Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting

Jun Fu[1,5], Xiaoying Bian[2,5], Shengbaio Hu[1,3], Hailong Wang[1,3], Fan Huang[1,3], Philipp M Seibert[1], Alberto Plaza[2], Liqiu Xia[3], Rolf Müller[2], A Francis Stewart[1] & Youming Zhang[4]

Functional analysis of genome sequences requires methods for cloning DNA of interest. However, existing methods, such as library cloning and screening, are too demanding or inefficient for high-throughput application to the wealth of genomic data being delivered by massively parallel sequencing. Here we describe direct DNA cloning based on the discovery that the full-length Rac prophage protein RecE and its partner RecT mediate highly efficient linear-linear homologous recombination mechanistically distinct from conventional recombineering mediated by Redαβ from lambda phage or truncated versions of RecET. We directly cloned all ten megasynthetase gene clusters (each 10–52 kb in length) from *Photorhabdus luminescens* into expression vectors and expressed two of them in a heterologous host to identify the metabolites luminmycin A and luminmide A/B. We also directly cloned cDNAs and exactly defined segments from bacterial artificial chromosomes. Direct cloning with full-length RecE expands the DNA engineering toolbox and will facilitate bioprospecting for natural products.

In the search for useful bioactive compounds, prokaryotic secondary metabolites have been a rich source of products including antibiotics (e.g., erythromycin from *Saccharopolyspora erythraea*), chemotherapeutics (e.g., epothilone from *Sorangium cellulosum*), immunosuppressants (e.g., rapamycin from *Streptomyces hygroscopicus*) and insecticides (e.g., spinosad from *Saccharopolyspora spinosa*)[1]. With the advent of massively parallel DNA sequencing, complete genomes are now being determined at a rapidly increasing rate[2,3] revealing thousands of secondary-metabolite pathways. However, many of these secondary metabolites cannot be determined because either the host cannot be cultivated or, if it can, the pathways are silent under laboratory conditions[1,4,5,6]. Prokaryotic secondary metabolites are mostly produced from polyketide synthase (PKS) and nonribosomal peptide synthetase (NRPS) pathways, which are encoded by large operons[7,8]. These pathways are typically only expressed under specific circumstances. It is largely guesswork to find suitable laboratory conditions that promote expression in the natural host. Consequently, methods that do not rely on cultivation of the natural host are required for efficient and routine investigation of secondary metabolites. Because the pathways encoding these metabolites are conserved by the host, although usually not expressed[1], it is likely that the encoded secondary metabolites are bioactive, a potential treasure trove for bioprospecting.

To facilitate functional investigation of sequence data, we aimed to develop a method for cloning a target DNA segment from a primary DNA preparation exactly and directly into an expression vector. Apart from short segments that can be obtained by PCR or *de novo* DNA synthesis, cloning DNA regions longer than ~5 kb depends upon the construction of DNA libraries large enough to ensure a good chance that the region can be found, followed by screening to identify it. Library cloning and screening are laborious, and often the desired region must be reconstructed from two or more pieces of DNA and/or subcloned into an expression vector for functional studies. Other alternatives include *in vitro* DNA capture promoted by RecA, protein nucleic acid or triplex affinity capture[9–11], recombination *in vivo* using transformation-associated recombination cloning in yeast[12], truncated RecE-RecT or Redα-Redβ in *E. coli*[13], or insertion of foreign DNA into the *Bacillus subtilis* genome[14]. However, none of these *in vitro* or *in vivo* approaches, including our own, has met with broad success as all appear to be too demanding or inefficient for general application. To address these limitations, we have been exploring unconventional aspects of recombineering.

Conventional recombineering refers to using proteins expressed by the lambda phage Red operon—Redα, Redβ and Redγ—to promote homologous recombination between a linear and circular DNA molecule in *E. coli*[15–17]. The Rac prophage also contains a pair of proteins that are functionally similar to Redα and Redβ. Previously we showed that a truncated version of RecE, consisting of just the C terminus starting at residue 588 (RecE588), along with RecT, can be used for recombineering[15]. Like Redα and Redβ, RecE588 and RecT

[1]Technische Universitaet Dresden, Genomics, BioInnovationsZentrum, Dresden, Germany. [2]Helmholtz Institut für Pharmazeutische Forschung, Helmholtz Zentrum für Infektionsforschung and Universität des Saarlandes, Department of Pharmaceutical Biotechnology, Saarbrücken, Germany. [3]Key Laboratory of Microbial Molecular Biology of Hunan Province, College of Life Science, Hunan Normal University, Changsha, Hunan Province, People's Republic of China. [4]Gene Bridges, University Campus, Saarbrücken, Germany. [5]These authors contributed equally to this work. Correspondence should be addressed to A.F.S. (stewart@biotec.tu-dresden.de) or R.M. (rom@mx.uni-saarland.de) or Y.Z. (youming.zhang@genebridges.com).

Received 24 October 2011; accepted 16 March 2012; published online 29 April 2012; doi:10.1038/nbt.2183

cooperate as a 5′-3′ exonuclease and single-stranded annealing protein (SSAP) pair[18]. The RecE588-RecT pair was less efficient than the Redα-Redβ (Redαβ) pair for recombineering, and subsequently we chose to use the Red system for most applications[16,19–22].

In comparison to Redα, which is a 226-amino-acid exonuclease[23], RecE is a much larger protein, 866 amino acids in size, of which the exonuclease domain encompasses the last 260 amino acids[24,25]. In this work, we discovered that full-length RecE along with RecT efficiently facilitates homologous recombination between two linear molecules. We used this pair of proteins to clone large regions of DNA directly from purified genomic DNA into *E. coli* expression vectors bypassing library construction and screening. To demonstrate utility, we focused on direct cloning of prokaryotic secondary-metabolite pathways for bioprospecting. We selected the genome sequence of *P. luminescens* TT01 (ref. 26) because it had ten unexplored secondary-metabolite pathways. All ten pathways were directly cloned and so far two products have been obtained after expression in heterologous hosts. These products were further characterized using genetics, mass spectrometry and nuclear magnetic resonance. We also report an optimized protocol that uses both Redα- and RecE-based techniques to efficiently clone large DNA fragments.

RESULTS

Previous recombineering approaches relied on homologous recombination occurring between linear and replicating circular DNA molecules[27]. In this work, we describe recombineering between two linear DNA molecules, a plasmid vector flanked by two short sequence regions ('homology arms') that define the ends of the target segment and the target DNA segment. For convenience, we term the latter LLHR (linear plus linear homologous recombination) and the former LCHR (linear plus circular homologous recombination).

Full-length RecET is more efficient at LLHR than Redαβ

Figure 1a,**b** illustrates a simple assay to compare LLHR and LCHR efficiencies. The substrates and products are identical except that for the LLHR assay, the backbone vector p15A-cm was linearized between the two 50-bp homology arms. Using this assay, we assessed various combinations of Rec and Red proteins expressed in *E. coli* to mediate homologous recombination (**Fig. 1c**,**d**). In the LCHR assay, the complete Red operon (γβα, gba) was superior to any other combination. As shown before[18], Redαβ was about twofold better than RecE602T at LCHR in the absence of Redγ[18] (**Fig. 1c**). Inclusion of Redγ, which inhibits endogenous RecBCD in *E. coli*[28,29], enhanced LCHR when added to both Redαβ and RecE602T, with a greater effect on Redαβ.

We then sought to compare the efficiency of LLHR using both truncated and full-length versions of RecE. Truncations beyond amino acid 602 destroyed homologous recombination activity, however RecE602 was better at LCHR than RecE588 (refs. 15,18). We found that full-length RecE was worse at LCHR than RecE602 (**Fig. 1c**) but much better at LLHR relative to either RecE602 or Red (**Fig. 1d**). Expressing Redγ along with full-length RecET (RecET) increased

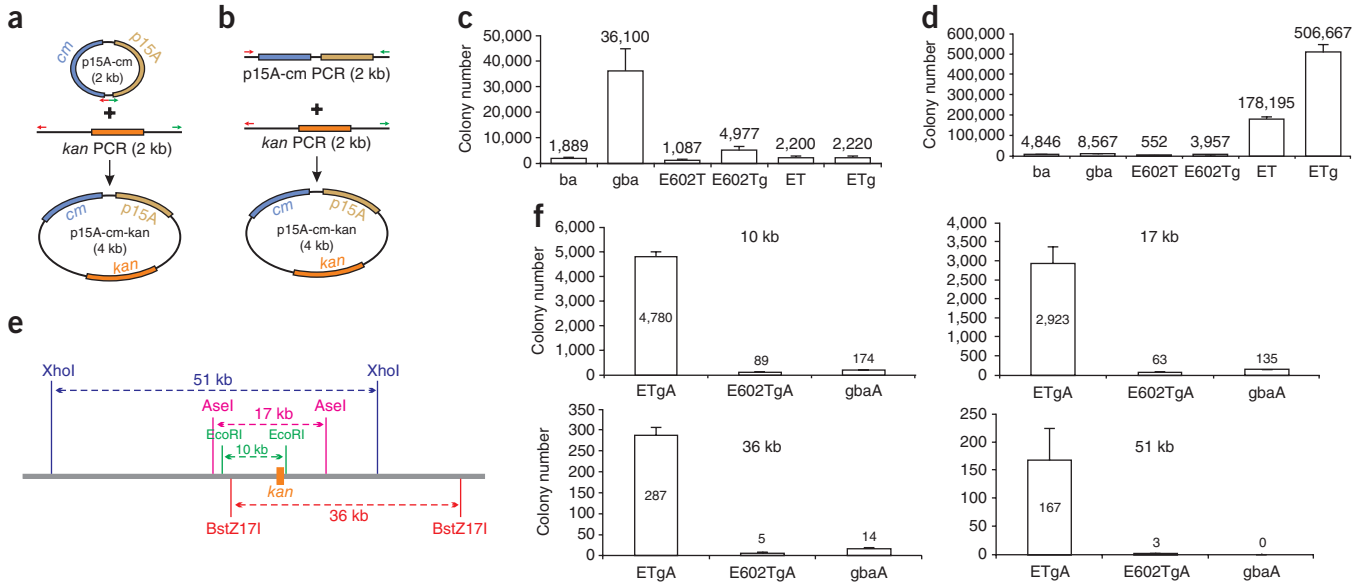

Figure 1 Full-length RecE mediates recombination between two substrates. (**a**) Linear plus circular recombination assay shows the plasmid p15A-cm and a PCR product carrying the kanamycin resistance gene (*kan*). Each end of the *kan*-PCR product has a 50-bp homology arm to the p15A-cm plasmid between the chloramphenicol gene (*cm*) and the p15A origin (indicated by red and green arrows). Electroporation of the circular plasmid and the PCR product together into *E. coli* expressing the protein combinations indicated in **c** generated the chloramphenicol plus kanamycin-resistant plasmid, p15A-cm-kan. (**b**) Like **a** except the p15A-cm plasmid was linearized between the two 50-bp homology arms. Electroporation of the linearized plasmid and the PCR product together into *E. coli* expressing the protein combinations indicated in **d** generated p15A-cm-kan. (**c**) Resulting number of colonies for the linear plus circular (LCHR) assay in **a** upon expression of Redβ and α (ba); Redγ, β and α (gba); RecE602 and T (E602T); RecE602, T and Redγ (E602Tg); full-length RecE and T (ET); and full-length RecE, T and Redγ (ETg). All genes are driven by the P_{BAD} promoter in pSC101-tet vector. (**d**) As in **c** except using the LLHR assay shown in **b**. (**e**) Comparison of full-length RecE, truncated RecE and Redα to mediate LLHR with respect to increasing size. Mouse genomic BAC (RP24-334D1 encoding Jarid1b) containing a kanamycin-resistance gene (*kan*) inserted by recombineering (LCHR) showing the restriction sites that were used to release EcoRI (10 kb), AseI (17 kb), BstZ17I (36 kb) and XhoI (51 kb) fragments. (**f**) Results from four PCR-amplified, 2.1-kb, p15A-amp subcloning vectors, which contained 50-bp terminal homology arms to the very ends of the 10-, 17-, 36- and 51-kb fragments, respectively. After electroporation of 5 μg restriction enzyme–digested BAC DNA and 1 μg of the corresponding p15A-amp PCR product into arabinose-induced *E. coli* harboring pSC101 plasmids encoding full-length ETg and RecA (ETgA); truncated E602Tg and RecA (E602TgA); or gba and RecA (gbaA) colonies were selected on ampicillin plus kanamycin plates and counted. Error bars, s.d.; $n = 3$.

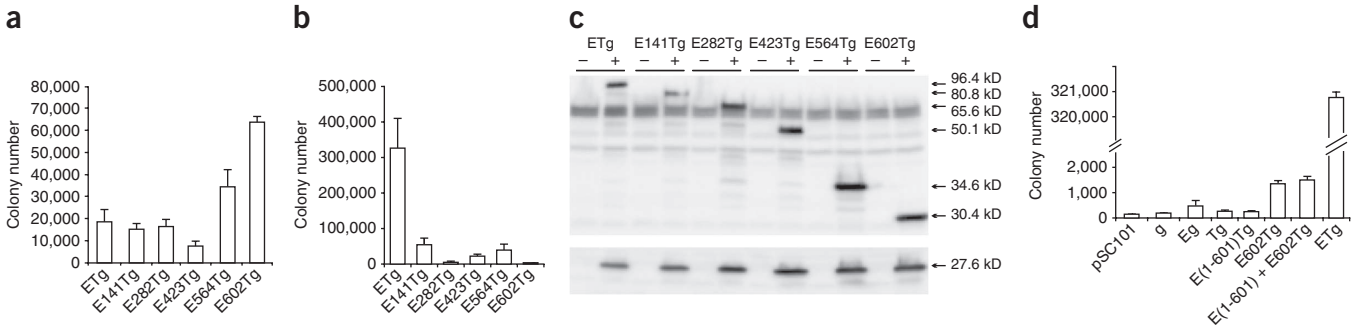

Figure 2 Full-length RecE plus RecT is required for enhanced LLHR. (**a**) Full-length RecE with RecT and Redγ (ETg) was compared to truncated versions as indicated using the LCHR assay of **Figure 1a**. (**b**) Same as **a** except using the LLHR assay of **Figure 1b**. (**c**) Expression of the RecE versions was evaluated before (–) and after (+) arabinose induction from the pSC101-P_{BAD} expression plasmids using an antibody to RecE602 for western blot analysis of full-length RecE (96.4 kD), RecE141 (the first 141 amino acids were deleted; 80.8 kD), RecE282 (65.6 kD), RecE423 (50.1 kD), RecE564 (34.6 kD) and RecE602 (30.4 kD). In the lower panel, RecT (27.6 kD) expression was evaluated from the same protein extracts. (**d**) Protein combinations as indicated were tested in the LLHR assay of **Figure 1b**; pSC101, empty vector; Redγ only (g); full-length RecE with Redγ (Eg); RecT with Redγ (Tg); RecE from a.a. 1 to 601 with RecT and Redγ (E(1-601)Tg); RecE602 with RecT and Redγ (E602Tg); RecE from a.a. 1 to 601 with RecE602, RecT and Redγ (E(1-601) + E602Tg); full-length RecE with RecT and Redγ (ETg). Error bars, s.d.; $n = 3$.

the LLHR yield threefold. To further explore the LLHR properties of full-length RecE, truncated RecE and Redα, we examined the impact of size on cloning efficiencies (**Fig. 1e**,**f**). In all instances, full-length RecE was more than 20 times better, which was critically important for the longest (51 kb), most difficult cloning exercise.

To determine how much of RecE is necessary for maximal LLHR activity, we evaluated a RecE deletion series in both the LLHR and LCHR assays (**Fig. 2**). Full-length RecE showed the most LLHR activity, whereas the shortest, exonuclease-only version, RecE602, showed the most LCHR activity. These differences were not due to differential protein expression levels (**Fig. 2c**). Furthermore, the LLHR activity resident in the N terminus of RecE cannot be provided *in trans* but must be connected to the exonuclease domain (**Fig. 2d**). These results indicate that LCHR and LLHR are mechanistically separable and suggest that efficiency at one comes at a price of inefficiency at the other.

A replication test distinguishes between LCHR and LLHR

LCHR recombination mediated by the Red proteins occurs at the replication fork and therefore requires ongoing replication[27,30–34]. To determine whether LLHR operates by a similar mechanism, we carried out LCHR and LLHR assays using a plasmid backbone that contains the *pir*–dependent R6K origin of replication, but which itself contains only half of the *pir* gene (**Fig. 3**). The other half of *pir* must be introduced by recombination. If recombination occurs before replication, *pir* will be expressed and the plasmid can replicate. However, if the plasmid needs to be replicating for recombination to occur, then no product will arise unless *pir* is expressed in *trans*. As shown previously[27], Red-mediated recombination in an LCHR assay requires *pir* expression *in trans*. This was also the case for RecET-Redγ mediated recombination in the LCHR assay (**Fig. 3a**). However, both the complete Red operon and RecET-Redγ mediated LLHR in the absence of *pir* expression, implying that replication is not required (**Fig. 3c**). This demonstrates that LCHR and LLHR are mechanistically different and again illustrates that the Red operon is more suited to LCHR, whereas RecET is more suited to LLHR. To further compare LCHR and LLHR, we examined the relationship between homologous recombination efficiency and homology arm length but found a very similar qualitative profile (**Supplementary Fig. 1**). Briefly, both require at least 30 bp of homology with efficiencies steadily increasing as the amounts of homology increase.

Direct cloning of secondary-metabolite pathways

To enhance the practical use of full-length RecE for direct cloning, we tested various means of expressing RecET-Redγ (**Fig. 4**) including inducible expression plasmids along with generating a strain of *E. coli* GB05-dir in which RecET-Redγ was stably integrated and under the control of the

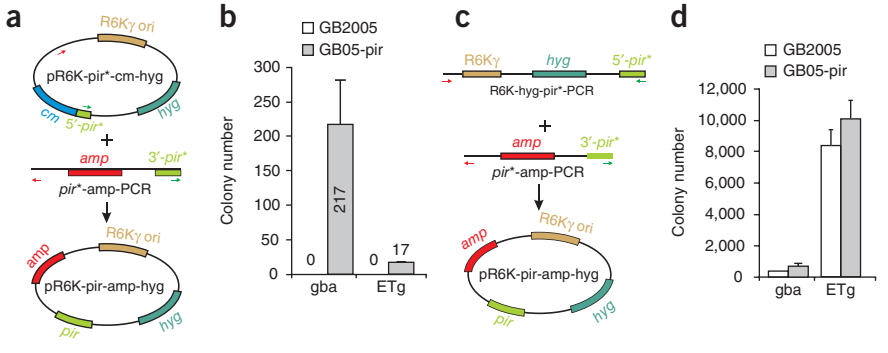

Figure 3 LCHR and LLHR are mechanistically different. (**a**) Recombination assay to see if DNA replication is required to initiate LCHR. Plasmid pR6K-pir*-cm-hyg has only the 5′ part of *pir*, which is required for the R6K origin of replication. This plasmid cannot replicate in a *pir*-negative strain. The PCR product of *pir**-amp has the 3′ part of the *pir* gene including a region of overlap that serves as a homology arm for recombination. Upon recombination, a complete copy of *pir* is generated and its product, Pi, will be expressed. The substrates were electroporated into either GB2005 or the derivative, GB05-pir, which expresses *pir* either with expression of Redγβα (gba) or full-length RecE, RecT and Redγ (ETg), as indicated. Red and green arrows indicate limits of homology arm. (**b**) No recombinants were observed in GB2005, indicated as '0', whereas 217 and 17 recombinant colonies were observed with gba and ETg, respectively. (**c**) As for **a**, except using a linearized version of pR6K-pir*-cm-hyg. Red and green arrows as in **a**. (**d**) As for **b** except results from the LLHR assay in **c** are shown. The product plasmids of **a** and **b** were identical after restriction analysis and subsequent testing for replication in GB2005 without additional Pi (data not shown). Error bars, s.d.; $n = 3$.

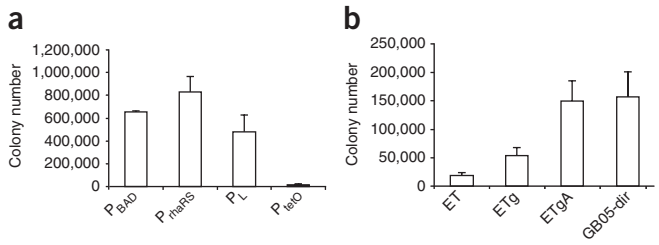

Figure 4 Evaluation of different configurations of inducible recombinase expression. (**a**) The LLHR assay (**Fig. 1b**) was used to evaluate the effect of inducing RecE, RecT and Redγ (ETg) from P_{BAD}, P_{rhaRS}, the lambda temperature-inducible promoter (P_L) or the tetracycline-inducible promoter (P_{tetO}). All constructs were identically configured in the pSC101 plasmid. (**b**) LLHR recombination efficiency in *GB2005* mediated by full-length RecE plus RecT alone (ET), with Redγ (ETg) or with RecA (ETgA), or in *E. coli* strain GB05-dir, in which an *araC-*P_{BAD}*-ETgA* operon has been integrated into the genome at the *ybcC* locus. Error bars, s.d.; $n = 3$.

P_{BAD} promoter (**Supplementary Fig. 2**). We found that the plasmids containing RecET-Redγ expressed by arabinose- and rhamnose-inducible promoters, P_{BAD} and P_{rhaRS}, respectively, were better than the tetO tetracycline-inducible promoter and the leaky temperature-sensitive promoter P_L. As previously reported, endogenous RecA is not required for LCHR[31–33] but transient expression of RecA in a *recA*-deficient host improves the transformation efficiency[34]. Therefore, we added *recA* to the *recET-redγ* expression plasmid and observed, as expected, about a threefold increase in the number of colonies obtained. Using these expression constructs, we developed methods to directly subclone (i) exactly defined segments from bacterial artificial chromosomes (BACs) (**Supplementary Fig. 3**), (ii) a full-length cDNA from a mRNA preparation directly into an expression vector (**Supplementary Fig. 4**) and (iii) secondary-metabolite pathways directly from a prokaryotic genome.

To demonstrate the potential of direct cloning with full-length RecE, we applied it to bioprospecting for prokaryotic secondary metabolites (**Supplementary Fig. 5a**). We chose to study *P. luminescens* TT01 (DSM15139), which has a 5.7-Mbp genome[26,35] containing ten unknown PKS-NRPS gene clusters (**Table 1** and **Supplementary Fig. 5b**), although two homologs have been partly characterized in a human pathogenic strain, *Photorhabdus asymbiotica*[36] (**Supplementary Tables 1–3**). To clone these ten PKS-NRPS gene clusters directly from the genome, we designed ten oligo PCR pairs to generate linear cloning vectors (pGB-amp-P_{tetO}) carrying terminal 70-bp homology arms to the target clusters (**Supplementary Table 4**). The 5′ homology arm was chosen to place the initiating ATG codon of the first gene in each cluster directly under the control of the tetracycline-inducible promoter P_{tetO}. The 3′ homology arm was chosen based on the location of suitable restriction sites flanking the clusters. Whereas the 5′ homology arms lay at variable distances from the 5′ end of the target restriction segments, the 3′ homology arms were positioned at the very 3′ end of the target restriction segments because terminal positioning of the homology arms enhances cloning efficiency (**Table 1** and **Supplementary Fig. 6**). We digested genomic DNA with the chosen restriction enzymes to release the clusters on single DNA segments (**Table 1**). Then we transformed L-arabinose–induced GB05-dir cells with ten independent linear DNA mixtures (linear vector plus digested genomic DNA). After antibiotic selection, 12 colonies from each transformation were picked and transferred to 96-deep-well plates, and plasmid DNA from these clones was verified by restriction analysis. Using this parallel cloning strategy, nine of the ten gene clusters were successfully cloned in one round of LLHR (**Table 1**).

Heterologous expression of two PKS-NRPS gene clusters

End-sequencing of the cloned clusters revealed that the constructs containing the *plu1210–plu1222* and *plu3123* clusters contained mutations at their 5′ ends, preventing expression (**Table 1** and **Supplementary Fig. 7**). These results suggested that these clusters are toxic to the *E. coli* cloning host even without tetracycline induction. For the seven clusters that were cloned without mutations, we sought to induce expression in a heterologous host. For simplicity, we first tried *E. coli* as the heterologous host using either *E. coli* Nissle1917, which carries the colibactin PKS-NRPS gene cluster and contains a phosphopantheteine transferase[37] required for post-translational activation of PKS-NRPS proteins, or a modified laboratory strain, GB05-MtaA, made by introducing a phosphopantheteine transferase from *Stigmatella aurantiaca*[38] into our standard recombineering host, GB05.

We transformed *E. coli* Nissle 1917 and GB05-MtaA with the seven cloned PKS-NRPS gene clusters, and two of them (*plu1881-plu1877* and *plu3263*) produced compounds at a detectable level (>0.1 mg/l; **Supplementary Results**). The *plu1881-plu1877* gene cluster has a similar gene structure to the glidobactin (*glb*; **Supplementary Fig. 8**) and syringolin mixed NRPS/PKS pathways[39,40] but no similar compound was observed in cultures of *P. luminescens* TT01 (**Fig. 5a**, top). After transformation of *E. coli* Nissle 1917 with the *plu1881-plu1877* direct clone (**Supplementary Fig. 6b**) and tetracycline induction, one major compound (1, *m/z* 505.3) was detected (**Fig. 5a**, middle). When *glbH*[40] was added into *plu1881-plu1877*, both glidobactin A (2, *m/z* 521.3)[41] and 1 (*m/z* 505.3) were detected (**Fig. 5a**, bottom). Together with mass spectrometry fragmentation analysis (**Fig. 5b** and **Supplementary Figs. 9** and **10**), we established the structure of 1 and designated it luminmycin A (**Fig. 5c**), also known as 10-deshydroxy glidobactin A. We propose that its biosynthetic pathway begins with the NRPS encoded by *plu1878* (C-A-T) followed by *plu1880* (C-A-T-C-A-T-KS-AT-DH-KR-T-Te) from the *plu1881-plu1877* gene cluster (**Fig. 5c**).

Table 1 Direct cloning of PKS-NRPS gene clusters from *P. luminescens*

PKS-NRPS	Genome position	Size (bp)	Digestion enzyme	Fragment (bp)	5′ to ATG (bp)	3′ to homology (bp)	Correct/checked
plu0897–plu0899	1021625...1035887	14,262	XbaI-SacI	15,095	512	0	5/12
plu1113–plu1115	1277270...1299002	21,732	ApaLI	29,834	7,181	0	3/12
plu1210–plu1222	1392054...1415717	23,663	BstZ17I	24,659	182	0	4/12[a]
plu1881–plu1877	2239952... 2221617	18,336	PacI	25,546	899	0	6/12
plu2318–plu2323	2716778...2740058	23,281	SwaI	29,471	1,565	0	3/12
plu2670	3173674...3124571	49,104	XbaI-XmaI	52,622	2,134	0	0/48
plu3123	3662492...3646119	16,374	AseI	16,858	184	1,294	3/12[a]
plu3130	3688463...3678528	9,936	AseI	10,233	144	138	6/12
plu3263	3880777...3865127	15,651	NdeI-PacI	16,920	552	0	4/12
plu3535–plu3532	4164946...4132061	32,886	XbaI	37,501	0	0	2/12

[a]All correct clones were mutated at the 5′ end of the operon but were otherwise intact (**Supplementary Fig. 7**).

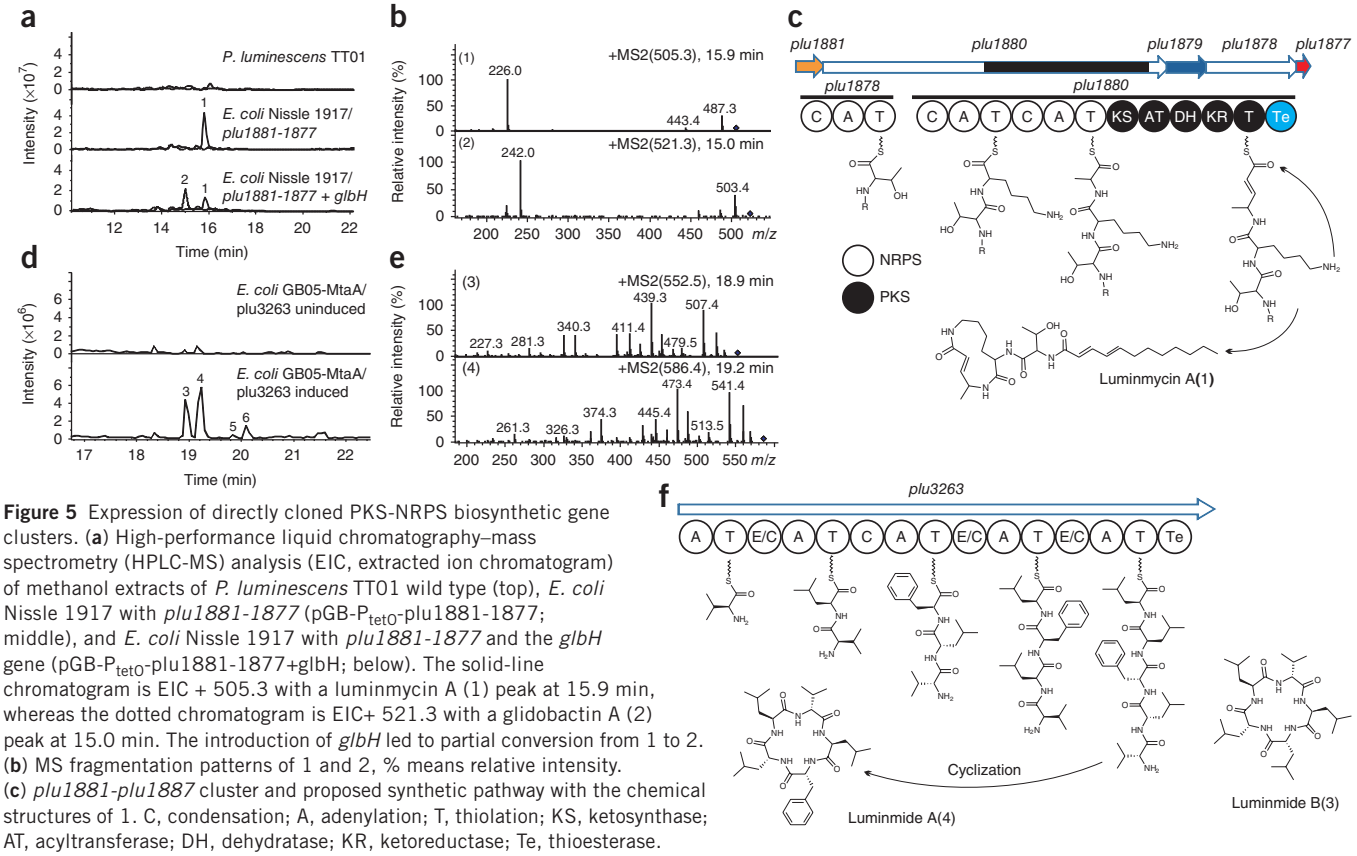

Figure 5 Expression of directly cloned PKS-NRPS biosynthetic gene clusters. (**a**) High-performance liquid chromatography–mass spectrometry (HPLC-MS) analysis (EIC, extracted ion chromatogram) of methanol extracts of *P. luminescens* TT01 wild type (top), *E. coli* Nissle 1917 with *plu1881-1877* (pGB-P_{tetO}-plu1881-1877; middle), and *E. coli* Nissle 1917 with *plu1881-1877* and the *glbH* gene (pGB-P_{tetO}-plu1881-1877+glbH; below). The solid-line chromatogram is EIC + 505.3 with a luminmycin A (1) peak at 15.9 min, whereas the dotted chromatogram is EIC+ 521.3 with a glidobactin A (2) peak at 15.0 min. The introduction of *glbH* led to partial conversion from 1 to 2. (**b**) MS fragmentation patterns of 1 and 2, % means relative intensity. (**c**) *plu1881-plu1887* cluster and proposed synthetic pathway with the chemical structures of 1. C, condensation; A, adenylation; T, thiolation; KS, ketosynthase; AT, acyltransferase; DH, dehydratase; KR, ketoreductase; Te, thioesterase. (**d**) HPLC-MS analysis (base peak chromatogram (BPC) in the mass range 550.0–600.0) of methanol extracts from *E. coli* GB05-MtaA containing pGB-P_{tetO}-plu3263, uninduced (above) and induced with tetracycline (below). Four major compounds (3, 4, 5 and 6) were detected as the products of NRPS *plu3263* by comparative metabolite profiles. (**e**) MS fragmentation patterns of luminmide A (4, bottom) and luminmide B (3, top). (**f**) *plu3263* cluster and proposed biosynthetic pathway with the chemical structure of 3 and 4. E/C, dual epimerization/condensation domain; other abbreviations as in **c**.

Plu3263 is a 15.6-kb gene containing five NRPS modules comprising A-T-E/C-A-T-C-A-T-E/C-A-T-E/C-A-T-Te domains. After transformation of *E. coli* GB05-MtaA with the *plu3263* direct clone and tetracycline induction, four major compounds (3, 4, 5 and 6) were produced (**Fig. 5d**). Feeding studies with labeling precursors, detailed mass spectrometry fragmentation analysis (**Supplementary Figs. 11** and **12**) combined with a phylogenetic analysis of condensation (C) domains from *plu3263* (**Supplementary Fig. 13**) confirmed assignments of the two most-prominent peaks (4 and 3, **Fig. 5e**) as luminmide A (cyclo (D-Val-L-Leu-D-Phe-D-Leu-L-Leu)) and B (cyclo (D-Val-L-Leu-D-Leu-D-Leu-L-Leu)). Both compounds were produced and isolated from the heterologous host and structures confirmed by NMR and advanced Marfey's analysis plus bioinformatics tools (**Supplementary Tables 5–7**), thereby confirming the proposed pathway (**Fig. 5f**).

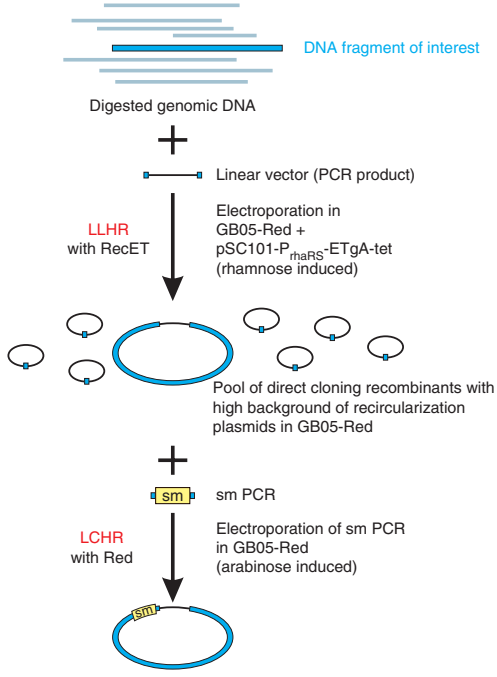

Figure 6 A two-step, double recombination 'fishing' strategy enhances the identification of desired products. The DNA is digested or sheared and electroporated with the PCR-amplified vector (e.g., with an ampicillin-resistance gene 'ampR') carrying the terminal homology arms into a host containing full-length RecET. After selection, the ampicillin-resistant colonies are pooled and electroporated with a linear DNA molecule encoding a selectable marker (sm) gene (e.g., chloramphenicol) flanked by homology arms corresponding to part of the intended cloned region. The desired colonies will grow after selection for the last selectable gene. To facilitate this strategy, which is essentially an LLHR step followed by an LCHR step, a combinatorial host was developed. This host, GB05-Red has the araC-P_{BAD}-gbaA operon integrated into the chromosome so that arabinose induces the expression of Red gbaA (**Supplementary Fig. 2**). The plasmid pSC101-P_{rhaRS}-ETgA-tet, in which the RecE, RecT, Redγ and RecA are expressed after rhamnose inducible (**Supplementary Fig. 2**), was also introduced.

Efficient cloning by combining LLHR with LHCR

In LLHR, the primary source of background colonies came from recircularization of the empty cloning vector by intramolecular recombination. This can be reduced by avoiding terminal repeats in the cloning vector, which can be deleterious even if there is only an 8 bp match[13]. Even with this precaution, we could not directly clone the largest *P. luminescens* PKS-NRPS clusters (*plu2670*; ~52 kb) using LLHR alone (**Table 1**). To extend the reach of these cloning approaches to include more difficult exercises, we made the strain GB05-red. This strain carries the arabinose-inducible P_{BAD}-*redγβα* operon integrated into the chromosome (**Supplementary Fig. 2c**). To express full-length RecE (along with RecT, Redγ and RecA) after rhamnose induction, we introduced pSC101-P_{rhaRS}-ETgA-tet (**Supplementary Fig. 2b**) into GB05-red. Then, to directly clone the 52-kb *plu2670*, we electroporated XbaI-XmaI digested *P. luminescens* genomic DNA with p15A-amp-BSD carrying homology arms for the *plu2670* gene (**Table 1**). After the LLHR step, which was done as before except that rhamnose was used to induce expression of RecET, Redγ and RecA, we pooled candidate colonies, added arabinose to induce expression of the Red operon and electroporated with a PCR product carrying homology arms flanking a selectable gene (chloramphenicol) to target the 3′ end of the *plu2670* cluster. Hence, the correct clones were identified by LCHR (**Fig. 6**). This two-step cloning method yielded 6/21 correct, 52-kb *plu2670* clones.

DISCUSSION

Very little sequence data were available when the original DNA library methods were first developed. Now direct, homology-based cloning approaches that exploit the increasing volumes of available sequence information are possible. Direct cloning using LLHR will be especially relevant for the increasing number of cases where a library of the entire genome is not wanted but where certain regions are potentially useful for further studies. Clearly it will be better to store primary DNA samples and directly clone the regions of interest on demand rather than store and screen entire libraries.

This work is based on the unexpected finding that full-length RecE (with RecT) is substantially more efficient at mediating homologous recombination by means of a classical annealing mechanism than either its truncated versions (RecE588 or RecE602) or Redαβ. Until now, it has been assumed that these phage-encoded, 5′-3′ exonuclease-SSAP pairs were functionally equivalent. Furthermore, standard models of homologous recombination have been based on two types of initiating steps, either strand invasion, which is a property of the RecA-RAD51 SSAP class, or strand annealing, which is a property of the RAD52-RecT-Redβ SSAP class[42–44]. The idea that strand annealing might encompass two mechanistically different processes has not gained any prominence, although several supporting observations have been made.

Strand-specific asymmetry in single-stranded, oligonucleotide-directed mutagenesis was discovered in yeast[45] and then as a property of Redαβ[34] and RecET recombination[30] in *E. coli*, where the asymmetry was related to the direction of replication. Single-stranded oligonucleotides that could prime Okazaki-like fragment synthesis are consistently more efficient at site-directed mutagenesis than their complementary oligonucleotides. Recently, biased annealing at the replication fork has been shown in other prokaryotes[46] and has also been shown to underpin efficient double-stranded DNA recombination mediated by Redαβ[27,32]. These observations imply that homologous recombination initiated by annealing can occur either by simple annealing between two single-stranded regions or by annealing at the replication fork. We show that these two annealing mechanisms are distinct and that different proteins can specialize at one or the other. More work is required to elucidate the mechanistic differences between the two ways that 5′-3′ exonuclease-SSAP protein pairs can initiate homologous recombination.

The rapidly growing numbers of PKS-NRPS gene clusters being discovered by genome sequencing projects represent a remarkable opportunity to discover natural products. We sought to develop a bioprospecting approach (**Supplementary Fig. 5**) for cloning, expression and further mutagenesis of these potentially valuable resources. Because they are modular, secondary-metabolite pathways usually contain highly repetitive DNA regions, which can be unstable and present a challenge for cloning and mutagenesis. Consequently *E. coli* is the preferred host for cloning and maintenance because endogenous recombination can be minimized more thoroughly than for any other host. A further advantage of our cloning strategy is the ability to position the first open reading frame of an operon precisely into an expression vector correctly positioned adjacent to a ribosomal binding site and under the control of an inducible promoter. This is particularly important for cloning PKS-NRPS gene clusters, which in our experience are often toxic to the host when constitutively expressed.

Rapid acquisition of expression clones for secondary-metabolite pathways is a major step toward our goal of evaluating unknown pathways at reasonable throughput. However, the successful expression of a pathway is unpredictable owing to the idiosyncrasies of both different heterologous hosts, which may not produce all necessary precursors, and different pathways. Here we employed *E. coli* for convenience and successfully expressed two out of seven cases (or four of nine; **Supplementary Fig. 7**). For other cases, successful expression may require several attempts with other heterologous hosts[47], such as *Pseudomonas putida*[48]. Further problems confronting heterologous expression include the possibility that the cluster contains an internal promoter(s). Pathway problems such as these, which are major difficulties for approaches to activate endogenous silent pathways by gene targeting, can be readily addressed by Red recombineering once a clone has been obtained[48,49]. Hence, we are confident that many more pathways can be activated by our approach given a few more expedient hosts and the ability to rapidly alter the expression construct using recombineering.

To increase the ability to clone directly from more complex DNA mixtures, we developed a two-step strategy incorporating both LLHR and LHCR (**Fig. 6**). This enabled the direct cloning of *plu2670* from the *P. luminescens* genome. Consequently, most secondary-metabolite pathways and many other contiguous regions of interest in prokaryotic genomes, such as pathogenicity islands, can be directly cloned. The increased homologous recombination efficiency of full-length RecE has other implications. Among a variety of possibilities, we are exploring its capacity to assemble multiple linear pieces for synthetic biology or more efficient full-length cDNA library construction.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturebiotechnology/.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

Research work in R&D in Gene Bridges (Y.Z.) was partially funded by the Bundesministerium für Bildung und Forschung (MiPro). This work was supported by funding to A.F.S. from the EU 6th and 7th Framework projects, EUCOMM and EUCOMMTOOLS. Research in the laboratory of R.M. was funded by the Deutsche Forschungsgemeinschaft and the Bundesministerium für Bildung und Forschung. X.B. is supported by China Scholarship Council. The authors thank J. Herrmann in the laboratory of R.M. for cytotoxic activity assay.

AUTHOR CONTRIBUTIONS

J.F., A.F.S. and Y.Z. discovered the LLHR activity of full-length RecE. J.F., X.B., S.H., H.W., F.H., P.M.S., L.X. and Y.Z. made the DNA constructs and *E. coli* strains. X.B., A.P., R.M. and Y.Z. analyzed the secondary metabolites. J.F., X.B.,L.X., R.M., A.F.S. and Y.Z. designed the experiments and wrote the paper.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturebiotechnology/.

Published online at http://www.nature.com/naturebiotechnology/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Bode, H.B. & Müller, R. The impact of bacterial genomics on natural product research. *Angew. Chem. Int. Ed. Engl.* **44**, 6828–6846 (2005).
2. Koonin, E.V. & Wolf, Y.I. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res.* **36**, 6688–6719 (2008).
3. Lagesen, K., Ussery, D.W. & Wassenaar, T.M. Genome update: the 1000th genome–a cautionary tale. *Microbiology* **156**, 603–608 (2010).
4. Banik, J.J. & Brady, S.F. Recent application of metagenomic approaches toward the discovery of antimicrobials and other bioactive small molecules. *Curr. Opin. Microbiol.* **13**, 603–609 (2010).
5. Simon, C. & Daniel, R. Metagenomic analyses: past and future trends. *Appl. Environ. Microbiol.* **77**, 1153–1161 (2011).
6. Donadio, S., Monciardini, P. & Sosio, M. Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics. *Nat. Prod. Rep.* **24**, 1073–1109 (2007).
7. Hertweck, C. The biosynthetic logic of polyketide diversity. *Angew. Chem. Int. Edn Engl.* **48**, 4688–4716 (2009).
8. Fischbach, M.A. & Walsh, C.T. Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. *Chem. Rev.* **106**, 3468–3496 (2006).
9. Zhumabayeva, B., Chenchik, A. & Siebert, P.D. RecA-mediated affinity capture: a method for full-length cDNA cloning. *Biotechniques* **27**, 834 (1999).
10. Demidov, V.V. *et al.* Kinetics and mechanism of the DNA double helix invasion by pseudocomplementary peptide nucleic acids. *Proc. Natl. Acad. Sci. USA* **99**, 5953–5958 (2002).
11. Ito, T., Smith, C.L. & Cantor, C.R. Sequence-specific DNA purification by triplex affinity capture. *Proc. Natl. Acad. Sci. USA* **89**, 495–498 (1992).
12. Kouprina, N. & Larionov, V. Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast *Saccharomyces cerevisiae*. *Nat. Protoc.* **3**, 371–377 (2008).
13. Zhang, Y., Muyrers, J.P.P., Testa, G. & Stewart, A.F. DNA cloning by homologous recombination in *Escherichia coli*. *Nat. Biotechnol.* **18**, 1314–1317 (2000).
14. Yonemura, I. *et al.* Direct cloning of full-length mouse mitochondrial DNA using a *Bacillus subtilis* genome vector. *Gene* **391**, 171–177 (2007).
15. Zhang, Y., Buchholz, F., Muyrers, J.P. & Stewart, F.A. A new logic for DNA engineering using recombination in *Escherichia coli*. *Nat. Genet.* **20**, 123–128 (1998).
16. Muyrers, J.P.P., Zhang, Y., Testa, G. & Stewart, A.F. Rapid modification of bacterial artificial chromosomes by ET-recombination. *Nucleic Acids Res.* **27**, 1555–1557 (1999).
17. Copeland, N.G., Jenkins, N.A. & Court, D.L. Recombineering: a powerful new tool for mouse functional genomics. *Nat. Rev. Genet.* **2**, 769–779 (2001).
18. Muyrers, J.P., Zhang, Y., Buchholz, F. & Stewart, A.F. RecE/RecT and Redalpha/Redbeta initiate double-stranded break repair by specifically interacting with their respective partners. *Genes Dev.* **14**, 1971–1982 (2000).
19. Testa, G. *et al.* Engineering the mouse genome with bacterial artificial chromosomes to create multipurpose alleles. *Nat. Biotechnol.* **21**, 443–447 (2003).
20. Wang, J. *et al.* An improved recombineering approach by adding RecA to lambda Red recombination. *Mol. Biotechnol.* **32**, 43–53 (2006).
21. Sarov, M. *et al.* A recombineering pipeline for functional genomics applied to *Caenorhabditis elegans*. *Nat. Methods* **3**, 839–844 (2006).
22. Fu, J., Teucher, M., Anastassiadis, K., Skarnes, W. & Stewart, A.F. A recombineering pipeline to make conditional targeting constructs. *Methods Enzymol.* **477**, 125–144 (2010).
23. Kovall, R. & Matthews, B.W. Toroidal structure of lambda-exonuclease. *Science* **277**, 1824–1827 (1997).
24. Chu, C.C., Templin, A. & Clark, A.J. Suppression of a frameshift mutation in the recE gene of *Escherichia coli* K-12 occurs by gene fusion. *J. Bacteriol.* **171**, 2101–2109 (1989).
25. Zhang, J.J., Xing, X., Herr, A.B. & Bell, C.E. Crystal structure of *E. coli* RecE protein reveals a toroidal tetramer for processing double-stranded DNA breaks. *Structure* **17**, 690–702 (2009).
26. Duchaud, E. *et al.* The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens*. *Nat. Biotechnol.* **21**, 1307–1313 (2003).
27. Maresca, M. *et al.* Single-stranded heteroduplex intermediates in lambda Red homologous recombination. *BMC Mol. Biol.* **11**, 54 (2010).
28. Murphy, K.C. The lambda Gam protein inhibits RecBCD binding to dsDNA ends. *J. Mol. Biol.* **371**, 19–24 (2007).
29. Court, R., Cook, N., Saikrishnan, K. & Wigley, D. The crystal structure of lambda-Gam protein suggests a model for RecBCD inhibition. *J. Mol. Biol.* **371**, 25–33 (2007).
30. Zhang, Y., Muyrers, J.P., Rientjes, J. & Stewart, A.F. Phage annealing proteins promote oligonucleotide-directed mutagenesis in *Escherichia coli* and mouse ES cells. *BMC Mol. Biol.* **4**, 1 (2003).
31. Court, D.L., Sawitzke, J.A. & Thomason, L.C. Genetic engineering using homologous recombination. *Annu. Rev. Genet.* **36**, 361–388 (2002).
32. Mosberg, J.A., Lajoie, M.J. & Church, G.M. Lambda Red recombineering in *Escherichia coli* occurs through a fully single-stranded intermediate. *Genetics* **186**, 791–799 (2010).
33. Poteete, A.R. Involvement of DNA replication in phage lambda Red-mediated homologous recombination. *Mol. Microbiol.* **68**, 66–74 (2008).
34. Ellis, H.M., Yu, D.G., DiTizio, T. & Court, D.L. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc. Natl. Acad. Sci. USA* **98**, 6742–6746 (2001).
35. Bowen, D. *et al.* Insecticidal toxins from the bacterium *Photorhabdus luminescens*. *Science* **280**, 2129–2132 (1998).
36. Waterfield, N.R. *et al.* Rapid Virulence Annotation (RVA): identification of virulence factors using a bacterial genome library and multiple invertebrate hosts. *Proc. Natl. Acad. Sci. USA* **105**, 15967–15972 (2008).
37. Homburg, S., Oswald, E., Hacker, J. & Dobrindt, U. Expression analysis of the colibactin gene cluster coding for a novel polyketide in *Escherichia coli*. *FEMS Microbiol. Lett.* **275**, 255–262 (2007).
38. Gaitatzis, N., Hans, A., Müller, R. & Beyer, S. The *mtaA* gene of the myxothiazol biosynthetic gene cluster from *Stigmatella aurantiaca* DW4/3–1 encodes a phosphopantetheinyl transferase that activates polyketide synthases and polypeptide synthetases. *J. Biochem.* **129**, 119–124 (2001).
39. Amrein, H. *et al.* Functional analysis of genes involved in the synthesis of syringolin A by *Pseudomonas syringae* pv. *syringae* B301D-R. *Mol. Plant Microbe Interact.* **17**, 90–97 (2004).
40. Schellenberg, B., Bigler, L. & Dudler, R. Identification of genes involved in the biosynthesis of the cytotoxic compound glidobactin from a soil bacterium. *Environ. Microbiol.* **9**, 1640–1650 (2007).
41. Oka, M. *et al.* Glidobactin-A, glidobactin-B and glidobactin-C, new antitumor antibiotics. II. Structure elucidation. *J. Antibiot.* **41**, 1338–1350 (1988).
42. Kuzminov, A. Recombinational repair of DNA damage in *Escherichia coli* and bacteriophage lambda. *Microbiol. Mol. Biol. Rev.* **63**, 751–813 (1999).
43. Iyer, L.M., Koonin, E.V. & Aravind, L. Classification and evolutionary history of the single-strand annealing proteins, RecT, Redbeta, ERF and RAD52. *BMC Genomics* **3**, 8 (2002).
44. Erler, A. *et al.* Conformational adaptability of Redbeta during DNA annealing and implications for its structural relationship with Rad52. *J. Mol. Biol.* **391**, 586–598 (2009).
45. Yamamoto, T., Moerschell, R.P., Wakem, L.P., Komarpanicucci, S. & Sherman, F. Strand-specificity in the transformation of yeast with synthetic oligonucleotides. *Genetics* **131**, 811–819 (1992).
46. Swingle, B. *et al.* Oligonucleotide recombination in Gram-negative bacteria. *Mol. Microbiol.* **75**, 138–148 (2010).
47. Craig, J.W., Chang, F.Y., Kim, J.H., Obiajulu, S.C. & Brady, S.F. Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid environmental DNA libraries in diverse proteobacteria. *Appl. Environ. Microbiol.* **76**, 1633–1641 (2010).
48. Wenzel, S.C. *et al.* Heterologous expression of a myxobacterial natural products assembly line in pseudomonads via Red/ET recombineering. *Chem. Biol.* **12**, 349–356 (2005).
49. Fu, J. *et al.* Efficient transfer of two large secondary metabolite pathway gene clusters into heterologous hosts by transposition. *Nucleic Acids Res.* **36**, e113 (2008).

ONLINE METHODS

Bacterial genomic DNA isolation. *P. luminescens* subsp. *laumondii* TTO1 was cultured in 2 ml Luria broth (LB) medium at 30 °C for 18 h. After centrifugation the cells were resuspended in 450 µl of water. 30 µl of 20% SDS and 30 µl of 20 mg/ml proteinase K were added and incubated at 40 °C for 3 h until the solution became clear. DNA was obtained from the lysis mix by phenol-chloroform extraction and ethanol precipitation. The DNA pellet was resuspended in ddH_2O.

Bacterial strains and culturing conditions. All recombineering was performed in *E. coli* GB2005 and its derivatives cultured in LB medium and antibiotics (chloramphenicol [Cm], 15 µg/ml; kanamycin [Km], 15 µg/ml; ampicillin [Amp], 100 µg/ml; and tetracycline [Tet], 5 µg/ml) or low salt LB and hygromycin B [Hyg], 30 µg/ml). The strains used were: GB05, derived from DH10B by deletion of *fhuA*, *ybcC* and *recET*[22,49]. GB05-dir, derived from GB2005 by the P_{BAD}-*ETgA* operon, was integrated into the *ybcC* locus in GB2005 to create GB05-dir. The integration ablates expression of *ybcC*, which encodes a putative exonuclease similar to that encoded by *Redα*. GB05-red; derived from GB2005 by insertion of the P_{BAD}-*gbaA* cassette at the *ybcC* locus[22,49]. GB05-pir was constructed by insertion of *pir* into the *ybcC* locus[27].

pSC101 expression plasmids and plasmid construction. All expression plasmids used here were based on pSC101-BAD-gba[20] with the relevant open reading frames replaced by recombineering using YZ2005 (ref. 13). Plasmids were constructed by lambda Red- or RecET-mediated recombineering or conventional DNA ligation with T4 DNA ligase. Genes encoding RecE and RecT were PCR amplified from DH10B genomic DNA. The rhaSR operon, rhaBAD promoter, tetR and tetO promoter are from pRedFlp[21].

Preparation of electrocompetent cells for recombineering. From overnight cultures containing the expression plasmid (OD_{600} around 3 to 4), 40 µl was diluted into 1.4 ml LB medium with appropriate antibiotics and grown at 30 °C, 1,000 r.p.m. for 2 h. After addition of the inducer (final concentrations 2.5 mg/ml L-(+)-arabinose for araBAD promoter, 2.5 mg/ml L-rhamnose for rhaBAD promoter or 0.2 µg/ml anhydrotetracycline (AHT; Sigma-Aldrich) for tetR repressed tetO promoter), the cells were grown at 37 °C at 1,000 r.p.m. for 45 min. Cells were then centrifuged for 30 s at 9,500 r.p.m. at 2 °C. The supernatant was discarded and the cell pellet resuspended in 1 ml ice-cold water and centrifuged. The ice-cold water washing was repeated once more. After that, cells were resuspended in 30 µl ice-cold water and DNA was added. 0.1 µg of PCR products plus 0.1 µg of circular vector or linear vector were used in experiments of **Figures 1**–**3**. 0.5 µg of linear plasmid plus ~5.0 µg of digested *P. luminescens* genomic DNA were used for the direct cloning experiments. Electroporation was performed using ice-cold cuvettes and an Eppendorf electroporator 2510 set at 1350 V. One milliliter LB medium was added after electroporation. The cells were incubated at 37 °C for 70 min with shaking and spread on appropriate antibiotic plates.

Expression, extraction and analysis of the compounds. The *E. coli* strains harboring the *plu* gene clusters were incubated in 15-ml glass tubes containing 5 ml LB plus ampicillin (100 µg/ml). The culture was inoculated from an overnight culture (1:50) with shaking for 4 h at 30 °C, 150 r.p.m. After induction with tetracycline (0.5 µg/ml, water as control), the culture was incubated for another 4 h and 2% of adsorber resin Amberlite XAD-16 was added, then cultivated for 1–2 d. The resin and biomass were harvested by centrifugation and then extracted with 0.5 ml acetone first and 1 ml methanol later. The extracts were evaporated to dryness and dissolved in 400 µl methanol before analysis by LC/MS. Standard analysis of crude extracts was measured on an HPLC-DAD system from the Agilent 1100 series, coupled to a Bruker Daltonics HCTultra ESI-MS ion trap instrument operating in positive and negative ionization mode. The chromatographic conditions were: Luna RP-C_{18} column, 100 by 2 mm, 2.5-µm particle size. Solvent gradient (with solvents A [water and 0.1% of formic acid] and B [acetonitrile and 0.1% of formic acid]) from 5% B at 2 min to 95% B within 20 min, followed by 3 min with 95% B at a flow rate of 0.4 ml/min. Detection was recorded by both diode array and ESI-MS. High-resolution mass spectrometry was done on Accela UPLC-system (Thermo-Fisher) coupled to a linear trap-FT-Orbitrap combination (LTQ-Orbitrap) in positive ionization mode with a Waters BEH RP-C_{18} column (50 × 2 mm; 1.7 µm particle diameter, flow rate 0.6 ml/min) with a mobile phase of water/acetonitrile each containing 0.1% of formic acid, using a gradient from 5–95% of acetonitrile over 9 min. The corresponding products of *plu* gene cluster were identified by manual comparative metabolic profiles between tetracycline-induced and uninduced HPLC/MS chromatograms.

Addition of *glbH* to the luminmycin gene cluster. The *glb-Cm* PCR product, used to introduce the *glbH* gene into pGB-amp-P_{tetO}-plu1881-1877, was generated using oligonucleotides glbH1880u and glbH1880b (**Supplementary Table 1c**) with p15A-P_{BAD}-glb-cm (X.B., Y.Z. and R.M., unpublished data) as a template. The targeting DNA fragment *glbH-Cm* was electroporated into GB05-Red competent cells in which the pGB-amp-P_{tetO}-plu1881-1877 was resident. Recombinants were selected on LB plates containing 15 µg/ml of chloramphenicol to get pGB-P_{tetO}-plu1881-1877-glbH-cm. In this construct, *tetR* was also deleted and this fusion gene cluster was under the control of the P_{tetO} constitutive promoter.

Feeding studies. Feeding experiments were performed in 5 ml LB medium plus corresponding commercially available d_8-L-Valine (Deutero GmbH), d_3-L-Leucine (Deutero GmbH), ^{15}N-L-isoleucine (Sigma-Aldrich) and $^{13}C_6$-L-Phenylalanine (Cambridge Isotope Laboratories, Inc.), d_1-L-alanine (Isotec) (2.5 mg), respectively. Stocks were dissolved in LB medium and sterile filtered.

Bioinformatics analysis of sequences. Annotations of the catalytic domains from the *plu* gene clusters were carried out using PKS-NRPS Analysis[50] and NRPS-PKS[51] in combination with NCBI Blast analysis[52] and the Pfam search engine[53]. The specificities of AT domains from PKS were predicted by alignments against known AT sequences[54]. The substrates of NRPS modules were predicted by comparison of binding pocket signature sequences extracted from A domains employing PKS-NRPS analysis[50] and NRPSpredictor[55]. All the sequences alignments were performed by Clustal W[56].

Extraction and purification of luminmide A (4) and B (3). The overnight *E. coli* GB05-MtaA/pGB-P_{tetO}-plu3263 culture was inoculated in 10.5 L LB medium (1.75 liters × 6) in the presence of 50 µg/ml ampicillin and 2% XAD-16 adsorber resin (added 8 h later) at 30 °C for 30 hours in shaking flasks (1.75 L/5 L; 150 r.p.m.). The XAD was separated from the supernatant by sieving, washed with 2 L dH_2O twice and extracted batchwise with a total volume of 3 liters methanol (1 liter × 3), 3 liters acetone (1 liter × 3) and 3 liters ethyl acetate (1 liter × 3). The resulting extracts were analyzed by LC/MS and most of the compounds were in the methanol extract (10.06 g). Then, this extract was partitioned with equal volume ethyl acetate and dH_2O to obtain 0.85 g extract from ethyl acetate phase. In parallel, the biomass was centrifuged and extracted by 2 liters methanol (1 liter × 2) to get 1.96 g extracts and then partitioned with *n*-butanol and dH_2O (1:1) to 0.7 g extract, followed by one more partition with ethyl acetate and dH_2O (1:1) to gain 56.7 mg extracts. The two ethyl acetate fractions were combined and loaded on a Sephadex LH-20 column and methanol as mobile phase. The samples were collected per 15 min one tube (around 10 mL). Fractions containing significant amounts of 4 were combined and evaporated to dryness. The further purification was performed on repeated semi-preparative HPLC instrument equipped with UV detector (220 nm, and 280 nm). An RP-C12 column (Jupiter, Proteo 90A, 250 × 10 mm, 4 µm) was used for further purification with a solvent system consisting of water (A) containing 0.05% TFA and acetonitrile (B). The gradient was as follows: 0–5 min 50% B, 5–15 min 50–70% B, 15–40 min 70% B. Around 2 mg luminmide A (4) was afforded at the retention time of 26.1–26.8 min as a white powder. For the NMR data of luminmide A, see **Supplementary Table 5**. For the isolation of luminmide B (3), 18 liters culture of *E. coli* GB05-MtaA/pGB-P_{tetO}-plu3263-A3-C278M (X.B., Y.Z. and R.M., unpublished data) was cultivated as before. The resins were washed with 2 liters dH_2O twice and 2 liters 20% MeOH twice, then extracted stepwise with a total volume of 2 liters hexane (2 liters × 1), 6 liters ethyl acetate (2 liters × 3). The EtOAc extract was reverse partitioned with EtOAc-H_2O (1:1) to afford a dried extract (1.3 g). The EtOAc extract was fractionated initially on a Sephadex LH-20 column (100 × 2.5 cm)

using MeOH as mobile phase, and more than 50 fractions were obtained. Fraction 16 (2.7 mg) and 17 (4.3 mg) were dissolved into DMSO and were further purified by repeated semi-preparative HPLC as for luminmide A. Around 0.8 mg luminmide B was obtained by collection at the retention time of 25.1–26 min as a white powder. The luminmide B was dissolved into DSMO-d_6 for NMR recording; for data see **Supplementary Table 6**.

Advanced Marfey's analysis. Approximately 0.2 mg of 3 and 4 were hydrolyzed with 6 N HCl (0.8 mL) at 90 °C for 16 h. These solutions were evaporated to dryness, and the residue was dissolved in 100 µL water. To each a half portion (50 µL) were added 1 N $NaHCO_3$ (20 µL) and 1% 1-fluoro-2, 4-dinitrophenyl-5-L-Leucinamide (L-FDLA or D-FDLA in acetone, 100 µL), and the mixture was vortexed and incubated at 37 °C for 60 min. After the reaction was quenched by the addition of 2 N HCl (20 µL) and evaporated to dryness. The residues were resuspended in 300 µL of acetonitrile and about 10 µL of each solution of FDLA derivatives was analyzed by LC/MS same to the general procedure (expression, extraction and analysis of the compounds). The retention times of the Marfey-derivatized amino acids are summarized in **Supplementary Table 7**. Retention times of the FDLA-derivatized authentic standards are L-Val 15.5 min, D-Val 17.1 min *m/z* 412.2 $[M+H]^+$; L-Phe16.4 min, D-Phe 17.7 min *m/z* 460.3 $[M+H]^+$; L-Leu 16.2 min, D-Leu 18.2 min *m/z* 426.2 $[M+H]^+$.

50. Bachmann, B.O. & Ravel, J. Methods for in silico prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data. *Methods Enzymol.* **458**, 181–217 (2009).
51. Ansari, M.Z., Yadav, G., Gokhale, R.S. & Mohanty, D. NRPS-PKS: a knowledge-based resource for analysis of NRPS/PKS megasynthases. *Nucleic Acids Res.* **32**, W405–W413 (2004).
52. 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).
53. Bateman, A. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **28**, 263–266 (2000).
54. Del Vecchio, F. *et al.* Active-site residue, domain and module swaps in modular polyketide synthases. *J. Ind. Microbiol. Biotechnol.* **30**, 489–494 (2003).
55. Rausch, C., Weber, T., Kohlbacher, O., Wohlleben, W. & Huson, D.H. Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). *Nucleic Acids Res.* **33**, 5799–5808 (2005).
56. Larkin, M.A. *et al.* Clustal W and clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).

doi:10.1038/nbt.2183