Identifying producers of antibacterial compounds by screening for antibiotic resistance

Maulik N Thaker[1], Wenliang Wang[1], Peter Spanogiannopoulos[1], Nicholas Waglechner[1], Andrew M King[1], Ricardo Medina[2] & Gerard D Wright[1]

Microbially derived natural products are major sources of antibiotics and other medicines, but discovering new antibiotic scaffolds and increasing the chemical diversity of existing ones are formidable challenges. We have designed a screen to exploit the self-protection mechanism of antibiotic producers to enrich microbial libraries for producers of selected antibiotic scaffolds. Using resistance as a discriminating criterion we increased the discovery rate of producers of both glycopeptide and ansamycin antibacterial compounds by several orders of magnitude in comparison with historical hit rates. Applying a phylogeny-based screening filter for biosynthetic genes enabled the binning of producers of distinct scaffolds and resulted in the discovery of a glycopeptide antibacterial compound, pekiskomycin, with an unusual peptide scaffold. This strategy provides a means to readily sample the chemical diversity available in microbes and offers an efficient strategy for rapid discovery of microbial natural products and their associated biosynthetic enzymes.

Natural products from microbes, and the actinomycetes in particular, have been the sources of some of the most important antimicrobial drug classes, including the β-lactams, tetracyclines, macrolides, aminoglycosides, rifamycins and glycopeptides. Indeed, four-fifths of antibiotic scaffolds approved for clinical use over the past decade have been microbial natural products[1]. Despite these successes, natural products have fallen out of favor in antibiotic drug discovery for several reasons, including the fact that traditional antibiotic screens of actinomycete extracts often yield known compounds. This has resulted in estimates that millions of actinomycetes need to be screened to identify new antimicrobial natural products[2], a daunting task even in the era of automated high-throughput screens. On the other hand, over two decades of experience in screening libraries of synthetic compounds have not met expectations either. Faced with no clear direction on where to find new antibiotic leads, the pharmaceutical sector has gradually phased out their antibiotic discovery programs[3] and moved to more tractable areas of clinical need.

At the same time that the pharmaceutical sector is abandoning natural products for antibiotics and other clinical indications, recently available genome sequences from multiple bacterial and environmental metagenomes have revealed a remarkably rich potential for the production of natural products[4,5]. Furthermore, techniques to genetically modify these organisms and their associated biosynthetic gene clusters are becoming more sophisticated, offering the possibility of expanding chemical diversity using synthetic biology approaches. There is an opportunity here to increase the chemical space of structurally complex and synthetically intractable molecular scaffolds using enzymes that modify core scaffold structures[6].

These modifications can affect pharmacology and target affinity and, in the case of antibiotics, mitigate resistance.

One of the main challenges in the field is the identification of microbes with the biosynthetic capacity to produce desired compound classes. It has been estimated, for example, that the frequency of specific scaffolds in antibacterial producers ranges over 6 orders of magnitude: from 1 in 10 for antibacterials such as streptothricin, to 1 in 10^7 for daptomycin[7]. A similar calculation for the glycopeptide antibacterials (GPA), the class that includes the clinically important drugs vancomycin and teicoplanin, revealed that ~150,000 actinomycete strains need to be screened to identify a glycopeptide producer[7].

The glycopeptide class is one where synthetic biology approaches could prove especially useful in expanding diversity through, for example, the known modifications by halogenation, oxidation, sulfation, acylation and glycosylation[8], and total synthesis is so far impractical on a commercial scale. The class includes two dominant, clinically relevant scaffolds, vancomycin and teicoplanin, each of which is composed of a highly cross-linked and modified heptapeptide; a third, related scaffold complestatin shares some structural features with GPAs but lacks antibacterial activity. Similarly, the ansamycin class of natural products includes the rifamycin antibacterials and antitumor agents such as geldanamycin, both of which present opportunities for expansion of chemical diversity through scaffold-modifying enzymes.

Here we report a systematic approach for the discovery of new antibacterials from actinomycetes using resistance as a trait for selecting GPA-producing organisms. By applying a screen for resistance, we greatly enrich isolated cultures for GPA producers, thereby improving efficiency of discovery by at least four orders of magnitude over

[1]Department of Biochemistry and Biomedical Sciences, M.G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, Ontario, Canada. [2]Department of Microbiology, Chemical Bioactive Center, Central University of Las Villas, Santa Clara, Villa Clara, Cuba. Correspondence should be addressed to G.D.W. (wrightge@mcmaster.ca).

Figure 1 Comparison of resistance-based GPA screening with other screening approaches and their rates of success. (**a**) In this culture-independent method, the soil metagenome is captured in cosmids to create a megalibrary (10^7 clones) in *E. coli*. The frequency of finding a vancomycin-group GPA is estimated at 1 in 2 million clones and for a teicoplanin-group GPA, 1 in 5–10 million clones. (**b**) The phenotype-based screening is the conventional and most commonly used method for new antibiotic discovery. The method involves isolation, followed by activity-guided selection of antimicrobial producers, followed by identification of the molecules yielding a GPA producer in every 150,000 strains screened. (**c**) The resistance-based screening is a hybrid approach involving initial isolation in the presence of GPA, permitting only a very small subset of strains from the soil. GPA fingerprinting analysis shows 1 in every 10 strains has the potential to produce a GPA.

a Metagenomic megalibrary–based screen

b Phenotype-based screen

c Resistance-based screen
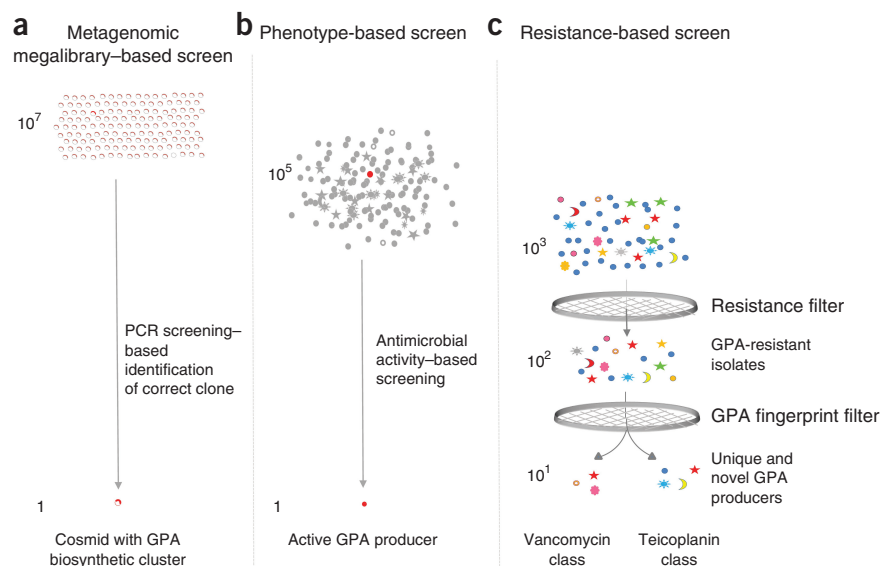
historical estimates (**Fig. 1**). Using this strategy, we isolated a new GPA, pekiskomycin, with a rare peptide scaffold and novel chemical modifications. Furthermore, we also demonstrated the versatility of our screening process in harnessing natural product diversity by identifying a producer of geldanamycin, an ansamycin, using a rifampin-resistance screen. This work offers an approach to access and expand the chemical diversity of antimicrobials as well as associated enzymes and other natural product scaffolds.

RESULTS

Resistance enrichment screen for putative GPA producers

It is well known that antibacterial producers need to evolve a self-protection mechanism to avoid suicide[9]. We reasoned that this would be an appropriate filter to identify antibiotic-producing organisms from a library of actinomycetes. GPA resistance is most frequently achieved in both producers and pathogens by a conserved *vanHAX* operon that modifies cell wall precursors to terminate in D-Ala-D-Lac rather than the canonical D-Ala-D-Ala[10,11]. The ability to recognize various GPAs varies, however, among the producers and the pathogens.

We randomly selected 1,000 actinomycete strains from our in-house library of environmental actinomycetes. These were screened for resistance to vancomycin at 20 µg/ml and 39 were found to be resistant (~4%). The resistance genotype of the 39 strains was verified by positive amplification of the *vanHAX* GPA-resistance genes.

The vancomycin-resistant strains were further analyzed for the presence of GPA biosynthetic genes, using a series of diagnostic primers (**Supplementary Table 1**), designed from conserved regions of

signature genes representing GPA biosynthetic machinery (**Fig. 2**). These genes offer a molecular fingerprint for the presence of glycopeptide biosynthetic clusters. In particular, the amplicons differentiate between vancomycin and teicoplanin GPAs as well as the structurally related but nonantibacterial natural product complestatin[12]. GPA fingerprint analysis of the 39 resistant strains identified one isolate WAC (Wright Actinomycetes Collection) 1227 as a likely vancomycin-group GPA producer (**Fig. 3a**). A control fingerprint screen of 200 vancomycin-sensitive strains from the same collection identified no GPA-producing candidates, demonstrating that resistance is a useful discriminating filter to identify antibacterial producers.

To investigate whether application of the resistance filter early in the process can also facilitate the discovery process from natural sources, we used vancomycin (20 µg/ml) for selective isolation of glycopeptide-resistant strains from soil samples. GPAs are produced by a number of diverse genera of the order Actinomycetales (**Supplementary Table 2**). The relatively fast-growing members of the *Streptomyces* genus readily outnumber rarer genera when isolated on common laboratory media. The introduction of the selection strategy ensured exclusive isolation of GPA-resistant strains, irrespective of their phylogeny, relative abundance or growth rate. We selected
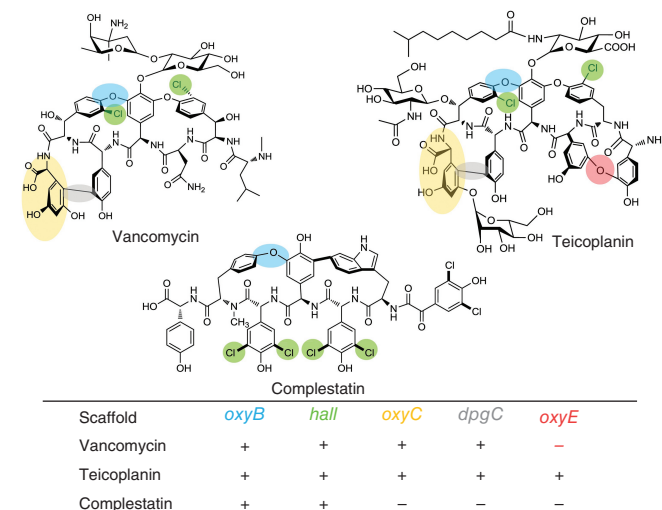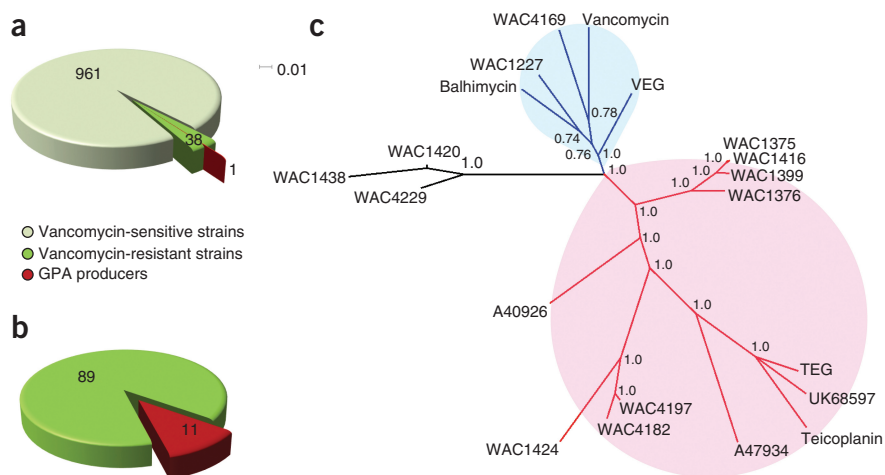
Figure 2 The heptapeptide core of the glycopeptides varies in amino acid composition and intra-strand cross-links, creating scaffold diversity. The nonantibiotic, anti-complement, complestatin core peptide is cross-linked to create a two-ring structure, whereas the antibiotic scaffolds represented by vancomycin and teicoplanin are cross-linked to create three- and four-ring core structures, respectively. The watermarks represent the site of action of monooxygenases (*oxyB*-blue, *oxyC*-gray, *oxyE*-red) for cross-linking of the heptapeptide, the conserved tailoring enzyme halogenase (*halI*-green) and the enzyme involved in biosynthesis of the nonproteogenic, amino acid dihydroxyphenyl glycine (*dpgC*-yellow). The panel of these conserved genes constitutes the GPA fingerprint PCR detection set for WAC strains. Shown in the table is the amplicon fingerprint key that serves to discriminate among different glycopeptide scaffolds.

Scaffold	*oxyB*	*halI*	*oxyC*	*dpgC*	*oxyE*
Vancomycin	+	+	+	+	−
Teicoplanin	+	+	+	+	+
Complestatin	+	+	−	−	−

Figure 3 The abundance and diversity of GPA producers. (**a**) In a nonselective screen of 10^3 actinomycetes, GPA resistance is infrequent and biosynthesis is a subset of resistance. (**b**) A selective isolation in the presence of vancomycin found 11% of strains had the potential to produce a GPA, a notable enrichment of the GPA producers. (**c**) The composite phylogenetic tree from the amplicon sequences of *oxyB*, *oxyC* and *dpgC* differentiates between the producers of GPAs with a vancomycin and teicoplanin scaffold as well as outliers. Follow-up studies with selected strains verified the prediction model with WAC4169 and WAC1375, showing the presence of the expected scaffold. WAC1420 and WAC4229 clustered close to each other but on an outlying branch, constituting a variant backbone.

a
961
38
1

○ Vancomycin-sensitive strains
○ Vancomycin-resistant strains
● GPA producers

b
89
11

c
0.01
WAC4169 Vancomycin
WAC1227
Balhimycin
VEG
0.78
0.74
0.76 1.0
1.0
WAC1420 1.0
WAC1438
WAC4229
WAC1375
WAC1416
WAC1399
WAC1376
1.0
1.0
1.0
1.0
1.0
1.0
A40926
1.0
1.0
TEG
UK68597
1.0
1.0
WAC4197
WAC4182
Teicoplanin
WAC1424
A47934

100 vancomycin-resistant strains in this fashion and explored their GPA biosynthesis potential.

Enrichment of the culture collection using vancomycin raised the number of potential GPA producers. Here, out of the 100 resistant strains, 11 strains tested positive using our fingerprinting method for GPA biosynthetic clusters (**Fig. 3b**). The unique identity of the producers was determined using the diagnostic BOX PCR approach, where amplicon patterns exclusive to each genome are compared[13]. We also identified several resistant strains that amplified singular biosynthetic genes or their combinations which were inconsistent with the expected GPA fingerprint pattern. We expect that either these are degraded GPA biosynthetic clusters or they represent other biosynthetic programs.

GPA biosynthesis fingerprints discriminates among scaffolds

Our PCR-based molecular fingerprint strategy enabled us to rapidly screen resistant organisms for their potential to produce GPAs. One of the challenges of natural product discovery is dereplication, the rediscovery of known molecules. Bacteria, even across different genera, can produce identical secondary metabolites[14]. We recognized that a phylogenetic analysis of our amplified, fingerprint biosynthetic genes would offer a new approach to dereplication.

The fingerprint GPA amplicons were sequenced and used to construct an unrooted Bayesian phylogenetic tree for the concatenated translated sequences of the three core GPA biosynthetic genes (*oxyB*, *oxyC* and *dpgC*) (**Fig. 3c**). The relatively short amplicons make estimation of phylogenetic trees for individual genes challenging; however when concatenated, the signal in the GPA biosynthetic sequences became strong enough that the major GPA types resolved into distinct branches. Based on this analysis we rapidly predicted the relatedness of each of the novel clusters to those previously reported and used these predictions to identify the strains most worthy of additional exploration without full genome sequencing or purification of the target molecules. This approach facilitates rapid screening of GPA producers for chemical diversity.

The known GPA producers branched into the separate vancomycin and teicoplanin groups (**Fig. 3c**). Among the new strains, only three (WAC1420, WAC4229 and WAC1438) did not group with known GPAs suggesting that these three may produce a novel scaffold.

Validation of the screening procedure

Having investigated the genetic potential of 12 resistant strains for GPA biosynthesis, we next examined their ability to produce a GPA. Only one of these strains, WAC4169, which we predicted to be a

vancomycin-group producer, yielded readily detectable quantities of a GPA. A molecule with mass 1,477.4674, as determined by high-resolution mass spectrometry (HRMS), was identified as *N,N*-dimethylvancomycin (**Fig. 4a** and **Supplementary Fig. 1**). This GPA, first reported in 1988 (ref. 15), has an unusual trimethylated primary amine on the N-terminal leucine, which was confirmed by NMR (**Supplementary Figs. 2** and **3**).

The phylogenetic tree also predicted a clade for strains comprising known, as well as putative, teicoplanin-group producers identified by the GPA-fingerprinting amplicon pattern. We selected WAC1375 as a representative of the clade for further verification. A draft sequence of the genome revealed a GPA biosynthetic cluster, including *oxyE* connecting 4-hydroxyphenylglycine-1 and 3,5-dihydroxyphenylglycine-3 of the heptapeptide in the teicoplanin group (**Fig. 5a** and **Supplementary Table 3**). A-domain prediction of the nonribosomal peptide synthase modules identified the teicoplanin heptapeptide core (**Supplementary Table 4** and **Fig. 4b**). Optimization of the fermentation conditions and subsequent purification of an associated GPA (**Supplementary Fig. 4**) identified it as teicoplanin aglycone, an unexpected result given the presence of several genes encoding predicted scaffold-modifying enzymes. Further optimization of production may yield additional GPA molecules from this cluster. The results of both WAC4169 and WAC1375 validate the potential of GPA fingerprinting and the associated phylogenetic analysis in identifying and differentiating GPA scaffolds.

Identification of the novel GPA pekiskomycin

We next investigated the GPA-producing potential of the strains that did not group with either vancomycin- or teicoplanin-group members in the phylogeny (**Fig. 3c**). Follow-up studies identified conditions for GPA production in WAC1420. These results were supported by RT-PCR amplification of the biosynthetic gene transcripts, using the GPA fingerprinting primers under the same conditions (**Supplementary Fig. 5**).

NMR, mass spectrometry and chemical degradation analysis of the GPA purified from WAC1420 identified a GPA not previously described, which we named pekiskomycin (**Fig. 4c**, **Supplementary Figs. 6–13** and **Supplementary Table 5**), in recognition of the collection site, Pekisko, Alberta, Canada. Pekiskomycin is an unusual GPA with several unique features and others shared with both canonical GPA scaffolds. The peptide has three intramolecular cross-links similar to those of vancomycin, consistent with the absence of *oxyE* in the GPA fingerprint. Unlike vancomycin, pekiskomycin lacks the

Figure 4 GPAs identified in the screen. (**a**) Structure of dimethylvancomycin isolated from WAC4169. The molecule has a trimethylated amino terminus instead of monomethylation in vancomycin. (**b**) Purification and analysis of the GPA from WAC1375 identified teicoplanin aglycone, confirming the predicted GPA group for the molecule. (**c**) Structure of pekiskomycin with its characteristic amino acids, alanine and glutamic acid (green) at position 1 and 3, respectively, in the heptapeptide core. The GPA is dimethylated (red) at N-terminal alanine at position 1 and is also sulfated at β-hydroxytyrosine (blue). A methylated glucose moiety (gray) is present on the hydroxyphenylglycine-4.

Dimethylvancomycin

Teicoplanin class GPA

Pekiskomycin
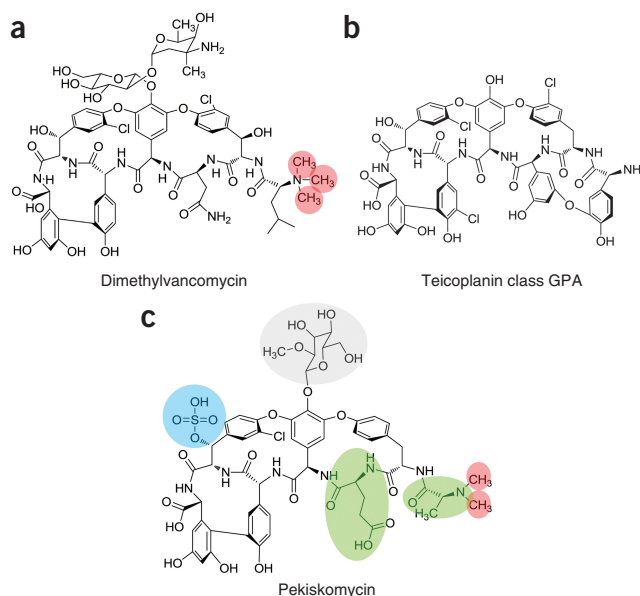
N-methyl-Leu at residue 1 and an asparagine at residue 3; rather it has *N,N*-dimethyl-Ala and glutamic acid, respectively, at these positions. On the other hand, tyrosine at residue two (instead of the β-hydroxy-Tyr characteristic of the vancomycin group) and sulfation are features shared with the teicoplanin group of molecules. Though pekiskomycin has modest antibacterial activity, it has some effect against vancomycin-resistant enterococci-type B (**Supplementary Fig. 14**). Consistent with the antibacterial spectrum, isothermal titration calorimetry measurements of pekiskomycin with the *N,N*-diacetyl-Lys-D-Ala-D-Ala tripeptide give a K_d of 66.6 µM, 36-fold weaker than the value for vancomycin (1.82 µM; **Supplementary Fig. 15**). Pekiskomycin therefore may act more as a signaling molecule than as an antibacterial[16].

A draft genome sequence of WAC1420 revealed an ~86-kb cluster containing 41 coding sequences, consistent with the predicted biosynthetic requirement, export, regulation and resistance to pekiskomycin (**Fig. 5b** and **Supplementary Table 6**). Among the scaffold-accessorizing genes are a glycosyltransferase, two methyltransferases and a sulfotransferase (*pek25*). Sulfotransferases are the least abundant of all the GPA-modifying enzymes and have been previously found only with teicoplanin-related antibacterials[17–20]. To the best of our knowledge, pekiskomycin is the first naturally occurring sulfated GPA that does not have a teicoplanin scaffold.

Biosynthesis phylogeny allows early dereplication
Our strategy of using concatenated fingerprint sequences for phylogeny distinguished between different GPA scaffolds with high confidence. We further postulated that the model could be equally useful in grouping different strains with identical or highly similar GPA

products, and thus can be used for early dereplication. To explore this hypothesis, we selected WAC4229, which is present on a branch very close to WAC1420. These two strains are distinguished by morphology, geography of isolation, BOX PCR amplicon pattern and 16S rRNA gene sequence (**Supplementary Fig. 16**). We purified the GPA from WAC4229 and confirmed it to be identical to pekiskomycin. We also generated a draft genome for WAC4229 and identified a cluster capable of pekiskomycin biosynthesis (**Supplementary Table 7**). The cluster shows 90% identity with its counterpart in WAC1420, except for one of the ends, which shows evidence of multiple recombination events (**Supplementary Fig. 17**).

Resistance-based discovery of ansamycins
Finally, we tested whether our approach of resistance-guided strain selection followed by phylogeny-based classification would also be applicable to other scaffolds. We chose the ansamycin natural products for follow-up study, owing to their known chemical diversity and importance in antibiotic therapy (e.g., the RNA polymerase inhibitor rifamycin) and anticancer therapy (e.g., the HSP90 inhibitor geldanamycin).
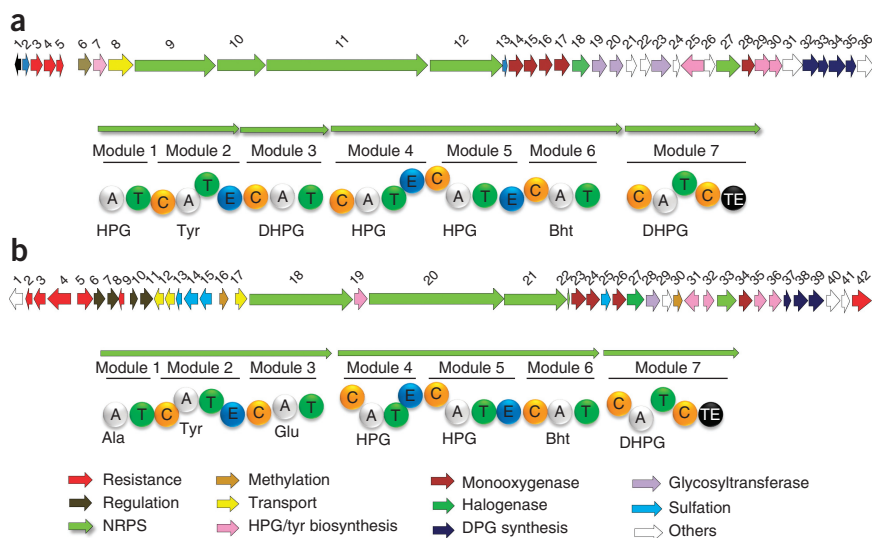

Figure 5 GPA biosynthetic clusters from the draft genomes of WAC strains. (**a**) Organization of the GPA biosynthetic cluster in WAC1375 showing a typical teicoplanin heptapeptide core and presence of *oxyE* (*orf29*). The cluster lacks the *vanRS* two-component system for regulation of self-protection. (**b**) Gene cluster for pekiskomycin in WAC1420 with the modular organization of the nonribosomal peptide synthase (NRPS) genes. The nonconserved amino acids coded by A-domains in modules 1 and 3 could not be predicted *in silico*. The cluster, flanked by genes for resistance on either side, has an operon (*pek11-pek15*), predicted to be involved in sulfate uptake and transfer, not found in any other sulfated GPA biosynthetic clusters. The ORFs are color coded based on their putative functions and the NRPS. Domains: condensation (C), adenylation (A), thiolation (T), epimerization (E) and a thioesterase (TE) for peptide chain termination. HPG, 4-hydroxyphenylglycine; DHPG, 3,5-dihydroxyphenylglycine; Bht, β-hydroxytyrosine.

Figure 6 Rifampin-assisted isolation of ansamycin producers. (**a**) A collection of 213 rifampin-resistant strains were screened by PCR using AHBA synthase–specific degenerate primers, which yielded 16 possible ansamycin producers and a hit rate of 7.5%. (**b**) Phylogenetic analysis of partial AHBA synthases from 7 known ansamycin biosynthetic (**Supplementary Table 9**) gene clusters and 16 identified from this study. (**c**) Production assays confirmed WAC5038, which clusters near the geldanamycin AHBA synthase, does produce geldanamycin (**Supplementary Fig. 19**).
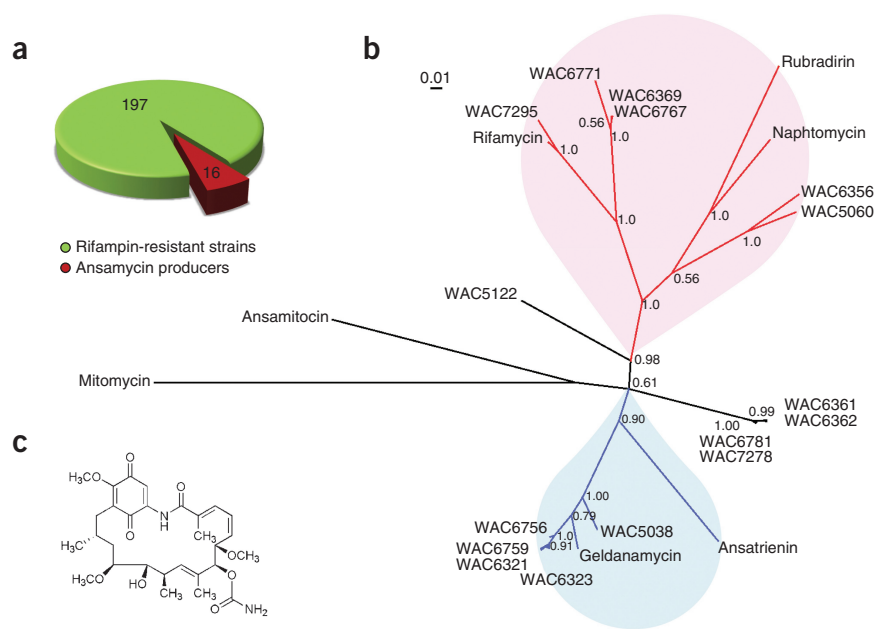
Using rifampin (20 μg/ml) we selectively identified 213 resistant actinomycete strains. These were then screened by PCR for the presence of the conserved aminohydroxybenzoic acid synthase (AHBA) gene (**Supplementary Table 1**), yielding 16 unique, putative, ansamycin producers (a 7.5% rate of success (**Fig. 6a** and **Supplementary Fig. 18**)). A phylogenetic tree based on AHBA sequence amplicons identified four strains branching with the known rifamycin producer (**Fig. 6b**). Bioassay of the fermentation products of these strains was consistent with the production of an RNA polymerase–targeted antibacterial (**Supplementary Fig. 19**). Out of the five strains that grouped in the geldanamycin group (**Fig. 6b**), we selected WAC5038 for follow-up studies and confirmed that it produced geldanamycin (**Fig. 6c**, **Supplementary Figs. 20** and **21**).

DISCUSSION

Natural products can provide useful leads for the development of therapeutic agents but have fallen out of favor in recent years largely because of their low hit rates (e.g., 1 in 150,000 for GPAs[7]) and their tendency to yield known molecules. The drug discovery sector therefore has turned increasingly to synthetic molecules, but these often lack the chemical complexity (e.g., number of stereocenters, H-bond donors and acceptors, ring systems) that track with the bioactivity of natural products. The emerging area of synthetic biology offers a potential solution to this challenge of chemical complexity[8], and here access to libraries of scaffolds and associated modifying enzymes is critical.

In recent years, the approach of screening metagenomic libraries of DNA from soil for natural product biosynthetic clusters has also evolved greatly. Although this strategy benefits from the ability to screen otherwise unculturable strains, it is technically demanding, requiring large cosmid libraries (10^7 clones) isolated from a substantial mass (hundreds of grams) of starting material, the isolation of overlapping clones to complete clusters and eventually expression in heterologous hosts for compound production.

Our approach of resistance-guided isolation overcomes many of these hurdles as it requires a small number of strains to be screened (~10 strains for a GPA) and also provides ease of sampling multiple environments because only small quantities of soil or other source material is needed. We note that such approaches have in the past been applied for the enrichment of producers of aminoglycosides and lincosamides[21,22]. Furthermore, the identification of the producing organism also enables ready access to the entire biosynthetic gene cluster and source of the produced molecule upon fermentation. The addition of phylogeny-based cluster analysis also offers a facile way to dereplicate known structures.

Using GPA resistance as a discriminating filter, we eliminated ~96% of strains not producing GPA. The up-front triage of GPA-sensitive strains by selection on vancomycin also provides an opportunity to enrich the library for slow-growing or less abundant actinomycetes, thereby increasing genetic diversity (**Supplementary Table 8**).

Our results offer an explanation for the historically low detection-rate estimates for GPA producers. Traditionally these were detected by cell-killing phenotype screens of cell-free extracts of actinomycetes grown on various media; these represent only a small subset of the total biosynthetic potential in the soil microbiome[4]. In our screen we addressed this limitation by using fingerprinting primers that detect putative GPA-producing strains even in the absence of antibacterial production. This approach increases the success rate by more than four log scale values over historical estimates.

In addition to improving the frequency of identification of antibacterial producers, we simultaneously addressed the dereplication issue by constructing a phylogenetic tree with amplicons from biosynthetic genes. The phylogenetic data highlight the diversity of hits from the screen exemplified by the strains bearing GPAs with different conserved cores; WAC4169 (vancomycin group) and WAC1375 (teicoplanin group) as well as WAC1420 and WAC4229 both with a noncanonical peptide core. This approach informs downstream exploration of various producers based on the desired outcomes: new scaffolds, variations of known scaffolds (and therefore likely sources of new accessorizing enzymes) and avoidance of similar clusters.

The results of the phylogenetic analysis also indicate that the use of vancomycin for isolation did not create a bias in favor of the vancomycin-group scaffold. This is because the sensitivity of the producers and resistant pathogens toward various GPA molecules is determined independently by the ability of their respective sensor kinases to recognize each GPA[23,24]. Thus, the use of vancomycin during isolation does not compromise the likelihood of discovering a GPA active against vancomycin-resistant enterococci.

Of all the candidate GPA producers we investigated, only WAC4169 (the producer of dimethylvancomycin) expressed readily detectable amounts of the GPA under our standard growth conditions. The novel heptapeptide pekiskomycin from WAC1420 was produced in low abundance and required purification of sufficient quantities for downstream analysis. This finding supports the hypothesis that any easily

detectable activity is likely to have been already discovered during earlier activity-based screens[2,25]. Because such screens rely exclusively on the ability of the bacterium to express the requisite biosynthetic genes under laboratory conditions, the inherent potential to produce molecules remains unsampled. These latent producers represent the pool of strains that escaped detection in earlier screens. Our results demonstrate that these clusters can in fact occur at much higher frequencies than previously anticipated. Once novel leads are identified by, for example, our biosynthetic fingerprint and associated phylogeny, the strain can be strategically modified or conditions manipulated to enhance bioproduction[26,27]. Also, we note that the GPA producers obtained in this study are obtained from resistance-based enrichment using only vancomycin. Similar screens in the presence of different GPAs from the same soils may yield more novel producers that were lost because of their sensitivity to vancomycin.

Finally, the application of the resistance-based screening approach for discovery of ansamycins using a rifamycin resistance filter not only confirmed the high hit rates obtained in the GPA screen but also yielded a diverse set of ansamycin producers. In fact, we found more putative ansamycin producers for nonantibiotic groups than rifamycin analogs. The easy identification of a geldanamycin producer shows that the scope of our strategy is not limited to antibiotics discovery but can be applied to the discovery of various diverse natural products with other pharmaceutical applications.

METHODS
Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

1. Wright, G.D. Antibiotics: a new hope. *Chem. Biol.* **19**, 3–10 (2012).
2. Anonymous. A call to arms. *Nat. Rev. Drug Discov.* **6**, 8–12 (2007).
3. Cooper, M.A. & Shlaes, D. Fix the antibiotics pipeline. *Nature* **472**, 32 (2011).
4. Reddy, B.V.B. *et al.* Natural product biosynthetic gene diversity in geographically distinct soil microbiomes. *Appl. Environ. Microbiol.* **78**, 3744–3752 (2012).
5. Nett, M., Ikeda, H. & Moore, B.S. Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat. Prod. Rep.* **26**, 1362–1384 (2009).
6. Lamb, S.S. & Wright, G.D. Accessorizing natural products: adding to nature's toolbox. *Proc. Natl. Acad. Sci. USA* **102**, 519–520 (2005).
7. Baltz, R.H. Marcel Faber Roundtable: is our antibiotic pipeline unproductive because of starvation, constipation or lack of inspiration? *J. Ind. Microbiol. Biotechnol.* **33**, 507–513 (2006).
8. Thaker, M.N. & Wright, G.D. Opportunities for synthetic biology in antibiotics: expanding glycopeptide chemical diversity. *ACS Synth. Biol.* doi:10.1021/sb300092n (17 December 2012).
9. Cundliffe, E. & Demain, A.L. Avoidance of suicide in antibiotic-producing microbes. *J. Ind. Microbiol. Biotechnol.* **37**, 643–672 (2010).
10. Wright, G.D. The antibiotic resistome: the nexus of chemical and genetic diversity. *Nat. Rev. Microbiol.* **5**, 175–186 (2007).
11. Courvalin, P. Vancomycin resistance in Gram-positive cocci. *Clin. Infect. Dis.* **42**, S25–S34 (2006).
12. Chiu, H.T. *et al.* Molecular cloning and sequence analysis of the complestatin biosynthetic gene cluster. *Proc. Natl. Acad. Sci. USA* **98**, 8548–8553 (2001).
13. Lanoot, B. *et al.* BOX-PCR fingerprinting as a powerful tool to reveal synonymous names in the genus Streptomyces. Emended descriptions are proposed for the species *Streptomyces cinereorectus, S. fradiae, S. tricolor, S. colombiensis, S. filamentosus, S. vinaceus* and *S. phaeopurpureus. Syst. Appl. Microbiol.* **27**, 84–92 (2004).
14. Thaker, M.N. *et al.* Biosynthetic gene cluster and antimicrobial activity of the elfamycin antibiotic factumycin. *Med. Chem. Commun.* **3**, 1020–1026 (2012).
15. Nagarajan, R. *et al.* M43 antibiotics: methylated vancomycins and unrearranged CDP-I analogs. *J. Am. Chem. Soc.* **110**, 7896–7897 (1988).
16. Yim, G., Wang, H.H. & Davies, J. Antibiotics as signalling molecules. *Phil. Trans. R. Soc. Lond. B* **362**, 1195–1200 (2007).
17. Pootoolal, J. *et al.* Assembling the glycopeptide antibiotic scaffold: The biosynthesis of A47934 from Streptomyces toyocaensis NRRL15009. *Proc. Natl. Acad. Sci. USA* **99**, 8962–8967 (2002).
18. Skelton, N.J., Williams, D.H., Monday, R.A. & Ruddock, J.C. Structure elucidation of the novel glycopeptide antibiotic UK-68,597. *J. Org. Chem.* **55**, 3718–3723 (1990).
19. Banik, J.J. & Brady, S.F. Cloning and characterization of new glycopeptide gene clusters found in an environmental DNA megalibrary. *Proc. Natl. Acad. Sci. USA* **105**, 17273–17277 (2008).
20. Banik, J.J., Craig, J.W., Calle, P.Y. & Brady, S.F. Tailoring enzyme-rich environmental DNA clones: a source of enzymes for generating libraries of unnatural natural products. *J. Am. Chem. Soc.* **132**, 15661–15670 (2010).
21. Bibikova, M.V., Ivanitskaia, L.P. & Singal, E.M. *Antibiotiki* Directed screening of aminoglycoside antibiotic producers on selective media with gentamycin. (Original in Russian.) **26**, 488–492 (1981).
22. Ivanitskaia, L.P., Bibikova, M.V., Gromova, M.N., Zhdanovich Iu, V. & Istratov, E.N. *Antibiotiki* Use of selective media with lincomycin for the directed screening of antibiotic producers. (Original in Russian.) **26**, 83–86 (1981).
23. Hong, H.J., Hutchings, M.I. & Buttner, M.J. Vancomycin resistance VanS/VanR two-component systems. *Adv. Exp. Med. Biol.* **631**, 200–213 (2008).
24. Koteva, K. *et al.* A vancomycin photoprobe identifies the histidine kinase VanSsc as a vancomycin receptor. *Nat. Chem. Biol.* **6**, 327–329 (2010).
25. Berdy, J. Bioactive microbial metabolites. *J. Antibiot. (Tokyo)* **58**, 1–26 (2005).
26. Hosaka, T. *et al.* Antibacterial discovery in actinomycetes strains with mutations in RNA polymerase or ribosomal protein S12. *Nat. Biotechnol.* **27**, 462–464 (2009).
27. Martín, J.-F. & Liras, P. Engineering of regulatory cascades and networks controlling antibiotic biosynthesis in Streptomyces. *Curr. Opin. Microbiol.* **13**, 263–273 (2010).

ONLINE METHODS

Isolation of actinomycetes and screening. The strains used in the study are a subset of the Wright Actinomycetes Collection (WAC), isolated from soil samples across Canada, Cuba, Nigeria and France. For isolation, 1 g dry weight of soil was treated with dry heat, phenol or rehydrated in 9 ml water. Serial dilutions of suspension or supernatant were spread on either Streptomyces isolation media[28] or Humic acid Vitamin Agar[29] supplemented with Nalidixic acid (100 µg/ml). Actinomycete colonies were isolated based on their colony morphology and restreaked on Bennett's Agar for further verification and purification before making the spore suspension. Antibiotic-resistant strains were isolated following the same protocol but supplemented with vancomycin (10 µg/ml) or rifampin (20 and 50 µg/ml).

Primer design and PCR conditions. The degenerate PCR primers for GPA biosynthesis fingerprinting were designed using the consensus sequences of the conserved regions in the respective genes (*oxyB, oxyC, halI, dpgC, oxyE*). Conditions for PCR amplification were standardized using available producer strains (**Supplementary Table 2**). A touchdown program with annealing temperature decreasing from 62 °C to 58 °C for initial four cycles followed by 26 cycles at 57 °C was used for all GPA fingerprinting primer sets except *oxyE*, for which a constant annealing temperature of 55 °C for 30 cycles was used. AHBA synthase gene–specific primers were used as previously described[30]. For amplification of 16S rRNA gene fragments, *vanHAX*[28] and BOX PCR[14], protocols standardized earlier were followed. *Phylogenetic analysis.* Translated amplicon sequences for each of *oxyB, oxyC* and *dpgC* from 12 candidates identified in this study were collected along with the corresponding sequences from the published A40926, teicoplanin, vancomycin, balhimycin, A47934, VEG and TEG clusters where available.

The AHBA synthase analysis was conducted using 16 newly identified translated amplicons and 7 previously identified AHBA synthases involved in ansamycin biosynthesis (**Supplementary Table 9**). The translated sequences for each gene were aligned with MUSCLE, using default parameters[31] and manual curation. Two sequences were marked with 'missing' symbols; unavailable *dpgC* from TEG cluster and *oxyB* from WAC1438 where the amplified monoxygenase did not align with the others. The final alignment consisted of 727 sites and 20 taxa.

This alignment served as input for Bayesian phylogenetic analysis in MrBayes 3.2.0 using a mixed amino acid model prior, allowing gamma rate variation and a fraction of invariant sites. Starting from random trees, two runs of five million generations with four chains of MCMC were sampled every 1,000 generations. The final average s.d. of split frequencies was 0.005662. A 50% majority rule consensus tree was estimated from these samples using a burn-in of 1,000.

Glycopeptide production and detection. For starter cultures, fresh spores from 12 putative GPA producers were inoculated in 3 ml Streptomyces Vegetative Media (SVM) and incubated at 30 °C and 250 r.p.m. for 48 h. A 1% inoculum for 50 ml production media (Bennett's, SAM, R2YE or V40P) was used and further incubated for 8 d. The cell pellet was lysed using equal volume (w/v) of 1% NH₄OH solution, on a gyratory shaker at 4 °C for 1 h. The lysate was centrifuged at 13,000 r.p.m. for 20 min and the supernatant was concentrated under reduced pressure. The cell mass were re-extracted using 10 volumes of acetone following the same procedure, followed by concentration using a rotary evaporator. The cultures on solid media were extracted from agar plugs following the same procedure as for cell pellets. The culture supernatant and the extracts were tested for antimicrobial activity against *Escherichia coli* DH5α and *B. subtilis* by Kirby Bauer disc diffusion assay.

Isolation and purification of the GPAs. For isolation, 1% of SVM starter culture of WAC4169 and WAC1375 or WAC1420 and WAC4229 were inoculated in 1 liter SAM broth or 8 liter modified Bennett's agar (10 g potato starch, 2 g casamino acids, 1.8 g yeast extract, 2 ml Czapek's mineral mix, 5 g NaCl, 0.5 g K₂HPO₄, 15 g agar, 1 g L-proline, 1,000 ml water), respectively. The cultures were incubated for 1 week (WAC4169 and WAC1375) or 3 weeks (WAC1420 and WAC4229) at 30 °C followed by extraction using 1% NH₄OH solution. The filtrate was adjusted to pH 7.0 with 5 M HCl and evaporated under reduced pressure with 2% (W/V) HP-20 (Diaion) resin to give a residue Fra-E. This was loaded on a 1 liter HP-20 column eluting with H₂O (2 liters), 10% methanol (2 liters), 20% methanol (2 liters) and 40% methanol (2 liters) to yield four fractions Fra-E-1, Fra-E-2, Fra-E-3 and Fra-E-4. For WAC4169 extracts, Fra-E-2 was concentrated under reduced pressure and was purified using

semi-preparative high-performance liquid chromatography (HPLC; Atlantis, prep T3, 10 × 100 mm) with elution at 8% acetonitrile in H₂0 (0.1% FA) to give 10 mg of dimethylvancomycin. For WAC1420 and WAC4229 extracts, the fraction Fra-E-3 containing the molecule was further purified by passing through a Sephadex LH-20 column (100 ml), eluting with 30% methanol, to give 26 subfractions. Subfractions Fra-E-3-7 to Fra-E-3-13 were combined and purified by semi-preparative HPLC (Atlantis, prep T3, 10 × 100 mm), eluting with 12% acetonitrile in H₂O (0.1% FA), to give pure pekiskomycin (~2.8 mg). The WAC1375 Fra-E was partially purified by FLASH chromatography using C18 reversed-phase column. Each fraction with absorbance at 220 nm and 280 nm was evaporated under reduced pressure and further purified by affinity chromatography[32]. The resin for the column was prepared by linking *N,N*-diacetyl Lys-D-Ala-D-Ala to NHS-activated Sepharose 4 Fast Flow (GE Healthcare Life Sciences, USA) following manufacturer's instructions.

Geldanamycin production. WAC5038 was grown on SM (2% soy, 2% mannitol) agar media for 7 d. Agar plates were then extracted with acetone overnight. The extract was filtered to remove agar pellets and the acetone extract was dried down under reduced pressure. The resulting extract was resuspended in a mixture of acetonitrile-water (1:1) and analyzed using liquid chromatography-electrospray ionization-mass spectrometry (LC-ESI-MS) with an Agilent 1100 series LC system (Agilent Technologies Canada, Inc.) and a QTrap LC/MS/MS system (Applied Biosystems/MDS SCIEX).

Isothermal titration calorimetry. For isothermal titration calorimetry of pekiskomycin binding to D-Ala-D-Ala, various dilutions of *N,N*-diacetyl Lys-D-Ala-D-Ala tripeptide (50, 100, 150 and 600 µM) were prepared in sodium citrate buffer (pH 5.1) with 0.22% DMSO final concentration with vancomycin at 10 µM, and pekiskomycin at 10 µM and 50 µM. Both the antibiotic stock solutions were in 100% DMSO, all working solutions had a final DMSO concentration of 0.22%.

Titrations were performed in a Nano Isothermal Titration Calorimeter Low Volume (Nano ITC LV) with a gold reaction vessel.

The tripeptide, at a concentration of 100 µM, was loaded into the 50 µl titration syringe and 300 µl of the vancomycin at 10 µm was loaded into the cell. This was followed by experiments of 100 µM tripeptide titrated into 10 µM pekiskomycin and 600 µM tripeptide into 50 µM pekiskomycin. For all the experiments, the binding affinity (K), enthalpy (ΔH) and entropy (ΔS) for binding of the tripeptide to the GPAs were measured. In addition to the binding study, a background titration of tripeptide into buffer with 0.22% DMSO was also collected. A 60-s baseline was collected before the first injection was delivered and was equilibrated to baseline before starting of each assay.

Genome sequencing of WAC strains. Genomic DNA was submitted for sequencing on the Illumina MiSeq platform at the Farncombe Genomics Facility, McMaster University. For WAC1375, ~1 Gbp of sequence data was produced consisting of paired 150 bp reads. WAC4229 was sequenced using paired 250-bp reads totaling ~1.33 Gbp. Prior to assembly, reads were trimmed 3′ to Q30 using the FASTX-toolkit (version 0.0.13; http://hannonlab.cshl.edu/fastx_toolkit/index.html). WAC1375 and WAC4229 reads were corrected, assembled and scaffolded into contigs using Fermi[33]. For WAC1420, Genomic DNA was submitted for sequencing on the 454 FLX platform using Titanium chemistry at the Farncombe Metagenomics Facility, McMaster University. 292 Mbp of data was produced and assembled with MIRA using the quickswitch parameters '–job = genome,denovo,454,accurate' and setting the number of passes to 8 (ref. 34).

Identification of biosynthetic clusters. The assembled contigs for WAC1375, WAC1420 and WAC4229 were checked for the presence of GPAbiosynthetic clusters using the A47934 cluster (accession U82965) as a query in BLAST[35]. For WAC4229 three large overlapping contigs were manually merged to produce the cluster. The contig for each cluster was submitted to antiSMASH for further exploration of the GPA biosynthetic clusters[36].

HRMS and NMR experiments. High-resolution mass spectra (HRMS) were obtained using Thermo Fisher-XL-Orbitrap Hybrid mass spectrometer (Thermo Fisher, Bremen, Germany) equipped with electrospray interface operated in positive ion modes. 1D and 2D NMR experiments were done using Bruker AVIII 700 MHz instrument equipped with a cryoprobe in DMSO-*d₆*. Chemical shifts are reported in parts per million relative to tetramethylsilane using the residual solvent signals at 2.50 p.p.m. (¹H NMR) and 39.5 p.p.m. (¹³C NMR) as internal signals.

28. D'Costa, V.M., McGrann, K.M., Hughes, D.W. & Wright, G.D. Sampling the antibiotic resistome. *Science* **311**, 374–377 (2006).

29. Hayakawa, M. & Nonomura, H. Humic acid-vitamin agar, a new medium for the selective isolation of soil actinomycetes. *J. Ferment.Technol.* **65**, 501–509 (1987).

30. He, W., Wu, L., Gao, Q., Du, Y. & Wang, Y. Identification of AHBA biosynthetic genes related to geldanamycin biosynthesis in *Streptomyces hygroscopicus* 17997. *Curr. Microbiol.* **52**, 197–203 (2006).

31. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

32. Folena-Wasserman, G., Sitrin, R.D., Chapin, F. & Snader, K.M. Affinity chromatography of glycopeptide antibiotics. *J. Chromatogr. A* **392**, 225–238 (1987).

33. Li, H. Exploring single-sample SNP and INDEL calling with whole-genome de novo assembly. *Bioinformatics* **28**, 1838–1844 (2012).

34. Chevreux, B., Wetter, T. & Suhai, S. in *Computer Science and Biology: Proceedings of the German Conference on Bioinformatics*, Hannover, Germany, 4–6 October, pp. 45–56 (GCB, 1999).

35. Altschul, S.F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).

36. Medema, M.H. *et al.* antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* **39**, W339–W346 (2011).

doi:10.1038/nbt.2685