors a gene highly similar to Cbeg12. Commendamide resembles long-chain *N*-acyl-amides that function as mammalian signaling molecules through activation of G-protein-coupled receptors (GPCRs), which led us to the observation that commendamide activates the GPCR G2A/GPR132. G2A has been implicated in disease models of autoimmunity and atherosclerosis. This study shows the utility of functional metagenomics for identifying potential mechanisms used by commensal bacteria for host interactions and outlines a functional metagenomics-based pipeline for the systematic identification of diverse commensal bacteria effectors that impact host cellular functions.

functional metagenomics | NF- $\kappa$ B | *N*-acyl amino acids | commendamide | Cbeg

The human body is home to hundreds of distinct bacterial species and trillions of individual bacteria (1). Sequencing of DNA extracted from patient samples is the most commonly used approach for studying the human microbiome (2). Human cohort sequencing studies have found strong correlations between changes in bacterial populations and human pathophysiology (3–7). Mouse models have been used to show that native bacterial ecology is necessary for normal physiologic functions and dysbiosis causes diseases like obesity, cancer, diabetes, and colitis among others (8–11). Despite evidence linking changes in commensal bacteria populations to disease in mice and correlative evidence in humans, it is still largely unknown how specific bacterial functions affect mammalian physiology (effector functions). Human commensal bacteria effector functions are programmed in the metagenome of the human microbiome, which is predicted to contain 100 times the number of unique genes than can be found in the human genome (12). Tremendous resources have been allocated to the sequencing and bioinformatic organization of genes within the human commensal metagenome; however, very few genes have been shown to encode either for proteins or indirectly for small molecules that affect the human host through specific cellular

receptors (e.g., effector molecules) (3, 13, 14). Here, we use functional metagenomics and high-content imaging to identify and characterize human microbiome effector genes and their products.

In functional metagenomic studies, large fragments of DNA extracted directly from an environmental sample are cloned into a model bacterial host, and the resulting clones are examined for phenotypes of interest. This approach circumvents the culture barrier and allows for the simultaneous identification of effectors from both cultured and uncultured microbes. In addition, it couples each observed phenotype to a single fragment of cloned metagenomic DNA, making it possible to simultaneously identify specific effector molecules and the specific effector genes that encode these molecules. Functional metagenomics has been used to isolate small molecules and proteins from soil metagenomes (15–24); however, its application to the human microbiome has to date been very limited (16, 25, 26). In this study, a set of arrayed, large-insert cosmid libraries hosted in *Escherichia coli* was created from DNA isolated from the stool of three patients who, based on their phenotype, were predicted to have different commensal bacteria cohorts. Bacterial populations in healthy patients are more similar to each other than to bacterial populations in patients with certain disease phenotypes including

## Significance

In this study, we demonstrate a method for rapidly identifying bacterial effector genes and gene products from human commensal bacteria. Identification of specific effector genes and small molecules improves our understanding of how bacteria might interact with human cells and contribute to both health and disease. The small molecules we isolated, *N*-acyl-3-hydroxyglycines, resemble endogenously produced *N*-acyl-amide signaling molecules and were found to activate the human G-protein-coupled receptor (GPCR) GPR132/G2A. G2A has potentially important implications for autoimmune disease and atherosclerosis. Finding commensal small molecules that appear to structurally mimic host signaling metabolites provides greater insight into how commensal bacteria may interact with human physiology and the methods required for future discovery of other commensal effectors.

Author contributions: L.J.C. and S.F.B. designed research; L.J.C., H.-S.K., J.C., Y.-H.H., E.A.G., B.V.B.R., M.A.T., and J.W.C. performed research; L.J.C., B.V.B.R., and S.F.B. analyzed data; and L.J.C. and S.F.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

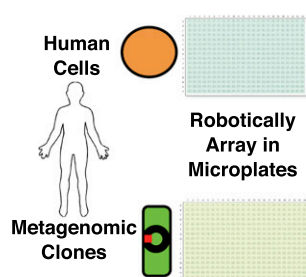
Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession nos. [KT336241](https://doi.org/10.1093/ncbi/kt336241)–[KT336282](https://doi.org/10.1093/ncbi/kt336282)).

<sup>1</sup>Present address: Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115.

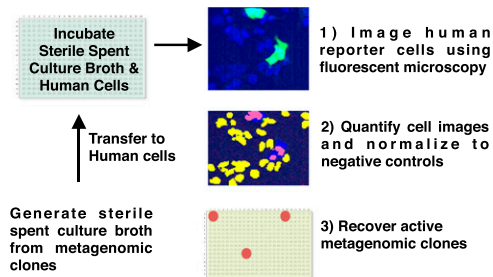
<sup>2</sup>To whom correspondence should be addressed. Email: [sbrady@rockefeller.edu](mailto:sbrady@rockefeller.edu).

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1508737112/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1508737112/-DCSupplemental).

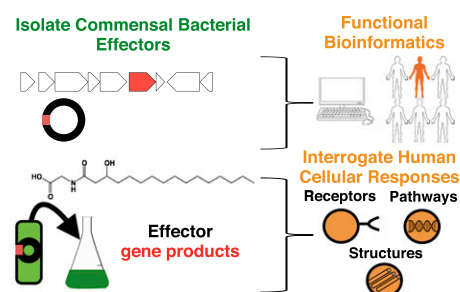
## 1. Samples



## 2. Assay



## 3. Host-Microbial Interactions



**Fig. 1.** Overview of metagenomic methods. Step 1: The human reporter cell line and individual metagenomic bacterial clones are robotically arrayed in separate 384-well microplates. Step 2: Mature bacterial cultures are filter sterilized, and sterile spent culture broth is then transferred to plates containing the human reporter cell line. Human reporter cells that have been exposed to spent culture broth are imaged by fluorescent microscopy to identify metagenomic clones that activate the reporter. Step 3: Once an active metagenomic clone is confirmed, the specific effector genes are identified through sequencing and transposon mutagenesis, and the effector molecules (proteins or small molecules) are characterized from large-scale cultures of the active metagenomic clone.

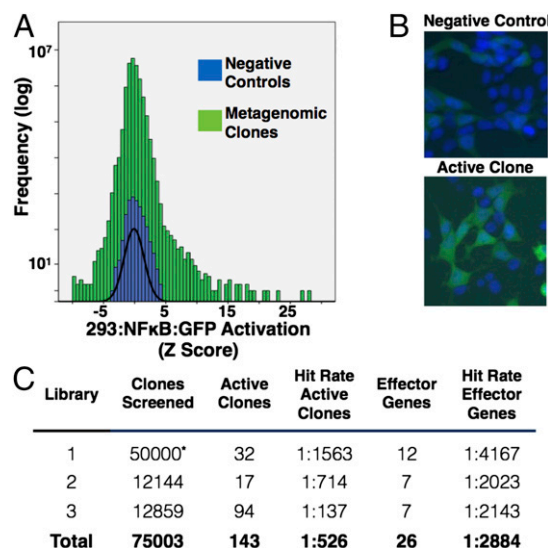
inflammatory bowel disease, obesity, cancer, cirrhosis, and diabetes among others. To ensure we capture the greatest diversity of bacterial species and their encoded effector functions in our metagenomic library collection, we created metagenomic libraries using DNA isolated from the stool of a healthy patient, a patient with Crohn's disease, and a patient with ulcerative colitis. Crohn's disease and ulcerative colitis are collectively known as inflammatory bowel disease (IBD), and bacterial populations in each of these diseases are unique relative to each other and to patients without IBD (12, 27–29).

As the specific roles bacterial effectors might play in human biology are still not clear, we initially sought to identify effectors using a screen that would report very broadly on the ability of metagenomic clones to perturb human cells. Nuclear factor- $\kappa$ B (NF- $\kappa$ B) is a “rapid-acting” broadly regulated transcription factor that is involved in a myriad of normal and disease cellular processes. Few signaling pathways have been found to respond to a more diverse set of extracellular inducers, making it a potentially useful reporter of a broad range of host–microbial interactions (30). To identify bacterial effector functions, sterile spent culture broth filtrates derived from individually arrayed metagenomic clones were screened for NF- $\kappa$ B activity using a human embryonic kidney (HEK293) cell line stably transfected with a GFP reporter construct under the control of NF- $\kappa$ B. From the screening of ~75,000 metagenomic clones, we identified 26 unique commensal bacteria effector genes. Taxonomic analysis of the effector genes indicates enrichment for genes from the phyla Bacteroidetes. Detailed bioinformatics analyses of these effector genes suggest they most frequently encode for proteins that interact with glycans or lipids either as binding partners or substrates.

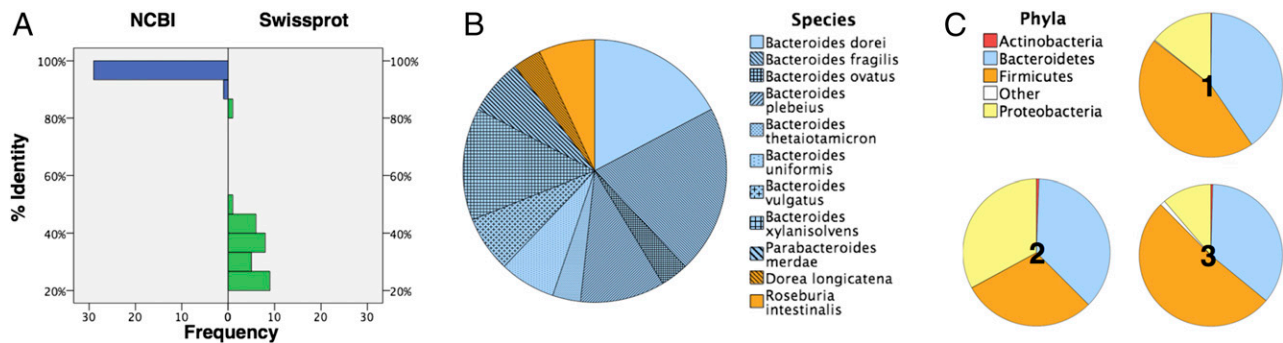
In-depth functional analysis of one of these effector gene families led to the isolation of the previously unknown natural product *N*-acyl-3-hydroxy palmitoyl glycine (**1**), which we have named commendamide. We show that this natural product is present in the culture broth of *Bacteroides vulgatus*, one of several related commensal bacteria that harbor a gene highly similar to Cbeg12. Commendamide resembles mammalian long-chain *N*-acyl-amide signaling molecules that are known to specifically activate diverse G-protein-coupled receptors (GPCRs). At an order of magnitude lower concentration than we see induction of NF- $\kappa$ B, commendamide activates the GPCR, G2A. G2A has been implicated in autoimmune disease and atherosclerosis, which are both diseases where host–microbial interactions are thought to play a role (31–33). The structural similarity between commendamide and a broad class of eukaryotic signaling molecules is striking and suggests that chemical mimicry may be a mechanism used by commensal bacteria to facilitate interactions with their human host (34, 35).

## Results and Discussion

**Selection of a Patient Cohort and Construction of Metagenomic DNA Cosmid Libraries.** Demographic details for the ulcerative colitis patient (UC) (library 1), Crohn's patient (CD) (library 2), and healthy control (HC) (library 3) from whom stool was collected for this study are shown in *SI Appendix*. The construction of metagenomic libraries from patient stool samples proceeded using methods developed for cloning DNA directly from soil (36). In brief, high-molecular-weight environmental DNA (eDNA) was



**Fig. 2.** Screen results. (A) Histogram of HEK293:Nf- $\kappa$ B:GFP reporter activation Z scores for all 75,003 metagenomic clones screened in this study. The percentage of activated cells was determined on a per-well basis and normalized to negative control wells on the assay plate to give a Z score. The Z-score distribution from metagenomic clones demonstrates a positive skew toward increased GFP activation compared with negative control wells (normal curve, black line). (B) Images of a representative negative control well and a well from an active metagenomic clone. Nuclei appear blue (Hoechst 33442 stain). HEK293:Nf- $\kappa$ B:GFP cells expressing GFP appear green. (C) Table of total metagenomic clones screened and hit rates for each library. Active clones have reproducible HEK293:Nf- $\kappa$ B:GFP cell activation in a secondary assay. Effector genes were identified by transposon mutagenesis of unique cosmids recovered from the active hits. All hit rates are expressed relative to total metagenomic clones screened. \*Library 1 was not robotically arrayed into a microplate but mechanically dispensed at a dilution of 0–3 metagenomic clones per well. Total number of clones screened in library 1 is therefore an estimate based on the number of wells screened and average number of clones per well.



**Fig. 3.** Bacteroidetes species appear to be enriched in effector genes. (A) The percent identity for the top hit identified in a BlastN search of each effector gene against either the NCBI reference genome dataset or the Swissprot dataset is tabulated. (B) The phylogenetic origin of the sequence from NCBI reference genome dataset to which each effector gene is most closely related is tabulated [Bacteroidetes (blue), Firmicutes (orange)]. (C) A 16S gene analysis (V4 region) of each metagenomic library was carried out to assess phylogenetic diversity in these libraries. For each library, the percentage of effector genes predicted to arise from Bacteroidetes spp. (88%; B) is greater than the percentage of 16S genes predicted to arise from Bacteroidetes spp. (36–40%; C).

extracted directly from 4 g of stool within 24 h of collection. Crude eDNA extracts were then gel purified, blunt ended, ligated with the broad host-range cosmid pJWC1, packaged into lambda phage, transfected into *Escherichia coli*, and the transformants were selected using tetracycline (15  $\mu$ g/mL) (20). Based on titers obtained from the original transfection plates, each of the resulting metagenomic libraries is predicted to contain in excess of  $1 \times 10^6$  unique cosmid clones, representing over 10,000 4-Mb genome equivalents of commensal bacteria DNA.

**High-Throughput Fluorescent Microscopy to Identify Bacterial Effector Genes That Activate NF- $\kappa$ B.** The reporter cell line (HEK293:NF- $\kappa$ B:GFP) that we used in this study consisted of HEK293-TN cells stably transfected with the pGreenFire lentiviral NF- $\kappa$ B GFP-luciferase plasmid (Systems Biosciences). This plasmid contains a minimal CMV promoter in conjunction with four copies of the NF- $\kappa$ B consensus transcriptional response element to control the expression of GFP (25). HEK293 cells were specifically selected for use in our functional metagenomic screen because they lack native TLR2 and TLR4 expression ensuring low baseline activation from *E. coli* host membrane lipids and glycans (37). An overview of the screening protocol used to identify clones producing effectors that activate NF- $\kappa$ B is depicted in Fig. 1. Briefly, clones from the three patient libraries were arrayed into 384-well microtiter plates (~50,000 clones library 1, ~12,500 clones library 2, ~12,500 clones library 3) and allowed to grow for 4 d in LB medium. Sterile spent culture broth was collected from each well by passage through a 0.45- $\mu$ m filter, and then a defined volume (20  $\mu$ L) transferred to wells containing the HEK293:NF- $\kappa$ B:GFP reporter cell line. After 24 h of incubation in the presence of the sterile spent culture broth filtrate, GFP expression in each well was measured by high-throughput fluorescent microscopy. The percentage of viable human cells showing GFP expression in each test well was normalized to data from negative control wells on the same assay plate, and the extent of GFP induction for each well was then reported as a Z score (Fig. 2 and *SI Appendix*).

Metagenomic clones from wells exhibiting Z scores greater than 3 were recovered from archived arrayed clone plates and rescreened in quadruplicate under the same assay conditions. A total of 143 clones with the ability to reproducibly activate GFP expression were identified from the screening of ~75,000 wells. This corresponds to 1 activating clone per 526 metagenomic clones or 1 active clone per 10–15 Mb of screened eDNA. De novo sequencing of cosmids from these activating clones identified 24 unique genomic regions. A representative cosmid from each unique genomic region identified in a patient library was selected for detailed analysis. This included 12 clones from library 1, 6 clones from library 2, and 6 clones from library 3 (*SI Appendix*). To identify specific bacterial effector genes,

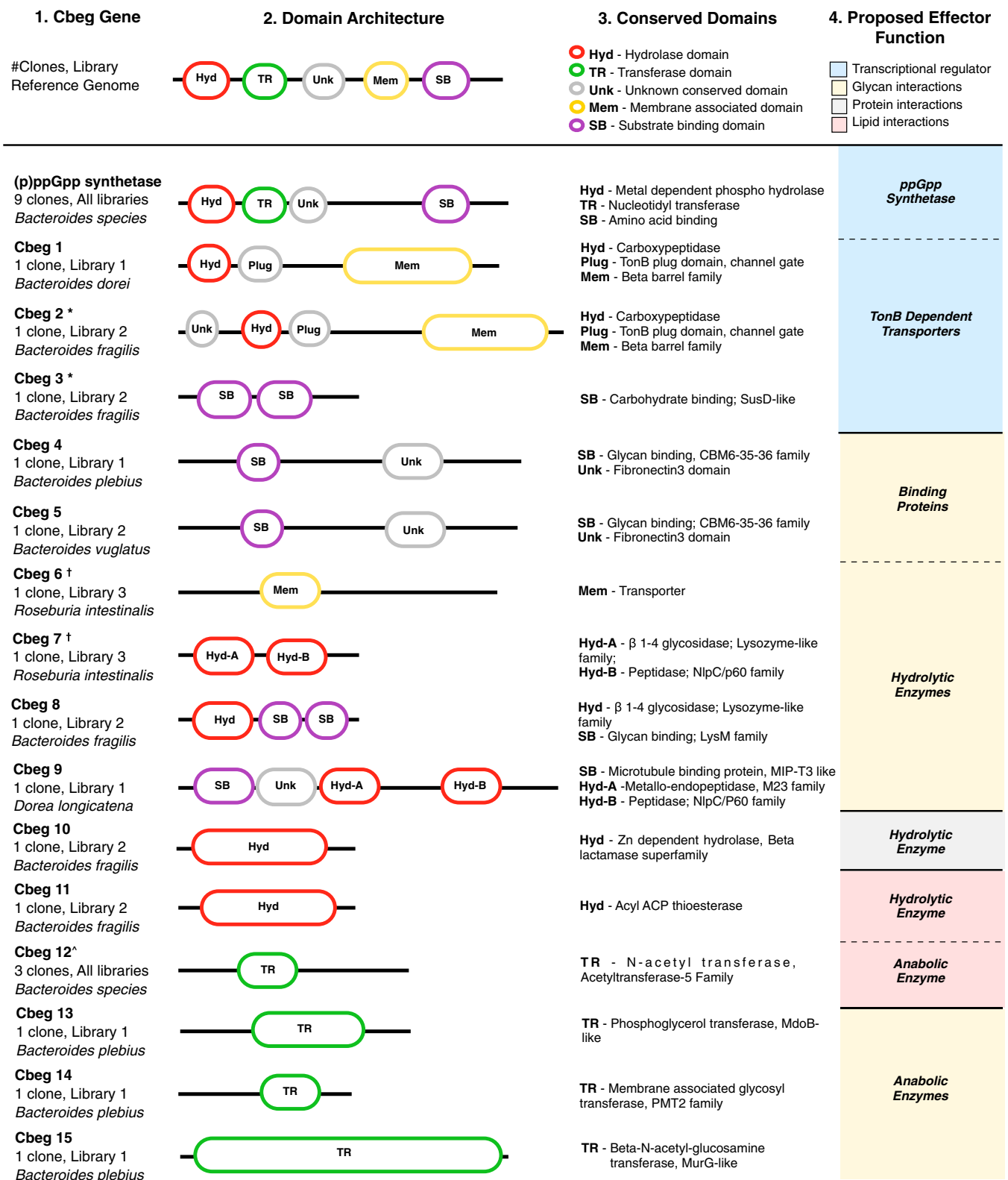
individual cosmids were transposon mutagenized, reintroduced into *Escherichia coli*, and screened for mutants that no longer induced GFP expression. The sequencing of loss-of-function mutants identified transposons in either one or two genes in each cosmid to give a total of 26 commensal bacteria effector genes (Cbegs).

#### Comparative Phylogenetic and Functional Analysis of Effector Genes.

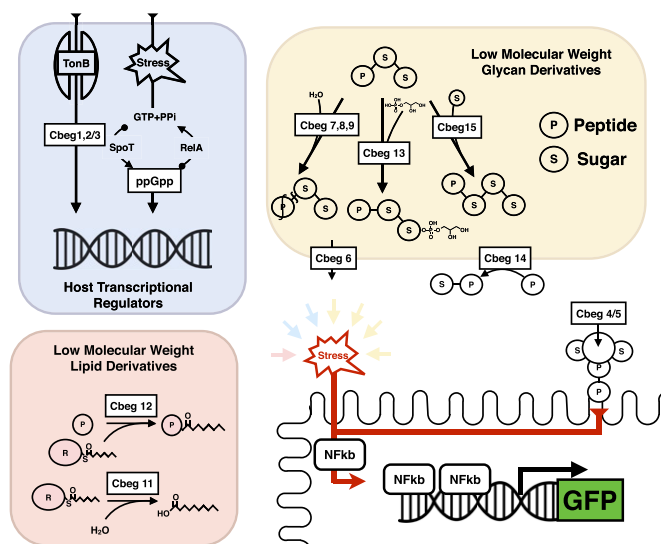
To identify the source of its closest relative, each effector gene was aligned (BLASTn) against the National Center for Biotechnology Information (NCBI) reference genome dataset, including 1,512 complete genomes from commensal bacteria species. All effector genes mapped to a commensal bacteria reference genome at >93% nucleotide identity (Fig. 3A and B, and *SI Appendix*). The most closely related sequences identified for each Cbeg derived from just two phyla, either Bacteroidetes (23 of 26 genes) or Firmicutes (3 of 26 genes). Even though all libraries were screened using the Proteobacterium *E. coli* as a host and Proteobacterial 16S sequences represent 11–33% of the 16S sequences observed in each library, none of the effector genes are most closely related to genes found in Proteobacterial reference genomes, which might have been expected if there were a host expression bias (Fig. 3B and C). The high proportion of genes predicted to derive from Bacteroidetes spp. (88%) is greater relative to the fraction of Bacteroidetes DNA predicted to be present in our metagenomic libraries based on 16S RNA gene analysis (36–40%) (38). The observation that Bacteroidetes species may be enriched for effectors that interact with human cells is not unexpected as Bacteroidetes are a dominant phyla in the intestine and have likely adapted to interact extensively with human hosts (2).

Although effector genes show high identity to sequences in the NCBI-curated commensal reference genome dataset, a BLASTp search of a similarly manually curated dataset of functionally characterized proteins, SwissProt, returned only a single hit with greater than 50% identity (Fig. 3A and *SI Appendix*). The discordance between sequence identity and specific functional annotation of commensal genes exemplifies the historical context in which the microbiome has been studied. In particular, a tremendous effort has been put into the sequencing and cataloging of genes from the human microbiome, yet the detailed functional characterization of these genes remains rare. A BLASTp search of conserved protein domains found in the broader NCBI nr database suggests that effector genes encode diverse transcriptional, catabolic, anabolic, and ligand-binding functions and that they most frequently interact with either sugars or lipids (Figs. 4 and 5, and *SI Appendix*).

**Effectors That Likely Induce Changes in the *E. coli* Host.** Nine effector genes including unique examples from all three patient libraries are homologs of the (p)ppGpp biosynthesis enzymes spoT/reIA



**Fig. 4.** Summary of commensal bacterial effector genes (Cbegs). Each effector gene was queried by BLASTn against the NCBI nr dataset to determine the reference genome from which the most closely related sequence arises (column 1) and to predict the domain architecture and makeup of each Cbeg protein (columns 2 and 3, respectively). The metagenomic library from which each Cbeg was recovered is shown in column 1. In column 4, we have grouped Cbegas based on their general predicted functions. \*Cbeg2 and Cbeg3 are found on the same unique cosmid. †Cbeg6 and Cbeg7 are found on the same unique cosmid. ^Cbeg12 is found at 98% identity in three unique cosmids, each in a different patient library (Cbeg12-1, Cbeg12-2, Cbeg12-3).



**Fig. 5.** General schematic of predicted Cbeg effector functions. Cbegs are predicted to encode proteins that result in activation of the HEK293:NfκB:GFP reporter through different mechanisms. This includes proteins that likely induce major transcriptional changes in the bacterial host as well proteins with diverse catabolic, anabolic, and ligand-binding functions. See the main text for a detailed discussion of each Cbeg.

(SI Appendix). This is the most common effector gene function detected. The alarmone, (p)ppGpp, controls the stringent response, an adaptive response to environmental stressors that involves changes in the expression level of hundreds of different genes across the bacterial genome (39, 40). Three other effector genes (Cbeg1 and 2/3) are structurally related to TonB-dependent transporters (TBDT). TonB-dependent transport is a mechanism for active transport of environmental chemicals across the Gram-negative outer membrane (41–43). This uptake process is frequently accompanied by changes in transcription in response to the chemicals taken up from the environment. At this point, we do not know whether spoT or TBDT expression induces the *E. coli* host to produce a single effector molecule or a large collection of effectors due to global transcriptional changes. Whether this type of host-specific response is functionally relevant to commensal human interactions or an artifact of the heterologous expression system remains to be seen. It is interesting to note that we only identified effector genes with high sequence identity to *Bacteroides* spp., despite the presence of ppGpp synthesis enzymes and TBDTs in most bacterial phyla. Although the reason for this finding not clear, it is possible that some *Bacteroides* spp. homologs of these proteins are either constitutively active or more sensitive to common environmental stimuli when introduced into this heterologous expression system.

**Effector Genes Lead to the Production of Low- and High-Molecular-Weight Products That Interact with Human Cells.** The remaining effectors are predicted to be ligand-binding proteins (Cbeg4, -5), hydrolytic enzymes (Cbeg7, -8, -9, -10, -11), and anabolic enzymes (Cbeg12, -13, -14, -15). Cbeg4 and -5 contain CBM6-35-36-like glycan binding domains but no obvious catalytic domains. They both have secretion signals, suggesting they could be secreted and activate NF-κB through direct interactions with epitopes on the human cell (SI Appendix). Among predicted hydrolytic effector enzymes, a conserved domain analysis suggests their substrates likely include glycans (Cbeg7, -8, -9), peptide bonds (Cbeg10), and acyl thioesters (Cbeg11). A number of predicted catabolic enzymes resemble peptidoglycan hydrolyzing enzymes with different arrangements of NlpC/p60-like cell wall peptidase domains and

lysozyme-like glycosyl hydrolase domains. For example, Cbeg7 has both an NlpC/p60 and a lysozyme-like domain, whereas Cbeg8 and -9 have either an NlpC/p60 (Cbeg9) or lysozyme-like (Cbeg8) domain in addition to either a M23 zinc metallopeptidase (Cbeg9) or LysM carbohydrate binding domain (Cbeg8) (44–46). The domain architecture seen in Cbeg8 is typical of bacterial autolysins that are associated with both reorganizing cell membranes for cell division and host invasion by pathogens (47). Cbeg7 was identified in a two-gene operon with Cbeg6. Cbeg6 is predicted to contain seven-transmembrane helices (48), suggesting an association with the membrane and potentially a role in exporting the catabolic products of Cbeg7 or even Cbeg7 itself (SI Appendix).

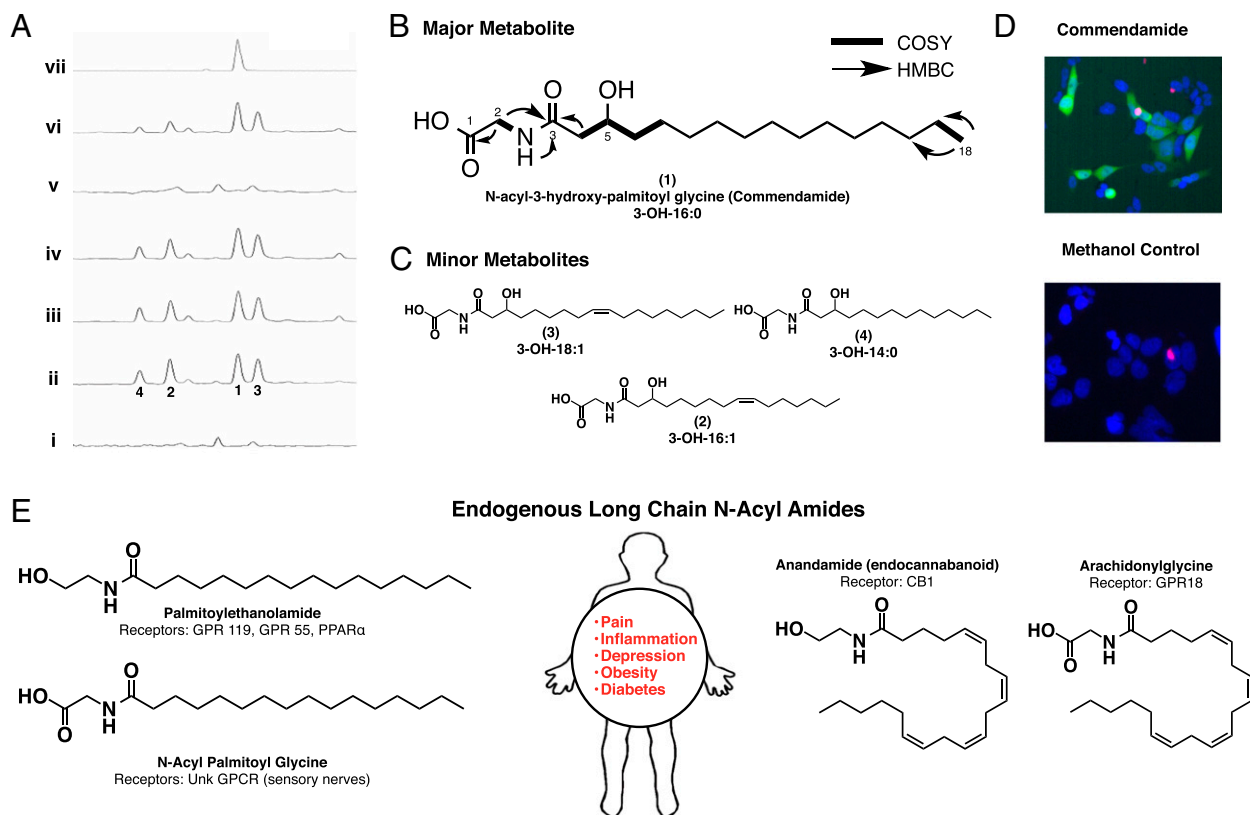
Cbegs predicted to be anabolic enzymes are expected to use either sugars (Cbeg13, -14, -15) or acyl groups (Cbeg12) as substrates. Cbeg13 and -15 are related to enzymes involved in the synthesis of bacterial cell wall glycans. Cbeg13 is related to MdoB, which catalyzes the transfer of phosphoglycerol to periplasmic glycans in *E. coli*, and Cbeg15 is related to MurG, which catalyzes the formation of the β1–4 glycosidic bonds between the peptidoglycan sugars GlcNAc and MurNAc in *E. coli* (49, 50). Cbeg14 has a small region of similarity to membrane associated glycosyl-transferases (PFAM 02366) (51). The fact that effector proteins are commonly predicted to interact with glycans may reflect the abundance of glycans in the intestine which provides both an accessible and diverse resource for facilitating host interactions.

#### Characterization of Small-Molecule Elicitors Produced by Cbeg12. Of

the predicted catabolic and anabolic Cbegs, only one gene family, Cbeg12, was recovered from all three patient libraries. The Cbeg12 family is composed of two distinct sequences that are 98% identical to one another. One Cbeg12 sequence was recovered from library 2 (Cbeg12-1), and the other Cbeg sequence was recovered from library 1 (Cbeg12-2) and library 3 (Cbeg12-3). Because of its potential common occurrence in diverse human microbiomes, we elected to study this Cbeg in more detail. Cbeg12 is predicted to be a member of the acetyltransferase-5 family of enzymes, which is part of the lysophospholipid acyltransferase superfamily (PFAM CL0257) common to every domain of life (52). Enzymes from the acetyltransferase-5 family are not known to produce commensal bacteria effectors, although acetyltransferase-5 enzymes from soil metagenomes are known to make *N*-acylated amino acids (15, 20, 21). A BLAST search of the NCBI database found that genes highly similar to Cbeg12 (>70% protein identity) were almost exclusively restricted to commensal Bacteroidetes species suggesting these effector genes may be specific to commensal host–microbial interactions (see Fig. 7). All closely related sequences are currently generically annotated as either acyl transferases [e.g., lysophospholipid acyltransferase (LPLAT)] or hemolysins (53). No small-molecule products have been associated with this Bacteroidetes-derived family of genes.

Our transposon mutagenesis studies indicated that Cbeg12 was necessary for NF-κB activation. To determine whether Cbeg12 was sufficient for NF-κB activation, one representative Cbeg12 gene (Cbeg12-1) and its promoter were subcloned and retransformed into *E. coli*. Spent culture broth from the resulting strain was subsequently rescreened in the HEK293:NfκB:GFP assay. As seen with the full-length cosmid clone, spent culture broth from the Cbeg12 subclone reproducibly induced GFP expression in the reporter assay (SI Appendix), indicating that Cbeg12 is both necessary and sufficient for activating the NF-κB reporter.

The first step in our characterization of the Cbeg12 effector was to determine whether the effector molecule present in Cbeg12 *E. coli* culture broth was a large molecule (i.e., Cbeg12 itself) or a small molecule (i.e., a product of Cbeg12). To do this, spent culture broth from Cbeg12 *E. coli* cultures was passed through a 10-kDa cutoff membrane and the flow-through was assayed for the ability to induce GFP expression in the HEK293:NfκB:GFP assay. We found that the low-molecular-weight flow-through



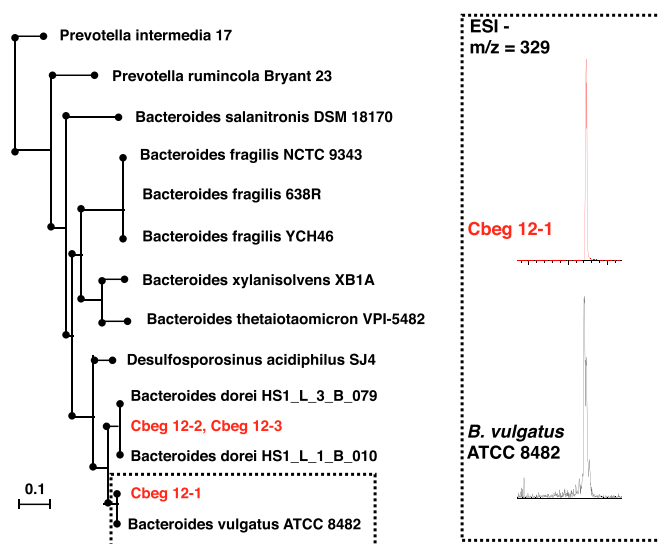
**Fig. 6.** Characterization of commendamide. (A) Electrospray ionization (ESI)-mass spectroscopy (MS) traces of culture broth extracts from *E. coli* transformed with an empty pJWC1 cosmid vector (i), cosmid Cbeg12-1 (ii), cosmid Cbeg12-2 (iii), cosmid Cbeg12-3 (iv), cosmid Cbeg12-1 with a transposon insertion in the Cbeg12-1 gene (v), Cbeg12-1 subcloned into pJWC1 (vi), and synthetic commendamide (vii). (B) Key NMR correlations used to define the structure of commendamide (1), the major clone-specific peak found in cultures of *E. coli* transformed with Cbeg12-1, are shown. (C) The structures for three minor clone-specific metabolites related to commendamide (compounds 2–4) were also determined using NMR and MS data. (D) Purified commendamide activates the HEK293:NF- $\kappa$ B GFP reporter assay. (E) Endogenous long-chain *N*-acyl-amides that are structurally related to commendamide are reported to function as agonists for numerous receptors, including many GPCRs. Such signaling systems have been therapeutically targeted for the treatment of pain, inflammation, depression, obesity, and diabetes.

retained all inducing activity present in spent culture broth, indicating that a small molecule, likely a *N*-acylated metabolite and not the Cbeg12 protein itself (37 kDa), was responsible for activating NF- $\kappa$ B.

Many *N*-acylated small molecules are hydrophobic and can therefore be extracted from spent culture broth using organic solvents. Ethyl acetate extracts of cultures of *E. coli* expressing Cbeg12 showed potent clone-specific NF- $\kappa$ B-inducing activity. Bioassay-guided fractionation of this activity from large-scale ethyl acetate extracts led to the isolation of four clone-specific metabolites that induce GFP expression in the HEK293:NF- $\kappa$ B:GFP assay (Fig. 6 and *SI Appendix*). When examined by liquid chromatography-mass spectrometry (LC-MS), extracts of *E. coli* cultures transformed with either Cbeg12-1, Cbeg12-2, or Cbeg12-3 gene-containing metagenomic cosmid clones showed the presence of the same set of four clone-specific molecules (Fig. 6). The structure of each metabolite was determined using a combination of 1D and 2D NMR and high-resolution mass spectrometry. The most active metabolite, which is also the major clone-specific metabolite produced by cultures transformed with Cbeg12, is *N*-acyl-3-hydroxypalmitoyl glycine (1) (Fig. 6). The three minor active clone-specific metabolites found in the culture broth extract are long-chain *N*-acyl glycine derivatives with different 3-hydroxy fatty acid side chains [C16:1 (2), C18:1 (3), C14 (4)]. A 3OH-C15:0 derivative is also predicted to be present based on the mass of another minor clone specific peak seen in the LC-MS trace. To confirm the structure of compound 1, synthetic *N*-acyl-3-hydroxypalmitoyl

glycine was produced from glycine and 3-hydroxypalmitic acid. Synthetic *N*-acyl-3-hydroxypalmitoyl glycine was found to be identical by both NMR and LC-MS to the natural metabolite (Fig. 6). To the best of our knowledge, *N*-acyl-3-hydroxypalmitoyl glycine has never been described as a natural product. We have given this compound the trivial name commendamide (commensal mimicking endogenous amide).

A small group of long-chain *N*-acyl amino acids have been characterized from eDNA clones found in soil DNA metagenomic libraries (15, 20, 21). The most common of these are long-chain tyrosine, phenylalanine, and tryptophan *N*-acyl derivatives, and in no cases are these reported to contain hydroxylated fatty acids. 3-Hydroxylated fatty acids are constituents of membrane phospholipids in a number of bacteria (54). *N*-3-(Acyloxyacyl)glycines have been identified as metabolites produced by the marine bacterium *Cytophaga* sp. In these compounds, a 3-hydroxy branched chained fatty acid is appended to the nitrogen of glycine and a second unsaturated branched chained fatty acid is appended to the hydroxyl group on the fatty acid. Although these metabolites activate N-type calcium channels, no N-type calcium channel activation was seen with simple *N*-acylated glycines (55). *N*-Acyl amino acid synthases (NAS) cloned from soil eDNA also fall into the acetyltransferase-5 family of *N*-acyl transferases, but they are only distantly related to one another (*SI Appendix*). Soil bacteria derived enzymes have been shown to use acyl carrier protein (ACP) linked fatty acids as a source of the acyl substituent (56). The source of the acyl groups used by Cbeg12 enzymes is not yet known.



**Fig. 7.** Phylogeny of the commendamide *N*-acyl synthase. A BLASTn search of the three Cbeg12 genes against the NCBI reference genome dataset. Genes similar to Cbeg12 at >70% identity are represented in a phylogenetic tree annotated by reference genome. All genes are from commensal bacteria except *Desulfosporosinus acidophilus* SJ4, which was isolated from an acid mine. *B. vulgatus* ATCC 8482 is 100% identical to Cbeg12-1, and *Bacteroides dorei* HS1 is 100% identical to Cbeg12-2 and Cbeg12-3. *B. vulgatus* ATCC 8482 was purchased from ATCC and grown for 14 d in LY-BHI medium under anaerobic conditions. The *B. vulgatus* culture was extracted 1:1 with ethyl acetate, and crude extracts were fractionated by flash column chromatography (water:methanol, 0.1% TFA). MS [electrospray ionization (ESI)] analysis of extract fractions identified a metabolite with the same mass and retention time as commendamide in the *B. vulgatus* culture broth.

**Commendamide Is Produced by *Bacteroides* spp.** In a BLAST search of publicly available sequenced genomes, three sequenced commensal *Bacteroides* spp. (*Bacteroides vulgatus*, *Bacteroides dorei*, and *Bacteroides massiliensis*) were found to contain genes that encode for proteins that are essentially identical to Cbeg12 (>98% amino acid identity). The full-length metagenomic cosmid clones on which the Cbeg12-1 and Cbeg12-2 genes are found are most similar to genomic sequences from *B. vulgatus* and *B. dorei*, respectively (*SI Appendix*). To determine whether commendamide is natively produced by commensal bacteria, ethyl acetate extracts from anaerobically grown cultures of *B. vulgatus* were analyzed by LC-MS. Although peaks corresponding to the minor clone-specific metabolites (2–4) were not detected in these extracts, a compound with the same retention time and mass as commendamide (1) was, indicating that commendamide is natively produced by commensal *Bacteroides* (Fig. 7).

**Commendamide and Structurally Related Endogenous Long-Chain *N*-Acyl-Amides Activate Discrete Human Receptors.** A number of long-chain *N*-acylated glycines and ethanolamines that are close structural relatives of commendamide have been described as endogenous metabolites in humans and other mammals. The best studied of these is *N*-arachidonyl ethanolamide (anandamide), an endogenous ligand for the cannabinoid GPCR CB1 (57). Other endogenous long-chain *N*-acyl glycines and ethanolamines are reported to activate diverse receptors including many GPCRs (34, 35, 58, 59). For example, *N*-arachidonyl-glycine activates GPR18 and T-type calcium channels; *N*-palmitoyl-ethanolamide activates GPR119, GPR55, and PPAR $\alpha$ ; and *N*-palmitoyl-glycine is believed to act through an unidentified GPCR in sensory nerves (Fig. 6).

Because of the structural relation between commendamide and endogenous GPCR activating metabolites, we explored the possibility that commendamide might also activate a specific human

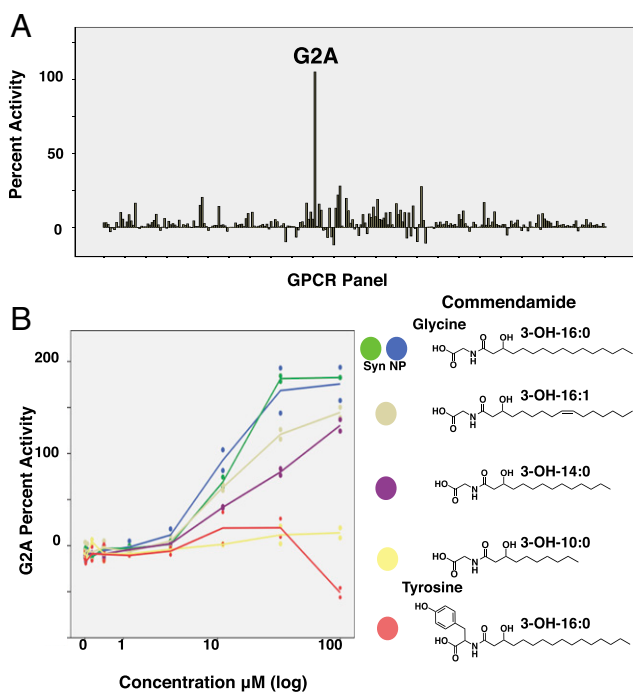
GPCR. When commendamide was screened at 10  $\mu$ M for agonist activity against a library of 242 GPCRs, including 73 orphan GPCRs, it was found to activate a single receptor, GPCR132/G2A (EC<sub>50</sub>, 11.8  $\mu$ M) (Fig. 8). This activity was confirmed using a synthetic sample of commendamide (Fig. 8). In addition to synthetic and natural commendamide, we also examined the GPR132 agonist activity of compounds with changes in fatty acid length (natural 2, 3-OH-C16:1; natural 4, 3-OH-C14:0; synthetic 3-OH-decanoyl-glycine) and changes in head group (synthetic 3-OH-palmitoyl-tyrosine) (Fig. 8). We elected to specifically look at a tyrosine analog because long-chain *N*-acyl tyrosine producing clones are frequently found in soil eDNA libraries (15, 20, 21, 36, 56, 60). All of the commendamide analogs tested, whether they contained changes in the head group or acyl chain, showed reduced GPR132/G2A activity. In fact, the tyrosine and decanoyl analogs were completely inactive toward the GPR132/G2A receptor at even the highest concentrations tested.

GPR132/G2A was first described as a transcriptional target of the BCR-ABL tyrosine kinase attenuating B-cell expansion in vitro and arresting cells at G<sub>2</sub> during mitosis (61). Oxidized long-chain fatty acids, lysophosphatidylcholine, 9(*S*)-HODE and ( $\pm$ )11-HETE have been reported to be potential endogenous ligands of G2A (62, 63). G2A-knockout mice have increased susceptibility to atherosclerosis and develop late-onset autoimmunity (31). These phenotypes are thought to arise from disruption of G2A-dependent modulation of immune cell function including differentiation and chemotaxis (62, 64–71). This suggests that *Bacteroides* spp. could influence host immune systems through G2A activation via the production of commendamide. Interestingly, GPCR-mediated signaling pathways modulated by endogenous *N*-acyl-amides that are structurally related to commendamide have been extensively explored as targets for the therapeutic modulation of pain sensation, depression, diabetes, and inflammation (72–75). The fact that commendamide is produced by a commensal bacterium suggests that bacteria may already represent a natural exogenous delivery system for these types of metabolites and thus a potential method for the therapeutic delivery of such metabolites. The characterization of small molecules produced by commensal bacteria will undoubtedly provide not only insights into how commensal bacteria and their human hosts interact but also ways of modulating host–microbial interactions as a means of improving human health and preventing disease.

## Conclusion

To our knowledge, commendamide is the first successful example of using functional metagenomics to identify a commensal metabolite involved in host–microbial interactions. The structural similarity of commendamide to the endogenous human signaling *N*-acyl-amides supports the potential use of chemical mimicry by commensal bacteria as a mechanism for mutualistic interactions. Our findings suggest that it will be interesting to examine whether commendamide might either induce tolerance to *Bacteroides* spp. or reduce the inflammatory response in the host through GPR132/G2A-mediated immune functions (31, 64, 66). Insights gained from the characterization of biologically active metabolites encoded by commensal bacteria should facilitate the development of more detailed mechanistic hypotheses about host commensal interactions that are important to human health and disease.

Many organisms throughout nature use mimicry to avoid detection, send signals, or gain access to nutrients (76). In the human microbiome, pathogenic bacteria and viruses are known to use chemical mimicry to subvert host defenses (77, 78). It would not be surprising that such a widely used adaptive strategy is part of how commensal bacteria might interact with their host environment. We predict that mimicry of endogenous eukaryotic signaling molecules and signaling through GPCRs are likely to be common themes among effectors produced by commensal bacteria.



**Fig. 8.** Commendamide activates the human GPCR G2A. (A) Using the  $\beta$ -arrestin PathHunter assay (DiscoverRx), commendamide was screened (10  $\mu$ M) for agonist activity against a panel of 242 GPCRs including 73 orphan GPCRs. For each GPCR, percent activity is expressed relative to baseline activity of the receptor. (B) Dose–response of natural and synthetic long-chain *N*-acyl-amides in the PathHunter assay using the receptor G2A. Synthetic (green) and natural product (blue) commendamide are equipotent. Two of the minor metabolites (compound 2: 3-OH-16:1; and compound 4: 3-OH-14:0) activate G2A but are slightly less potent than commendamide. Dramatic changes in either the amino acid head group (glycine to tyrosine) or fatty acid tail (3-OH-16:0 to 3-OH-10:0) result in analogs that do not appreciably activate the G2A receptor.

As with Cbeg12 encoding the production of commendamide, the 23 additional unique commensal bacteria effector genes identified in this study represent starting points for the more specific interrogation of commensal bacteria–host interactions. Bioinformatics analyses of these other Cbegs suggests the frequent use of glycans as anabolic or catabolic substrates or as direct binding partners for the purpose of host interaction. This is not surprising in light of the abundance of carbohydrates in the human intestine and the frequent appearance of genes predicted to interact with carbohydrates in the genomes of commensal bacteria (2, 5). When considering a common function such as glycan hydrolysis, numerous closely related genes that are currently indistinguishable by bioinformatics are found throughout commensal genomes. A key advantage of functional metagenomics is the ability to differentiate specific effector genes from related genes involved in bacterial functions or processes independent of host interaction.

This study highlights important differences between functional metagenomic approaches and the sequencing and bioinformatics approaches that currently dominate the study of the human microbiome. In particular, because functional metagenomics couples each observed phenotype to a small fragment of cloned DNA, it is possible to not only identify effector molecules (e.g., proteins, small molecules) but also identify functionally important genes from among the tremendous biosynthetic diversity that is found in the human microbiome. Functional metagenomics can be easily adapted to identify a wider variety of bacterial effectors from the human microbiome by simply changing the source of metagenomic

DNA, changing the host used for heterologous expression or expanding the repertoire of reporter assays used.

## Materials and Methods

**Study Population and Sample Collection.** Patients were recruited at Mount Sinai and Rockefeller Hospitals (New York, NY) from July 2011 to September 2011 under institutional review board-approved consents (#11-0716 Mount Sinai; SBR 0752 Rockefeller). Patients with a well-established diagnosis of UC or CD were identified by their primary gastroenterologists at Mount Sinai. All patients were over 18 y old, free of intestinal surgery including cholecystectomy, and without antibiotic use for more than 6 mo. Basic demographic and medical information was collected and for patients with CD or UC a disease activity index calculated—the Harvey Bradshaw index (79) for CD and simple clinical colitis activity index for UC (80). Patients with IBD had active disease as defined by a Harvey Bradshaw index of greater than 5 for CD and simple clinical colitis activity index of greater than 4 for UC. After informed consent, patients provided a single stool sample into a sterile container. Six grams of the sample was immediately placed at  $-20^{\circ}\text{C}$ . DNA was extracted from the frozen sample within 12 h.

**eDNA Isolation and Metagenomic Library Construction.** Stools from one patient with active UC, one patient with active CD, and one healthy control were resuspended in 0.9% NaCl to a total volume of 40 mL and then centrifuged ( $800 \times g$ , 5 min,  $4^{\circ}\text{C}$ ). The resulting pellets were resuspended in 40 mL of 0.9% NaCl and pelleted once again ( $3,200 \times g$ , 30 min,  $4^{\circ}\text{C}$ ). Washed stool samples were resuspended in 5 mL of 0.9% NaCl. A volume of 750  $\mu$ L of this sample was layered on 500  $\mu$ L of Nycodenz in a 1.5-mL Eppendorf tube (20). After centrifugation ( $21,130 \times g$ , 10 min, room temperature), the top layer was removed and mixed with 500  $\mu$ L of 0.9% NaCl in a 15-mL conical tube, and bacteria were collected by centrifugation ( $5,800 \times g$ , 10 min, room temperature). This pellet was resuspended in 8 mL of bacterial lysis buffer [100 mM Tris-HCl, 100 mM Na EDTA, 1.5 M NaCl, 1% (wt/vol) cetrimerium bromide, 2% (wt/vol) SDS, pH 8.0] and incubated at  $70^{\circ}\text{C}$  for 2 h. Crude eDNA was precipitated with addition of 0.7 vol of isopropanol, collected by centrifugation ( $4,000 \times g$ , 30 min), washed with 70% (vol/vol) ethanol, and then resuspended in 50  $\mu$ L of TE buffer (10 mM Tris-HCl, 1 mM Na EDTA, pH 8.0). eDNA was separated from the crude extract by preparative agarose (0.7% agarose) gel electrophoresis (3 h, 100 V). The gel slice containing high-molecular-weight DNA was excised from this gel and collected by electroelution (100 V, 2 h). Purified high-molecular-weight eDNA was then blunt ended (End-It; Epicentre), ligated into Scal-digested pJWC1 cosmid vector, packaged into lambda phage in vitro (MaxPlax Packaging Extracts; Epicentre), and transfected into *E. coli* EC100 (20). Titers were determined for packaging reaction, and the three libraries were expanded until they each contained  $\sim 1.5$  million clones. In total, 20–50,000 individual clones from each library were robotically arrayed into 384-well microplates and stored as frozen glycerol stocks.

**Generation of Sterile Spent Culture Broth Filtrates.** Arrayed metagenomic clones were pin transferred into sterile deep-well 384-well microplates containing 150  $\mu$ L of LB (tetracycline, 15  $\mu$ g/mL) per well. Arrayed clones were grown for 18 h at  $37^{\circ}\text{C}$ , and then subcultured (40 nL) into microplates containing LB (tetracycline, 15  $\mu$ g/mL). These plates were incubated at  $30^{\circ}\text{C}$  without shaking for 4 d. Mature culture plates were centrifuged at  $1,600 \times g$  for 15 min to pellet the *E. coli*. Eighty microliters of supernatant was then transferred to a Pall 384-well filter plate (0.45- $\mu$ m membrane, low binding GHP membrane), and sterile spent culture broth filtrate collected by centrifugation ( $800 \times g$ , 2 min).

**High-Throughput Fluorescent Microscopy Screening of Metagenomic Clone for NF- $\kappa$ B Activators.** HEK293-TN cells stably transfected with the pGreenFire lentiviral NF- $\kappa$ B GFP-luciferase plasmid from Systems Biosciences were used as a fluorescent reporter cell line (HEK293-NF- $\kappa$ B:GFP). Cells were grown at  $37^{\circ}\text{C}$  5%  $\text{CO}_2$  in 75-cm<sup>2</sup> cell culture flasks with DMEM (10% FBS, 2 mM glutamine, 1% penicillin/streptomycin). For the purpose of the assay, cells were grown in a single large culture and then frozen as 1-mL aliquots (growth media with 10% DMSO) in liquid nitrogen. For each assay, cells were thawed, media replaced at 24 h, and cells grown to 85% confluence. Cells were then trypsinized, counted (Countess; Invitrogen), diluted to 2,000 cells per 25  $\mu$ L, and dispensed into wells of 384-well clear-bottom microplates (uClear; Grenier) using a liquid dispensing robot (Multidrop ThermoScientific). To identify clones that produce effector molecules, 20  $\mu$ L of sterile spent culture broth obtained from each arrayed metagenomic clone was transferred to a unique well in a 384-well plate seeded with HEK293-NF- $\kappa$ B:GFP reporter. HEK293-NF- $\kappa$ B:GFP



were exposed to the spent culture broth for 24 h (37 °C 5% CO<sub>2</sub>). Two hours before imaging, 10 μL of a 1:10,000 dilution of Hoechst 33442 (10 mg/mL; Invitrogen) and a 1:150 dilution of propidium iodide (1 mg/mL; Sigma) were added to each well. Each plate was then transferred to an ImageXpress XLS Widefield High Content Microscope (Molecular Devices) and imaged with a 10× objective. The following fluorescent filters were used for imaging: propidium iodide: Tx red (562/624); GFP: FITC 482/536; Hoechst 33442: DAPI (377/447) (excitation/emission).

Images were analyzed using MetaXpress High Content Image Acquisition Software (Molecular Devices). Nuclei were identified based on Hoechst 33442 staining. Cells were considered dead if they costained with propidium iodide. GFP expression was calculated for live cells and used as a measure of NF-κB activation. For each well, the total number of nuclei, dead cells, and GFP-expressing live cells was recorded. NF-κB activation was then expressed as a percentage of total live cells expressing GFP. Each microplate contained two columns of negative control wells (columns 2, 12). Population statistics, including the mean and SD of NF-κB active cells (SPSS Statistics IBM) were calculated for the negative control wells, and each test well was then given a Z score based on the population statistics from the negative control wells on the same plate:  $Z \text{ score} = (\text{MEAN}_{\text{test}} - \text{MEAN}_{\text{control}}) / (\text{SD}_{\text{control}})$ . Wells on the outer edge of the plate (column 1, 24, and row A, P) were not used due to poor reproducibility of these wells.

Based on the distribution of negative control and test population Z scores, we chose a Z score greater than 3 to be considered an initial screening hit. Each hit was isolated from its corresponding glycerol stocked freezer plates and arrayed into a 96-well microplate. This plate was grown overnight, and then using a 96- to 384-well plate replicator, each well from the 96-well was inoculated into four corresponding wells in a 384-well plate (i.e., four replicates were created). Replicate plates were then grown and assayed as described above. A clone was only carried forward from this reassy step if at least two of the four replicate wells showed a Z score of greater than 2.5.

- Ley RE, Lozupone CA, Hamady M, Knight R, Gordon JI (2008) Worlds within worlds: Evolution of the vertebrate gut microbiota. *Nat Rev Microbiol* 6(10):776–788.
- Human Microbiome Project Consortium (2012) Structure, function and diversity of the healthy human microbiome. *Nature* 486(7402):207–214.
- Kamada N, Seo SU, Chen GY, Núñez G (2013) Role of the gut microbiota in immunity and inflammatory disease. *Nat Rev Immunol* 13(5):321–335.
- Gagliani N, Hu B, Huber S, Elinav E, Flavell RA (2014) The fire within: Microbes inflame tumors. *Cell* 157(4):776–783.
- Turnbaugh PJ, et al. (2009) A core gut microbiome in obese and lean twins. *Nature* 457(7228):480–484.
- Qin N, et al. (2014) Alterations of the human gut microbiome in liver cirrhosis. *Nature* 513(7516):59–64.
- Qin J, et al. (2012) A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* 490(7418):55–60.
- Smith MI, et al. (2013) Gut microbiomes of Malawian twin pairs discordant for kwashiorkor. *Science* 339(6119):548–554.
- Ridaura VK, et al. (2013) Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science* 341(6150):1241214.
- Bongers G, et al. (2014) Interplay of host microbiota, genetic perturbations, and inflammation promotes local development of intestinal neoplasms in mice. *J Exp Med* 211(3):457–472.
- Garrett WS, et al. (2007) Communicable ulcerative colitis induced by T-bet deficiency in the innate immune system. *Cell* 131(1):33–45.
- Qin J, et al.; MetaHIT Consortium (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464(7285):59–65.
- Ivanov II, Honda K (2012) Intestinal commensal microbes as immune modulators. *Cell Host Microbe* 12(4):496–508.
- Holmes E, et al. (2012) Therapeutic modulation of microbiota-host metabolic interactions. *Sci Transl Med* 4(137):137rv6.
- Brady SF, Chao CJ, Clardy J (2004) Long-chain N-acetyltyrosine synthases from environmental DNA. *Appl Environ Microbiol* 70(11):6865–6870.
- Brady SF, Simmons L, Kim JH, Schmidt EW (2009) Metagenomic approaches to natural products from free-living and symbiotic organisms. *Nat Prod Rep* 26(11):1488–1503.
- Chang FY, Brady SF (2013) Discovery of indolotryptoline antiproliferative agents by homology-guided metagenomic screening. *Proc Natl Acad Sci USA* 110(7):2478–2483.
- Charlop-Powers Z, Banik JJ, Owen JG, Craig JW, Brady SF (2013) Selective enrichment of environmental DNA libraries for genes encoding nonribosomal peptides and polyketides by phosphopantetheine transferase-dependent complementation of siderophore biosynthesis. *ACS Chem Biol* 8(1):138–143.
- Charlop-Powers Z, Milshteyn A, Brady SF (2014) Metagenomic small molecule discovery methods. *Curr Opin Microbiol* 19:70–75.
- Craig JW, Chang FY, Brady SF (2009) Natural products from environmental DNA hosted in *Ralstonia metallidurans*. *ACS Chem Biol* 4(1):23–28.
- Craig JW, Chang FY, Kim JH, Obiajulu SC, Brady SF (2010) Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid environmental DNA libraries in diverse proteobacteria. *Appl Environ Microbiol* 76(5):1633–1641.
- Feng Z, Chakraborty D, Dewell SB, Reddy BV, Brady SF (2012) Environmental DNA-encoded antibiotics fusamycins A and B inhibit FabF in type II fatty acid biosynthesis. *J Am Chem Soc* 134(6):2981–2987.
- Feng Z, Kallifidas D, Brady SF (2011) Functional analysis of environmental DNA-derived type II polyketide synthases reveals structurally diverse secondary metabolites. *Proc Natl Acad Sci USA* 108(31):12629–12634.
- Kallifidas D, Kang HS, Brady SF (2012) Tetarimycin A, an MRSA-active antibiotic identified through induced expression of environmental DNA gene clusters. *J Am Chem Soc* 134(48):19552–19555.
- Lakhdari O, et al. (2010) Functional metagenomics: A high throughput screening method to decipher microbiota-driven NF-κB modulation in the human gut. *PLoS One* 5(9):e13092.
- Culligan EP, Sleanor RD, Marchesi JR, Hill C (2012) Functional metagenomics reveals novel salt tolerance loci from the human gut microbiome. *ISME J* 6(10):1916–1925.
- Gevers D, et al. (2014) The treatment-naïve microbiome in new-onset Crohn's disease. *Cell Host Microbe* 15(3):382–392.
- Hansen R, et al. (2012) Microbiota of de-novo pediatric IBD: Increased *Faecalibacterium prausnitzii* and reduced bacterial diversity in Crohn's but not in ulcerative colitis. *Am J Gastroenterol* 107(12):1913–1922.
- Morgan XC, et al. (2012) Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol* 13(9):R79.
- Baltimore D (2011) NF-κB is 25. *Nat Immunol* 12(8):683–685.
- Le LQ, et al. (2010) Mice lacking the orphan G protein-coupled receptor G2A develop a late-onset autoimmune syndrome. *Immunity* 14(5):561–571.
- Wang Z, et al. (2011) Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* 472(7341):57–63.
- Tlaskalová-Hogenová H, et al. (2011) The role of gut microbiota (commensal bacteria) and the mucosal barrier in the pathogenesis of inflammatory and autoimmune diseases and cancer: Contribution of germ-free and gnotobiotic animal models of human diseases. *Cell Mol Immunol* 8(2):110–120.
- Rimmerman N, et al. (2008) N-Palmitoyl glycine, a novel endogenous lipid that acts as a modulator of calcium influx and nitric oxide production in sensory neurons. *Mol Pharmacol* 74(1):213–224.
- Hanuš L, Shohami E, Bab I, Mechoulam R (2014) N-Acyl amino acids and their impact on biological processes. *Biofactors* 40(4):381–388.
- Brady SF (2007) Construction of soil environmental DNA cosmid libraries and screening for clones that produce biologically active small molecules. *Nat Protoc* 2(5):1297–1305.
- Divanovic S, et al. (2005) Negative regulation of Toll-like receptor 4 signaling by the Toll-like receptor homolog RP105. *Nat Immunol* 6(6):571–578.
- Nayfach S, Pollard KS (2015) Average genome size estimation improves comparative metagenomics and sheds light on the functional ecology of the human microbiome. *Genome Biol* 16(1):51.
- Atkinson GC, Tenson T, Hauryliuk V (2011) The RelA/SpoT homolog (RSH) superfamily: Distribution and functional evolution of ppGpp synthetases and hydrolases across the tree of life. *PLoS One* 6(8):e23479.
- Srivatsan A, Wang JD (2008) Control of bacterial transcription, translation and replication by (p)ppGpp. *Curr Opin Microbiol* 11(2):100–105.

## Sequencing and Determination of Effector Genes by Transposon Mutagenesis.

Cosmid DNA was isolated from each reproducibly active clone and sequenced by PGM IonTorrent. Gaps were closed by primer walking. Final assemblies were annotated using CloVR (81). When two or more clones containing overlapping sequences were identified, only one representative clone was chosen for downstream analysis. Cosmid DNA from each unique clone was transposon mutagenized using the EZ TN-5 Kan Transposon (Epicentre). Transposon mutants were selected for using kanamycin (50 μg/mL) and tetracycline (15 μg/mL), and individual clones were inoculated into 96-well microplates and assayed for GFP-inducing activity as outlined above. Each microplate contained the original unmutagenized clone and negative control in individual columns. Transposon mutants that varied by at least 3 SDs from the positive control were selected as primary hits and then reassayed (28 replicates) to confirm loss of the ability to activate the reporter cell line. The locations of transposons in knockout mutants were confirmed by Sanger sequencing using transposon-specific sequencing primers. Transposon mutagenesis was not performed in three instances due to the presence within a given cosmid of a gene showing high similarity (100%, 95%, 99%) to a gene already identified by transposon mutagenesis (8j18, 37b14, 40k7).

**Supporting Information.** Please see *SI Appendix* for methods concerning the detailed characterization of commendamide.

**ACKNOWLEDGMENTS.** This work was in part supported by the Center for Basic and Translational Research on Disorders of the Digestive System through the generosity of the Leona M. and Harry B. Helmsley Charitable Trust; Saferstein Family Charitable Foundation; Grant UL1 TR000043 from the National Center for Advancing Translational Sciences, National Institutes of Health Clinical and Translational Science Award Program; RAININ Foundation; and NIH Grants GM077516 and T32GM07739.

41. Bolam DN, Koropatkin NM (2012) Glycan recognition by the Bacteroidetes Sus-like systems. *Curr Opin Struct Biol* 22(5):563–569.
42. Hall LMC, et al. (2005) Sequence diversity and antigenic variation at the rag locus of *Porphyromonas gingivalis*. *Infect Immun* 73(7):4253–4262.
43. Noinaj N, Guillier M, Barnard TJ, Buchanan SK (2010) TonB-dependent transporters: Regulation, structure, and function. *Annu Rev Microbiol* 64:43–60.
44. Anantharaman V, Aravind L (2003) Evolutionary history, structural features and biochemical diversity of the NlpC/P60 superfamily of enzymes. *Genome Biol* 4(2):R11.
45. Xu Q, et al. (2010) Structure of the  $\gamma$ -D-glutamyl-L-diamino acid endopeptidase Ykfc from *Bacillus cereus* in complex with L-Ala- $\gamma$ -D-Glu: Insights into substrate recognition by NlpC/P60 cysteine peptidases. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 66(Pt 10):1354–1364.
46. Xu Q, et al. (2014) Structure-guided functional characterization of DUF1460 reveals a highly specific NlpC/P60 amidase family. *Structure* 22(12):1799–1809.
47. Lenz LL, Mohammadi S, Geissler A, Portnoy DA (2003) SecA2-dependent secretion of autolytic enzymes promotes *Listeria monocytogenes* pathogenesis. *Proc Natl Acad Sci USA* 100(21):12432–12437.
48. Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J Mol Biol* 305(3):567–580.
49. Mohammadi T, et al. (2007) The essential peptidoglycan glycosyltransferase MurG forms a complex with proteins involved in lateral envelope growth as well as with proteins involved in cell division in *Escherichia coli*. *Mol Microbiol* 65(4):1106–1121.
50. Fiedler W, Rotering H (1985) Characterization of an *Escherichia coli* mdoB mutant strain unable to transfer sn-1-phosphoglycerol to membrane-derived oligosaccharides. *J Biol Chem* 260(8):4799–4806.
51. Oriol R, Martinez-Duncker I, Chantret I, Mollicone R, Codogno P (2002) Common origin and evolution of glycosyltransferases using Dol-P-monosaccharides as donor substrate. *Mol Biol Evol* 19(9):1451–1463.
52. Shindou H, Hishikawa D, Harayama T, Yuki K, Shimizu T (2009) Recent progress on acyl CoA: Lysophospholipid acyltransferase research. *J Lipid Res* 50(Suppl):S46–S51.
53. Robertson KP, Smith JC, Gough AM, Rocha ER (2006) Characterization of *Bacteroides fragilis* hemolysins and regulation and synergistic interactions of HlyA and HlyB. *Infect Immun* 74(4):2304–2316.
54. Mayberry WR (1980) Cellular distribution and linkage of  $\alpha$ -(-)-3-hydroxy fatty acids in *Bacteroides* species. *J Bacteriol* 144(1):200–204.
55. Morishita T, et al. (1997) N-type calcium channel blockers from a marine bacterium, *Cytophaga* sp. SANK 71996. *J Antibiot (Tokyo)* 50(6):457–468.
56. Brady SF, Chao CJ, Clardy J (2002) New natural product families from an environmental DNA (eDNA) gene cluster. *J Am Chem Soc* 124(34):9968–9969.
57. Felder CC, et al. (1993) Anandamide, an endogenous cannabimimetic eicosanoid, binds to the cloned human cannabinoid receptor and stimulates receptor-mediated signal transduction. *Proc Natl Acad Sci USA* 90(16):7656–7660.
58. Bradshaw HB, Rimmerman N, Hu SS-J, Burstein S, Walker JM (2009) Novel endogenous N-acyl glycines identification and characterization. *Vitam Horm* 81:191–205.
59. Wellner N, Diep TA, Janfelt C, Hansen HS (2013) N-Acylation of phosphatidylethanolamine and its biological functions in mammals. *Biochim Biophys Acta* 1831(3):652–662.
60. Brady SF, Clardy J (2005) N-Acyl derivatives of arginine and tryptophan isolated from environmental DNA expressed in *Escherichia coli*. *Org Lett* 7(17):3613–3616.
61. Weng Z, et al. (1998) A DNA damage and stress inducible G protein-coupled receptor blocks cells in G2/M. *Proc Natl Acad Sci USA* 95(21):12334–12339.
62. Kabarowski JH (2009) G2A and LPC: Regulatory functions in immunity. *Prostaglandins Other Lipid Mediat* 89(3–4):73–81.
63. Obinata H, Hattori T, Nakane S, Tatei K, Izumi T (2005) Identification of 9-hydroxyoctadecadienoic acid and other oxidized free fatty acids as ligands of the G protein-coupled receptor G2A. *J Biol Chem* 280(49):40676–40683.
64. Bolick DT, et al. (2009) G2A deficiency in mice promotes macrophage activation and atherosclerosis. *Circ Res* 104(3):318–327.
65. Frasch SC, et al. (2007) Lysophospholipids of different classes mobilize neutrophil secretory vesicles and induce redundant signaling through G2A. *J Immunol* 178(10):6540–6548.
66. Hasegawa H, et al. (2011) Lysophosphatidylcholine enhances the suppressive function of human naturally occurring regulatory T cells through TGF- $\beta$  production. *Biochem Biophys Res Commun* 415(3):526–531.
67. Khan SY, et al. (2010) Lysophosphatidylcholines activate G2A inducing  $G_{\alpha_{i1}}$ -/ $G_{\alpha_{q/11}}$ - $Ca^{2+}$  flux,  $G_{\beta\gamma}$ -Hck activation and clathrin/ $\beta$ -arrestin-1/GRK6 recruitment in PMNs. *Biochem J* 432(1):35–45.
68. Murakami N, et al. (2009) Transcriptional regulation of human G2A in monocytes/macrophages: Involvement of  $\alpha$ EBPs, Runx and Pu.1. *Genes Cells* 14(12):1441–1455.
69. Radu CG, Yang LV, Riedinger M, Au M, Witte ON (2004) T cell chemotaxis to lysophosphatidylcholine through the G2A receptor. *Proc Natl Acad Sci USA* 101(1):245–250.
70. Yan J-J, et al. (2004) Therapeutic effects of lysophosphatidylcholine in experimental sepsis. *Nat Med* 10(2):161–167.
71. Yang LV, Radu CG, Wang L, Riedinger M, Witte ON (2005) Gi-independent macrophage chemotaxis to lysophosphatidylcholine via the immunoregulatory GPCR G2A. *Blood* 105(3):1127–1134.
72. Marini I, Bartolucci ML, Bortolotti F, Gatto MR, Bonetti GA (2012) Palmitoylethanolamide versus a nonsteroidal anti-inflammatory drug in the treatment of temporomandibular joint inflammatory pain. *J Orofac Pain* 26(2):99–104.
73. Sheerin AH, Zhang X, Saucier DM, Corcoran ME (2004) Selective antiepileptic effects of N-palmitoylethanolamide, a putative endocannabinoid. *Epilepsia* 45(10):1184–1188.
74. Hesselink JM, Hekker TA (2012) Therapeutic utility of palmitoylethanolamide in the treatment of neuropathic pain associated with various pathological conditions: A case series. *J Pain Res* 5:437–442.
75. Ning Y, et al. (2008) Endogenous and synthetic agonists of GPR119 differ in signalling pathways and their effects on insulin secretion in MIN6c4 insulinoma cells. *Br J Pharmacol* 155(7):1056–1065.
76. Vereecken NJ, McNeil JN (2010) Cheaters and liars: Chemical mimicry at its finest. *Can J Zool* 88(7):725–752.
77. Stebbins CE, Gala JE (2001) Structural mimicry in bacterial virulence. *Nature* 412(6848):701–705.
78. Murphy PM (2001) Viral exploitation and subversion of the immune system through chemokine mimicry. *Nat Immunol* 2(2):116–122.
79. Harvey RF, Bradshaw JM (1980) A simple index of Crohn's-disease activity. *Lancet* 1(8167):514.
80. Rachmilewitz D (1989) Coated mesalazine (5-aminosalicylic acid) versus sulphasalazine in the treatment of active ulcerative colitis: A randomised trial. *BMJ* 298(6666):82–86.
81. Angiuoli SV, et al. (2011) CloVR: A virtual machine for automated and portable sequence analysis from the desktop using cloud computing. *BMC Bioinformatics* 12:356.