Discovery of phosphonic acid natural products by mining the genomes of 10,000 actinomycetes

Kou-San Ju[a], Jiangtao Gao[a], James R. Doroghazi[a], Kwo-Kwang A. Wang[a], Christopher J. Thibodeaux[a], Steven Li[a], Emily Metzger[a], John Fudala[a], Joleen Su[a], Jun Kai Zhang[a,b], Jaeheon Lee[a], Joel P. Cioni[a,b], Bradley S. Evans[a], Ryuichi Hirota[a,c], David P. Labeda[d], Wilfred A. van der Donk[a,e,1], and William W. Metcalf[a,b,1]

[a]Carl R. Woese Institute for Genomic Biology, University of Illinois, Urbana, IL 61801; [b]Department of Microbiology, University of Illinois, Urbana, IL 61801; [c]Department of Molecular Biotechnology, Graduate School of Advanced Sciences of Matter, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8530, Japan; [d]Bacterial Foodborne Pathogens and Mycology Research, US Department of Agriculture, Agricultural Research Service, National Center for Agricultural Utilization Research, Peoria, IL 61604; and [e]Department of Chemistry and Howard Hughes Medical Institute, University of Illinois, Urbana, IL 61081

Edited by Jerrold Meinwald, Cornell University, Ithaca, NY, and approved July 31, 2015 (received for review January 14, 2015)

Although natural products have been a particularly rich source of human medicines, activity-based screening results in a very high rate of rediscovery of known molecules. Based on the large number of natural product biosynthetic genes in microbial genomes, many have proposed "genome mining" as an alternative approach for discovery efforts; however, this idea has yet to be performed experimentally on a large scale. Here, we demonstrate the feasibility of large-scale, high-throughput genome mining by screening a collection of over 10,000 actinomycetes for the genetic potential to make phosphonic acids, a class of natural products with diverse and useful bioactivities. Genome sequencing identified a diverse collection of phosphonate biosynthetic gene clusters within 278 strains. These clusters were classified into 64 distinct groups, of which 55 are likely to direct the synthesis of unknown compounds. Characterization of strains within five of these groups resulted in the discovery of a new archetypical pathway for phosphonate biosynthesis, the first (to our knowledge) dedicated pathway for H-phosphinates, and 11 previously undescribed phosphonic acid natural products. Among these compounds are argolaphos, a broad-spectrum antibacterial phosphonopeptide composed of aminomethylphosphonate in peptide linkage to a rare amino acid N^5-hydroxyarginine; valinophos, an *N*-acetyl L-Val ester of 2,3-dihydroxypropylphosphonate; and phosphonocystoximate, an unusual thiohydroximate-containing molecule representing a new chemotype of sulfur-containing phosphonate natural products. Analysis of the genome sequences from the remaining strains suggests that the majority of the phosphonate biosynthetic repertoire of Actinobacteria has been captured at the gene level. This dereplicated strain collection now provides a reservoir of numerous, as yet undiscovered, phosphonate natural products.

natural products | genome mining | phosphonic acid | antibiotic

The use of genomic data to enable discovery of novel biological processes, often referred to as genome mining, has the potential to revolutionize numerous areas of modern biology, including the field of natural product discovery. These biologically produced small molecules have been the source of, or inspiration for, nearly two-thirds of all human medicines (1), yet research in this area has dwindled in recent years due to, among other reasons, high costs and increasing rates of rediscovery (2–4). Although a number of solutions to this dilemma have been proposed, many within the natural product biosynthesis field have suggested that the solution lies in genomics (5, 6). Thus, by focusing research efforts on strains that encode genes for the biosynthesis of uncharacterized natural products, one can dereplicate, streamline, and accelerate the discovery process. Indeed, genome mining has led to the discovery of several novel natural products, but most efforts have been limited to individual strains or small collections (5–9). Mining of metagenomic libraries has also produced a number of new compounds (10–12), but it has not been generally adopted. If we hope to revitalize the use of natural products in the pharmaceutical industry, genome mining must be shown to be a high-throughput discovery process complementary or superior to existing methods (13, 14). Here, we demonstrate the feasibility of this approach in a campaign to identify the full genetic repertoire of phosphonic acid natural products encoded by a collection of over 10,000 actinomycetes.

Phosphonic acid natural products possess several traits that make them ideal candidates for large-scale genome mining. First and foremost, phosphonates have great pharmaceutical potential, with a commercialization rate of 15% [three of 20 isolated compounds (15)]. This percentage is much higher than the 0.1% average estimated for natural products as a whole (16). The potent bioactivity of phosphonates derives from their chemical mimicry of essential metabolites, including phosphate esters and anhydrides, as well as carboxylate reaction intermediates (15). Given the ubiquitous presence of these chemical moieties in biology, phosphonates are unrivaled in the range of targets they can potentially affect. Consistent with this idea, phosphonate natural products with herbicidal, insecticidal, antibacterial, antiparasitic, antiviral, and antihypertensive activities are known (15, 17). Notable examples include fosfomycin (Monurol), clinically prescribed for

Significance

The discovery of natural products, an important source of human medicines, is critical for the development of new therapeutics against health threats, including cancer and multidrug-resistant pathogens. Yet, in recent years, industrial development of pharmaceuticals from natural products has been stymied due to a variety of reasons, including the repeated discovery of previously known compounds. Here, we demonstrate large-scale genomics as one potential solution to this problem by mining a collection of 10,000 actinomycetes for novel phosphonic acids, an important class of natural products with antimicrobial, antiviral, antimalarial, and herbicidal activities. The framework described here provides a foundation for rapid, large-scale discovery of other classes of natural products and their use as lead compounds in the pharmaceutical industry.

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

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in the BioProject database, ncbi.nlm.nih.gov/bioproject (accession no. PRJNA238534).

[1]To whom correspondence may be addressed. Email: vddonk@illinois.edu or metcalf@illinois.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1500873112/-/DCSupplemental.

acute cystitis; FR-900098 and fosmidomycin, antimicrobials undergoing clinical trials for malaria; and phosphinothricin, the active component in several commercial herbicides (Liberty, Basta, and Rely). Second, the methodology needed for gene-based discovery of phosphonate biosynthetic loci has been rigorously established (17–23). This method relies on the fact that all but two characterized phosphonate biosynthetic pathways begin with the enzyme phosphoenolpyruvate (PEP) mutase (encoded by *pepM*) (14). Thus, amplification of an internal fragment of *pepM* with degenerate PCR primers is sufficient to identify strains or plasmid clones that encode phosphonate biosynthetic pathways. Third, gene-based surveys have proven that phosphonate biosynthesis is relatively common in microorganisms, with the largest reservoir of unexplored pathways observed within Actinobacteria (23). Finally, the carbon-phosphorus bond endows phosphonates with unique chemical properties that enable their chemispecific detection from complex mixtures of metabolites using MS and NMR spectroscopy, even when the structure of the molecule in question is unknown (24, 25). These methods facilitate the isolation and characterization of phosphonate natural products by enabling their direct detection during production screening and purification.

Results and Discussion

Identification and Genome Sequencing of Actinobacteria That Contain Phosphonate Biosynthetic Pathways. A collection of over 10,000 actinomycetes, composed of *ca.* 2,500 isolated in our laboratory (23) and nearly 7,500 from the Agricultural Research Service (ARS) Culture Collection (NRRL) at the National Center for Agricultural Utilization Research, were screened for the presence of *pepM*. The NRRL collection houses the largest public repository of Actinobacteria, encompassing species from 117 different genera, important industrial strains, and the retired strain collections of several prominent streptomycete researchers (e.g., S. A. Waksman, D. Gottlieb, E. B. Shirling). To enable rapid genetic screening, genomic DNA was isolated from all strains and arrayed into a 96-well format. Blind PCR screening of this library identified 403 putative phosphonate producers. Subsequent draft genome sequencing of all candidate strains confirmed the presence of *pepM* in 278 strains, including 244 from the NRRL collection; 30 from our laboratory collection; and four additional strains specifically added to the study because of their known ability to produce phosphonothrixin (26), FR-900098 (27), and fosmidomycin (28) (*SI Appendix*, Table S1). The majority (87%) of the *pepM*$^+$ strains are *Streptomyces*, with the remainder scattered among 24 genera, a distribution that mirrors the composition of the strain collection. A single copy of *pepM* was found in all but five strains, each of which encodes two distinct *pepM* homologs. Significantly, all known phosphonate producers deposited in the NRRL collection, including producers of fosfomycin, dehydrophos, and phosalacine, were rediscovered in our blind screening, validating the sensitivity and thoroughness of our *pepM*-based approach.

Dereplication of Phosphonate Biosynthetic Pathways and Production Screening. Although recent years have seen great improvements in subtractive MS- and NMR-based screening methods for new natural products (29, 30), in our hands, ^{31}P NMR spectroscopy has proven more reliable than MS-based methods for detection of phosphonates of unknown structure. However, NMR screening of 10,000 strains is impractical. Therefore, the genomic data were used to identify, classify, and dereplicate phosphonate biosynthetic pathways within the candidate strains before characterization of the natural products they produce. For this process, we used two established methods for linking genetic data to phosphonate natural product diversity (23). Initially, we examined the relatedness of the PepM proteins encoded by the 278 strains. This analysis revealed 168 unique sequences that distributed into 78 groups at 80% sequence identity (Fig. 1*A*). Because natural product biosynthetic genes are almost always clustered together in a single chromosomal locus, we also examined the similarity of the genes surrounding *pepM*. This parallel approach uncovered 64 discrete gene neighborhoods (Fig. 1*B*). Known phosphonate producers map to nine of these groups, most of which also include additional strains. Thus, our genome mining effort revealed 55 new potential producers of hydroxynitrilaphos (cyanohydroxymethylphosphonate), 8 of plumbemycin, 7 of fosfazinomycin, 2 of dehydrophos, and 1 of fosfomycin. We also identified the putative biosynthetic gene cluster for phosphonothrixin in the genome sequence of the only known producer, *Saccharothrix* ST-888 (26). Our observation of phosphonothrixin in the spent media of four of the 13 strains that carry this gene cluster (*SI Appendix*, Fig. S2 and Table S2) provides strong evidence that these genes direct the synthesis of this herbicidal compound (31). After eliminating strains linked to known products, we were left with a dereplicated collection of 192 strains grouped into 69 PepM clades and 55 *pepM* gene cluster families. By both measures of diversity, greater than 85% of the pathways reported here direct synthesis of new phosphonate natural products.

Production screening of all 244 *pepM*$^+$ strains from the NRRL collection by ^{31}P NMR spectroscopy suggested that at least 45 produced phosphonates. Three of these strains, *Streptomyces monomycini* NRRL B-24309, *Streptomyces* sp. NRRL S-474, and *Streptomyces* sp. NRRL S-481, were of particular interest because they elicited positive responses in a phosphonate-specific bioassay (18) and because the genomic information illustrated that their biosynthetic gene clusters were novel.

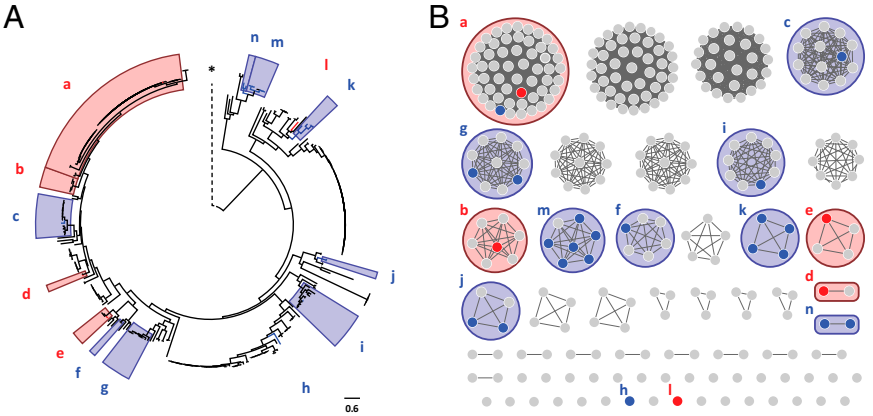

Fig. 1. Diversity of phosphonate biosynthesis in Actinobacteria. (*A*) Maximum-likelihood phylogeny of actinomycete PepMs using methylisocitrate lyase as the outgroup (*). Complete labeling of strain names is provided in *SI Appendix*, Fig. S1. (*B*) Network diagram of phosphonate gene clusters. In both diagrams, highlighting indicates clades and strains containing pathways for previously described phosphonates (blue) and those phosphonates newly isolated in this study (red). Groups encoding known phosphonate natural products are as follows: a, hydroxynitrilaphos; b, phosphonocystoximate; c, phosphonothrixin; d, valinophos; e, argolaphos; f, dehydrophos; g, fosfazinomycin; h, fosmidomycin; i, plumbemycins and phosacetamycin; j, FR-900098; k, phosphinothricin peptides; l, DMPT; m, fosfomycin; n, 2-hydroxyethylphosphonate polyglycans. Phosphonate gene clusters are available (www.igb.illinois.edu/labs/metcalf/gcf/Phosphonates.html).

Argolaphos A and B, Broad-Spectrum Antibacterial Phosphonopeptides. Two phosphonopeptides, which we have designated argolaphos A and B, were purified from spent media of *S. monomycini* NRRL B-24309. Argolaphos B is composed of aminomethylphosphonate (AmPn) in an amide linkage with the rare amino acid N^5-hydroxyarginine (N5-OH-Arg), whereas argolaphos A has an additional peptide linkage to Val (Fig. 2*A*). Although AmPn is not a known natural product, it has been reported as an off-pathway metabolite in mutants blocked for phosphinothricin production and as a product of glyphosate degradation (32, 33). N5-OH-Arg has previously been isolated as a free amino acid from strains of *Bacillus cereus* (34, 35) and the fungus *Nannüzia gypsea* (*Arthroderma gypseum*) (36). The *pepM* gene cluster of *S. monomycini* NRRL B-24309 suggests (Fig. 2*B*) that the AmPn core is synthesized from PEP by homologs of PhpA–PhpE from phosphinothricin biosynthesis (32). Putative ATP-grasp ligase and GCN5-related *N*-acetyltransferases (GNAT; FemXAB nonribosomal peptidyltransferase) encoded by neighboring genes may catalyze peptide bond formation. Similar enzymes are used in the biosynthesis of the phosphonopeptides rhizocticin (21) (ATP-grasp ligases) and dehydrophos (37) (GNAT), and a combination of both has been suggested to account for the peptide linkages in fosfazinomycin (38).

Argolaphos has broad-spectrum antibacterial activity, with the highest efficacy against *Salmonella typhimurium*, *Escherichia coli*, and *Staphylococcus aureus*, and it is weakly inhibitory against *Mycobacterium smegmatis* (*SI Appendix*, Fig. S5). Because AmPn and N5-OH-Arg are inhibitors with different modes of action (34, 39), the combination of both in argolaphos may serve to enhance its bioactivity. It is clear that both moieties contribute to toxicity, because argolaphos is a more potent antibiotic than free AmPn against *E. coli* strains expressing a broad-specificity phosphonate transport system (Fig. 2*C*).

Phosphonocystoximates, a New Class of Sulfur-Containing Phosphonate Natural Products. Two other bioactive strains, *Streptomyces* sp. NRRL S-474 and NRRL S-481, belong to a group of seven strains with a *pepM* gene cluster similar, but not identical, to the one found in *Streptomyces regensis* NRRL WC-3744, the known producer of hydroxynitrilaphos (40) (Fig. 3*A*). In preliminary experiments, both strains produced several compounds with ^{31}P NMR signals in the range consistent with phosphonic acids. Only one of these compounds was made in sufficient amounts to allow purification and structural characterization. This molecule, purified from spent media of *Streptomyces* sp. NRRL S-481 by ^{31}P NMR-guided fractionation and named phosphonocystoximate, represents a new class of sulfur-containing phosphonate natural products (Fig. 3*B*). The *S*-alkyl thiohydroximate and *N*-acetyl-Cys moieties of phosphonocystoximate are chemically similar to biosynthetic intermediates of glucosinolates, which are plant natural products with potential antioxidant and anticancer properties, most commonly known as the molecules responsible for the pungent taste of broccoli, Brussels sprouts, and wasabi (41). Unfortunately, neither of the two strains produced sufficient levels of phosphonocystoximate to allow bioactivity testing of the purified molecule. Thus, the identity of the bioactive phosphonate produced by these strains remains to be determined.

Phosphonocystoximate and Hydroxynitrilaphos Share Common Biosynthetic Origins. Comparisons between the phosphonocystoximate and hydroxynitrilaphos biosynthetic gene clusters show conservation of all but three genes (Fig. 3*A*). Thus, some of the differences between the gene clusters can be attributed to differences in the enzymatic steps involved in the synthesis of the likely early intermediate 2-aminoethylphosphonate (2AEPn). The phosphonocystoximate gene cluster lacks homologs of the genes encoding phosphonopyruvate (PnPy) decarboxylase and homocitrate synthase, which are required to drive the thermodynamically unfavorable PepM reaction in other phosphonate biosynthetic pathways (17). This finding suggests a novel driving reaction is used in phosphonocystoximate biosynthesis. A second difference lies in the presence of an additional α-ketoglutarate–dependent dioxygenase within the hydroxynitrilaphos gene cluster, suggesting strains containing the two groups of gene clusters might produce similar molecules that differ by hydroxylation. To test this hypothesis, we examined spent media from several strains in both groups using high resolution MS (*SI Appendix*, Fig. S3 and Table S3). Indeed, strains encoding the related gene cluster families produced a suite of molecules that differ solely by hydroxylation at the α-carbon, relative to the phosphonate moiety (Fig. 3*B*). Among these molecules are several notable natural products, including 1-hydroxy-2-aminoethylphosphonate (1H2AEPn) and hydroxyphosphonocystoximate, made by strains encoding the dioxygenase, and nitrilaphos (cyanomethylphosphonate), made by strains lacking this gene. Additionally, 1H2AEPn was found to be an effective antibiotic against *E. coli* upon induction of phosphonate uptake machinery (*SI Appendix*, Fig. S5).

Based on these data, we present a model where 2AEPn serves as the common biosynthetic intermediate for both pathways (Fig. 3*B*). The putative FAD-NAD(P)H–dependent *N*-hydroxylase encoded within both gene clusters may oxidize the amine to the oxime. Mycothiol-dependent transferase and peptidase genes suggest mycothiol as the probable source of *N*-acetylcysteine in phosphonocystoximate formation, similar to the use of glutathione as the source of Cys in glucosinolates from plants (41) (Fig. 3*C*). Accordingly, inositolglucosamine conjugants of phosphonocystoximate and hydroxyphosphonocystoximate were detected from extracts (Fig. 3*B* and *SI Appendix*, Fig. S3 and Table S3). Subsequent hydrolysis of the amide bond of the conjugated mycothiol by the encoded peptidases would release the observed phosphonocystoximates. Consistent with the metabolism of glucosinolates in plants, these compounds may be biosynthetic intermediates en route to the formation of a yet unidentified phosphonosinolate. Whether the nitrilaphos and hydroxynitrilaphos produced by these

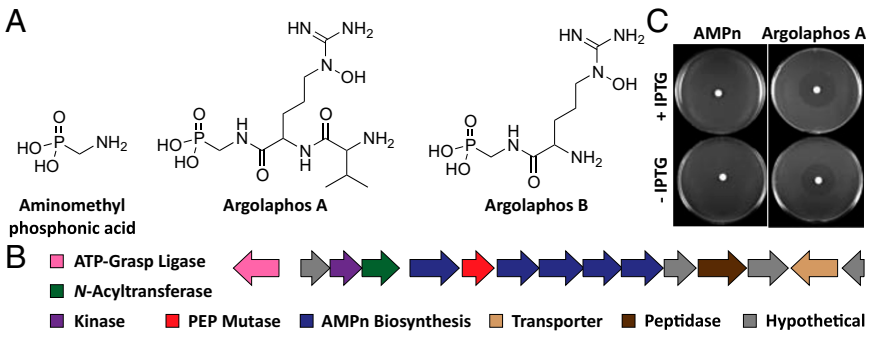

Fig. 2. *S. monomycini* B-24309. (*A*) Isolated compounds. (*B*) Putative argolaphos gene cluster. Core genes encoding for enzymes predicted for AmPn biosynthesis are shown in dark blue (*pepM* in red). (*C*) Growth inhibition of *E. coli* WM6242 by 40 nmol of AmPn and argolaphos in a disk diffusion assay, with (*Top*) and without (*Bottom*) isopropyl-β-D-thiogalactopyranoside (IPTG) induction of phosphonate-specific transporters.

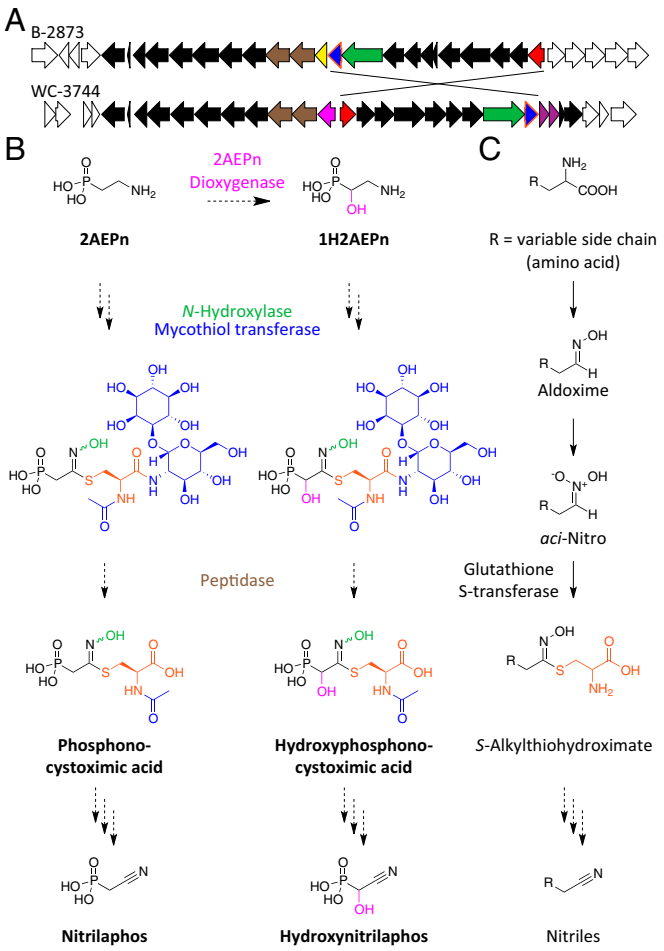

Fig. 3. Comparisons between phosphonocystoximic acid and hydroxynitrilaphos biosynthesis. (*A*) Putative gene clusters. The majority of genes are shared between the two pathways, with some notable differences. Genes for α-ketoglutarate–dependent dioxygenase (magenta) and PnPy decarboxylase (purple) are absent from the phosphonocystoximate gene cluster, which instead contains a gene for pyridoxamine-phosphate oxidase (yellow). Half of the shared genes encoding the two pathways are inverted relative to each other. Other labeled genes putatively encode for PepM (red), mycothiol transferase (blue, outlined in red), *N*-hydroxylase (green), and peptidases (brown). (*B*) Proposed biosynthetic pathways for phosphonocystoximic acids and cyanomethylphosphonic acids. (*C*) Proposed biosynthetic pathway for glucosinolates, adapted from Halkier and Gershenzon (41). Several intermediates in glucosinolate biosynthesis and metabolism share similarities to the compounds from these pathways. Glutathione is the source of Cys (orange) in glucosinolate biosynthesis, whereas mycothiol (orange and blue) may contribute *N*-acetylcysteine in phosphonocystoximic acid biosynthesis.

strains are biosynthetic intermediates or the products of degradation remains to be determined.

Valinophos, a Phosphonopeptide Produced by a New Archetypical Pathway. *Streptomyces durhamensis* NRRL B-3309 was the next strain selected for further analysis because of the unusual content of its *pepM* gene neighborhood. Two novel phosphonate natural products, 2-hydroxy-3-phosphonopropanoic acid and (*R*)-2,3-dihydroxypropylphosphonic acid (DHPPA), were purified from spent media (Fig. 4*A*). These compounds are likely early intermediates in the formation of a fourth new phosphonate that we name valinophos, an *N*-acetyl L-Val ester of DHPPA that was also produced by *S. durhamensis*. Only DHPPA was purified in sufficient quantity to assess bioactivity, which slightly inhibited growth of *M. smegmatis* (*SI Appendix*, Fig. S5).

We propose a biosynthetic pathway for valinophos that is supported by the content of the *pepM* gene cluster (Fig. 4*B*) and the predicted biochemistry of the encoded proteins. Like the phosphonocystoximate cluster, a known example of an enzyme that could be coupled with PEP mutase (VlpA) to drive the thermodynamically unfavorable rearrangement of PEP to PnPy is absent. Instead, VlpB, a putative short-chain dehydrogenase, may reduce PnPy to 2-hydroxy-3-phosphonopropanoic acid (phosphonolactate) in a reaction analogous to the highly favorable reduction of pyruvate to lactate by NADH-dependent lactate dehydrogenases (equilibrium constant, $K_{eq} = 1 \times 10^5$) (42). Analogous to the proposed formation of 1-oxo-2-phosphorylethylphosphonic acid by the phosphoglycerate kinase DhpB in dehydrophos biosynthesis (37), phosphorylation of phosphonolactate by VlpD would yield 2-hydroxy-3-oxo-3-phosphoxypropylphosphonic acid. Reduction by VlpC, homologous to *N*-acetyl γ-glutamyl phosphate (AGP) reductase, results in 2-hydroxy-3-oxopropylphosphonic acid that could be converted to DHPPA via another reduction by VlpE (alcohol dehydrogenase). Although ATP-grasp ligases typically adjoin substrates by amide linkages, these enzymes have also been shown to form ester bonds (43). Here, ATP-grasp ligases VlpF and VlpG are proposed to install L-Val onto DHPPA sequentially and then to acetylate the terminal amine to yield valinophos. Interestingly, *vlpBCDE* appears as a conserved cassette within the *pepM* gene clusters for phosphonothrixin. This observation suggests the early steps for valinophos are the archetype of a new branch of phosphonate biosynthetic pathways, from which phosphonothrixin also originates.

Dedicated Pathway for H-Phosphinate Biosynthesis. *Nonomureae* sp. NRRL B-24552 was the final strain chosen for study due to the similarity of its *pepM* gene cluster to the gene cluster involved in the synthesis of phosphinothricin tripeptide (PTT), the only known phosphinate natural product. However, unlike the well-characterized PTT biosynthetic gene clusters, *Nonomureae* sp. NRRL B-24552 lacks the genes needed for tripeptide biosynthesis (*phsABC*; nonribosomal peptide synthetases) and formation of the C-P-C moiety of phosphinothricin (*phpK*; P-methylase) (19), suggesting that this organism may instead produce an H-phosphinate (H-P-C bond motif) (Fig. 5*A*). Consistent with this idea, we purified 2-phosphinomethylmalic acid and desmethylphosphinothricin (DMPT) from extracts of this strain (Fig. 5*B*). DMPT

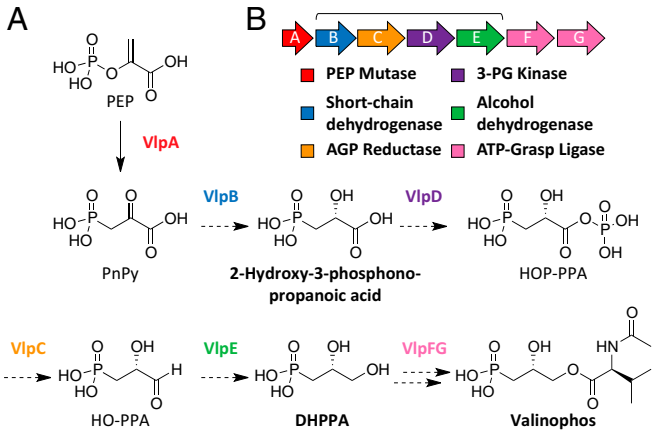

Fig. 4. *S. durhamensis* B-3309. (*A*) Isolated compounds and proposed pathway for valinophos biosynthesis. (*B*) Putative Vlp gene cluster, with the genetic cassette conserved in phosphonothrixin biosynthesis indicated by brackets. HO-PPA, 2-hydroxy-3-oxopropylphosphonic acid; HOP-PPA, 2-hydroxy-3-oxo-3-phosphoxypropylphosphonic acid; 3-PG, 3-phosphoglycerate.

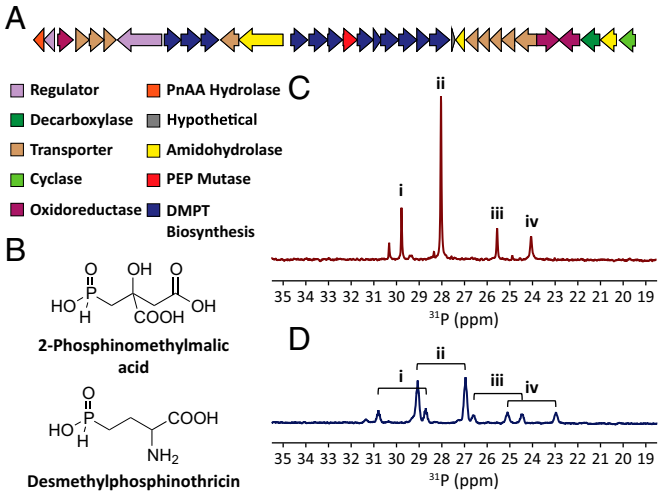

Fig. 5. *Nonomuraea* B-24552. (*A*) Putative H-phosphinate biosynthetic genes. Core genes encoding for predicted enzymes in DMPT biosynthesis are shown in dark blue (*pepM* is shown in red). (*B*) Isolated compounds. (*C*) ^{1}H-decoupled ^{31}P NMR analysis of spent medium revealing several compounds with chemical shifts characteristic of H-phosphinates. (*D*) ^{1}H-coupled ^{31}P NMR spectra of concentrated extract. The four largest signals (i–iv) exhibit distinctive splitting patterns and coupling constants consistent with H-phosphinates (i, $J_{H\text{-}P}$ = 505 Hz; ii, $J_{H\text{-}P}$ = 512 Hz; iii, $J_{H\text{-}P}$ = 518 Hz; iv, $J_{H\text{-}P}$ = 519 Hz).

is not a known natural product; it has been observed only as a byproduct in *Streptomyces hygroscopicus* strains with mutations in the PTT pathway (44). Based on the content of the *Nonomureae* sp. NRRL B-24552 *pepM* gene cluster and the results of ^{31}P NMR analyses of spent media (Fig. 5*C*), we believe it is likely DMPT is further modified to produce additional H-phosphinates in this organism.

Summary. Our successful campaign for phosphonates validates the utility of large-scale genome mining for discovery of new natural products, fulfilling a promise of genomics-enabled drug discovery made 20 y earlier. Including the eight additional molecules we recently reported elsewhere (24, 38, 40, 45) (*SI Appendix*, Fig. S4), this program has yielded 19 new phosphonate natural products to date. In addition to increasing the number of actinomycete phosphonate gene cluster groups by over eightfold, our efforts doubled the total diversity of phosphonate pathways identified to date. Moreover, rarefaction and extrapolation analyses suggest that the majority of the remaining phosphonate biosynthetic repertoire of Actinobacteria can be found within the dereplicated strain collection (Fig. 6). Accordingly, using our previously published approach (46), we predict that Actinobacteria have a total capacity for *ca.* 125 distinct pathways for phosphonic acid natural products. The genome mining of 10,000 strains reported here has already identified 78 discrete biosynthetic gene clusters, with multiple producing organisms available for most groups. Lastly, we note that the actinobacterial DNA library and hundreds of genome sequences reported in this study will inform and enable rapid, large-scale genome mining for other natural product classes. We suggest that the future of natural products research and antibiotic development is bright, with many of the best discoveries yet to come.

Materials and Methods

The full description of experimental techniques is provided in *SI Appendix*.

Construction of the Actinomycete DNA Library. Strains from the US Department of Agriculture ARS Actinobacteria collection were revived from lyophilized vials or slants by inoculation into 150-mm × 18-mm test tubes containing 5 mL of 172 medium. Cultures were grown at 30 °C on an angled New Brunswick Scientific roller drum (75 rpm) for 6 d and then arrayed by aliquoting (2 mL) into 96-well deep blocks. Supernatants were removed after centrifugation, and the remaining cell pellets were frozen at −80 °C for 24 h. Cell pellets were then allowed to thaw completely at room temperature, and they were immediately refrozen and stored at −20 °C.

DNA was isolated using a 96-well UltraClean-HTP Microbial DNA Kit (MoBio) according to the manufacturer's instructions but with the following modifications. Cell pellets were thawed, resuspended with 300 μL of MicroBead solution containing lysozyme (1 mg/mL) and DNase-free RNase A (100 μg/mL), and incubated at 30 °C for 1 h with gentle rocking. Treated cells were then transferred into the 96-well bead plate. After addition of solution MD1, 96-well blocks were firmly sealed with the provided silicon covers, inverted several times to mix, pulsed in a tabletop centrifuge to draw all liquid to the bottom, and incubated in a 70 °C water bath for 20 min. After cooling to room temperature, blocks were attached onto a Restech bead mill and cells were lysed by running at setting 20 for 5 min. Blocks were removed, reoriented so that the side closest to the machine body was now furthest away, and shaken again at setting 20 for an additional 5 min. Before DNA elution, spin plates were centrifuged twice to remove residual ethanol and air-dried for 20 min at room temperature. DNA was eluted from spin plates by adding 120 μL of preheated sterile water (70 °C) to wells and waiting 5 min before centrifugation into microplates. DNA stocks were kept at −80 °C for long-term storage. The completed library contains genomic DNA isolated from 7,488 actinomycetes.

Genetic Screen for *pepM*$^{+}$ Strains. The genomic DNA library was screened by PCR using four pairs of degenerate primers designed to amplify a 406-bp conserved fragment within bacterial *pepM* genes (forward: pepMF, CGCCGGCGTCTGCNTNGARGAYAA; reverse: pepMR, GGCGCGCATCATGTGRTTNGCVYA; pepMX, GCGCGCATCATGTGGTTNGCCCADAT; pepMW, GCGCGCATCATGTGGTTNGCRTADAT; pepMZ, GCGCGCATCATGTTGTTNCCCCADAT). Promega Go Taq was used in 25-μL reactions containing each primer (5 μM), 0.5 μL of DMSO, and 0.5 μL of template DNA. Thermocyclers were operated under the following program: (*i*) initial denaturation for 3 min at 98 °C; (*ii*) 30 sec at 98 °C; (*iii*) 30 sec at 58 °C; (*iv*) 30 sec at 72 °C; (*v*) 3 min at 72 °C; and (*vi*) hold at 12 °C. Steps *ii–iv* were repeated for a total of 35 cycles. PCR amplification of the *pepM* fragment using the genomic DNA of *Streptomyces viridochromogenes* DSM 40736 (PTT producer), *Streptomyces luridus* NRRL 15101, and *Streptomyces rubellomurinus* American Type Culture Collection (ATCC) 31215 was routinely performed to verify the integrity of reaction master mixes. PCR products were separated by electrophoresis through 1% agarose gels with positive controls on each side. The

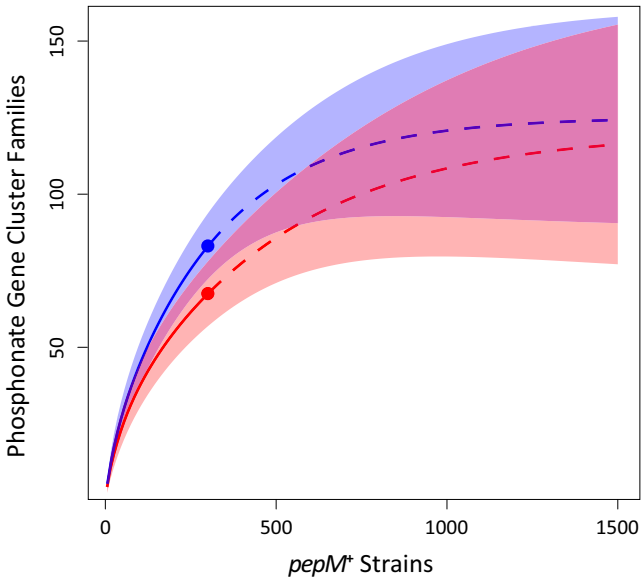

Fig. 6. Extrapolation of unique phosphonate natural product gene cluster families. Calculations were based on amino acid (blue) or gene cluster family (red) data. The current extent of sampling is shown with filled circles. Extrapolation is based on discovery of *pepM*$^{+}$ strains, equivalent to sampling of an additional 40,000 actinomycetes in the same manner as the 10,000 screened for this work. Shaded areas show the 95% confidence intervals for the analyses.

presence of an ~400-bp band matching the size of the positive controls indicated a positive result.

Genome Sequencing and Assembly. *Saccharothrix* ST-888, *Streptomyces lavendulae* Fujisawa 8006, *S. rubellomurinus* subsp. *indigoferus* ATCC 31304, *S. rubellomurinus* ATCC 31215, and all PCR-positive strains from the NRRL collection were revived from freezer stocks on 172 agar plates and regrown as described above. Genomic DNA was isolated using an UltraClean-HTP Microbial DNA kit as described above, except mycelia aggregates were first homogenized in microfuge tubes using sterile pestles. Paired-end, multiplexed sequencing libraries were constructed using Nextera DNA Sample Preparation Kits (Illumina). Samples were sequenced in multiple batches at either the University of Wisconsin Biotechnology Center (Illumina Genome Analyzer IIX) or the University of Illinois at Urbana–Champaign Roy Carver Biotechnology Center (Illumina Hi-Seq 2000 or MiSeq platform). Sequencing reads produced from the Genome Analyzer IIX or the Hi-Seq 2000 with version 2 chemistry were assembled as previously described (47). ORF prediction was performed with Prodigal version 2.50 (48). The genomes sequenced using the Illumina Hi-Seq 2000 with version 3 chemistry were assembled using IDBA-UD version 1.0.9 (49), and ORF prediction was performed using Prodigal version 2.60. All sequenced strains used in this study are listed in *SI Appendix*, Table S1. DNA sequences have been deposited in GenBank under BioProject PRJNA238534.

ACKNOWLEDGMENTS. We thank S. A. Borisova, H. A. Cooke, A. Erb, S. Y. Kim, S. C. Peck, G. Rao, J. E. Vélasquez, X. Yu, X. Guan, and L. Zhu for their assistance with screening and NMR experiments. This work was funded by NIH Grant P01 GM077596 (to W.A.v.d.D. and W.W.M.). NMR instrumentation was purchased with support from NIH Grant S10 RR028833. K.-S.J. was supported by NIH Ruth L. Kirchstein National Service Award F32GM100658, and J.R.D. was supported by an Institute for Genomic Biology Postdoctoral Fellowship. D.P.L. and the ARS Culture Collection Current Research Information System project are funded through ARS National Program 301.

1. Cragg GM, Newman DJ (2013) Natural products: A continuing source of novel drug leads. *Biochim Biophys Acta* 1830(6):3670–3695.
2. Koehn FE, Carter GT (2005) The evolving role of natural products in drug discovery. *Nat Rev Drug Discov* 4(3):206–220.
3. Brötz-Oesterhelt H, Sass P (2010) Postgenomic strategies in antibacterial drug discovery. *Future Microbiol* 5(10):1553–1579.
4. Baltz RH (2006) Marcel Faber Roundtable: Is our antibiotic pipeline unproductive because of starvation, constipation or lack of inspiration? *J Ind Microbiol Biotechnol* 33(7):507–513.
5. Bachmann BO, Van Lanen SG, Baltz RH (2014) Microbial genome mining for accelerated natural products discovery: Is a renaissance in the making? *J Ind Microbiol Biotechnol* 41(2):175–184.
6. Challis GL (2008) Genome mining for novel natural product discovery. *J Med Chem* 51(9):2618–2628.
7. Bode HB, Müller R (2005) The impact of bacterial genomics on natural product research. *Angew Chem Int Ed Engl* 44(42):6828–6846.
8. Zettler J, et al. (2014) New aminocoumarins from the rare actinomycete *Catenulispora acidiphila* DSM 44928: Identification, structure elucidation, and heterologous production. *ChemBioChem* 15(4):612–621.
9. Pidot S, Ishida K, Cyrulies M, Hertweck C (2014) Discovery of clostrubin, an exceptional polyphenolic polyketide antibiotic from a strictly anaerobic bacterium. *Angew Chem Int Ed Engl* 53(30):7856–7859.
10. 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.
11. Owen JG, et al. (2013) Mapping gene clusters within arrayed metagenomic libraries to expand the structural diversity of biomedically relevant natural products. *Proc Natl Acad Sci USA* 110(29):11797–11802.
12. Chang FY, Brady SF (2013) Discovery of indolotryptoline antiproliferative agents by homology-guided metagenomic screening. *Proc Natl Acad Sci USA* 110(7):2478–2483.
13. Abdelmohsen UR, et al. (2015) Elicitation of secondary metabolism in actinomycetes. *Biotechnol Adv*, 10.1016/j.biotechadv.2015.06.003.
14. Luo Y, Cobb RE, Zhao H (2014) Recent advances in natural product discovery. *Curr Opin Biotechnol* 30:230–237.
15. Metcalf WW, van der Donk WA (2009) Biosynthesis of phosphonic and phosphinic acid natural products. *Annu Rev Biochem* 78:65–94.
16. Bérdy J (2012) Thoughts and facts about antibiotics: Where we are now and where we are heading. *J Antibiot (Tokyo)* 65(8):385–395.
17. Ju KS, Doroghazi JR, Metcalf WW (2014) Genomics-enabled discovery of phosphonate natural products and their biosynthetic pathways. *J Ind Microbiol Biotechnol* 41(2):345–356.
18. Eliot AC, et al. (2008) Cloning, expression, and biochemical characterization of *Streptomyces rubellomurinus* genes required for biosynthesis of antimalarial compound FR900098. *Chem Biol* 15(8):765–770.
19. Blodgett JA, Zhang JK, Metcalf WW (2005) Molecular cloning, sequence analysis, and heterologous expression of the phosphinothricin tripeptide biosynthetic gene cluster from *Streptomyces viridochromogenes* DSM 40736. *Antimicrob Agents Chemother* 49(1):230–240.
20. Woodyer RD, et al. (2006) Heterologous production of fosfomycin and identification of the minimal biosynthetic gene cluster. *Chem Biol* 13(11):1171–1182.
21. Borisova SA, Circello BT, Zhang JK, van der Donk WA, Metcalf WW (2010) Biosynthesis of rhizocticins, antifungal phosphonate oligopeptides produced by *Bacillus subtilis* ATCC6633. *Chem Biol* 17(1):28–37.
22. Circello BT, Miller CG, Lee JH, van der Donk WA, Metcalf WW (2011) The antibiotic dehydrophos is converted to a toxic pyruvate analog by peptide bond cleavage in *Salmonella enterica*. *Antimicrob Agents Chemother* 55(7):3357–3362.
23. Yu X, et al. (2013) Diversity and abundance of phosphonate biosynthetic genes in nature. *Proc Natl Acad Sci USA* 110(51):20759–20764.
24. Evans BS, et al. (2013) Discovery of the antibiotic phosacetamycin via a new mass spectrometry-based method for phosphonic acid detection. *ACS Chem Biol* 8(5):908–913.
25. Nielsen ML, Pustinger JV, Strobel J (1964) Phosphorus-31 nuclear magnetic resonance chemical shifts of phosphorus compounds. *J Chem Eng Data* 9:167–170.
26. Takahashi E, Kimura T, Nakamura K, Arahira M, Iida M (1995) Phosphonothrixin, a novel herbicidal antibiotic produced by *Saccharothrix* sp. ST-888. I. Taxonomy, fermentation, isolation and biological properties. *J Antibiot (Tokyo)* 48(10):1124–1129.
27. Okuhara M, et al. (1980) Studies on new phosphonic acid antibiotics. I. FR-900098, isolation and characterization. *J Antibiot (Tokyo)* 33(1):13–17.
28. Okuhara M, et al. (1980) Studies on new phosphonic acid antibiotics. III. Isolation and characterization of FR-31564, FR-32863 and FR-33289. *J Antibiot (Tokyo)* 33(1):24–28.
29. Krug D, Müller R (2014) Secondary metabolomics: The impact of mass spectrometry-based approaches on the discovery and characterization of microbial natural products. *Nat Prod Rep* 31(6):768–783.
30. Forseth RR, Schroeder FC (2011) NMR-spectroscopic analysis of mixtures: From structure to function. *Curr Opin Chem Biol* 15(1):38–47.
31. Lin J, Nishiyama M, Kuzuyama T (2015) Identification of the biosynthetic gene cluster for the herbicide phosphonothrixin in Saccharothrix sp. ST-888. *J Antibiot (Tokyo)* 68(5):357–359.
32. Blodgett JA, et al. (2007) Unusual transformations in the biosynthesis of the antibiotic phosphinothricin tripeptide. *Nat Chem Biol* 3(8):480–485.
33. Hove-Jensen B, Zechel DL, Jochimsen B (2014) Utilization of glyphosate as phosphate source: Biochemistry and genetics of bacterial carbon-phosphorus lyase. *Microbiol Mol Biol Rev* 78(1):176–197.
34. Maehr H, Blount JF, Pruess DL, Yarmchuk L, Kellett M (1973) Antimetabolites produced by microorganisms. 8. N5-hydroxy-L-arginine, a new naturally occurring amino acid. *J Antibiot (Tokyo)* 26(5):284–288.
35. Perlman D, Vlietinck AJ, Matthews HW, Lo FF (1974) Microbial production of vitamin B12 antimetabolites. I. N5-hydroxy-L-arginine from *Bacillus cereus* 439. *J Antibiot (Tokyo)* 27(11):826–832.
36. Fischer B, et al. (1973) [Metabolic products of microorganisms. 118. Delta-N-hydroxy-L-arginine, an amino acid antagonist from Nannizzia gypsea]. *Arch Mikrobiol* 91(3):203–220. German.
37. Bougioukou DJ, Mukherjee S, van der Donk WA (2013) Revisiting the biosynthesis of dehydrophos reveals a tRNA-dependent pathway. *Proc Natl Acad Sci USA* 110(27):10952–10957.
38. Gao J, et al. (2014) Use of a phosphonate methyltransferase in the identification of the fosfazinomycin biosynthetic gene cluster. *Angew Chem Int Ed Engl* 53(5):1334–1337.
39. Atherton FR, et al. (1982) Antibacterial activity and mechanism of action of phosphonopeptides based on aminomethylphosphonic acid. *Antimicrob Agents Chemother* 22(4):571–578.
40. Cioni JP, et al. (2014) Cyanohydrin phosphonate natural product from *Streptomyces regensis*. *J Nat Prod* 77(2):243–249.
41. Halkier BA, Gershenzon J (2006) Biology and biochemistry of glucosinolates. *Annu Rev Plant Biol* 57:303–333.
42. Hakala MT, Glaid AJ, Schwert GW (1956) Lactic dehydrogenase. II. Variation of kinetic and equilibrium constants with temperature. *J Biol Chem* 221(1):191–209.
43. Philmus B, Guerrette JP, Hemscheidt TK (2009) Substrate specificity and scope of MvdD, a GRASP-like ligase from the microviridin biosynthetic gene cluster. *ACS Chem Biol* 4(6):429–434.
44. Seto H, et al. (1983) Studies on the biosynthesis of bialaphos (SF-1293). 2. Isolation of the first natural products with a C-P-H bond and their involvement in the C-P-C bond formation. *J Antibiot (Tokyo)* 36(1):96–98.
45. Yu X, Price NP, Evans BS, Metcalf WW (2014) Purification and characterization of phosphonoglycans from *Glycomyces* sp. strain NRRL B-16210 and *Stackebrandtia nassauensis* NRRL B-16338. *J Bacteriol* 196(9):1768–1779.
46. Doroghazi JR, et al. (2014) A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nat Chem Biol* 10(11):963–968.
47. Doroghazi JR, et al. (2011) Genome sequences of three tunicamycin-producing Streptomyces Strains, *S. chartreusis* NRRL 12338, *S. chartreusis* NRRL 3882, and *S. lysosuperificus* ATCC 31396. *J Bacteriol* 193(24):7021–7022.
48. Hyatt D, et al. (2010) Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119.
49. Peng Y, Leung HC, Yiu SM, Chin FY (2012) IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28(11):1420–1428.