Thousands of Novel Endolysins Discovered in Uncultured Phage Genomes

Iris Fernández-Ruiz[†], Felipe H. Coutinho[†] and Francisco Rodriguez-Valera*

Evolutionary Genomics Group, Departamento de Producción Vegetal y Microbiología, Universidad Miguel Hernández de Elche, San Juan de Alicante, Spain

Bacteriophages express endolysins toward the end of their replication cycle to degrade the microbial cell wall from within, allowing viral progeny to be released. Endolysins can also degrade the prokaryotic cell wall from the outside, thus have potential to be used for biotechnological and medical purposes. Multiple endolysins have been identified within the genomes of isolated phages, but their diversity in uncultured phages has been overlooked. We used a bioinformatics pipeline to identify novel endolysins from nearly 200,000 uncultured viruses. We report the discovery of 2,628 putative endolysins, many of which displayed novel domain architectures. In addition, several of the identified proteins are predicted to be active against genera that include pathogenic bacteria. These discoveries enhance the diversity of known endolysins and are a stepping stone for developing medical and biotechnological applications that rely on bacteriophages, the most diverse biological entities on Earth.

Keywords: bacteriophage, endolysins, metagenomics, biotechnology, protein discovery

INTRODUCTION

Bacteriophages (Phages) have evolved a multitude of endolysins for the purpose of degrading the complex cell wall structures of their bacterial hosts (Pimentel, 2014). Despite sharing a common function, endolysins are a diverse class of enzymes with a multitude of action mechanisms and protein architectures (Oliveira et al., 2013; Vidová et al., 2014). Protein domains within endolysins grant them their host specificity and efficiency (Hermoso et al., 2007). Multiple classes of endolysins have been described and classified according to the specific enzymatic activities of their catalytic domains, including glycosidases, amidases, and carboxy/endo-peptidase (Schmelcher et al., 2012). Endolysins of phages that infect Gram-positive bacteria often have a modular structure that includes one enzymatic catalytic domain (ECD) at the N-terminal and at least one cell wall binding domain (CBD) at the C-terminal portion of the protein connected by a flexible interdomain linker (Nelson et al., 2012). Meanwhile endolysins of phages that infect Gram-negative bacteria often exhibit a globular structure that contains a single catalytic domain (Pohane and Jain, 2015).

Some endolysins can degrade the cell wall from the outside, which grants them potential to be used for biotechnological and medical applications (Nelson et al., 2012; Gutiérrez et al., 2018). The ongoing crisis of antibiotic based treatments calls for alternative strategies for fighting bacterial infections (Ventola, 2015) and endolysins have multiple advantages over antibiotics: they have higher target specificity and so far no forms of resistance have been reported (Nelson et al., 2012). Furthermore, endolysins can be engineered to alter their host range and efficiency (Díez-Martínez et al., 2014; Blázquez et al., 2016). Several recombinant endolysins are now in preliminary study phase for use in human and veterinary medicine, with promising results

for the treatment against both Gram-positive and Gram-negative bacteria (Briers et al., 2014; Cooper et al., 2016). Some endolysins can even combat intracellular human pathogens (Shen et al., 2016), demonstrating that their potential applicabilities are wider than originally conceived. Finally, these enzymes can be cloned into expression vectors for large scale synthesis. Therefore the potential applications of endolysins are not only of medical but also of industrial (e.g., detecting food-borne pathogens) and agricultural relevance (e.g., treatment against phytopathogens) (Schmelcher and Loessner, 2016).

Despite the recognized diversity and potential of endolysins, our current understanding of these enzymes is limited. Analysis of sequence repositories suggests that less than 1,000 of these proteins are currently known (Oliveira et al., 2013). Although previous studies have characterized the diversity of endolysins through bioinformatic approaches, they focused on reference genomes of isolated phages and prophages, and overlooked environmental phages that have not yet been isolated or cultured (Oliveira et al., 2013; Vidová et al., 2014). Thus, currently available catalogs of endolysins do not cover the diversity of enzymes encoded in the genomes of the many uncultured phages spread across Earth's many ecosystems. Culture independent approaches have revolutionized our understanding of phage genetic diversity, revealing thousands of phage genomes and entirely novel evolutionary lineages at an unprecedented scale (Mizuno et al., 2013; Roux et al., 2015; Paez-Espino et al., 2016; Yutin et al., 2017). These novel genomes are a rich resource for the discovery of endolysins that could have unique domain architectures and target hosts for which no endolysins are currently known. Thus, we sought to screen the genomes of bacteriophages discovered through culture independent approaches to expand the known repertoire of endolysins, asses their structural diversity, and determine how this diversity changes across targeted hosts and ecosystems.

MATERIALS AND METHODS

A database of uncultured viral genomes was compiled from publications aimed at large scale discovery of phages without culturing (Mizuno et al., 2013, 2016; Roux et al., 2015, 2016; Paez-Espino et al., 2016; Coutinho et al., 2017). This dataset comprised 183,298 genomic sequences (**Supplementary File S1**) of uncultured viral genomes adding up to 2.9 Gbp of raw data (**Supplementary Table S1**). We also retrieved available metadata associated with those sequences regarding the ecosystems from which they originated, and the predicted hosts reported in the original publications. These studies used multiple strategies to assign hosts to metagenomic contigs which included: high genomic similarity to reference phage genomes; homology matches between phage and prokaryotic genomes; CRISPR spacers from prokaryotic genomes matching metagenomic contigs; similarity between phage and prokaryotic tRNAs; and co-occurrence of phage genome pairs indicative of a shared host. Prophages described by Roux et al. (2015) were assigned hosts according to the prokaryotic genomes in which they were identified.

A dataset of *bona fide* endolysin sequences encoded in genomes of double stranded DNA phages was compiled to be used as a reference database. This database comprised 629 proteins from NCBI RefSeq phages and was manually curated so to exclude structural lysins (i.e., exolysins) (Oliveira et al., 2013). Prodigal (Hyatt et al., 2010) was run in metagenomic mode to identify protein encoding genes of uncultured phage genomes. All predicted protein sequences were queried against the reference endolysin database using Diamond (Buchfink et al., 2015). Proteins that had hits to the reference database were classified as putative endolysins if matches were within the following thresholds: identity $\geq$ 50%, e-value $\leq$ 0.001, query coverage $\geq$ 30%, and alignment length $\geq$ 50 amino acids. Protein domains of putative endolysins were identified by querying sequences against the Pfam database using HMMER version v3.1b2 (Finn et al., 2015) with default parameters. Additionally, putative endolysins were analyzed through SignalIP (Petersen et al., 2011) to detect signal peptide sequences. Finally, both the putative and reference endolysins were clustered into orthologous groups (OGs) using OrthoMCL (Li et al., 2003) within the GET_HOMOLOGUES pipeline (Contreras-Moreira and Vinuesa, 2013) by setting an inflation factor of 1 and all other parameters set to default. Multiple sequence alignments were constructed through Clustal Omega (Sievers et al., 2014) for each OG represented by at least 10 proteins. Alignments were used to perform phylogenetic reconstructions through FastTree (Price et al., 2010) using default parameters (Amino acid distances BLOSUM45 and Jones-Taylor-Thorton model and support value calculation). An additional tree was built based on a multiple alignment of all proteins assigned to OGs of 10 or more proteins.

RESULTS AND DISCUSSION

A total of 2,628 putative endolysins were identified (**Supplementary File S2**). Homolog identification clustered these proteins into 297 OGs. We focused downstream analysis on 46 OGs represented by 10 or more proteins. Each of these OGs was manually validated as true endolysins by inspecting for: presence of *bona fide* endolysins from the reference database; prevalence of typical endolysins domains (e.g., Lysozyme, Amidase, and Glycosyl Hydrolase); and close phylogenetic relationship between putative and *bona fide* endolysins as assessed by inspecting phylogenetic trees generated for each OG. Considering the degree of diversity among phage genomes and the rate in which their genes evolve, we used very conservative thresholds to identify putative endolysins (identity $\geq$ 50%, e-value $\leq$ 0.001, query coverage $\geq$ 30% and alignment length $\geq$ 50 amino acids). Yet, due to high levels of structural similarities between structural lysins and endolysins, it is possible that some of the putative endolysins identified are actually structural lysins (enzymes used by phages to breach the cell wall at the beginning of the infection process). Yet, using a curated database of *bona fide* endolysins from reference phage genomes, conservative thresholds, and manual validation of the OGs should minimize the occurrence of false positives in our dataset.

Proteins assigned to the same OG often displayed identical domain architectures, although some exceptions were observed (**Figure 1** and **Supplementary File S3**). A total of 62 domains were identified across 46 OGs (**Table 1**), including 26 types of ECDs, 12 CBDs, and 24 domains of unknown function. For 27 OGs only a single catalytic domain was observed. Meanwhile, 19 OGs harbored at least one protein with both cell wall binding and ECDs. Signal peptide sequences were detected only in six OGs, suggesting that most of the discovered endolysins rely on other mechanisms for membrane translocation such as holin-dependent transportation. The most frequent domain observed in the putative endolysins was phage_lysozyme (PF00959) a Glycoside Hydrolase, followed by Peptidase_M15 (PF08291), and Amidase_2 (PF01510).

We investigated associations between OGs and predicted hosts of uncultured phages. A total of 32 OGs had at least one endolysin derived from a phage genome for which host prediction was available. Endolysins from OGs predicted to target Gram-negatives were often composed of a single ECD, while those derived from OGs predicted to target Gram-positive bacteria were often composed of both a CBD and an ECD. Often a single bacterial genus could be targeted

by multiple OGs (**Figure 2**). For example, enzymes targeting *Streptococcus* were present in nine distinct OGs and those targeting *Bacillus* were found in four OGs. Most OGs are predicted to be active against multiple bacterial genera with only two exceptions: OG44 and OG26 which respectively target *Streptococcus* and *Bacillus*. In general, OGs predicted to target multiple taxa were restricted to organisms with the same gram staining patterns. Interestingly endolysins from phages predicted to infect *Mycobacterium* were all assigned to a single group (OG6), likely due to the unique cell wall composition of these organisms.

A total of 34 OGs harbor at least one endolysin predicted to be active against genera that include potentially pathogenic bacteria (e.g., *Staphylococcus*, *Clostridium*, and *Klebsiella*). Ten of those OGs have at least one protein predicted to target bacteria currently classified by the World Health Organization as critical priority for the development of alternative therapies. In addition, five OGs include proteins predicted to be active against phytopathogens (e.g., *Xylella, Erwinina,* and *Burkholderia*) (Mansfield et al., 2012; Blomme et al., 2017). *In vivo* experiments are necessary to confirm the efficiency of these proteins for fighting pathogens, since different species or strains of bacteria
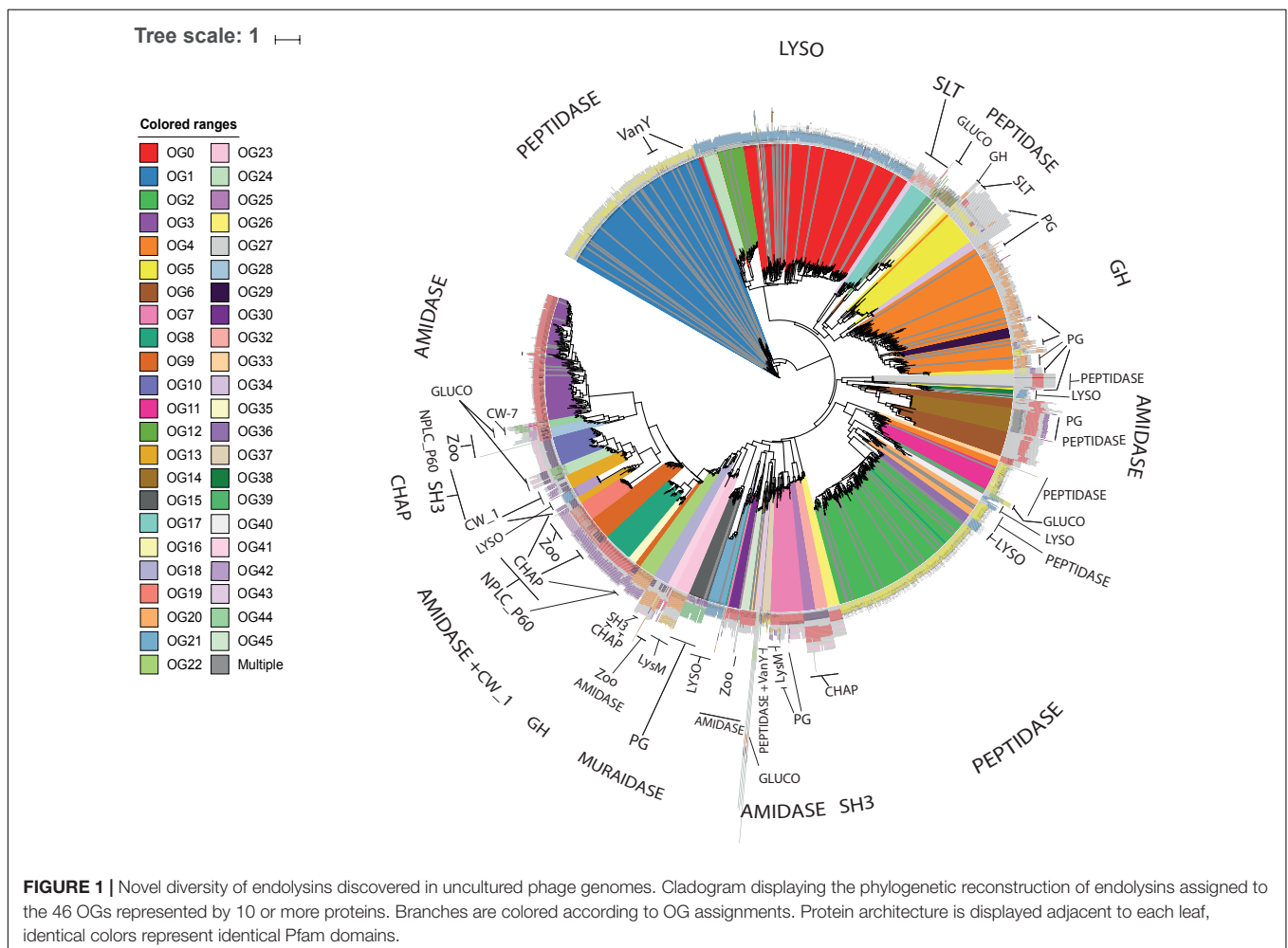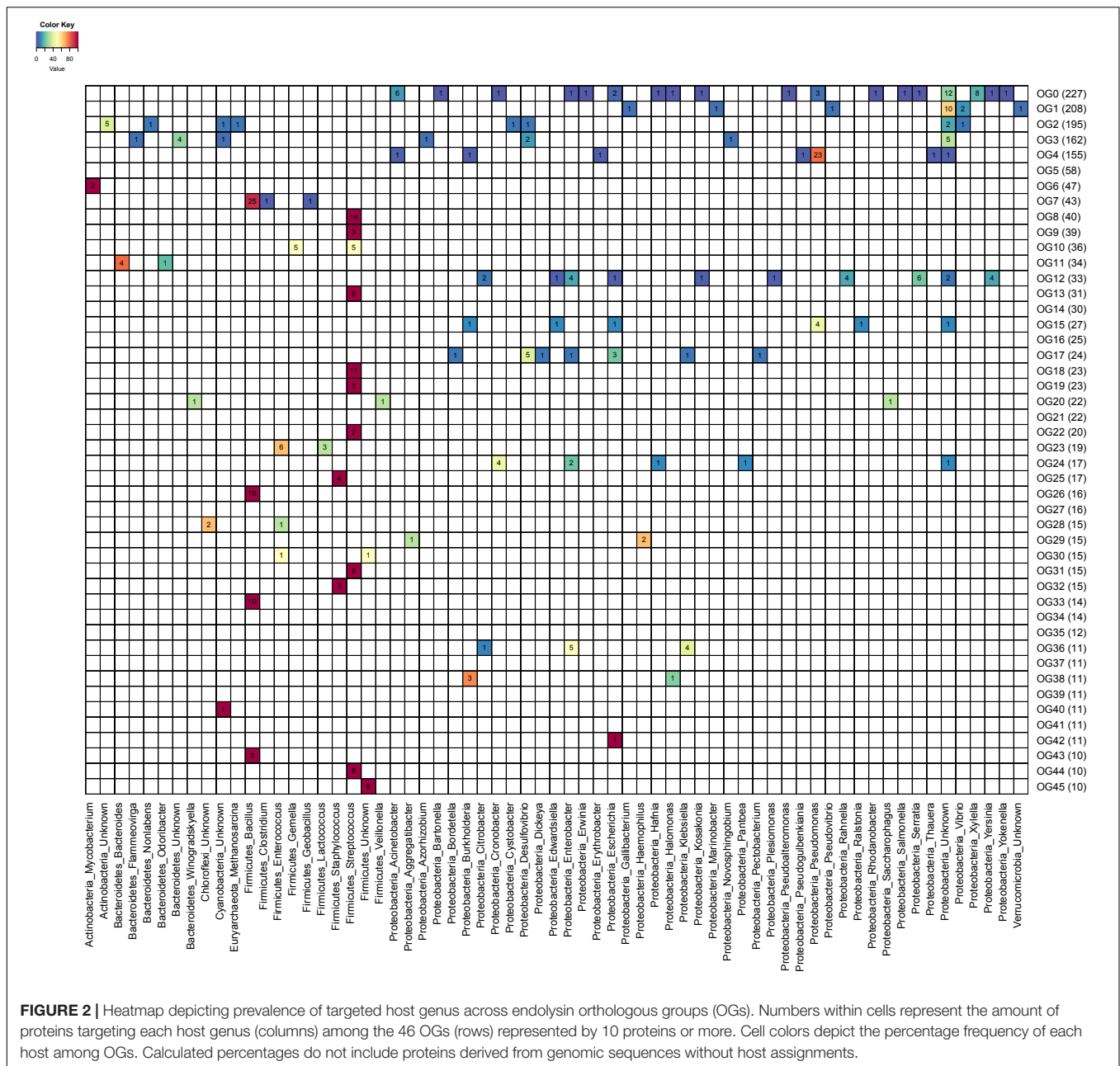

FIGURE 1 | Novel diversity of endolysins discovered in uncultured phage genomes. Cladogram displaying the phylogenetic reconstruction of endolysins assigned to the 46 OGs represented by 10 or more proteins. Branches are colored according to OG assignments. Protein architecture is displayed adjacent to each leaf, identical colors represent identical Pfam domains.

TABLE 1 | Prevalence of domains, signal peptide and inferred cell wall target of endolysin orthologous groups.

FIGURE 2 | Heatmap depicting prevalence of targeted host genus across endolysin orthologous groups (OGs). Numbers within cells represent the amount of proteins targeting each host genus (columns) among the 46 OGs (rows) represented by 10 proteins or more. Cell colors depict the percentage frequency of each host among OGs. Calculated percentages do not include proteins derived from genomic sequences without host assignments.
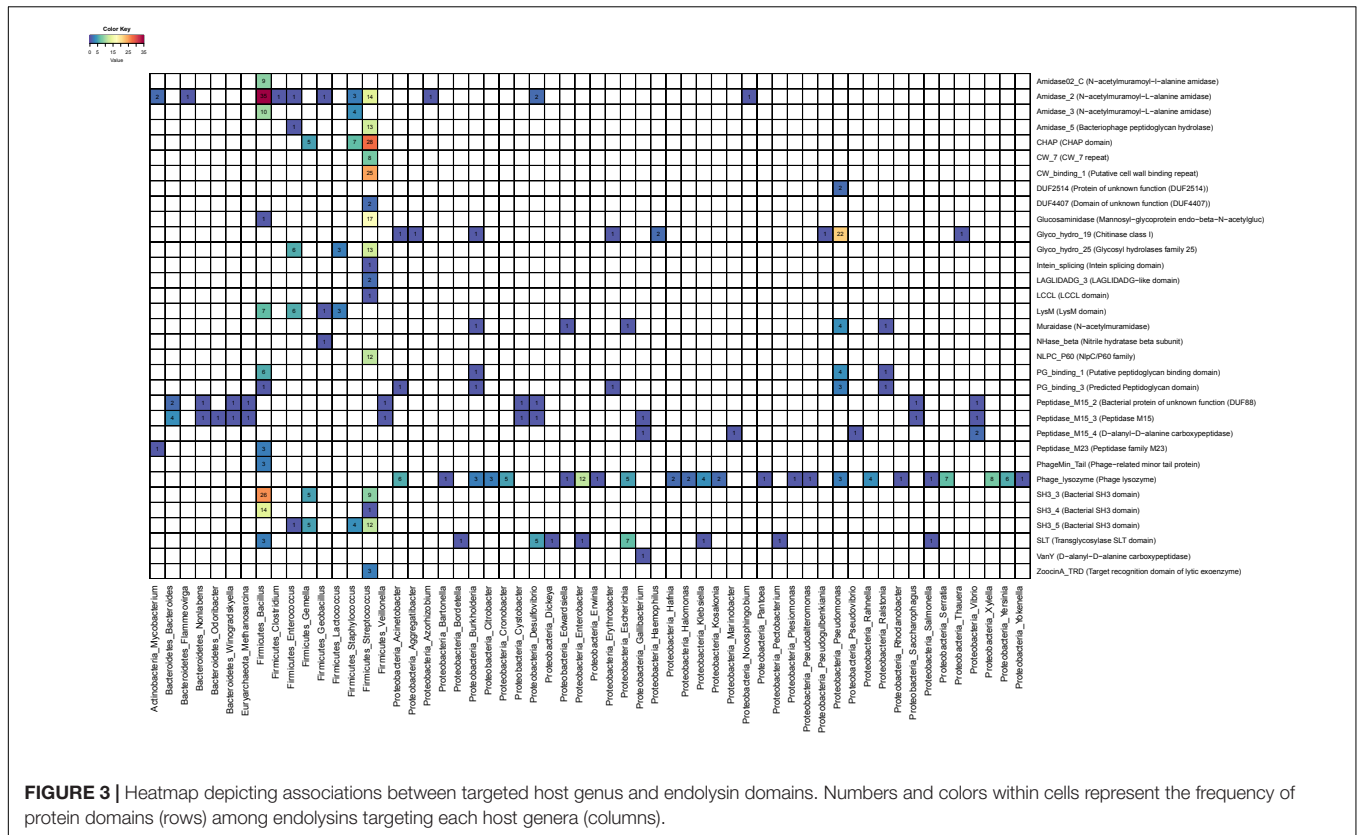
that belong to the same genus may differ in their pathogenic potential.

Next we investigated the frequency of associations between endolysin domains and targeted bacteria (**Figure 3**). On the one hand, some genera were targeted by few domain categories. For example, most of the enzymes that targeted *Bacillus* have amidase (PF12123/PF01510/PF01520) activity in their catalytic domain and SH3 (PF08239/PF06347) on their CBDs, suggesting that this might be the most efficient mechanism to degrade the cell wall of the members of this clade. On the other hand, genera such as *Streptococcus* were targeted by endolysins that rely on multiple types of catalytic activities such as Amidase (PF01510/PF05382), Cysteine Histidine-dependent Amidohydrolases/Peptidases

(CHAP) domain (PF05257), Glucosaminidase (PF01832), Glycosyl hydrolases family 25 (PF01183), as well as multiple CBDs such as CW1 (PF01473), CW7 (PF08230), and SH3 (PF08239/PF06347). This pattern suggests that multiple domain architectures might be efficient for targeting the cell walls of *Streptococcus*. Finally, the Phage_lysozyme (PF00959) and Soluble Lytic Transglycosylase (SLT, PF01464/PF13406) domains were often detected in endolysins predicted to target Proteobacteria, demonstrating that these domains are tuned to degrade the cell walls of this diverse phylum of Gram-negatives.

Host assignments for phage genomes derived from metagenomic datasets are based on bioinformatic predictions which vary on their degree of precision and recall. These

FIGURE 3 | Heatmap depicting associations between targeted host genus and endolysin domains. Numbers and colors within cells represent the frequency of protein domains (rows) among endolysins targeting each host genera (columns).

approaches have high accuracy for higher taxonomic ranks such as phylum or class but sometimes lead to incorrect predictions at the lower ranks such as genus or species (Edwards et al., 2015; Coutinho et al., 2017). Thus, the associations between OGs/domains and targeted hosts inferred from metagenomic data should be interpreted as putative and in need of experimental validation (which is currently out of our scope). Yet the validity of our findings is corroborated by: (1) Use of conservative thresholds to maximize accuracy during the host prediction analysis described in the original publications that first reported phage genomes obtained from metagenomic samples. (2) Endolysin domains identified in uncultured phages match those previously described for isolated and cultured bacteriophages (Oliveira et al., 2013). (3) Approximately, 74% of host assignments obtained for the 46 OGs analyzed in depth were derived from prophage sequences integrated into host genomes that were obtained from pure cultures (instead of metagenomic data). Thus the bulk of the host/OG and host/domain associations are based on high confidence host assignments.

Several of the identified protein domains are not typically present in endolysins from reference phage genomes. Those include domains that are likely associated with peptidoglycan lysis such as Melibiase_2 (PF16499; Glycosyl hydrolases family 27) and Glyco_hydro_cc (PF11790; Glycosyl hydrolase catalytic core), but also domains with no obvious role in this process, such as: Methyltransf_16 (PF10294; Lysine methyltransferase),

Lipase_GDSL_2 (PF13472; GDSL-like Lipase/Acylhydrolase family), and PhageMin_Tail (PF10145; Phage-related minor tail protein). Presence of the latter could be either the result of protein fusions or represent novel functions or action mechanisms of endolysins. Since computational domain identification may be subjected to errors, experimental validation is necessary to corroborate the presence of these domains and to determine their molecular functions. Nevertheless, our results demonstrate that the molecular versatility of endolysins is still poorly understood considering the lack of information available for so many of the identified domains.

Finally, we explored associations between endolysin architecture and their ecosystem of origin. Most of the putative endolysins were derived from aquatic ecosystems or human microbiome samples, which are the habitats most often sampled by metagenomic studies (**Supplementary Figure S1**). This pattern highlights the potential for endolysin discovery in these sites but also the need for investigations of other ecosystems such as plant associated and terrestrial habitats, that also yielded novel endolysins. Aquatic phages appear to be a rich resource for endolysins containing amidase, peptidase and glycosidase domains, which points to them as a source of endolysins active against both Gram-positive and Gram-negative bacteria. Meanwhile, endolysins derived from phages of the human microbiome most often have amidase and peptidase domains, typical of enzymes acting on Gram-positive hosts.

CONCLUSION

We were able to notably expand the known diversity of endolysins and describe proteins with novel and often complex domain architectures, thus challenging the current understanding of the diversity of those enzymes. Our findings show that environmental phage genomes, specially those from aquatic and human associated microbiomes, are a rich resource for endolysins discovery. Since several of the identified endolysins are predicted to be effective against plant and animal pathogens, those are ideal candidates for purification and further characterization, specially those predicted to act on Gram-positive bacteria, against which endolysin therapy has showed the most promising results so far (Gerstmans et al., 2016). Our findings regarding associations between bacterial targets and domains provide insights for the engineering of recombinant proteins that have higher efficiency or an extended host spectrum. Further experimental research will be necessary to corroborate our findings regarding the endolytic activity, domain architecture, and target spectrum of these proteins and to evaluate their potential applications as antimicrobial agents. Culture independent approaches will continue to expand the genetic diversity of phages, and our strategy represents a simple, fast, and scalable approach for discovering endolysins encoded in their genomes.

AUTHOR CONTRIBUTIONS

IF-R, FC, and FR-V conceived the study and designed the experiments. IF-R and FC performed the experiments and analyzed the data. All authors contributed to writing the manuscript.

REFERENCES

Blázquez, B., Fresco-Taboada, A., Iglesias-Bexiga, M., Menéndez, M., and García, P. (2016). PL3 amidase, a tailor-made lysin constructed by domain shuffling with potent killing activity against pneumococci and related species. *Front. Microbiol.* 7:1156. doi: 10.3389/fmicb.2016.01156

Blomme, G., Dita, M., Jacobsen, K. S., Pérez Vicente, L., Molina, A., Ocimati, W., et al. (2017). Bacterial diseases of bananas and enset: current state of knowledge and integrated approaches toward sustainable management. *Front. Plant Sci.* 8:1290. doi: 10.3389/fpls.2017.01290

Briers, Y., Walmagh, M., Van Puyenbroeck, V., Cornelissen, A., Cenens, W., Aertsen, A., et al. (2014). Engineered endolysin-based "Artilysins" to combat multidrug-resistant gram-negative pathogens. *mBio* 5:e01379-14. doi: 10.1128/mBio.01379-14.Editor

Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176

Contreras-Moreira, B., and Vinuesa, P. (2013). GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl. Environ. Microbiol.* 79, 7696–7701. doi: 10.1128/AEM.02411-13

Cooper, C. J., Khan Mirzaei, M., and Nilsson, A. S. (2016). Adapting drug approval pathways for bacteriophage-based therapeutics. *Front. Microbiol.* 7:1209. doi: 10.3389/fmicb.2016.01209

Coutinho, F. H., Silveira, C. B., Gregoracci, G. B., Edwards, R. A., Brussaard, C. P. D., Dutilh, B. E., et al. (2017). Marine viruses discovered through metagenomics shed light on viral strategies throughout the oceans. *Nat. Commun.* 8, 1–12. doi: 10.1038/ncomms15955

Díez-Martínez, R., De Paz, H. D., García-Fernández, E., Bustamante, N., Euler, C. W., Fischetti, V. A., et al. (2014). A novel chimeric phage lysin with high in vitro and in vivo bactericidal activity against *Streptococcus pneumoniae*. *J. Antimicrob. Chemother.* 70, 1763–1773. doi: 10.1093/jac/dkv038

Edwards, R. A., McNair, K., Faust, K., Raes, J., and Dutilh, B. E. (2015). Computational approaches to predict bacteriophage-host relationships. *FEMS Microbiol. Rev.* 40, 258–272. doi: 10.1093/femsre/fuv048

Finn, R. D., Clements, J., Arndt, W., Miller, B. L., Wheeler, T. J., Schreiber, F., et al. (2015). HMMER web server: 2015 update. *Nucleic Acids Res.* 43, W30–W38. doi: 10.1093/nar/gkv397

Gerstmans, H., Rodriguez-Rubio, L., Lavigne, R., and Briers, Y. (2016). From endolysins to Artilysin(R)s: novel enzyme-based approaches to kill drug-resistant bacteria. *Biochem. Soc. Trans.* 44, 123–128. doi: 10.1042/BST20150192

Gutiérrez, D., Fernández, L., Rodríguez, A., and García, P. (2018). Are phage lytic proteins the secret weapon to kill *Staphylococcus aureus*? *mBio* 9, e1923-17. doi: 10.1128/mBio.01923-17

Hermoso, J. A., García, J. L., and García, P. (2007). Taking aim on bacterial pathogens: from phage therapy to enzybiotics. *Curr. Opin. Microbiol.* 10, 461–472. doi: 10.1016/j.mib.2007.08.002

Hyatt, D., Chen, G.-L., Locascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119

Li, L., Stoeckert, C. J. Jr., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189. doi: 10.1101/gr.1224503

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.01033/full#supplementary-material

FIGURE S1 | Heatmap depicting associations between ecosystem sources and endolysin domains. Numbers and colors within cells represent the frequency of protein domains (rows) among endolysins identified in phage genomes from each ecosystem category (columns).

TABLE S1 | Detailed metadata of the genomic sequences of uncultured phages analyzed in this study: including sequence identifier, original dataset, sequence length, number of identified protein encoding genes, ecosystem source and host taxonomic classification.

FILE S1 | Multifasta file containing the 183,6298 genomic sequences of uncultured viruses analyzed in this study.
File can be downloaded here: https://drive.google.com/open?id=15rRJMfguPGhfiuQor7puIrllr2J02tJQ

FILE S2 | Multifasta file containing the amino acid sequences of 2,628 putative endolysins identified in the genomes of uncultured viruses.

FILE S3 | Newick format file of the phylogenetic reconstruction of endolysin proteins displayed in **Figure 1**.

Mansfield, J., Genin, S., Magori, S., Citovsky, V., Sriariyanum, M., Ronald, P., et al. (2012). Top 10 plant pathogenic bacteria in molecular plant pathology. *Mol. Plant Pathol.* 13, 614–629. doi: 10.1111/j.1364-3703.2012.00804.x

Mizuno, C. M., Ghai, R., Saghaï, A., López-García, P., and Rodriguez-Valera, F. (2016). Genomes of abundant and widespread viruses from the deep ocean. *mBio* 7:e00805-16. doi: 10.1128/mBio.00805-16

Mizuno, C. M., Rodriguez-Valera, F., Kimes, N. E., and Ghai, R. (2013). Expanding the marine virosphere using metagenomics. *PLoS Genet.* 9:e1003987. doi: 10.1371/journal.pgen.1003987

Nelson, D. C., Schmelcher, M., Rodriguez-Rubio, L., Klumpp, J., Pritchard, D. G., Dong, S., et al. (2012). *Endolysins as Antimicrobials*, 1st Edn. New York, NY: Elsevier Inc. doi: 10.1016/B978-0-12-394438-2.00007-4

Oliveira, H., Melo, L. D., Santos, S. B., Nobrega, F. L., Ferreira, E. C., Cerca, N., et al. (2013). Molecular aspects and comparative genomics of bacteriophage endolysins. *J. Virol.* 87, 4558–4570. doi: 10.1128/JVI.03277-12

Paez-Espino, D., Eloe-Fadrosh, E. A., Pavlopoulos, G. A., Thomas, A. D., Huntemann, M., Mikhailova, N., et al. (2016). Uncovering Earth's virome. *Nature* 536, 425–430. doi: 10.1038/nature19094

Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8:785. doi: 10.1038/nmeth.1701

Pimentel, M. (2014). "Genetics of phage lysis," in *Molecular Genetics of Mycobacteria*, 2nd Edn, eds G. F. Hatfull and W. R. Jacobs Jr. (Washington, DC: ASM Press), 121–133.

Pohane, A. A., and Jain, V. (2015). Insights into the regulation of bacteriophage endolysin: multiple means to the same end. *Microbiology* 161, 2269–2276. doi: 10.1099/mic.0.000190

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. doi: 10.1371/journal.pone.0009490

Roux, S., Brum, J. R., Dutilh, B. E., Sunagawa, S., Duhaime, M. B., Loy, A., et al. (2016). Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* 537, 689–693. doi: 10.1101/053090

Roux, S., Hallam, S. J., Woyke, T., and Sullivan, M. B. (2015). Viral dark matter and virus-host interactions resolved from publicly available microbial genomes. *Elife* 4:e08490. doi: 10.7554/eLife.08490

Schmelcher, M., Donovan, D. M., and Loessner, M. J. (2012). Bacteriophage endolysins as novel antimicrobials. *Future Microbiol.* 7, 1147–1171. doi: 10.2217/fmb.12.97

Schmelcher, M., and Loessner, M. J. (2016). Bacteriophage endolysins: applications for food safety. *Curr. Opin. Biotechnol.* 37, 76–87. doi: 10.1016/j.copbio.2015.10.005

Shen, Y., Barros, M., Vennemann, T., Gallagher, D. T., Yin, Y., Linden, S. B., et al. (2016). A bacteriophage endolysin that eliminates intracellular streptococci. *Elife* 5, 1–26. doi: 10.7554/eLife.13152

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2014). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539–539. doi: 10.1038/msb.2011.75

Ventola, C. L. (2015). The antibiotic resistance crisis: part 2: management strategies and new agents. *Pharm. Ther.* 40, 344–352.

Vidová, B., Šramková, Z., Tišáková, L., Oravkinová, M., and Godány, A. (2014). Bioinformatics analysis of bacteriophage and prophage endolysin domains. *Biologia* 69, 541–556. doi: 10.2478/s11756-014-0358-8

Yutin, N., Makarova, K. S., Gussow, A. B., Krupovic, M., Segall, A., Edwards, R. A., et al. (2017). Discovery of an expansive bacteriophage family that includes the most abundant viruses from the human gut. *Nat. Microbiol.* 3, 38–46. doi: 10.1038/s41564-017-0053-y