luxR Homolog-Linked Biosynthetic Gene Clusters in Proteobacteria

Carolyn A. Brotherton, a Marnix H. Medema, b E. Peter Greenberg a

a Department of Microbiology, University of Washington, Seattle, Washington, USA
b Bioinformatics Group, Wageningen University, Wageningen, The Netherlands

ABSTRACT  Microbes are a major source of antibiotics, pharmaceuticals, and other bioactive compounds. The production of many specialized microbial metabolites is encoded in biosynthetic gene clusters (BGCs). A challenge associated with natural product discovery is that many BGCs are not expressed under laboratory growth conditions. Here we report a genome-mining approach to discover BGCs with luxR-type quorum sensing (QS) genes, which code for regulatory proteins that control gene expression. Our results show that BGCs linked to genes coding for LuxR-like proteins are widespread in Proteobacteria. In addition, we show that associations between luxR homolog genes and BGCs have evolved independently many times, with functionally diverse gene clusters. Overall, these clusters may provide a source of new natural products for which there is some understanding about how to elicit production.

IMPORTANCE  Bacteria biosynthesize specialized metabolites with a variety of ecological functions, including defense against other microbes. Genes that code for specialized metabolite biosynthetic enzymes are frequently clustered together. These BGCs are often regulated by a transcription factor encoded within the cluster itself. These pathway-specific regulators respond to a signal or indirectly through other means of environmental sensing. Many specialized metabolites are not produced under laboratory growth conditions, and one reason for this issue is that laboratory growth media lack environmental cues necessary for BGC expression. Here, we report a bioinformatics study that reveals that BGCs are frequently linked to genes coding for LuxR family QS-responsive transcription factors in the phylum Proteobacteria. The products of these luxR homolog-associated gene clusters may serve as a practical source of bioactive metabolites.

KEYWORDS  quorum sensing, secondary metabolism

Specialized bacterial metabolites have been and continue to be a major source of antibiotics and other bioactive compounds used in medicine (1, 2). The rise in antibiotic-resistant bacterial pathogens has made the discovery of novel antibiotics a pressing public health issue (3). A challenge in the field of natural-product discovery is that many biosynthetic gene clusters (BGCs) are not expressed under laboratory growth conditions (4). Although researchers have found various means to activate the expression of silent clusters, an understanding of the regulatory circuits that control BGCs of interest would provide a major practical advantage to the discovery and study of the encoded product(s) (5). Regulatory elements associated with defense from other microbes might help identify BGCs enriched in antimicrobial activity, and knowledge of the biology of those regulatory elements could guide studies aimed at “waking up” silent BGCs in laboratory culture.

We hypothesized that quorum sensing (QS)-regulated BGCs might be a rich source of novel bioactive compounds for which the regulatory components are well under...
stood in many organisms, such as *Pseudomonas aeruginosa*. QS is a process through which bacteria sense cell density and regulate gene expression in response (6). QS allows bacteria to modify their environment in a coordinated fashion and perform cooperative metabolic activities. QS-regulated products include exopolysaccharides, extracellular enzymes, and specialized metabolites, such as antibiotics. QS-based gene regulation has been studied in many organisms, including Gram-negative and Gram-positive bacteria, and involves a variety of signal molecules (6, 7).

One type of QS is mediated by acyl-homoserine lactone (AHL) signals. In AHL QS, genes are regulated by a member of the LuxR family of transcription factors. Generally, LuxR family members bind and respond to an AHL synthetized by a homolog of the signal synthase protein LuxI (7). The AHLS can diffuse in and out of cells, and thus the concentration of AHLS within the cell serves as a proxy for cell density (8, 9). Canonical AHL QS circuits involve a LuxR family member, which responds to the specific AHL produced by a cognate LuxI family member (10). There has been relatively recent interest in a subset of LuxR-like transcription factors without a cognate LuxI-type AHL synthase. These have been termed solos or orphans. Some solos respond to AHLS produced by other bacteria, and others respond to a non-AHL-based signal produced by the bacterium harboring the LuxR-type solo or a signal from another species (11–14).

LuxR-type QS has been shown to regulate secreted metabolites with an array of ecological functions, including siderophores, redox-active molecules, and antibiotics (15–17). The LuxR-type QS-regulated products carbapenem and mupirocin are antibiotics used in clinics today (18, 19). In addition, other natural products regulated by LuxR-type QS, such as bactobolin and enacyloxin, have antibiotic activity (20, 21). In many cases where BGCs are regulated by QS, the characterized QS-regulated BGCs are located adjacent to the luxR gene; this is the case for the BGCs encoding mupirocin, carbapenem, bactobolin, enacyloxin, and other QS-regulated natural products discussed in more detail below (20–23). However, there are known BGCs regulated by LuxR-type regulators that are not encoded within the BGC itself but rather elsewhere in the genome, such as in the case of the pigment metabolite violacein (24).

Several known antibiotics are QS regulated, and chemical genetics methods that have been developed to study LuxR-based QS might aid in metabolite discovery and isolation by overcoming issues of BGCs that are silent under laboratory conditions. To illustrate this approach, AHL QS was first discovered to control light production in the marine bacterium *Vibrio fischeri*, where luxI codes for an AHL synthase and LuxR responds to the LuxI-generated AHL. Addition of *V. fischeri* AHL (3-oxohexanoyl-L-homoserine lactone) exogenously leads to early-onset and dramatically higher levels of QS-regulated bioluminescence (25). This same strategy can be used to elicit the expression of LuxR homolog-regulated BGCs, which may be silent under laboratory growth conditions due to the lack of an unknown environmental stimulus or insufficient AHL production. Another approach may involve the use of AHL-degrading lactonases, such as AiiA, which effectively degrade AHL signals. AiiA or other lactonases can be added exogenously to cultures to provide a QS-negative control without the need of doing genetic analyses (26). Finally, rich literature on LuxR-type QS provides guidance on growth conditions that activate QS in various organisms, and this literature provides a starting point for studying the expression of BGCs activated by LuxR-like proteins (27).

To fulfill the potential of this QS-based genome-mining approach, computational identification and exploration of QS-regulated BGCs is first required. Here, we performed a systematic analysis of BGCs associated with luxR-type genes across publicly available bacterial genome sequences. Our results show that luxR homolog-associated BGCs are both common and widespread in bacteria in the phylum *Proteobacteria* and that such associations have evolved multiple times within diverse BGC classes.

**RESULTS AND DISCUSSION**

We used two Pfam domains that define LuxR-like proteins (Pfam accession numbers PF00196 and PF03472, which correspond to the DNA-binding and signal-binding
domains, respectively) to search BGCs identified by antiSMASH and obtained a list of 2,081 luxR homolog-associated BGCs. We performed redundancy filtering on gene clusters that harbored LuxR homologs with greater than 90% mutual identity at the amino acid level; one representative BGC from each cluster of LuxR homologs was retained. Thus, we obtained a list of 137 BGCs (see Data Set S1 in the supplemental material).

Our screen identified all known, characterized luxR homolog-associated, QS-regulated BGCs, including those that biosynthesize bactobolin, enacyloxin, mupirocin, carbapenem, malleilactone, thailandamide, corrugation, nunamycin, and phenazine (20–23, 28–33). These clusters have been shown to be regulated by associated LuxR homolog regulators through various experimental approaches, including natural-product isolation and transcriptomics experiments with QS mutants. Representatives of these characterized hits and their corresponding products are shown in Fig. 1. To date, all known LuxR-type QS-regulated BGCs, including those listed above, are restricted to bacteria in the phylum Proteobacteria. These organisms were isolated from a range of sources, including plant-associated (e.g., carbapenem, produced by the plant pathogen Erwinia carotovora), environmental (e.g., bactobolin, produced by soil-dwelling Burkholderia thailandensis E264), and host-associated (e.g., malleilactone, produced by the human-associated pathogen Burkholderia pseudomallei) sources. The identification of these known QS-regulated BGCs validates our bioinformatics method and confirms that this approach allows us to identify QS-regulated BGCs. These data suggest that many of the LuxR homolog-associated BGCs in our list are indeed regulated by QS systems in the producing organisms.

The types of BGCs identified in our screen were diverse (Fig. 2A). A majority of these BGCs have not been characterized, and many of the clusters likely code for the biosynthetic pathways of unknown products. The most common types of biosynthetic pathways represented were nonribosomal peptide synthetase (NRPS) (26%), polyketide synthase (PKS) (10%), and PKS-NRPS hybrid clusters (11%). Various other types of biosynthetic genes made up the remainder of the clusters, including those involved in
bacteriocin, lanthipeptide, thiopeptide, terpene, siderophore, and beta-lactam biosynthesis. Forty-seven percent of the BGCs contained a gene encoding a LuxI homolog, and the remainder have no luxI-type gene contained within the BGC. BGCs encoding LuxI-like proteins came from every class of gene cluster type except for beta-lactam BGCs.

In terms of the ecology of these organisms, the most highly represented members of this list were plant-associated bacteria, such as Erwinia carotovora (Fig. 2B). We identified several BGC-associated LuxR-like proteins encoded by plant-associated bacteria that contained amino acids grouping them with the recently discovered class of LuxR homologs that respond to plant signals rather than AHL signals (34) (Fig. S1). Environmental bacteria, defined as free-living organisms isolated from soil, water, or other environmental sources, were the next most common. Finally, human-associated bacteria, for example Acinetobacter baumannii AYE, were the next-largest group, followed by bacteria from other habitats that did not fit into the above three categories, such as insect symbionts, like Acetobacter malorum strain DmCS_005.

To date, LuxR-type regulators have been characterized only from bacteria within the phylum Proteobacteria. All but one of the BGCs identified here are encoded by bacteria in the phylum Proteobacteria. One hit comes from Streptomyces canus ATCC 12646, which is in the phylum Actinobacteria; its LuxR family member-associated BGC codes for biosynthesis of the antibiotic telomycin (35). According to the Pfam database, there are nine protein sequences from Actinobacteria that contain the LuxR signal-binding domain (Pfam accession number PF03472) (41). Although no LuxR-like protein has been characterized within Actinobacteria, a multiple sequence alignment of this protein with reference LuxR family sequences reveals that it differs at residues identified as amino acid positions that differentiate AHL-responsive LuxR family members and those that respond to plant-associated signals (Fig. S2) (34). The variation of amino acid residues from those in canonical, AHL-responsive as well as plant-responsive LuxR regulators suggests that this LuxR homolog may bind some novel QS signal. Researchers have isolated telomycin from broth cultures of S. canus ATCC 12646, which has had only its telomycin BGC sequenced; the telomycin producer S. canus C-509, which has had its genome fully sequenced, does not harbor a luxI-type gene but does harbor the luxR homolog containing telomycin BGC (37). Together, these data suggest that the LuxR family protein encoded in the telomycin BGC may bind and respond to a novel non-AHL signal molecule, but its role in the regulation of this antibiotic’s synthetic genes is unclear and worthy of future inquiry, especially given that no functional LuxR family protein has been characterized from any organisms in the phylum Actinobacteria.

We found that luxR-type QS-associated BGCs make up about 2.9% of BGCs encoded by bacteria in the phylum Proteobacteria; the antiSMASH database contains 72,178 BGCs encoded by Proteobacteria, and our unfiltered list of luxR homolog-associated BGCs contained 2,081 hits. We identified luxR family member-linked BGCs in members...
of four of the six classes in the phylum Proteobacteria. Forty-one percent of the linked BGCs were in the class Gammaproteobacteria, followed by the BGCs in the Alphaproteobacteria (31%) and Betaproteobacteria (26%). Interestingly, we discovered luxR homolog-associated BGCs in three organisms from the Deltaproteobacteria (Desulfo- capsasulfexigens DSM 10523, Geobacter uraniireducens RF4, and Haliangium ochraceum DSM 14365). This result intrigued us, as QS has not been well studied in organisms belonging to the Deltaproteobacteria, and our results suggest that AHL-based QS systems may regulate specialized metabolite production in some of these organisms. We did not identify any hits in bacteria belonging to the classes Epsilonproteobacteria, Oligoflexia, and Acidithiobacillia. No luxR or luxI homolog genes have been identified to date from Epsilonproteobacteria and Oligoflexia. Some organisms within the Acidithiobacillia class, for example Acidithiobacillus ferrooxidans ATCC 53993, harbor luxR- and luxI-type genes, but so far none has been linked to a BGC.

We examined the genetic context of the BGCs identified in our screen. Manual examination of several BGCs showed that they had GC contents significantly different from those in the organism’s genome (Fig. S3). Altered GC content can be an indicator of horizontal gene transfer events (38). We wondered whether phylogenetic analysis of the LuxR family proteins in our list would reveal any information regarding the evolution of these proteins. We were particularly interested to see whether LuxR family proteins grouped together according to the type of BGC with which they were associated.

Surprisingly, in the resulting phylogenetic analysis of the BGC-associated LuxR family amino acid sequences (Fig. S4), LuxR-type proteins did not group according to the type of BGC with which they were associated. We then examined the organization of the BGCs in groups of closely related LuxR-type proteins (Fig. S5). There were several examples of closely related LuxR-type proteins in which the associated LuxI family protein appeared in one BGC, but not in another. In addition, several closely related LuxR family proteins from related organisms were associated with different types of BGCs. These data suggest that the loss of the AHL synthase gene is a common evolutionary trajectory. In addition, these data suggest that LuxR homologs have dynamic associations with different types of BGCs, which have evolved independently.

Back-of-the-envelope calculations suggest that the associations of luxR homolog genes with BGCs are not random and may reflect functional significance. For example, our data set of 219,499 BGCs included 52,433 NRPS BGCs, of which 1,019, or about 2%, were luxR homolog associated. In contrast, there were 2,000 ectoine BGCs in the data set, out of which only 1, or 0.05%, was LuxR associated. These data suggest that luxR homolog genes more commonly associate with NRPS-type BGCs than with ectoine BGCs. The predicted function(s) of the products made by ectoine (an osmolyte) and NRPS-type BGCs (often antibiotics or other defense metabolites) underlines these differences in frequency of luxR association.

Conclusions. Overall, our results suggest that luxR homolog-associated BGCs are common and widely distributed within bacteria belonging to the phylum Proteobacteria. Our data suggest that luxR homolog-associated BGCs are especially prevalent within plant-associated bacteria.

Based on our bioinformatics method, it is possible that the hits in our list include BGCs that are encoded near or contain a luxR homolog gene but are not, in fact, regulated by QS. Our method identified known LuxR-type QS-regulated BGCs, increasing our confidence that many of the hits identified in our screen are truly QS regulated. Further in silico analysis, such as analysis of possible LuxR-binding boxes upstream of the start codons of the biosynthetic genes, might provide additional confidence regarding the transcriptional regulation of these clusters. However, to conclude that these gene clusters are indeed regulated by QS, in vivo experiments will be necessary. In addition to the luxR homolog-associated BGCs that we identified here, there are likely other BGCs encoded within Proteobacteria that are regulated by LuxR-type regulators that are not encoded within the BGCs but rather elsewhere in the genome. For
example, in the bacterium *Chromobacterium violaceum*, production of the pigment violacein is regulated by a LuxR-type protein that is not linked to the violacein BGC (24).

In conclusion, we anticipate that the products biosynthesized by the clusters identified here will prove to be interesting targets for further study. This is an area of active inquiry in our lab. Our list of luxR homolog-associated BGCs supports the view that there is still much unexplored biosynthetic chemistry within bacteria in the phylum *Proteobacteria*. Finally, the depth of understanding of the biochemical and genetic aspects of LuxR-type QS provides a practical handle on how to elicit production of the metabolites encoded by the BGCs identified here.

**MATERIALS AND METHODS**

**Bioinformatic identification of LuxR-associated BGCs.** To search for BGCs with a luxR homolog, we queried a database of 219,499 BGCs predicted by antiSMASH v3.0 (39) across all nucleotide sequences available in GenBank (40) in December 2015. The hmmsearch tool from the HMMER package (36) was then used to search for genes encoding both Pfam (41) protein domains corresponding to LuxR (Pfam accession numbers PF00196 [DNA binding domain] and PF03472 [signal-binding domain]). BGCs predicted by antiSMASH based solely on the presence of the luxI homolog, itself a biosynthetic gene, were filtered out (i.e., BGCs in which the only biosynthetic gene was the luxI homolog signal synthase were removed from our list), leaving 2,081 hits. Out of these, 1,164 contained a luxI homolog (Pfam accession number PF00765). Finally, CD-HIT (42, 43) was utilized to remove redundancy by using LuxR homolog protein sequences to form homologous clusters with a sequence identity cutoff of 0.9. One LuxR sequence and its corresponding BGC were retained from each cluster, leaving 137 hits.

**Phylogenetic analysis.** LuxR homolog protein sequences were aligned with MUSCLE using a gap open penalty of −2.9, a gap extend penalty of 0, and a hydrophobicity multiplier of 1.2 (44), and the edges of the alignment were trimmed with the alignment editor of MEGA7 (45). This phylogenetic tree is shown in Fig. S4 in the supplemental material. The evolutionary history was inferred using the neighbor-joining method (46). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) are shown next to the branches, and only bootstrap values over 50% are shown (47). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The optimal tree with the sum of branch lengths of 29.94 is shown. The analysis involved 152 amino acid sequences. There were a total of 66 positions in the final data set. The evolutionary distances were computed using the p-distance method (48) and are in the units of the number of amino acid differences per site. All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA7 (45). GerE (NCBI database accession number WP_000659484.1), which possesses a LuxR-like DNA binding domain (Pfam accession number PF00196) but not an AHL-binding domain, was used as an outgroup.

**SUPPLEMENTAL MATERIAL**

Supplemental material for this article may be found at https://doi.org/10.1128/mSystems.00208-17.

**FIG S1**, PDF file, 0.2 MB.
**FIG S2**, PDF file, 0.1 MB.
**FIG S3**, PDF file, 0.3 MB.
**FIG S4**, PDF file, 0.5 MB.
**FIG S5**, PDF file, 2.1 MB.
**DATA SET S1**, PDF file, 0.3 MB.

**ACKNOWLEDGMENTS**

We acknowledge Michael Fischbach for helpful discussions at the beginning of this project.

This work was supported by U.S. Public Health Service (USPHS) grant GM59026 (to E.P.G.). M.H.M. is supported by a Veni grant (863.15.002 to M.H.M.) from The Netherlands Organization for Scientific Research (NWO).

**REFERENCES**

1. Bérdy J. 2005. Bioactive microbial metabolites. J Antibiot 58:1–26. https://doi.org/10.1038/ja.2005.1.
2. Lewis K. 2013. Platforms for antibiotic discovery. Nat Rev Drug Discov 12:371–387. https://doi.org/10.1038/nrd3975.
3. Spellberg B, Shlaes D. 2014. Prioritized current unmet needs for antibacterial therapies. Clin Pharmacol Ther 96:151–153. https://doi.org/10.1038/clpt.2014.106.
4. Rutledge PJ, Challis GL. 2015. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. Nat Rev Microbiol 13:509–523. https://doi.org/10.1038/nrmicro3496.
5. Seyedsayamdost MR. 2014. High-throughput platform for the discovery of elicitors of silent bacterial gene clusters. Proc Natl Acad Sci U S A 111:7266–7271. https://doi.org/10.1073/pnas.1400019111.
6. Waters CM, Bassler BL. 2005. Quorum sensing: cell-to-cell communica-
tion in bacteria. Annu Rev Cell Dev Biol 21:319–346. https://doi.org/10.1146/annurev.cellbio.21.021703.130101
7. Schuster M, Sexton DJ, Diggle SP, Greenberg EP. 2013. Acyl-homoserine lactone quorum sensing: from evolution to application. Annu Rev Microbiol 67:43–63. https://doi.org/10.1146/annurev-micro-092412-155635
8. Pearson JP, Van Delden C, Iglewski BH. 1999. Active efflux and diffusion are involved in transport of Pseudomonas aeruginosa cell-to-cell signals. J Bacteriol 181:1203–1210.
9. Fujita WC, Winters SC, Greenberg EP. 1994. Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators. J Bacteriol 176:269–275. https://doi.org/10.1128/JB.176.2.269-275.1994.
10. Schaefer AL, Hanzelka BL, Eberhard A, Greenberg EP. 1996. Quorum sensing in Vibrio fischeri: probing autoinducer-LuxR interactions with autoinducer analogs. J Bacteriol 178:2897–2901. https://doi.org/10.1128/JB.178.9.2897-2901.1996.
11. Brameyer S, Kresovic D, Bode HB, Heermann R. 2015. Dialkylresorcinols are involved in transport of Pseudomonas aeruginosa cell-to-cell signals. J Bacteriol 197:2618–2627. https://doi.org/10.1128/JB.00099-15.
12. Brachmann AO, Brameyer S, Kresovic D, Hitkova I, Kopp Y, Manske C, Bode HB. 2016. Dialkylresorcinols are involved in transport of Pseudomonas aeruginosa cell-to-cell signals. J Bacteriol 198:1412–1422. https://doi.org/10.1128/JB.00200-16.
13. Sperandio V. 2010. SdiA sensing of acyl-homoserine lactones by enterobacterial LuxR transcriptional activators. Microbiology 141:541–550.
14. Parsek MR, Nierman WC, Greenberg EP. 2009. Quorum-sensing control of biofilm structures and phenotypic states. Annu Rev Microbiol 63:4579–4588. https://doi.org/10.1146/annurev-micro-092412-155635.
15. Haas D. 1997. The global activator GacA of E. coli (H3900) and pathogen virulence. Annu Rev Genet 31:439–458. https://doi.org/10.1146/annurev.genet.31.1.439.
16. Price-Whelan A, Dietrich LE, Newman DK. 2006. Rethinking “secondary” metabolism: physiological roles for phenazine antibiotics. Nat Chem Biol 2:71–78. https://doi.org/10.1038/nchembio764.
17. Papp-Wallace KM, Endimiani A, Taracila MA, Bonomo RA. 2011. Rethinking “secondary” metabolism: physiological roles for phenazine antibiotics. Nat Chem Biol 2:71–78. https://doi.org/10.1038/nchembio764.
18. Majerczyk C, Brittnacher M, Jacobs M, Armour CD, Radey M, Schneider E, Oinuma KI, Greenberg EP, Clardy J. 2010. Quorum-sensing-regulated metabolism in Burkholderia ambifaria. Proc Natl Acad Sci U S A 107:6398–6403. https://doi.org/10.1073/pnas.0913781107.
19. Danese V, Ferriani R, Puglisi E, Masiello G, Loper JE, Keel C, Venturi V, Catara V. 2007. Pseudomonas corrugata contains a conserved n-acyl homoserine lactone quorum sensing system; its role in tomato pathogenicity and tobacco hypersensitivity response. FEMS Microbiol Ecol 61:222–234. https://doi.org/10.1111/j.1574-6941.2007.00338.x.
20. Seyedsayamdost MR, Chandler JR, Blodgett JA, Lima PS, Duerkop BA, Fischbach MA, Müller R, Wohlleben W, Breitling R, Takano E, Medema MH. 2015. antiSMASH 3.0 —a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res 43:1021–1030. https://doi.org/10.1093/nar/gkv1344.
21. Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing sequences without alignment. J Comp Biol 13:2153–2177. https://doi.org/10.1089/cmb.2006.0215.
22. Chomczynski P, Sacchi N. 2000. Single-step method of RNA isolation by guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 228:150–154. https://doi.org/10.1006/abio.2000.5172.
23. Greenberg EP, Harwood CS. 2016. A LuxR homolog in a cottonwood tree symbiont regulates cellulose synthesis. Proc Natl Acad Sci U S A 113:13966–13971. https://doi.org/10.1073/pnas.1613966113.
24. Parsek MR, Nierman WC, Greenberg EP. 2009. Quorum-sensing control of biofilm structures and phenotypic states. Annu Rev Microbiol 63:4579–4588. https://doi.org/10.1146/annurev-micro-092412-155635.
25. Pfaffl MW. 2001. A new method for the accurate quantification of mRNA levels with real-time PCR. Nucleic Acids Res 29:402–408. https://doi.org/10.1093/nar/29.2.402.
26. Parsek MR, Nierman WC, Greenberg EP. 2009. Quorum-sensing control of biofilm structures and phenotypic states. Annu Rev Microbiol 63:4579–4588. https://doi.org/10.1146/annurev-micro-092412-155635.
racy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340.
45. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874. https://doi.org/10.1093/molbev/msw054.
46. Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454.
47. Felsenstein J. 1985. Confidence-limits on phylogenies—an approach using the bootstrap. Evolution 39:783–791. https://doi.org/10.1111/j.1558-5646.1985.tb00420.x.
48. Nei M, Kumar S. 2000. Molecular evolution and phylogenetics. Oxford University Press, New York, NY.
49. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539. https://doi.org/10.1038/msb.2011.75.