CRAGE enables rapid activation of biosynthetic gene clusters in undomesticated bacteria

Gaoyan Wang[1,10], Zhiying Zhao[1,10], Jing Ke[1,10], Yvonne Engel[2,10], Yi-Ming Shi[2,10], David Robinson[1], Kerem Bingol[3], Zheyun Zhang[1], Benjamin Bowen[1,4], Katherine Louie[1], Bing Wang[1], Robert Evans[1], Yu Miyamoto[1], Kelly Cheng[1], Suzanne Kosina[4], Markus De Raad[4], Leslie Silva[1], Alicia Luhrs[5], Andrea Lubbe[5], David W. Hoyt[3], Charles Francavilla[5], Hiroshi Otani[1,4], Samuel Deutsch[1,4,6], Nancy M. Washton[3], Edward M. Rubin[1], Nigel J. Mouncey[1,4], Axel Visel[1,4], Trent Northen[1,4], Jan-Fang Cheng[1,4], Helge B. Bode[1,7*] and Yasuo Yoshikuni[1,4,6,8,9*]

It is generally believed that exchange of secondary metabolite biosynthetic gene clusters (BGCs) among closely related bacteria is an important driver of BGC evolution and diversification. Applying this idea may help researchers efficiently connect many BGCs to their products and characterize the products' roles in various environments. However, existing genetic tools support only a small fraction of these efforts. Here, we present the development of chassis-independent recombinase-assisted genome engineering (CRAGE), which enables single-step integration of large, complex BGC constructs directly into the chromosomes of diverse bacteria with high accuracy and efficiency. To demonstrate the efficacy of CRAGE, we expressed three known and six previously identified but experimentally elusive non-ribosomal peptide synthetase (NRPS) and NRPS-polyketide synthase (PKS) hybrid BGCs from *Photorhabdus luminescens* in 25 diverse γ-Proteobacteria species. Successful activation of six BGCs identified 22 products for which diversity and yield were greater when the BGCs were expressed in strains closely related to the native strain than when they were expressed in either native or more distantly related strains. Activation of these BGCs demonstrates the feasibility of exploiting their underlying catalytic activity and plasticity, and provides evidence that systematic approaches based on CRAGE will be useful for discovering and identifying previously uncharacterized metabolites.

Microbial secondary metabolites (also called natural prod-ucts or specialized metabolites) play important roles in modulating biological systems and their host environments[1–3]. Their roles include intra- and inter-cellular communication[4], host defence[5] and pathogenicity[6]. Systematic discovery and characterization of secondary metabolites is therefore an essential step towards understanding diverse ecosystems and exploiting their chemical diversity for applications in health, food, agriculture and the environment[7,8]. Recent advances in computational tools (for example, antiSMASH[9] and PRISM[10]) have provided researchers with unprecedented opportunities to mine massive sequence-data spaces and to identify biosynthetic gene clusters (BGCs) responsible for producing these secondary metabolites, and therefore a possible blueprint for characterizing the biosynthesis of chemically diverse secondary metabolites.

These genome analyses revealed that individual bacterial strains may possess 20 to 50 BGCs[11,12]. However, around 90% of BGCs are not usually expressed under standard conditions, and the regulatory triggers required to activate them are typically unknown[13]. Although synthetic biology can, in principle, uncouple BGCs from their native regulatory constraints[8,14–18], current efforts have primarily focused on transcriptional and translational optimization in model chassis strains. However, successful activation of BGCs also requires that all translated products fold properly and undergo appropriate post-translational modifications, all substrates and co-factors be available, and all intermediates and products be tolerated by the chassis strain used. Significant advances in host strain engineering are needed to meet these complex requirements.

Evolutionary studies have often guided scientists in developing successful strategies for designing biological function[19,20]. Large-scale comparative genome studies are revealing that BGCs are probably evolved through horizontal gene transfer and subsequent modifications among closely related bacteria[21–26]. These BGC exchanges are thought to be beneficial because they offer opportunities to rapidly test the fitness effects of secondary metabolites[21]. This idea also implies that recipient bacteria could utilize the BGCs transferred from closely related bacteria more efficiently, as they may share physiological conditions suitable for BGC activation, but with greatly relaxed regulations[27]. Therefore, use of multiple closely related bacteria as chassis for BGC expression could, in theory, improve the efficiency of BGC activation. Although a lack of systematic studies leaves some uncertainty, the general conclusion from empirical studies in natural product research also supports this assumption[28,29].

[1]US Department of Energy Joint Genome Institute, Berkeley, CA, USA. [2]Molecular Biotechnology, Department of Biosciences and Buchmann Institute for Molecular Life Sciences, Goethe Universität Frankfurt, Frankfurt am Main, Germany. [3]Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA, USA. [4]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [5]Emery Pharma, Alameda, CA, USA. [6]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [7]LOEWE Centre for Translational Biodiversity Genomics, Frankfurt, Germany. [8]Center for Advanced Bioenergy and Bioproducts Innovation, Urbana, IL, USA. [9]Global Institution for Collaborative Research and Education, Hokkaido University, Hokkaido, Japan. [10]These authors contributed equally: G. Wang, Z. Zhao, J. Ke, Y. Engel, Y.-M. Shi. *e-mail: h.bode@bio.uni-frankfurt.de; yyoshikuni@lbl.gov

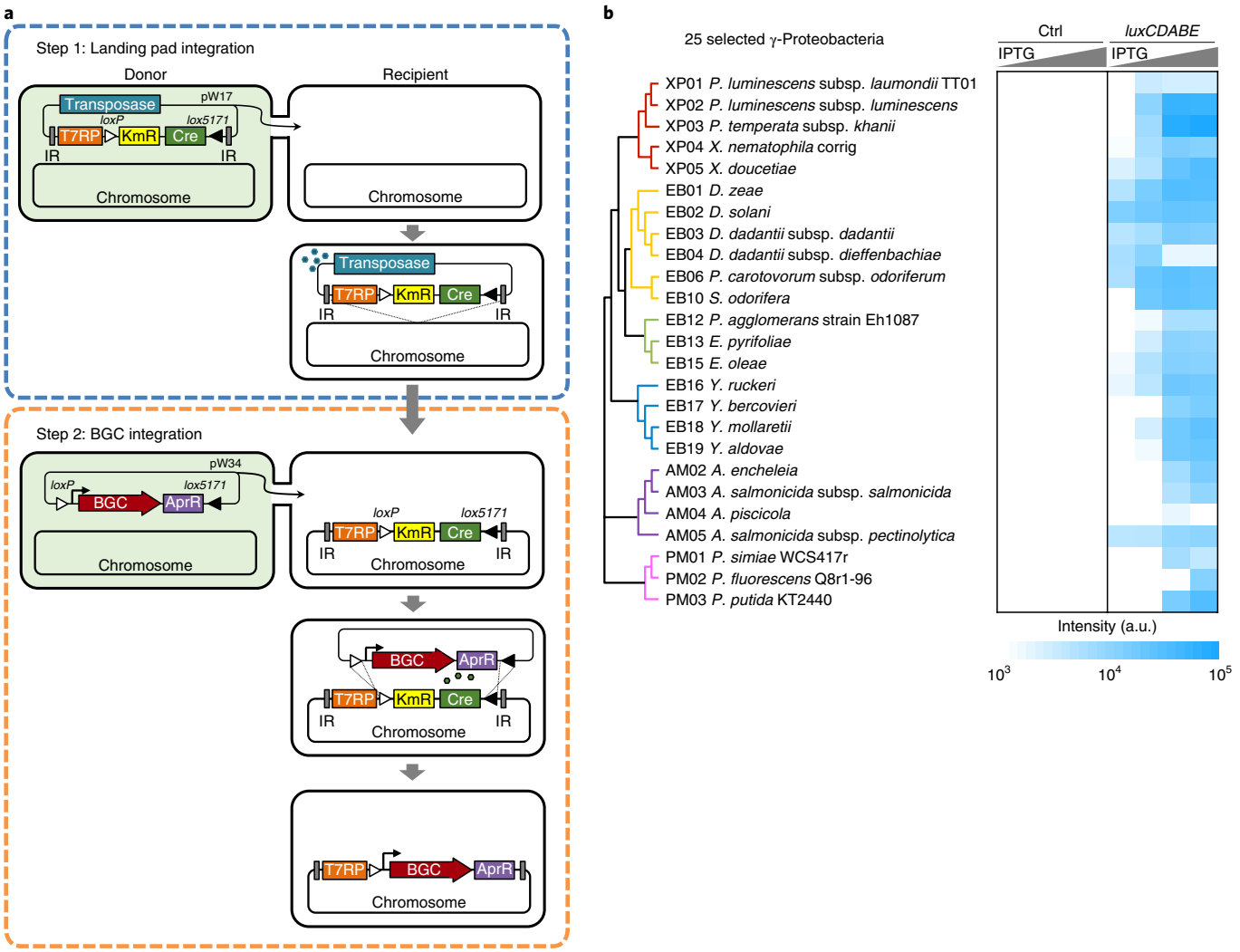

Fig. 1 | Chromosomal integration of BGCs mediated through CRAGE. a, Schematic for CRAGE, a genome engineering technology that allows complex biological systems to be implemented in a broad range of microbial strains. We primarily focused on formulating the design principle combining all existing technologies that have proven to work in a wide range of organisms. Step 1: A pW17 plasmid containing a mariner transposon and transposase was generated. The transposon contained a Cre recombinase gene and a kanamycin-resistant gene (KmR) flanked by two mutually exclusive lox sites (*loxP* and *lox5171*). In addition, a T7-RNA polymerase (T7RP) gene under the control of a *lacUV5* regulon was incorporated into the transposon. The pW17 plasmid was conjugated from donor *E. coli* strain BW29427 into the panel of recipient bacterial strains (Supplementary Tables 1 and 2), and the transposon was integrated into their chromosomes. Step 2: A different plasmid, pW34 (R6Kr ori) or pW5 (BAC-based), encoding a BGC under control of the T7 promoter and an apramycin-resistant gene (AprR) flanked by the two mutually exclusive lox sites, was conjugated into the recipient strain containing the LP. BGC integration into the chromosome of these chassis strains was mediated through Cre recombinase activity. **b**, For 25 chassis strains containing only an LP (control) and *luxCDABE*, luminescence activity was induced with four different IPTG concentrations (0, 0.01, 0.1 and 1 mM) and measured. All data were generated from biological triplicates. The standard deviations were generally less than 10%. The colours used for **b** are coordinated to represent strains classified in the same phylogenetic branches. a.u., arbitrary units.

To enable the multi-chassis approach, genome engineering tools such as phage integrases[30–33] can be useful. However, the utility of these tools is limited to a small group of bacteria (Supplementary Fig. 1). Additionally, integration efficiency significantly declines as insert size increases[32]. To address these issues, we attempted to develop chassis-independent recombinase-assisted genome engineering (CRAGE) applicable to diverse bacterial species across multiple phyla (Fig. 1a). CRAGE is based on recombinase-assisted genome engineering (RAGE) technology[34,35]. Use of RAGE has allowed single-step integration of pathways comprising 48 kb directly into the *Escherichia coli* chromosome without compromising integration efficiency. After simple antibiotic counter-selection, the integration yield reached 100%. CRAGE extends RAGE by enabling researchers to domesticate previously undomesticated microbes, and substantially increases the chance of successful BGC expression and discovery of previously uncharacterized secondary metabolites.

Results

Developing a design principle for CRAGE. In RAGE, a landing pad (LP) containing two mutually exclusive *lox* sites is first integrated into a chromosome. The DNA constructs flanked by these *lox* sites are then integrated into the LP, catalysed by Cre recombinase. We modified RAGE[34,35] in three major ways to establish the CRAGE design principle (Fig. 1a). First, we used transposon systems[36,37] (mariner system for Proteobacteria and Tn5 system for

Table 1 | BGCs used in this study

BGC	Gene(s)	Genomic location	Size (kb)	Products	Refs.
1	*plu0897–plu0899*	1021652–1035887	14.2	Known (**7**,**8**), Unknown (**9**,**10**)	50,57
2	*plu1113–plu1115*	1277282–1299002	21.7	Unknown	
3	*plu1210–plu1222*	1392054–1414406	22.4	Unknown	
4	*plu1881–plu1877*	2239952–2221617	18.3	Known (**1**–**3**)	18,53
5	*plu2316–plu2325*	2713457–2744350	30.9	Unknown (**19**–**21**)	
6	*plu2670*	3173674–3124571	49.1	Known (kolossins A–C)	58
7	*plu3123*	3662492–3646119	16.4	Unknown (**4**–**6**)	
8	*plu3130*	3688463–3678528	9.9	Unknown (**22**)	
9	*plu3263*	3880777–3865127	15.6	Known (**11**–**18**)	18,54–56
10	*plu3526–plu3538*	4167664–4122969	44.7	Unknown	

Actinobacteria) to insert the LP into recipient bacteria chromosomes. Because transposon systems are commonly used in engineering of diverse species ranging from prokaryotes to eukaryotes, they are suitable for the first step of domestication. Although the transposon is randomly integrated into the chromosome of the recipient strain, we can screen and select the transformants with the LP integrated into the location minimally affecting the host strain's physiology.

Second, we used a Lac-T7 expression system[38] to control the expression of BGCs. This system is orthogonal to native transcription; genes under the control of a T7 promoter are not transcribed unless a T7 RNA polymerase (T7RNAP) is present. Using this system, we can minimize the expression of BGCs whose products may be toxic to *E. coli* while we are assembling the BGC constructs. Additionally, although codon usage and ribosome binding sites may need to be redesigned for each chassis strain to obtain more optimal results, this design principle in general allows a high degree of flexibility, so that any single construct in any CRAGE strain can be expressed without re-cloning as long as the T7RNAP is expressed under the control of promoters that function in the recipient strains.

Several studies suggest that genomic integration location can also affect the expression of integrated genes[34,39–41]. Therefore, investigating different integration locations is another viable approach for exploring the effect of different expression levels on BGC activity. However, our previous study, as well as others, suggested that the effect of different integration locations on enzyme and pathway activity was generally small (at most less than about seven- to eightfold, usually two- to threefold for enzyme activity and less than two fold for pathway activity)[34,39–41]. In contrast, the Lac-T7 system allows us to explore a much wider range of pathway activity (10- to 1,000-fold) than we could if we explored according to integration location. Additionally, the approach of investigating different integration locations would complicate our workflow and make our approach less attractive. Therefore, we specifically chose to use the Lac-T7 system to explore the effect of different expression ranges on BGC activity.

Third, we incorporated the origin of transfer (oriT) into both the LP and BGCs carrying plasmids to use conjugation as a primary transformation method. Conjugation systems have been used for transformation of a wide spectrum of bacterial species including those in the Proteobacteria, Actinobacteria, Firmicutes, Bacteroidetes and Cyanobacteria phyla[42–44]. Although the present study focuses on γ-Proteobacteria, our preliminary results demonstrate that CRAGE can engineer bacteria across multiple phyla (for example, α-Proteobacteria, β-Proteobacteria and Actinobacteria) (Supplementary Figs. 1–3).

Selecting model biosynthetic gene clusters. Members of the genus *Photorhabdus*, as well as the related genus *Xenorhabdus*, are endosymbionts of soil-borne nematodes and are known to have potent bioactivity toward a wide range of insects and insect larvae; some of these entomopathogenic complexes are used as biological insecticides in agriculture[45,46]. These bacteria also produce numerous secondary metabolites to inhibit the growth of competing microbes within their hosts[47,48]. Genomes of ~50 *Photorhabdus* and *Xenorhabdus* species have been sequenced and are accessible in public databases. Computational analyses suggest that each of these genomes contains ~20 to 50 putative BGCs[11,12], and that many BGCs divergently evolved among species within these two genera[49,50]. The combination of extensive genomic resources and rich metabolic potential makes these species ideal test cases for a purpose-engineered multi-chassis strategy for BGC characterization.

We selected a model set of 10 NRPS and NRPS–PKS hybrid BGCs from *Photorhabdus luminescens* subsp. *laumondii* TTO1 (Table 1, Fig. 2 and Supplementary Fig. 4)[18,51]. Except for BGC6, these BGCs had been previously cloned and heterologously expressed in *E. coli*. However, only two of those nine BGCs had been successfully activated, making this study an ideal benchmark to test the efficacy of the multi-chassis approach mediated by CRAGE. The selected panel of pathways included four BGCs as controls that had been previously studied successfully by single-chassis approaches (BGCs 4 and 9)[18,52–56] or by promoter replacement approaches in a native strain (BGCs 1 and 6)[50,57,58], as well as six putative BGCs that were not functional using conventional chassis approaches (BGCs 2, 3, 5, 7, 8 and 10)[18].

Preparing phylogenetically diverse chassis strain panels using CRAGE. We selected 31 species of γ-Proteobacteria representing 10 different genera (Fig. 1, Supplementary Fig. 4 and Supplementary Tables 1 and 2). The panel consisted of several *Xenorhabdus* and *Photorhabdus* species (XPs) and many other species that are evolutionarily slightly distant from XPs (other Enterobacteria (EB) and bacteria in the *Aeromonas* (AM) and *Pseudomonas* (PM) genera). The panel also allowed us to systematically investigate correlations between evolutionary relatedness and physiological compatibilities of different chassis:BGC combinations. Selection criteria also included the availability of complete or draft genome sequences and classification as biosafety level 1. Furthermore, all the selected strains had EntD- and/or Sfp-type phosphopantetheinyl transferases (PPTases), required for modifying NRPS and PKS proteins to convert their inactive apo forms into the enzymatically functional holo forms[59].

The LP on a transposon was first randomly integrated into the chromosome of each strain and the integration site was determined

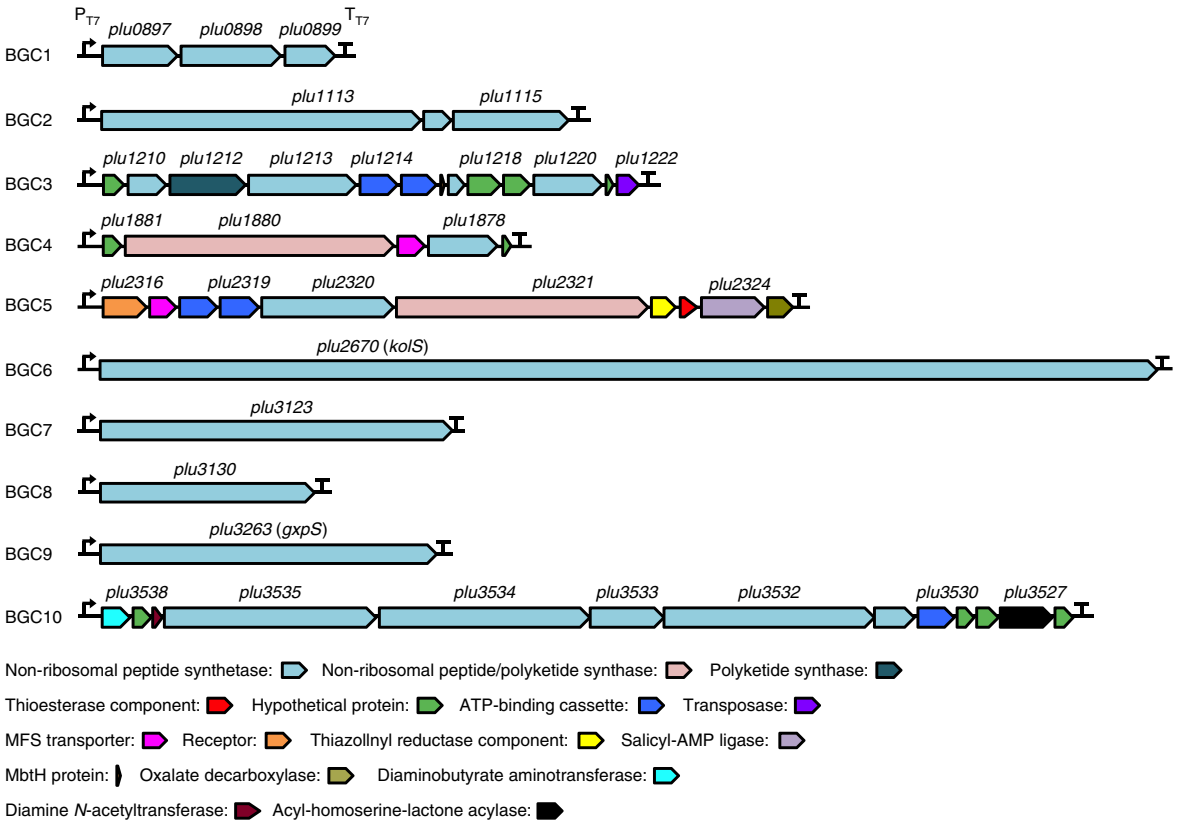

Fig. 2 | Design and architecture of BGC constructs. The design and gene architecture of each BGC construct. P_{T7} and T_{T7} correspond to a T7 promoter and a T7 terminator, respectively.

by whole genome sequencing. This purpose-engineering procedure was successful in 27 species representing all 10 genera selected. To demonstrate the general potential of this collection of chassis strains to efficiently integrate and express heterologous pathways, we selected a 7 kbp bacterial luciferase (*lux*) operon composed of five genes (*luxCDABE*) originally derived from *P. luminescens*[60]. The operon, flanked by mutually exclusive *lox* sites, was conjugated into the panel of recipient strains and stably integrated into the chromosome. Following antibiotic counter-selection, successful integration and operon activity were assessed by quantifying luciferase activity. In 25 species from all 10 genera, chromosomal integration was successful and luciferase activity was inducible with isopropyl β-D-1-thiogalactopyranoside (IPTG; Fig. 1b). Importantly, antibiotic counter-selection resulted in an integration efficiency of 100% (Supplementary Table 2).

For several strains, we obtained two to three variants with the LP inserted into different genomic locations. To investigate the effect of different integration locations on BGC activity, we integrated the *lux* operon into these strain variants. As expected from our previous study and others[34,39–41], we found that integration location had little effect on luminescence activity (Supplementary Fig. 5). Indeed, our strategy of using a Lac-T7 expression system covered a broader range of luminescence activity. Therefore, we decided not to consider the effect of integration location on BGC activity further.

Genomic integration of BGCs across diverse chassis strains. To introduce the selected NRPS and NRPS–PKS hybrid BGCs (Table 1, Fig. 2 and Supplementary Fig. 4) into the phylogenetically diverse set of chassis strains, BGC constructs were assembled. This process was successful for 9 of 10 BGCs. We were unable to clone BGC6 because it has extensive internal repeats reflected by its overall NRPS organization, which hampered homologous recombination in yeast. BGC6 was also not cloned in a previous study conducted by Fu and others[18]. The remaining nine BGCs were subsequently transferred to the panel of recipient strains via conjugation. This process was successful for almost all chassis:BGC combinations in 24 of 25 chassis strains, with an average integration efficiency of 57% as measured by simple antibiotics counter-selection and subsequent colony PCR to confirm successful BGC integration (Supplementary Tables 3 and 4). As expected, even the largest clusters tested were integrated without compromising the success rate. Across the BGC size range evaluated in this series (10–48 kbp), integration efficiency remained nearly identical, demonstrating the efficacy of CRAGE technology.

Using the purpose-engineered multi-chassis approach to rapidly identify known products. We cultivated the panels of 24 chassis strains harbouring each BGC at four different IPTG concentrations. We then used liquid chromatography–high resolution mass spectrometry (LC-HRMS) to measure secondary metabolites produced by each chassis:BGC combination. Controls included a conventional heterologous expression system, the *E. coli* BL21(DE3) strain with a genome-integrated *Bacillus subtilis* PPTase (*sfp*)[61] (Supplementary Fig. 4).

Targeted analysis of LC-HRMS data for chassis strains expressing BGCs 1, 4 and 9 revealed successful production of every previously described metabolite in at least one of the chassis strains examined here[18,50,52–57] (Fig. 3 and Supplementary Figs. 6–8). For example, BGC4 comprises five genes (*plu1877–plu1881*) encoding NRPS and NRPS–PKS hybrid (Fig. 3a). Previous studies have implicated the expression of BGC4 in the production of luminmycin A (**1**), glidobactin A (**2**) and cepafungin I (**3**), all of which show

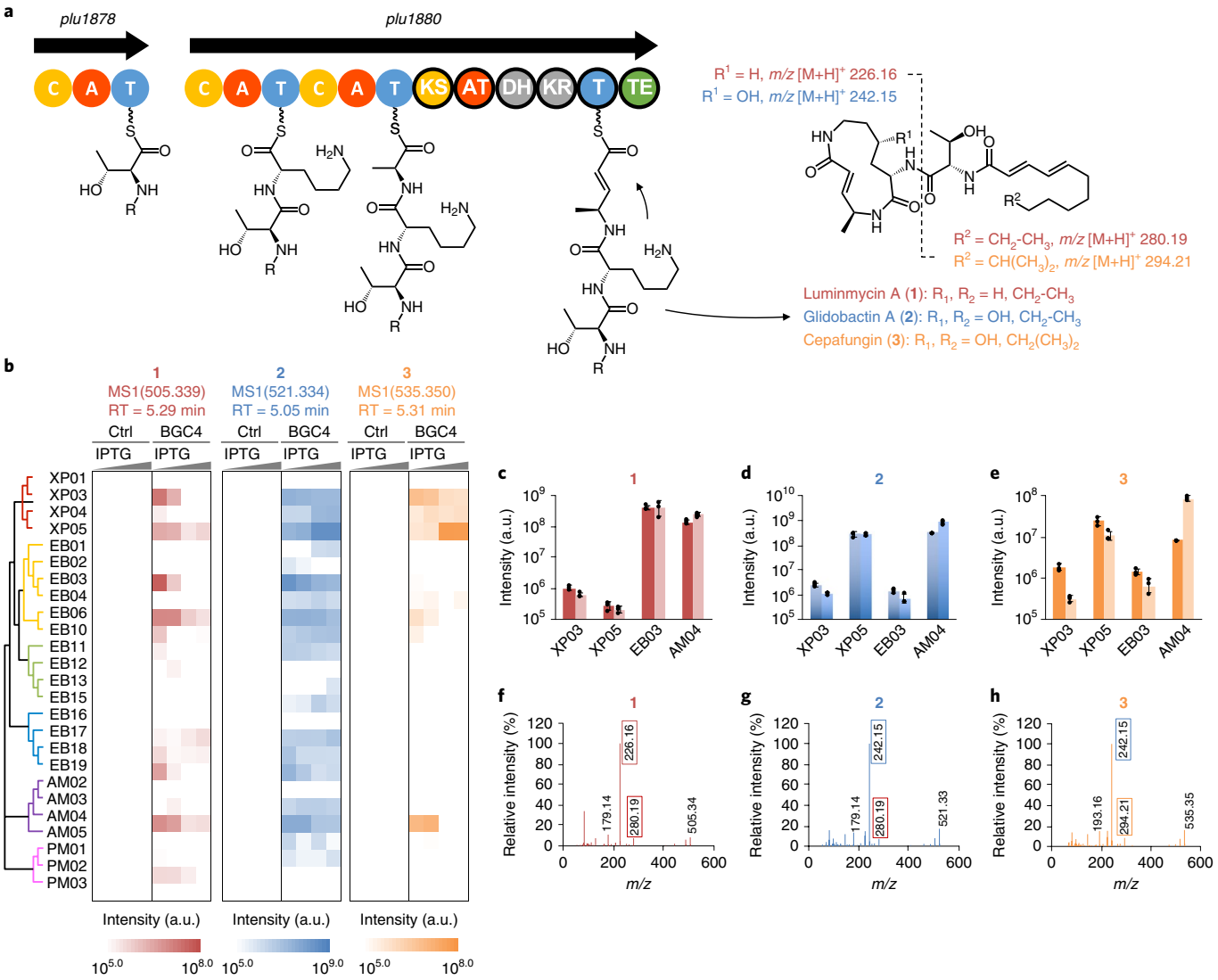

Fig. 3 | Expression of BGC4 in a panel of phylogenetically diverse chassis strains. a, A proposed scheme of secondary metabolite biosynthesis catalysed by BGC4 (*plu1877–plu1881*). NRPS and PKS domains are shown as circles (PKS modules have black borders); letters represent condensation (C), adenylation (A), thiolation (T), ketosynthase (KS), acyltransferase (AT), dehydratase (DH), ketoreductase (KR) and thioesterase (TE) domains. Signature fragment ions are also shown. **b**, Measured MS1 intensity for luminmycin A (**1**) (*m/z* 505.341 $[M+H]^+$), glidobactin A (**2**) (*m/z* 521.335 $[M+H]^+$) and cepafungin I (**3**) (*m/z* 535.350 $[M+H]^+$) produced from the panel of phylogenetically diverse chassis strains containing only an LP (control) and BGC4 induced with four different IPTG concentrations (0, 0.01, 0.1 and 1.0 mM; $n=1$ biologically independent experiment). The secondary metabolites were extracted from cell pellets and analysed by LC-HRMS. Products **1**–**3** were produced from multiple strains expressing BGC4, but not from the controls. RT, retention time of each metabolite on LC-HRMS analysis. **c**–**e**, For the top four producers (XP03, XP05, EB03 and AM04 induced with 0, 1.0, 0 and 0.01 mM of IPTG, respectively), the experiments were repeated. Production of **1** (**c**), **2** (**d**) and **3** (**e**) was reproduced. In these figures, dark and light colours represent measured MS1 intensity for each metabolite extracted from the cell pellet and freeze-dried supernatant samples, respectively. Error bars represent s.d.; $n=3$ biologically independent experiments (three BGC transformants). **f**–**h**, MS2 spectra for **1** (**f**), **2** (**g**) and **3** (**h**). The signature fragment ions are also marked.

strong cytotoxicity, proteasome inhibition and anti-fungal activity[18,52,53,62,63]. Although **1** and **2** have been produced by heterologous expression of BGC4 in model chassis strains[18,52,53], **3** has only been detected upon release of wild-type *P. luminescens* into the haemocoel of insect larvae, and the molecular triggers for in vitro production in native *P. luminescens* are unknown[62].

Expression of BGC4 across the panel of CRAGE strains yielded **1**–**3** (Fig. 3b–h). Consistent with a previous report[18], BGC4 was inactive in its native strain (Fig. 3b), suggesting strict regulatory control of this potent bioactive pathway. Homologous expression of BGC4 in XP01 injected into wax moth larvae, where native regulations are probably desensitized, resulted in production of **1** and **2** (Supplementary Fig. 9). In contrast, most heterologous strains expressing BGC4 produced at least one of **1**–**3** while they were cultured in synthetic media, indicating that this strict repression is not conserved in heterologous chassis strains, not even those closest to *P. luminescens* (for example, XP03–05).

Products **1**–**3** were also produced at different ratios depending on the strain, indicating that product specificity is determined by the background physiology of each strain. Remarkably, extended quantification of products in the four strongest producers (XP03, XP05, EB03 and AM04) revealed that the ratio of **2** to **1** and **3** to **1** production shifted by more than five and four orders of magnitude, respectively, among these strains (Fig. 3c–e). These results

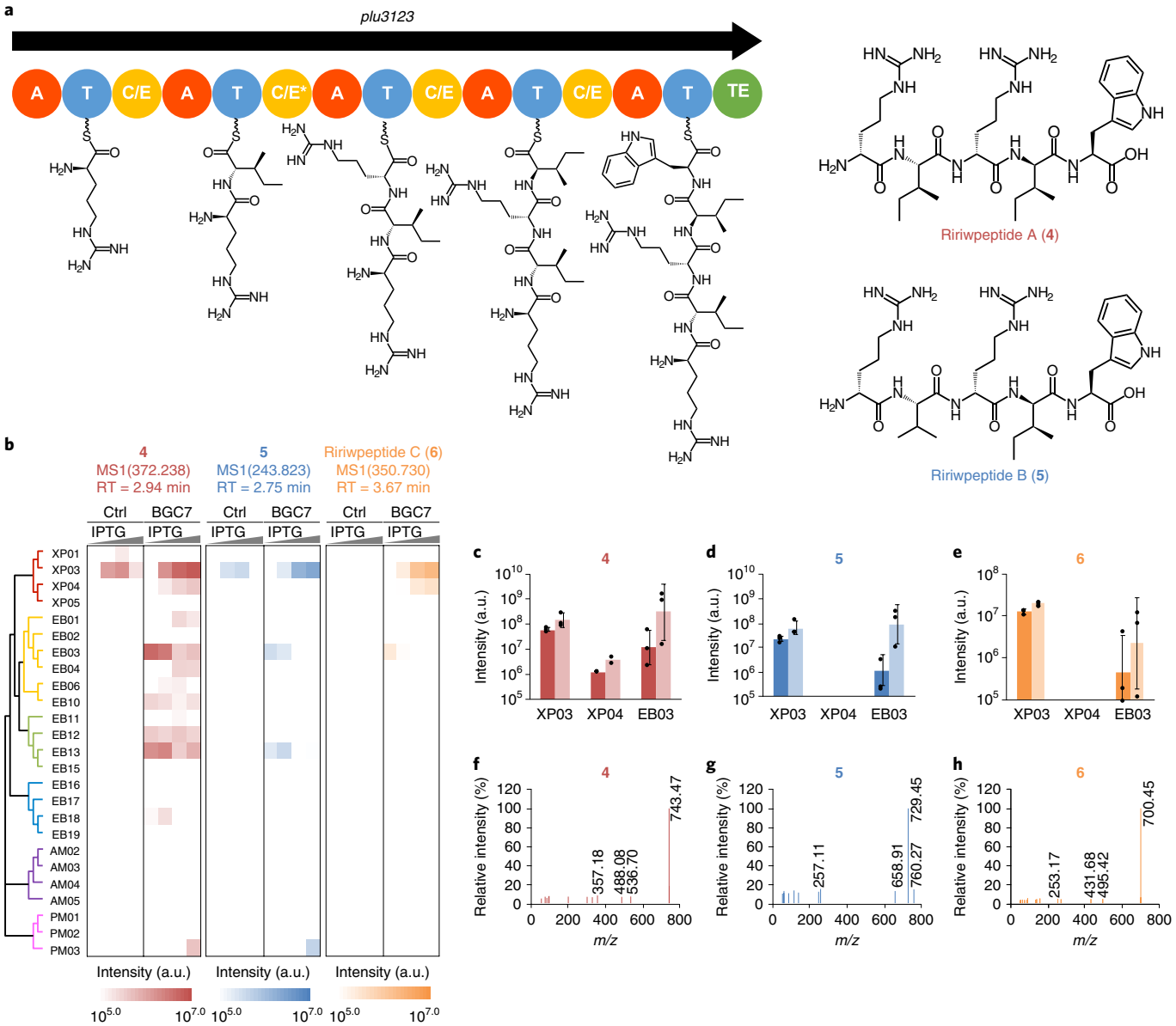

Fig. 4 | Expression of BGC7 in a panel of phylogenetically diverse chassis strains. a, A proposed scheme of secondary metabolite biosynthesis catalysed by BGC7 (*plu3213*). NRPS domains are shown as circles; letters represent adenylation (A), thiolation (T), dual condensation and epimerization (C/E) and thioesterase (TE) domains. C/E* indicates dual condensation and epimerization domains with non-functional epimerization domain. The structure of the main ririwpeptide A (**4**) was confirmed by synthesis and labelling experiments (Supplementary Figs. 11–15). **b**, Measured MS1 intensity for **4** (*m/z* 372.238 $[M+2H]^{2+}$), ririwpeptide B (**5**) (*m/z* 243.238 $[M+3H]^{3+}$) and ririwpeptide C (**6**) (*m/z* 350.730 $[M+2H]^{2+}$) produced from the panel of phylogenetically diverse chassis strains containing only an LP (control) and BGC7 induced with four different IPTG concentrations (0, 0.01, 0.1 and 1.0 mM) ($n=1$ biologically independent sample). The secondary metabolites were extracted from cell pellets and analysed by LC-HRMS. Products **4**–**6** were produced from multiple strains expressing BGC7, but not from the controls except for XP01 and XP03. **c**–**e**, For the top three producers (XP03, XP04 and EB03 induced with 1.0, 1.0 and 0 mM of IPTG, respectively), the experiments were repeated. Production of **4** (**c**), **5** (**d**) and **6** (**e**) was reproduced. In these figures, dark and light colours represent measured MS1 intensity for each metabolite extracted from the cell pellet and freeze-dried supernatant samples, respectively. Error bars represent s.d.; $n=3$ biologically independent experiments (three BGC transformants). **f**–**h**, MS2 spectra for **4** (*m/z* 743.464 $[M+H]^+$) (**f**), **5** (*m/z* 729.454 $[M+H]^+$) (**g**) and **6** (*m/z* 700.451 $[M+H]^+$) (**h**).

suggest that XP05 had strengthened monooxygenase activity for conversion of **1** to **2** and increased iso-fatty acids in its fatty acid pool. Additionally, **2** production titre in culture reached as high as 177 mg l^{-1}, which is more than 700-fold higher than the previously reported titre of **1** in *E. coli* where BGC4 was heterologously expressed[53]. Similar observations for BGCs 1 and 9 are reported in the Supplementary Results. The robust retrieval of all previously described metabolites for these pathways in a single experimental series demonstrates the power of the purpose-engineered multi-chassis approach enabled by CRAGE.

Using a purpose-engineered multi-chassis approach to reveal previously uncharacterized secondary metabolites. We evaluated the potential of CRAGE strains for discovering hitherto

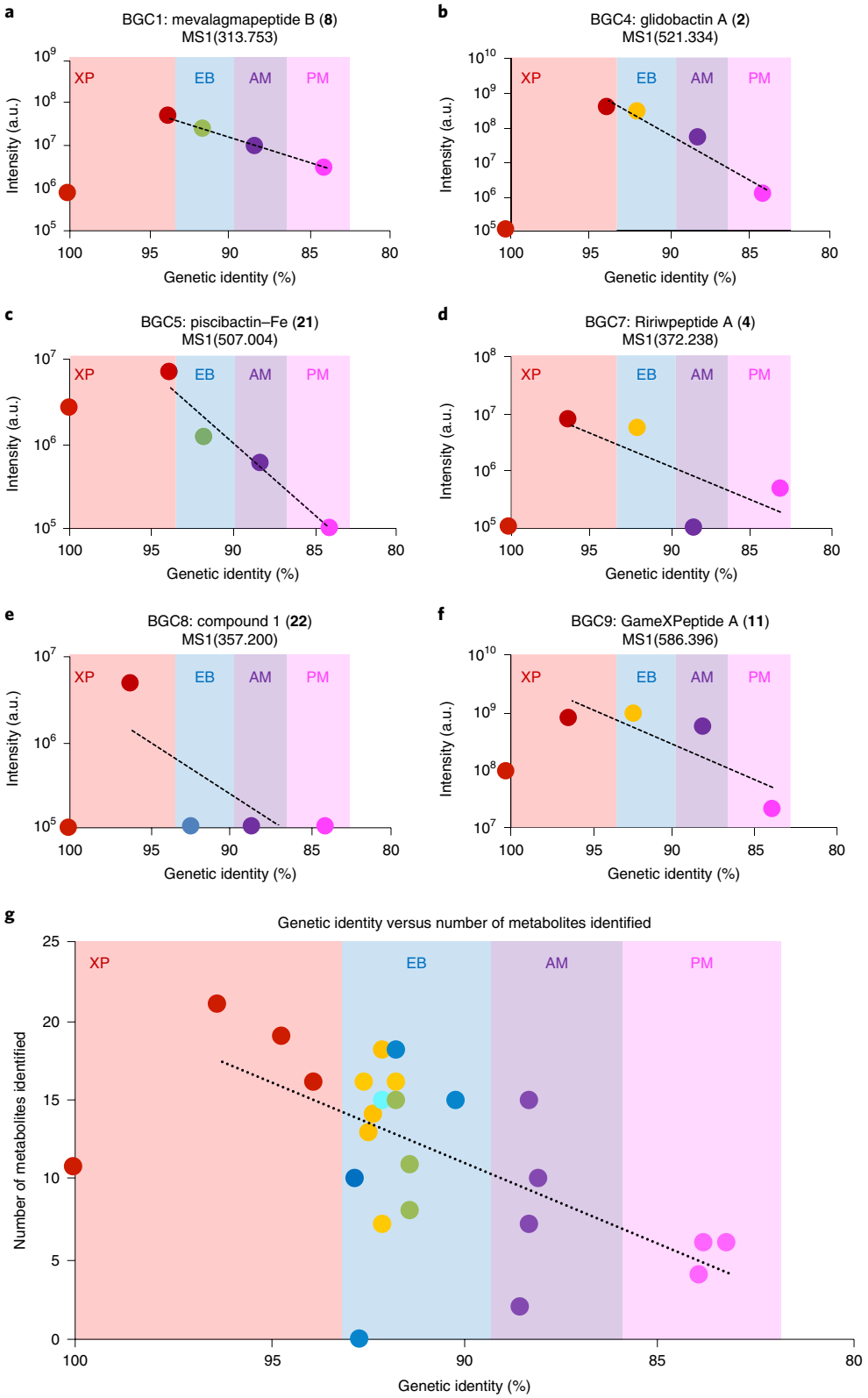

Fig. 5 | Relationship between secondary metabolite production and genetic identity for each representative chassis strain and *P. luminescens* subsp. *laumondii* TTO1. a–**f**, Maximum attained production titres of representative secondary metabolites from a strain in each group—mevalagmapeptide B (**8**) (*m/z* 313.753 $[M+2H]^{2+}$) from BGC1 (**a**), glidobactin A (**2**) (*m/z* 521.334 $[M+H]^{+}$) from BGC4 (**b**), piscibactin–Fe complex (**21**) (*m/z* 507.004 $[M+H]^{+}$) from BGC5 (**c**), ririwpeptide A (**4**) from BGC7 (*m/z* 372.238 $[M+2H]^{2+}$) (**d**), *m/z* 357.200 $[M+H]^{+}$ from BGC8 (**e**) and GameXPeptide A (**11**) (*m/z* 586.396 $[M+H]^{+}$) from BGC9 (**f**)—are plotted against the genetic identity for *P. luminescens* subsp. *laumondii* TTO1 and the strain. **g**, Number of metabolites produced from each strain, plotted against the genetic identity between *P. luminescens* subsp. *laumondii* TTO1 and the strain. Genetic identity represents the identity of 16S rRNA sequences between *P. luminescens* subsp. *laumondii* TTO1 and each strain. Each dot represents a strain and is colour-coordinated the same as in Fig. 1. The red dot at genetic identity 100 represents XP01 (*P. luminescens* subsp. *laumondii* TTO1).

uncharacterized *P. luminescens* secondary metabolites. Untargeted metabolite searches by LC-HRMS tend to be challenging and yield high rates of false-positive findings because of the large number of compounds and their adducts present in cell culture extracts[64,65]. However, when the same pathway is expressed simultaneously in multiple heterologous hosts, we can use independent occurrence of the same feature under different physiological conditions to narrow down the search space considerably and identify compounds reproducibly associated with the same pathway (Supplementary Fig. 10). This concept is synergistic to the fungal artificial chromosomes and metabolic scoring, discussed by Clevenger and others[66], where FAC score is defined as the likelihood of each feature to be uniquely associated with each FAC containing putative BGCs. Clevenger et al.[66] analysed 56 extracts of *Aspergillus nidulans* containing different FACs to filter out false positive features and successdully identified 15 previously uncharacterized secondary metabolites. The multi-chassis approach can further provide another criterion to filter out false positives and facilitate the untargeted analysis to identify unique features associated with each BGC.

First, we used an untargeted approach to identify possible additional metabolites produced by BGCs 1, 4 and 9. This analysis re-identified all 13 previously characterized products. In addition, untargeted analysis also identified two additional previously uncharacterized products from BGC1, which we named mevalagmapeptides C (**9**) and D (**10**) (Supplementary Fig. 6). The LC-HRMS analysis suggests that while these products share a core amino-acid structure with mevalagmapeptides A (**7**) and B (**8**), they carry a terminal cadaverine instead of the usual agmatine and putrescine modifications. **10** is also extended by an additional *N*-methyl valine unit, suggesting the catalytic activity of BGC1 might have changed when it was expressed in some strains, for reasons unknown but in accordance with the flexible biosynthesis of this natural product class[50,67]. Of note, an identical untargeted LC-HRMS analysis of extracts from *E. coli* alone confirmed only 3 of 13 previously characterized products, and none of the previously uncharacterized compounds (Supplementary Tables 5–7).

Next, we applied the same untargeted metabolite analysis approach to BGCs 2, 3, 5, 7, 8 and 10, whose products had not been previously characterized. This analysis identified six unique metabolites from strains expressing BGCs 5, 7 or 8 (Fig. 4 and Supplementary Figs. 11–26). For example, we discovered three previously uncharacterized peptides (*m*/*z* 372.238 $[M+2H]^{2+}$, *m*/*z* 243.238 $[M+3H]^{3+}$ and *m*/*z* 350.730 $[M+2H]^{2+}$), which we named ririwpeptides A–C (**4**–**6**) according to their peptide sequence produced by BGC7 (Fig. 4, Supplementary Figs. 11–19 and Supplementary Table 8). A detailed MS/MS analysis of **4** (*m*/*z* 372.238 $[M+2H]^{2+}$) and **5** (*m*/*z* 243.238 $[M+3H]^{3+}$) from BGC7 could confirm the composition and order (Supplementary Figs. 11–13) of amino acids, as suggested by the NRPS specificity of the adenylation domains as R-I-R-I-W and R-I-R-V-W. The presence of dual condensation/epimerization domains (Fig. 4a) suggested that all amino acids except the terminal Trp have D-configuration. However, chemical synthesis of a library of derivatives and co-injection of synthetic and natural derivatives (Supplementary Figs. 14 and 15) revealed that the E domain in the second C/E domain responsible for the epimerization of the first Ile is inactive, leading to this Ile appearing in the final peptide in L-configuration. The structure of **6** (*m*/*z* 350.730 $[M+2H]^{2+}$) could not be solved because of low production titres, but the HRMS data suggest that an Arg is exchanged with an Ile, which is also suggested by the shift in retention time. Metabolites were not produced when BGC7 was expressed in the native strain, suggesting complex regulatory controls over these pathways (Fig. 4b).

The results for expression of BGCs 5 and 8 are reported in the Supplementary Results. Together, these results demonstrate that the purpose-engineered multi-chassis approach using CRAGE is an effective strategy for activating BGCs and identifying previously uncharacterized metabolites produced from these pathways. The methods used to characterize these metabolites are summarized in Supplementary Table 11. For each BGC, we also examined secondary metabolite production in both the cell pellet and supernatant samples from three to five top producers, looking for trends (Figs. 3 and 4, and Supplementary Figs. 6, 7, 20 and 26). These results are summarized in the Supplementary Results.

Examining the correlation between evolutionary relation and physiological compatibility of diverse chassis:BGC combinations. The extensive matrix of BGCs and phylogenetically diverse hosts allowed us to examine the correlations between evolutionary distance and physiological compatibility of chassis:BGC combinations. Experimental work performed to understand such relationships can give general guidance for choosing strains to use to characterize a given BGC. Because pathway efficiency could have been significantly compromised if even one reaction step were suboptimal, we compared the maximum attained production titre from a representative strain in each group with the strain's genetic similarity to XP01 (Fig. 5 and Supplementary Fig. 28). Although we also explored the codon bias and GC content of each strain as alternatives, these parameters failed to provide high enough resolution between strains closely related and distantly related to XP01 (Supplementary Fig. 29). In addition, we compared the number of products from each chassis strain with the strain's genetic similarity to XP01. Because we can at least detect the products even if their production levels are suboptimal, we included all data points for this analysis. These results suggested some remarkable trends, as described in the following.

First, the native BGC strain (XP01) was, in general, a poor secondary metabolite producer compared with other XP strains. Even after homologous expression of each BGC in XP01, production titres of secondary metabolites did not improve, except for that of BGC5. This might be because tight regulatory controls at other than the transcription level are probably present in XP01. Alternatively, the T7 expression platform might not work well in XP01, as XP01 also showed poor luminescence activity (Fig. 1b). To test these ideas, we injected XP01 harbouring an extra copy of each BGC into wax moth larvae and measured the production titre of each metabolite[68]. Because XPs are entomopathogens, they can produce secondary metabolites when insect larvae are infected with them. In almost all cases, homologous expression of BGCs yielded 10-fold better production titres (Supplementary Fig. 9) than their control, suggesting that while transcriptional refactoring was working, there also exist other layers of regulatory mechanisms controlling secondary metabolite production in XP01.

Second, expression of BGCs in XPs (other than XP01) generally resulted in both substantially higher product yield and substantially higher number of products than did expression in the other strains. This result suggests that, with slightly increasing evolutionary distance from the native host, relaxation of regulatory restrictions tends to precede the enhanced physiological compatibility of chassis:BGC combinations. Both production yield and diversity of secondary metabolites dramatically decreased with further increased phylogenetic distance between chassis (EBs, AMs and PMs) and the native BGC strain. We interpreted these observations to mean that with increasing evolutionary distance from the XP family, the existence of optimal (that is, the native BGC strain's) physiological conditions to support functional BGC expression and/or precursors required for secondary metabolite production declines.

Third, although most distantly related strains were not as adaptable as closely related strains, some distantly related strains were highly adaptable. For example, *Aeromonas piscicola* (AM04) activated four BGCs and produced 14 different products, while other AMs did not perform as well. Some species of Enterobacteria such as *Dickeya dadantii* subsp. *dadantii* (EB03), *Pantoea agglomerans*

(EB12), *Erwinia oleae* (EB15) and *Yersinia aldovae* (EB19) activated at least one BGC and produced metabolites to one of the highest levels among the suite of chassis strains. Additionally, some of the chassis strains, *Serratia odorifera* (EB10), EB12, *Yersinia mollaretii* (EB18) and EB19, produced a previously uncharacterized rhabdopeptide/xenortide peptide (**10**). Although the exact physiological features distinguishing versatile highly adaptive strains from other strains remain unknown, these physiological traits may be useful in further expanding chemical innovations.

Discussion

Host selection is one of the important design principles for successful heterologous expression. However, domestication of non-model strains is not the primary strategic choice in synthetic biology unless there is a specific interest, because this process is often extremely laborious and time-consuming. To standardize this effort, we developed CRAGE, a powerful and versatile strain engineering technology. We demonstrated that using CRAGE enabled the purpose-engineered multi-chassis approach and greatly facilitated function characterization of BGCs. Systematic characterization of diverse BGC:chassis combinations provided evidence that BGC expression may be more successful if the panel of chassis strains is selected from strains closely related to the native BGC strain; these strains are likely to share similar physiology, but have much relaxed native regulations. A similar conclusion may also be inferred from evolutionary studies, which have shown that horizontal gene transfer events occur significantly more frequently among closely related strains than among distantly related strains[23–26].

Including our preliminary study, CRAGE technology has been extended to 30 species, including α-, β- and γ-Proteobacteria and a few species of Actinobacteria. However, we found that some strains (for example, *Pectobacterium* species) were resistant to CRAGE for reasons unknown (Supplementary Table 2). We are also testing CRAGE for four model *Streptomyces* species but have not yet succeeded, partly because we are not experienced with these microbes. Because both transposase and Cre work in many organisms including those[34,69–71], we believe CRAGE can be extended to them with careful adaptation. For example, we may need to optimize codons and refactor transposase and Cre. Each strain group may require unique transformation protocols. For some strains, circumventing host restriction barriers is extremely important. Using a donor *E. coli* strain that expresses a specific methylase improves transformation efficiency for these strains[72]. Additionally, we observed dramatic reduction of lux activity in strains outside γ-Proteobacteria (Supplementary Figs. 2 and 3), suggesting that our expression strategy needs to be carefully tailored for each strain group.

Although CRAGE needs further improvement to extend its utility, the efficiency, accuracy and speed of CRAGE pave the way for an advanced design paradigm for stable and optimal integration of complex pathways into purpose-engineered diverse non-model bacteria. Although CRAGE may have the potential to provide exciting opportunities to many scientists, we are also aware of its risks and the responsibility to consider ethical, legal, safety and environmental aspects of this technology. The responsibilities that accompany use of this innovation are discussed in the Supplementary perspectives.

Methods

Strains. *E. coli* TOP10 (Thermo Fisher Scientific) was used for plasmid assembly. *Saccharomyces cerevisiae* CEN.PK2 (Euroscarf) was used for assembly of the nine BGCs listed in Table 1. *E. coli* BW29427 (also known as WM3064, personal communication, W. Metcalf, University of Illinois at Urbana–Champaign) was used as a conjugal donor strain to transfer plasmids into the panel of the selected α-, β- and γ-Proteobacteria and Actinobacteria strains listed in Supplementary Table 1 and Supplementary Figs. 2 and 3. *E. coli* BL21(DE3) sfp[73] was used as a control for the model chassis strain.

Plasmids. *Construction of pW17.* Plasmid pW17 (Supplementary Fig. 30) was assembled from three DNA fragments comprising (1) loxP, Cre, KmR, lox5171, mariner IR oriT and Marinar transposase; (2) *parB*, *par*A, *repE*, oriS, oriV and *redF*; (3) a marinar IR, T7 RNA polymerase and a *lac*UV5 regulon. These three DNA fragments were assembled using a Gibson Assembly HiFi kit (SGI-DNA) to yield pW17. The sequence for pW17 was verified using a PacBio RSII system (Pacific Biosciences).

Construction of pW34. Plasmid pW34 (Supplementary Fig. 31) was generated to express a *lux* operon in a panel of strains. pW34 was assembled from two DNA fragments comprising (1) the *luxCDABE* operon flanked by the T7 promoter and T7 terminator and (2) an apramycin resistance gene, R6K replication origin, and oriT flanked by two mutually exclusive loxP and *lox*5171 sites. These two DNA fragments were assembled using a Gibson Assembly HiFi kit (SGI-DNA) to yield pW34. The sequence for pW34 was verified using a PacBio RSII system (Pacific Biosciences).

Cloning of 10 BGCs into the pW5Y-AprR accessory plasmid. We cloned BGCs into a single-copy-number plasmid containing oriT for conjugal transfer using yeast TAR cloning. We first modified a single-copy vector, pBeloBac11 (New England Biolabs), to yield pW5Y-AprR. The pW5Y-AprR plasmid (Supplementary Fig. 32) was assembled from five DNA fragments comprising (1) AprR flanked by the mutually exclusive *lox* sites *lox*P and *lox5171*, (2) oriT amplified from plasmid pKMW2[74], (3) the *lac*UV5 regulon, (4) a *Cre* recombinase amplified from plasmid pSC101-BAD-Cre[75] and (5) a plasmid backbone amplified from pBeloBAC11 (New England Biolabs). These five DNA fragments were assembled using the Gibson Assembly HiFi kit (SGI-DNA) to yield pW5-AprR. The sequence for pW5-AprR was verified using the PacBio RSII system (Pacific Biosciences). The pW5-AprR plasmid was then linearized using *Eco47III* and inserted with a fragment consisting of *CEN/ARS* and a *URA3* marker gene to yield pW5Y-AprR. The sequence of the pW5Y-AprR plasmid was verified by the PacBio RSII system (Pacific Biosciences).

We obtained 10 synthetic DNA fragments (Thermo Fisher Scientific), each containing a T7 promoter followed by a *lac* operator, a ribosome binding site (RBS), 400–500 bp of flanking sequences of each BGC (Table 1 and Supplementary Table 13) and a T7 terminator (Supplementary Fig. 32). The flanking sequences of each BGC were used as hooks to clone the rest of the BGC using yeast TAR cloning. These 10 synthetic DNA fragments were cloned into the *NotI* site of the pW5Y-AprR plasmid using the Gibson Assembly HiFi kit (SGI-DNA). The sequences of the resulted plasmids were verified using the PacBio RSII system. The RBS sequences were generated by the RBS calculator[76]. The first 40 bases of the left boundary and the last 40 bases of the right boundary of the 10 BGCs are shown in Supplementary Table 13.

All BGC constructs were assembled and cloned into pW5Y-AprR, a bacterial artificial chromosome (BAC)-based accessory plasmid, using yeast transformation-associated recombination (TAR) cloning[77]. The 10 plasmids containing the left and right boundary sequences of the targeted BGCs were amplified using the pair of hook primers listed in Supplementary Table 14. The rest of the BGCs were amplified as 3–6 kb fragments from *P. luminescens* genomic DNA using the fragment primer pairs (frag01, frag02, frag03 and so on), also listed in Supplementary Table 14. All of these amplified DNA fragments contain at least 70 bp of overlap with their flanking pieces as well as with the left and right boundary hooks. All amplified DNA fragments were gel purified. Approximately 50 ng of the plasmid DNA were mixed with the rest of the BGC fragments in 1 to 2 molar ratios and transformed into *S. cerevisiae* CEN.PK2 using the PEG-LiAc method (Merck). The resulting URA3-positive colonies (100–1,000) were pooled together, and plasmids were harvested using the lysis buffer and beads from a GeneArt High-Order Genetic Assembly kit (Thermo Fisher Scientific). The lysed cells were directly used to transform into *E. coli* DH10B electrocompetent cells (New England Biolabs). Colonies were picked and screened using a pair of primers amplifying the internal fragment, and DNA from positive colonies was sequence-verified using the PacBio RSII system. We were able to clone and sequence-verify 9 of the 10 BGCs. The exception was BGC6 *plu2670*, which contained multiple long repeats; we could not assemble that entire BGC using the yeast TAR-cloning approach. Given that the function of *plu2670* was recently characterized as kolossin A–C synthase[58] by promoter replacement in the native strains, we decided not to pursue cloning of this BGC further.

Chromosomal integration of the LP. We used conjugation to transform pW17 into a panel of the diverse bacterial strains. *E. coli* BW29427 harbouring pW17 was used as a conjugal donor strain. This donor strain was inoculated in Luria–Bertani (LB) medium containing 0.3 mM diaminopimelic acid (DAP) and 50 μg ml^{-1} kanamycin and was grown at 37 °C in an incubation shaker at 200 r.p.m. overnight. All recipient bacteria were inoculated into LB medium and were grown at 28 °C in an incubation shaker at 200 r.p.m. until they reached the late log phase. Donor and recipient cells were washed three times with LB medium containing 0.3 mM DAP and were mixed 4:1 by optical density at 600 nm (OD_{600}). This mixture was then pelleted and transferred on a nitrocellulose filter membrane on top of an LB agar plate containing 0.3 mM DAP, and was incubated at 28 °C for 5–12 h. The bacterial mixture grown on the membrane was scraped off, resuspended into the LB liquid

medium, and spread on an LB agar plate containing an appropriate concentration of kanamycin (see Supplementary Table 1 for the concentration of kanamycin used for each strain). Chromosomal integration of the LP into *Actinobacteria* species was performed using electroporation of the PCR-amplified LP (flanking Tn5 IR sequences were added during the PCR amplification) following the Ez Tn5 transposase protocol (http://www.epibio.com/docs/default-source/protocols/ez-tn5-transposase.pdf?sfvrsn=4).

Chromosomal integration of the LP was confirmed by colony PCR using two pairs of primers; pair 1 was 99_pW17detection and SBP0572_pW17detection, and pair 2 was 202_LPdetection and 203_LPdetection (Supplementary Table 14). The first and second pairs of primers can amplify sequences in the LP and backbone, respectively. Transformants positive to only the first pair of primers were selected for additional confirmation using whole genome sequencing. In some cases, pW17 remained as a plasmid. In these cases, we repeatedly cultured the transformants and screened them for successful LP integration.

Confirmation of integration location of the LP. Insertion location of the LP in each chromosome was confirmed by low-coverage whole genome sequencing using an Illumina MiSeq platform (Illumina). The raw sequencing reads were mapped to the LP sequence using Geneious software (Biomatters). The sequences flanking the LP were searched in the bacterial chromosome to identify the insertion location (these results are summarized in Supplementary Table 2 and Supplementary Figs. 2 and 3).

Chromosomal integration of BGCs. We used conjugation to transform both pW34-based and pW5Y-AprR-based BGC constructs into a panel of the chassis strains. Either *E. coli* BW29427 harbouring pW34 or each of the pW5Y-AprR-based BGC constructs were used as a conjugal donor strain. This donor strain was inoculated in LB medium containing 0.3 mM DAP and 10 μg ml^{-1} apramycin and was grown at 37 °C in the incubation shaker at 200 r.p.m. overnight. All recipient bacteria were inoculated in LB medium containing 10 μg ml^{-1} kanamycin and were grown at 28 °C in the incubation shaker at 200 r.p.m. until the late log phase. Donor and recipient cells were washed three times with LB medium containing 0.3 mM DAP and were mixed at a ratio of 4:1 by OD_{600}. The mixture was then pelleted and placed on a nitrocellulose filter membrane on top of an LB agar plate containing 0.3 mM DAP and was incubated at 28 °C for 5–12 h. The bacterial mixture grown on the membrane was scraped off, resuspended into the LB medium and spread on an LB agar plate containing the appropriate concentration of apramycin (see Supplementary Table 1 for the concentration of apramycin used for each strain). Because successful integration of a BGC makes the recipient strains also sensitive to kanamycin, we counter-selected using an LB agar plate containing the appropriate concentration of kanamycin to screen for successful BGC integrants. We performed colony PCR on three integrants to confirm successful integration of the BGCs into the LP in the genome using primers 230 to 249 (Supplementary Tables 4 and 14).

Bioluminescence assay. For each strain, a single colony was inoculated in 1 ml of LB medium with appropriate antibiotics and grown at 28 °C in an incubation shaker overnight. This culture was inoculated into fresh LB medium containing appropriate antibiotics to a final OD_{600} of 0.1–0.2 and was aliquoted into a 96-well black plate with a clear bottom. IPTG was added to these cultures at final concentrations of 0, 0.01, 0.1 and 1.0 mM ($n = 3$ biologically independent experiments for each IPTG concentration), and these cultures were incubated using a Synergy H1 microplate reader (Biotek). Bioluminescence of these cultures was measured at intervals of 12 min for 8 h.

Production of secondary metabolites. For each strain, a single colony was inoculated into 2 ml of LB medium containing 10 μg ml^{-1} apramycin. This culture was grown overnight. An aliquot of this culture was washed with M9-based medium[78] and inoculated into 25 ml of the medium to a final OD_{600} of 0.1. These cultures were then incubated for 4–5 h and distributed as four aliquots of 5 ml each. The aliquots were induced by IPTG at final concentrations of 0, 0.001, 0.1 and 1.0 mM, respectively. The cultures were grown at 28 °C in an incubation shaker at 200 r.p.m. for three days.

Extraction of secondary metabolites. For 5 min, 2 ml of each culture was centrifuged at 10,000 r.p.m. The supernatant was transferred to another tube and the pellet was resuspended in 1 ml acetone. This suspension was sonicated for 15 min, vortexed at 1,500 r.p.m. for 20 min, and centrifuged at 10,000 r.p.m. at 4 °C for 5 min. The supernatant was collected and dried using a SAVANT SPD111 SpeedVac concentrator (Thermo Scientific). The pellet was again resuspended, this time with 1 ml of ethyl acetate. This suspension was processed as described above. The collected supernatant was combined with the dried acetone extract and again dried using a SAVANT SPD111 SpeedVac concentrator (Thermo Scientific). Finally, the dried solid was resuspended in 150 μl of methanol containing 1 μg ml^{-1} of an internal standard (2-amino-3-bromo-5-methylbenzoic acid (ABMBA)), and was filtered through a 0.22 μm PVDF membrane (Millipore Ultrafree-MC) to prepare a sample for LC-MS analysis. The culture supernatant was freeze-dried using FreeZone 12 plus (Labconco). The secondary metabolites were extracted from the dried powder following the same protocol used for extracting from pellets.

Production of secondary metabolites in the haemocoel of *Galleria mellonella* larvae. Last-instar *Galleria mellonella* larvae weighing between 200 and 300 mg were used in all experiments within seven days of shipment. The larvae were kept at room temperature in darkness before use. The single colonies of XP01 harbouring the LP and BGCs were inoculated into 2 ml LB medium and were grown overnight at 28 °C. These cultures were diluted with 0.9% NaCl to a cell density of 8×10^7 cells ml^{-1}. Aliquots of these cultures were centrifuged at 3,200 r.p.m. for 10 min and then re-suspended in 1 ml 0.9% NaCl solution.

An automated syringe pump (Cole-Parmer) was used to precisely control the volume injected into the haemocoel of the larvae. The pump was set to a volume of 11 μl and an injecting rate of 66 μl min^{-1}. A 1 ml BD syringe (Becton Dickinson) was filled with 300 μl bacterial solution for injection into three larvae. Bubbles were removed by carefully tapping the syringe and ejecting the air. A 30-gauge 1/2 BD needle (Becton Dickinson) was then attached to the syringe. After the syringe was placed in the injector, 10 μl was discarded. Each cell suspension was then injected into the haemocoels of the larvae. After injection, larvae were incubated in Petri dishes at 25 °C in the dark for 72 h, then stored at −80 °C. Mortality was monitored daily. Larvae were considered dead when they did not move when touched.

LC-MS analysis. Reverse-phase chromatography was performed using an Agilent 1290 LC stack with a C18 column (Agilent ZORBAX Eclipse Plus C18, Rapid Resolution HD, 2.1 × 50 mm, 1.8 μm) at a flow rate of 0.4 ml min^{-1} with a 2–3 μl injection volume. To detect compounds, samples were run on the C18 column at 60 °C and equilibrated with 100% buffer A (100% H_2O with 0.1% formic acid) for 1 min, diluting buffer A down to 0% with buffer B (100% ACN with 0.1% formic acid) over 8 min, followed by isocratic elution in 100% buffer B for 1.5 min. MS1 and MS2 data were collected using a Q Exactive Orbitrap mass spectrometer (Thermo Scientific). Full MS spectra were collected from m/z 80 to m/z 1,200 or m/z 135 to m/z 2,000 at 70,000 resolution, with MS2 fragmentation data acquired using 10, 20 and 30 V collision energies at 17,500 resolution. The exact mass and retention time coupled with MS2 fragmentation spectra were used to identify compounds.

Untargeted and targeted metabolite analyses. Untargeted metabolite analysis was performed using MAVEN, as described previously[79]. The minimal peak height was set to 10^7 or 10^6 for MS1 and the minimal ratio of peak height between MS1s for the culture extracts of engineered and control strains was set to 100, 50 or 10. We subsequently used MAVEN, MZmine and Thermo Xcalibur (Thermo Scientific) to extract the peak height for all identified/targeted MS1s and created heatmaps. We accepted MS1 features as metabolite ions unique to BGC expression only when those MS1 features appeared on extracts from strains expressing BGCs except for cases in which the genome analysis suggested that chassis strains contained the BGC homologues (Supplementary Table 12). The high-resolution MS2 data for the identified MS1 features were manually curated using MZmine[80] and Thermo Xcalibur (Thermo Scientific).

Quantification of product titre. Because Glidobactin A (**2**) and GameXPeptide A (**11**) are the most abundantly produced metabolites from BGC4 and BGC9, respectively, **2** and **11** were chosen as standards for quantification. The standard compounds were prepared at different concentrations (100, 50, 25, 12.5, 6.25, 3.125, 1.56, 0.78, 0.39, 0.039 and 0.19 mg l^{-1}) and these samples were measured by LC-MS. The peak area for each compound at different concentrations was calculated using the Bruker Compass Data Analysis program and was plotted against the concentration. Standard curves for **2** and **11** are given by:

$$y = 3 \times 10^7 x - 1 \times 10^7 \left(R^2 = 0.996\right) \quad (1)$$

$$y = 8 \times 10^7 x \left(R^2 = 0.994\right) \quad (2)$$

where y and x represent the peak area and concentration, respectively. The culture supernatants (1.5 ml) were freeze-dried and re-suspended in 200 μl of methanol. Then, 50 μl of the re-suspended solution was diluted to 150 μl for LC-MS analysis. The peak area of each compound was obtained by extracting an ion chromatogram, and the corresponding production titre was calculated using the standard curve equation.

Synthesis and purification of ririwpeptides. 2-Cl trityl-resin was reactivated with 3 equiv. $SOCl_2$ in dichloromethane (DCM). Tryptophan was loaded onto the resin overnight with the addition of DIPEA in DCM. Reactive groups were deactivated with 17:2:1 DCM/MeOH/DIPEA. All possible pentapeptides (15 total) that had L-configurations instead of the expected D-configurations at the α-carbons were automatically synthesized with a SYRO peptide synthesizer using Fmoc-protected amino acids, as described previously[55,81,82].

To compare retention times of synthetic and natural ririwpeptide A (**4**), HPLC-MS analysis of both synthetic and natural peptides was performed. For

the natural peptide, 1 ml of the cell culture was mixed with the same amount of methanol and centrifuged for 15 min at 17,000 r.p.m. The supernatant was directly analysed on an UltiMate 3000 LC system (Dionex) coupled to an amaZon X ion trap with electrospray ionization (Bruker Daltonics). A water/acetonitrile gradient with 0.1% formic acid as mobile phase with 0.4 ml min^{-1} flow separated the compounds on a C18 column (ACQUITY UPLC BEH, 1.7 mm, Waters). Data were analysed using DataAnalysis 4.3 (Bruker). The identities of synthetic and natural derivatives were confirmed by coinjection and HPLC-MS analysis.

Large-scale production of a metabolite with *m/z* 307. The metabolite with *m/z* 307 was produced at the 10 l scale using the protocol described in the Production of secondary metabolites section. The culture of XP04 containing BGC5 was induced with 1.0 mM IPTG when the culture reached an OD_{600} of 0.6. The cells were harvested by centrifugation at 5,000 r.p.m. for 10 min, and the cell pellet was frozen at −80 °C.

Approximately 12 g of wet cell pellet was extracted with 200 ml of acetone using sonication for 15 min. This suspension was centrifuged for 10 min at 3,700 r.p.m. at 4 °C. The supernatant was dried using a rotary evaporator to yield ~200 mg of extract. The extract was dissolved in 5 ml of methanol, and 1 ml was injected on an LC-8A Shimadzu preparative liquid chromatograph equipped with a SPD-20A UV–vis detector (Shimadzu Scientific Instruments) fitted with a C18 Pinnacle II column (5 µm, 50 × 21.2 mm (Restek)). The mobile phase for the preparative run consisted of water with 0.1% acetic acid (mobile phase A) and acetonitrile (mobile phase B), and the flow rate was 30 ml min^{-1}. The column was held at 5% mobile phase B for 2 min, increased to 95% B over 38 min, and held at 95% B for 3 min. Conditions were returned to 5% B over 2 min, and held at these conditions for 5 min to give a total run time of 50 min. The chromatogram was monitored by UV detection at 215 nm and 228 nm, and fractions were collected containing the peak of interest. Fractions were analysed by LC-MS using a Shimadzu 2010 EV LC-MS with a SPD-M20A diode array detector (Shimadzu Scientific Instruments). Pure fractions of the target compound (*m/z* 307) were pooled, dried, purged with N_2 and stored at −20 °C. This preparative procedure was repeated twice to obtain 2.1 mg of purified material.

Sample preparation for NMR analysis. After the compound with *m/z* 307 was isolated from the bacterial cell extract, 0.5 mg of it was dissolved in 180 µl D_2O with 20 mM phosphate buffer and 0.5 mM DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid). The sample was transferred to a 3 mm tube for NMR measurement.

NMR experiments and data processing. The NMR spectra of the sample were collected on a Varian (VNMRS) 600 MHz solution state NMR spectrometer (Agilent Technologies) equipped with a Varian z-gradient triple resonance HCN cold probe. 2D ^{13}C-^{1}H heteronuclear single quantum coherence (HSQC)[83] was collected with a spectral width of 160 and 12 ppm along the indirect and direct dimensions, with N1 = 256 and N2 = 1,437 complex points. The number of scans per t1 increment was 64. 2D ^{13}C-^{1}H heteronuclear multiple bond correlation (HMBC)[84] was collected with a spectral width of 240 and 12 ppm along the indirect and direct dimensions, with N1 = 128 and N2 = 1,024 complex points. The number of scans per t1 increment was 512. By setting up the total correlated spectroscopy (TOCSY)[85] mixing time to 90 ms, a 2D ^{13}C-^{1}H HSQC-TOCSY[86] was collected with a spectral width of 160 and 12 ppm along the indirect and direct dimensions, with N1 = 128 and N2 = 1,024 complex points. The number of scans per t1 increment was 192. Finally, a 2D ^{1}H-^{1}H COSY[87] was collected with a spectral width of 12 ppm along the indirect and direct dimensions, with N1 = 512 and N2 = 1,024 complex points. The number of scans per t1 increment was 32. All spectra were collected at 25 °C. All data were zero-filled, Fourier transformed, and phase and baseline corrected using NMRPipe[88].

Data availability

The data that support the findings of this study are available from the corresponding author upon request.

References

1. Donia, M. S. & Fischbach, M. A. Human microbiota. Small molecules from the human microbiota. *Science* **349**, 1254766 (2015).
2. Keller, N. P. Translating biosynthetic gene clusters into fungal armor and weaponry. *Nat. Chem. Biol.* **11**, 671–677 (2015).
3. Vorholt, J. A. Microbial life in the phyllosphere. *Nat. Rev. Microbiol.* **10**, 828–840 (2012).
4. Waters, C. M. & Bassler, B. L. Quorum sensing: cell-to-cell communication in bacteria. *Annu. Rev. Cell Dev. Biol.* **21**, 319–346 (2005).
5. Blacher, E., Levy, M., Tatirovsky, E. & Elinav, E. Microbiome-modulated metabolites at the interface of host immunity. *J. Immunol.* **198**, 572–580 (2017).
6. Pusztahelyi, T., Holb, I. J. & Pocsi, I. Secondary metabolites in fungus–plant interactions. *Front. Plant Sci.* **6**, 573 (2015).
7. Kim, E., Moore, B. S. & Yoon, Y. J. Reinvigorating natural product combinatorial biosynthesis with synthetic biology. *Nat. Chem. Biol.* **11**, 649–659 (2015).
8. Smanski, M. J. et al. Synthetic biology to access and expand nature's chemical diversity. *Nat. Rev. Microbiol.* **14**, 135–149 (2016).
9. Weber, T. et al. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* **43**, W237–W243 (2015).
10. Skinnider, M. A. et al. Genomic charting of ribosomally synthesized natural product chemical space facilitates targeted mining. *Proc. Natl Acad. Sci. USA* **113**, E6343–E6351 (2016).
11. Mukherjee, S. et al. 1,003 reference genomes of bacterial and archaeal isolates expand coverage of the tree of life. *Nat. Biotechnol.* **35**, 676–683 (2017).
12. Hadjithomas, M. et al. IMG-ABC: a knowledge base to fuel discovery of biosynthetic gene clusters and novel secondary metabolites. *mBio* **6**, e00932 (2015).
13. Rutledge, P. J. & Challis, G. L. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat. Rev. Microbiol.* **13**, 509–523 (2015).
14. Cohen, L. J. et al. Commensal bacteria make GPCR ligands that mimic human signalling molecules. *Nature* **549**, 48–53 (2017).
15. Guo, C. J. et al. Discovery of reactive microbiota-derived metabolites that inhibit host proteases. *Cell* **168**, 517–526 (2017).
16. Shao, Z. et al. Refactoring the silent spectinabilin gene cluster using a plug-and-play scaffold. *ACS Synth. Biol.* **2**, 662–669 (2013).
17. Wenzel, S. C. & Muller, R. Recent developments towards the heterologous expression of complex bacterial natural product biosynthetic pathways. *Curr. Opin. Biotechnol.* **16**, 594–606 (2005).
18. Fu, J. et al. Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat. Biotechnol.* **30**, 440–446 (2012).
19. Yoshikuni, Y., Ferrin, T. E. & Keasling, J. D. Designed divergent evolution of enzyme function. *Nature* **440**, 1078–1082 (2006).
20. Yoshikuni, Y., Dietrich, J. A., Nowroozi, F. F., Babbitt, P. C. & Keasling, J. D. Redesigning enzymes based on adaptive evolution for optimal function in synthetic metabolic pathways. *Chem. Biol.* **15**, 607–618 (2008).
21. Ziemert, N. et al. Diversity and evolution of secondary metabolism in the marine actinomycete genus *Salinispora*. *Proc. Natl Acad. Sci. USA* **111**, E1130–E1139 (2014).
22. Jensen, P. R. Natural products and the gene cluster revolution. *Trends Microbiol.* **24**, 968–977 (2016).
23. Soucy, S. M., Huang, J. & Gogarten, J. P. Horizontal gene transfer: building the web of life. *Nat. Rev. Genet.* **16**, 472–482 (2015).
24. Williams, K. P. et al. Phylogeny of Gammaproteobacteria. *J. Bacteriol.* **192**, 2305–2314 (2010).
25. McDonald, B. R. & Currie, C. R. Lateral gene transfer dynamics in the ancient bacterial genus *Streptomyces*. *mBio* **8**, e00644–12 (2017).
26. Williams, D., Gogarten, J. P. & Papke, R. T. Quantifying homologous replacement of loci between Haloarchaeal species. *Genome Biol. Evol.* **4**, 1223–1244 (2012).
27. Engel, Y., Windhorst, C., Lu, X., Goodrich-Blair, H. & Bode, H. B. The global regulators Lrp, LeuO and HexA control secondary metabolism in entomopathogenic bacteria. *Front. Microbiol.* **8**, 209 (2017).
28. Iqbal, H. A., Low-Beinart, L., Obiajulu, J. U. & Brady, S. F. Natural product discovery through improved functional metagenomics in *Streptomyces*. *J. Am. Chem. Soc.* **138**, 9341–9344 (2016).
29. Nah, H. J., Pyeon, H. R., Kang, S. H., Choi, S. S. & Kim, E. S. Cloning and heterologous expression of a large-sized natural product biosynthetic gene cluster in *Streptomyces* species. *Front. Microbiol.* **8**, 394 (2017).
30. Bierman, M. et al. Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. *Gene* **116**, 43–49 (1992).
31. Gregory, M. A., Till, R. & Smith, M. C. Integration site for *Streptomyces* phage phiBT1 and development of site-specific integrating vectors. *J. Bacteriol.* **185**, 5320–5323 (2003).
32. Brophy, J. A. N. et al. Engineered integrative and conjugative elements for efficient and inducible DNA transfer to undomesticated bacteria. *Nat. Microbiol.* **3**, 1043–1053 (2018).
33. Wozniak, R. A. & Waldor, M. K. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat. Rev. Microbiol.* **8**, 552–563 (2010).
34. Santos, C. N., Regitsky, D. D. & Yoshikuni, Y. Implementation of stable and complex biological systems through recombinase-assisted genome engineering. *Nat. Commun.* **4**, 2503 (2013).
35. Santos, C. N. & Yoshikuni, Y. Engineering complex biological systems in bacteria through recombinase-assisted genome engineering. *Nat. Protoc.* **9**, 1320–1336 (2014).

36. Lampe, D. J., Akerley, B. J., Rubin, E. J., Mekalanos, J. J. & Robertson, H. M. Hyperactive transposase mutants of the Himar1 mariner transposon. *Proc. Natl Acad. Sci. USA* **96**, 11428–11433 (1999).
37. Hickman, A. B., Chandler, M. & Dyda, F. Integrating prokaryotes and eukaryotes: DNA transposases in light of structure. *Crit. Rev. Biochem. Mol. Biol.* **45**, 50–69 (2010).
38. Dubendorff, J. W. & Studier, F. W. Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *J. Mol. Biol.* **219**, 45–59 (1991).
39. Bilyk, B., Horbal, L. & Luzhetskyy, A. Chromosomal position effect influences the heterologous expression of genes and biosynthetic gene clusters in *Streptomyces albus* J1074. *Micro. Cell Fact.* **16**, 5 (2017).
40. Englaender, J. A. et al. Effect of genomic integration location on heterologous protein expression and metabolic engineering in *E. coli*. *ACS Synth. Biol.* **6**, 710–720 (2017).
41. Sousa, C., de Lorenzo, V. & Cebolla, A. Modulation of gene expression through chromosomal positioning in *Escherichia coli*. *Microbiology* **143**, 2071–2078 (1997).
42. Moriguchi, K., Yamamoto, S., Ohmine, Y. & Suzuki, K. A fast and practical yeast transformation method mediated by *Escherichia coli* based on a trans-kingdom conjugal transfer system: just mix two cultures and wait one hour. *PLoS ONE* **11**, e0148989 (2016).
43. Frost, L. S. in *Encyclopedia of Microbiology* 3rd edn (Ed. Baldauf, S. L. et al.) 517–531 (2009).
44. Trieu-Cuot, P., Carlier, C., Martin, P. & Courvalin, P. Plasmid transfer by conjugation from *Escherichia coli* to Gram-positive bacteria. *FEMS Microbiol. Lett.* **48**, 289–294 (1987).
45. Forst, S., Dowds, B., Boemare, N. & Stackebrandt, E. *Xenorhabdus* and *Photorhabdus* spp.: bugs that kill bugs. *Annu. Rev. Microbiol.* **51**, 47–72 (1997).
46. Somvanshi, V. S. et al. A single promoter inversion switches *Photorhabdus* between pathogenic and mutualistic states. *Science* **337**, 88–93 (2012).
47. Stock, S. P., Kusakabe, A. & Orozco, R. A. Secondary metabolites produced by *Heterorhabditis* symbionts and their application in agriculture: what we know and what to do next. *J. Nematol.* **49**, 373–383 (2017).
48. Bode, H. B. Entomopathogenic bacteria as a source of secondary metabolites. *Curr. Opin. Chem. Biol.* **13**, 224–230 (2009).
49. Tobias, N. J. et al. Natural product diversity associated with the nematode symbionts *Photorhabdus* and *Xenorhabdus*. *Nat. Microbiol.* **2**, 1676–1685 (2017).
50. Cai, X. et al. Entomopathogenic bacteria use multiple mechanisms for bioactive peptide library design. *Nat. Chem.* **9**, 379–386 (2017).
51. Duchaud, E. et al. The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens*. *Nat. Biotechnol.* **21**, 1307–1313 (2003).
52. Dudnik, A., Bigler, L. & Dudler, R. Heterologous expression of a *Photorhabdus luminescens* syrbactin-like gene cluster results in production of the potent proteasome inhibitor glidobactin A. *Microbiol. Res.* **168**, 73–76 (2013).
53. Bian, X., Plaza, A., Zhang, Y. & Muller, R. Luminmycins A–C, cryptic natural products from *Photorhabdus luminescens* identified by heterologous expression in *Escherichia coli*. *J. Nat. Prod.* **75**, 1652–1655 (2012).
54. Schimming, O., Fleischhacker, F., Nollmann, F. I. & Bode, H. B. Yeast homologous recombination cloning leading to the novel peptides ambactin and xenolindicin. *Chembiochem* **15**, 1290–1294 (2014).
55. Nollmann, F. I. et al. Insect-specific production of new GameXPeptides in *Photorhabdus luminescens* TTO1, widespread natural products in entomopathogenic bacteria. *Chembiochem* **16**, 205–208 (2015).
56. Bian, X., Plaza, A., Yan, F., Zhang, Y. & Muller, R. Rational and efficient site-directed mutagenesis of adenylation domain alters relative yields of luminmide derivatives in vivo. *Biotechnol. Bioeng.* **112**, 1343–1353 (2015).
57. Bode, E. et al. Simple 'on-demand' production of bioactive natural products. *Chembiochem* **16**, 1115–1119 (2015).
58. Bode, H. B. et al. Structure elucidation and activity of kolossin A, the D-/L-pentadecapeptide product of a giant nonribosomal peptide synthetase. *Angew. Chem. Int. Ed.* **54**, 10352–10355 (2015).
59. Lambalot, R. H. et al. A new enzyme superfamily—the phosphopantetheinyl transferases. *Chem. Biol.* **3**, 923–936 (1996).
60. Winson, M. K. et al. Engineering the luxCDABE genes from *Photorhabdus luminescens* to provide a bioluminescent reporter for constitutive and promoter probe plasmids and mini-Tn5 constructs. *FEMS Microbiol. Lett.* **163**, 193–202 (1998).
61. Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D. E. & Khosla, C. Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science* **291**, 1790–1792 (2001).
62. Stein, M. L. et al. One-shot NMR analysis of microbial secretions identifies highly potent proteasome inhibitor. *Proc. Natl Acad. Sci. USA* **109**, 18367–18371 (2012).
63. Theodore, C. M., King, J. B., You, J. & Cichewicz, R. H. Production of cytotoxic glidobactins/luminmycins by *Photorhabdus asymbiotica* in liquid media and live crickets. *J. Nat. Prod.* **75**, 2007–2011 (2012).
64. DeFelice, B. C. et al. Mass spectral feature list optimizer (MS-FLO): a tool to minimize false positive peak reports in untargeted liquid chromatography-mass spectroscopy (LC-MS) data processing. *Anal. Chem.* **89**, 3250–3255 (2017).
65. Myers, O. D., Sumner, S. J., Li, S., Barnes, S. & Du, X. One step forward for reducing false positive and false negative compound identifications from mass spectrometry metabolomics data: new algorithms for constructing extracted ion chromatograms and detecting chromatographic peaks. *Anal. Chem.* **89**, 8696–8703 (2017).
66. Clevenger, K. D. et al. A scalable platform to identify fungal secondary metabolites and their gene clusters. *Nat. Chem. Biol.* **13**, 895–901 (2017).
67. Zhao, L., Cai, X., Kaiser, M. & Bode, H. B. Methionine-containing rhabdopeptide/xenortide-like peptides from heterologous expression of the biosynthetic gene cluster kj12ABC in *Escherichia coli*. *J. Nat. Prod.* **81**, 2292–2295 (2018).
68. Harding, C. R., Schroeder, G. N., Collins, J. W. & Frankel, G. Use of *Galleria mellonella* as a model organism to study *Legionella pneumophila* infection. *J. Vis. Exp.* **2013**, e50964 (2013).
69. Louwerse, J. D. et al. Stable recombinase-mediated cassette exchange in arabidopsis using *Agrobacterium tumefaciens*. *Plant Physiol.* **145**, 1282–1293 (2007).
70. Glaser, S., Anastassiadis, K. & Stewart, A. F. Current issues in mouse genome engineering. *Nat. Genet.* **37**, 1187–1193 (2005).
71. Xu, Z. et al. Large-scale transposition mutagenesis of *Streptomyces coelicolor* identifies hundreds of genes influencing antibiotic biosynthesi. *Appl. Environ. Microbiol.* **83**, e02889-16 (2017).
72. Suzuki, H., Takahashi, S., Osada, H. & Yoshida, K. Improvement of transformation efficiency by strategic circumvention of restriction barriers in *Streptomyces griseus*. *J. Microbiol. Biotechnol.* **21**, 675–678 (2011).
73. Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D. E. & Khosla, C. Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science* **291**, 1790–1792 (2001).
74. Wetmore, K. M. et al. Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *mBio* **6**, e00306–e00315 (2015).
75. Anastassiadis, K. et al. Dre recombinase, like Cre, is a highly efficient site-specific recombinase in *E. coli*, mammalian cells and mice. *Dis. Model. Mech.* **2**, 508–515 (2009).
76. Salis, H. M. The ribosome binding site calculator. *Method Enzym.* **498**, 19–42 (2011).
77. Kouprina, N. & Larionov, V. Transformation-associated recombination (TAR) cloning for genomics studies and synthetic biology. *Chromosoma* **125**, 621–632 (2016).
78. Wargacki, A. J. et al. An engineered microbial platform for direct biofuel production from brown macroalgae. *Science* **335**, 308–313 (2012).
79. Clasquin, M. F., Melamud, E. & Rabinowitz, J. D. LC-MS data processing with MAVEN: a metabolomic analysis and visualization engine. *Curr. Protoc. Bioinformatics* **14**, 14.11.1–14.11.23 (2012).
80. Pluskal, T., Castillo, S., Villar-Briones, A. & Oresic, M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11**, 395 (2010).
81. Wesche, F. et al. Combined approach of backbone amide linking and on-resin N-methylation for the synthesis of bioactive and metabolically stable peptides. *J. Med. Chem.* **61**, 3930–3938 (2018).
82. Nollmann, F. I. et al. Synthesis of szentiamide, a depsipeptide from entomopathogenic *Xenorhabdus szentirmaii* with activity against *Plasmodium falciparum*. *Beilstein J. Org. Chem.* **8**, 528–533 (2012).
83. Bodenhausen, G. & Ruben, D. J. Natural abundance N-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.* **69**, 185–189 (1980).
84. Bax, A. & Summers, M. F. ^{1}H and ^{13}C assignments from sensitivity-enhanced detection of heteronuclear multiple-bond connectivity by 2D multiple quantum NMR. *J. Am. Chem. Soc.* **108**, 2093–2094 (1986).
85. Braunschweiler, L. & Ernst, R. R. Coherence transfer by isotropic mixing—application to proton correlation spectroscopy. *J. Magn. Reson.* **53**, 521–528 (1983).
86. Lerner, L. & Bax, A. Sensitivity-enhanced two-dimensional heteronuclear relayed coherence transfer NMR-spectroscopy. *J. Magn. Reson.* **69**, 375–380 (1986).
87. Rance, M. et al. Improved spectral resolution in Cosy H-1-NMR spectra of proteins via double quantum filtering. *Biochem. Biophys. Res. Commun.* **117**, 479–485 (1983).
88. Delaglio, F. et al. NMRpipe—a multidimensional spectral processing system based on unix pipes. *J. Biomol. NMR* **6**, 277–293 (1995).

Acknowledgements

The work conducted by the US Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported under contract no. DE-AC02-05CH11231. The work performed by the Environmental Molecular Sciences Laboratory, a National

Scientific User Facility sponsored by DOE's Office of Biological and Environmental Research and located at PNNL, is operated by Battelle for the DOE under contract no. DE-AC05-76RL01830. Work in the Bode laboratory was supported by the DFG within the priority programme SPP1617 and the LOEWE Center for Translational Biodiversity Genomics (LOEWE TBG). We thank A. Wahler for professional editing support and P. Jensen for reading and commenting on our manuscript. We thank A. Deutschbauer for providing the pKMW2 plasmid and W.W. Metcalf for providing the *E. coli* BW29427 strain.

Author contributions

G.W., Z. Zhao, D.R., R.E., S.D., J.-F.C. and Y.Y. designed and built all CRAGE and BGC constructs. G.W., Z. Zhao, Y.M. and K.C. performed QA/QC of all CRAGE transformation and analysed *lux* expression. G.W. and J.K. performed and analysed BGC expression and metabolite production in multiple chassis. K.L., S.K., M.D.R., L.S. and T.N. performed LC-HRMS analyses. G.W., J.K., Z. Zhang, Y.E., Y.-M.S., B.B. and L.S. performed both targeted and untargeted metabolite analyses. Y.E., Y.-M.S. and H.B.B. performed structural characterization of metabolites from BGC7. Y.E., Y.-M.S., K.B., D.W.H., N.M.W., C.F., A. Luhrs, A. Lubbe and H.B.B. performed structural characterization of metabolites from BGC5. G.W., Z. Zhao and B.W. extended the utility of CRAGE to α- and β-Proteobacteria and Actinobacteria. B.W. and H.O. tried to extend the utility of CRAGE to *Streptomyces* sp. G.W., J.K., Y.-M.S., E.M.R., N.J.M., A.V., H.B.B. and Y.Y. wrote the manuscript. H.B.B. and Y.Y. supervised the study.

Competing interests

Lawrence Berkeley National Laboratory filed a United States patent application for CRAGE technology (US patent 20190048354). The patent is currently pending. The application lists Y.Y., G.W., Z. Zhao., J.F.C. and D.R. as inventors.