Corrections

MICROBIOLOGY

The authors request that Guowei Yang, Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, United Kingdom, be added to the author list, between Andrea Dowling and Paul Wilkinson, and be credited with performing research. The author line has been corrected online. The corrected author and affiliation lines, and related footnotes, appear below.


*Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, United Kingdom; §Department of Biological Sciences, University of Exeter in Cornwall, Penryn TR10 9EZ, United Kingdom; ¶Pathogen Sequencing Group, Wellcome Trust Sanger Institute, Cambridge CB10 1SA, United Kingdom; and †Department of Pharmaceutical Biotechnology, Saarland University, 66123 Saarbrücken, Germany


†N.R.W and M.S.-C. contributed equally to this work.

‡To whom correspondence should be addressed. E-mail: bssnw@bath.ac.uk.

www.pnas.org/cgi/doi/10.1073/pnas.0811993106

GENETICS

The authors note that on page 1934, left column, line 12, “70 nM” should instead appear as “70 mM.” This error does not affect the conclusions of the article.

www.pnas.org/cgi/doi/10.1073/pnas.0811896106

www.pnas.org PNAS
February 10, 2009 vol. 106 no. 6 2083
Correction

MICROBIOLOGY

The authors request that Guowei Yang, Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, United Kingdom, be added to the author list, between Andrea Dowling and Paul Wilkinson, and be credited with performing research. The author line has been corrected online. The corrected author and affiliation lines, and related footnotes, appear below.


*Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, United Kingdom; §Department of Biological Sciences, University of Exeter in Cornwall, Penryn TR10 9EZ, United Kingdom; ¶Pathogen Sequencing Group, Wellcome Trust Sanger Institute, Cambridge CB10 1SA, United Kingdom; and ‖Department of Pharmaceutical Biotechnology, Saarland University, 66123 Saarbrücken, Germany


†N.R.W. and M.S.-C. contributed equally to this work.

‡To whom correspondence should be addressed. E-mail: bssnw@bath.ac.uk.

www.pnas.org/cgi/doi/10.1073/pnas.0811896106
Rapid Virulence Annotation (RVA): Identification of virulence factors using a bacterial genome library and multiple invertebrate hosts


*Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, United Kingdom; †Department of Biological Sciences, University of Exeter in Cornwall, Penryn TR10 9EZ, United Kingdom; ‡Pathogen sequencing group, Wellcome Trust Sanger Institute, Cambridge CB10 1SA, United Kingdom; and §Department of Pharmaceutical Biotechnology, Saarland University, 66123 Saarbrücken, Germany

Edited by Frederick M. Ausubel, Harvard Medical School, Boston, MA, and approved July 25, 2008 (received for review November 23, 2007)

Current sequence databases now contain numerous whole genome sequences of pathogenic bacteria. However, many of the predicted genes lack any functional annotation. We describe an assumption-free approach, Rapid Virulence Annotation (RVA), for the high-throughput parallel screening of genomic libraries against four different taxa: insects, nematodes, amoeba, and mammalian macrophages. These hosts represent different aspects of both the vertebrate and invertebrate immune system. Here, we apply RVA to the emerging human pathogen Photorhabdus asymbiotica using “gain of toxicity” assays of recombinant Escherichia coli clones. We describe a wealth of potential virulence loci and attribute biological function to several putative genomic islands, which may then be further characterized using conventional molecular techniques. The application of RVA to other pathogen genomes promises to ascribe biological function to otherwise uncharacterized virulence genes.

The growing speed with which the genomes of bacteria can be sequenced is producing an ever expanding knowledge gap between sequence data and its functional annotation. In the case of bacterial pathogens, the identification of virulence factors has typically relied on genetic knock-out and demonstration that the bacteria and leave in search of new prey (7). The requirement to overcome the insect immune system and to keep the insect cadaver free of saprophytic organisms in the soil means that all Photorhabdus produce a range of bioactive molecules including immune inhibitors, toxins, and powerful antimicrobials. The fully annotated genome sequence of the insect-only pathogen Photorhabdus luminescens strain TT01 is available (8) and the genome sequence of the clinical isolate Photorhabdus asymbiotica ATCC43949 is almost completed (http://www.sanger.ac.uk/Projects/P.asymbiotica/). The availability of the two genomes allows us to correlate the RVA data with regions unique to the human pathogenic P. asymbiotica.

For RVA analysis we sheared the P. asymbiotica ATCC43949 genome and cloned it into recombinant Escherichia coli. The library, covering 91.4% of the estimated 5.0 Mb genome with an average insert size of 37 kb, was arrayed into 16 96-well plates. All 1,536 clones were sequenced at both ends and end-sequences were assembled onto the genomic scaffold. This cosmid library was screened for gain of toxicity (GOT) assays against the nematode Caenorhabditis elegans (nGOT), serving as an oral route model; the single-cell protozoa Acanthamoeba polyphaga (aGOT), used as a phagocytosis model; and two caterpillar models (iGOT), the tobacco hornworm Manduca sexta and the Waxmoth Galleria mellonella, both of which represent the more complex insect immune systems. Numerous examples of the use of invertebrates as model hosts can be found in the literature (1, 9–12). Finally, we used the mouse BALB/c macrophage cell line J774–2 (mGOT) to represent the phagocytic component of the vertebrate immune system. The use of GOT studies in E. coli, in which the model host is challenged with individual clones, has the advantage over chromosomal mutagenesis of “unmasking” any virulence factors that would otherwise be hidden due to toxin redundancy or the presence of potent dominant toxins. To define the virulence-related regions (RVA regions), the end sequences of cosmids showing an effect were assembled onto a

Author contributions: N.R.W. and R.f.-C. designed research; M.S.-C., I.E., A.D., and G.Y. performed research; N.R.W., P.W., I.P., N.T., S.E.R., H.B.B., and S.D. analyzed data; and N.R.W., M.S.-C., and R.f.-C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in the EMBL database (accession nos. FM211043–FM211060).

†These authors contributed equally.

© 2008 by The National Academy of Sciences of the USA
Figs. S1–S25, and Tables S1 and S2. Figs. S2–S22 show corre-
quences of the 21 RVA regions identified are presented in Table
complete RVA dataset and more detailed diagrammatic sum-
maries of the 21 RVA regions identified are presented in Table
and in supporting information (SI) Materials and Methods,
Figs. S1–S25, and Tables S1 and S2. Figs. S2–S22 show corre-
lations between gene clusters and the effects on the mammalian
and invertebrate model targets. For brevity, here we will only
discuss specific examples of the different classes of virulence
factors detected, illustrating the robust nature of RVA-based
screening.

Fig. 1. RVA functional genomics map for P. asymbiotica. Boxes on the top layer represent the RVA regions (see Figs. S2–S22). The two large contigs of the
P. asymbiotica ongoing genome sequencing project are shown. “Pas” regions contain genes that do not identify homologues in P. luminescens TT01 (at 75%
identity level). The lower panel represents the actual genomic locations of the cosmids from the library analyzed by RVA screen. The dotted line represents a
sequencing gap, however, the order and orientation of the two large contigs is confirmed by the correct end sequence alignment of five cosmids across this
region. Note that a sequence gap of unknown size also exists so the other ends of the contigs are not contiguous.

Results

The application of RVA to P. asymbiotica gave a high detection
rate of candidate gene clusters, encoding virulence factors, which is a reflection of both the sensitivity of RVA method and also the
high level of redundancy in encoded pathogenesis factors in
Photorhabdus bacteria (7, 8, 13). The RVA regions were aligned
against the current assembly of the P. asymbiotica genome (Fig.
1), each region representing a cluster of cosmids identified either
in a single screen (with a minimum of two overlapping cosmids)
or via a number of parallel screens in combination (Fig. 2).

The complete RVA dataset and more detailed diagrammatic summaries of the 21 RVA regions identified are presented in Table
1 and in supporting information (SI) Materials and Methods,
Figs. S1–S25, and Tables S1 and S2. Figs. S2–S22 show corre-

Fig. 2. Biologically active cosmids in the RVA20 region. Boxes above and below the top central line represent ORFs on the forward and reverse strands. RVA
positive cosmids are represented below as bars. Hatched bars represent cosmids containing only intact mcf1. Black bars represent cosmids containing intact mcf1
and at least one of the up- or downstream NRPS clusters. Open bars represent cosmids encoding the NRPS downstream of mcf1 but not the intact gene.

Re-Identification of Known Virulence Factors. The effectiveness of
RVA is confirmed by the reidentification of previously known virulence factors, also providing an excellent “internal” positive
control. Thus, as anticipated, RVA detected mcf1 (RVA20) that
encodes a dominant insecticidal toxin (14) and also the phage-
related toxin delivery system PVCpfn (RVA21) (15). Some cosmids containing part or the full PVCpfn cluster, detected in
iGOT, aGOT, and mGOT assays, also contain a hemagglutinin-
like gene. However, one iGOT cosmid (Fig. S22) carries a
complete copy of PVCpfn along with a truncated nonribosomal
peptide synthase (NRPS), indicating that PVCpfn alone exhibits
toxicity to insects as previously demonstrated (15). Cosmids
containing mcf1 were identified in all types of parallel screens,
consistent with the ability of Mcf1 to cause apoptosis in both
insect and mammalian tissues (16). Fig. 2 illustrates how a
detailed examination of the cosmid locations relative to the
genome can reveal the contribution of different gene clusters to
toxicity. In several of the mcf1-containing cosmids, the further
involvement of tightly linked upstream and downstream NRPS
genes in the nGOT and iGOT screens could not be ruled out.
Conversely, several cosmids were identified in the mGOT and
aGOT assays, which contained only intact mcf1 and one cosmid
from each assay that contained only the downstream NRPS
cluster. Cosmid 3AG4, carrying only mcf1 (identified by aGOT),
was tested against the other models resulting in toxic phenotypes,
severe feeding delay in the nematodes, insect death at 24 h, and
an exhibition of the typical “floppy” phenotype consistent with
intoxication by Mcf1 (14).

Identification of Virulence Factor Homologues. Further confidence
in RVA comes from the detection of homologues of virulence factors from other organisms. A “type VI” virulence secretion
system described in Vibrio cholerae V52 is proposed to provide a
target for vaccines and therapeutic agents (17). The P. asym-
biotica RVA screen identified a highly homologous gene cluster
(RVA11) detected in the iGOT assay that is 46–90% similar at
the predicted amino acid level (Figs. S12 and S23). Indeed,
subsequent work has confirmed the importance of a homologous
operon in Burkholderia mallei (18). Interestingly, cosmids toxic
to macrophages (mGOT) also span an adjacent region that contains the putative vgrG toxin unique to the human pathogen *P. asymbiotica* and absent from *P. luminescens*, potentially representing a genomic island specialized for mammalian pathogenicity. Further, an RVA region showing activity in all assays (locus tag PA-RVAs–3205 to 3199, Fig. S6) is homologous (48–72% similar at the amino acid level) to the type VI like evp operon involved in virulence in the fish pathogen *Edwardsiella tarda* (19, 20).

**Functional Annotation of Secondary Metabolite Gene Clusters.** It is notoriously difficult to ascribe biological function to gene clusters responsible for the synthesis of secondary metabolites such as polyketide synthesis (PKS) and non-ribosomal peptide synthesis (NRPS) enzyme complexes. Like many pathogens, *Photorhabdus* dedicates a large amount of coding sequence to the manufacture of small molecules but their biological roles and structures remain unknown. RVA identified a number of regions predicted to encode NRPS/PKS-like complexes and Fig. 3A illustrates four examples of identified cosmids that exhibit a range of activities including the ability of the recombinant *E. coli* to kill *G. mellonella* upon injection (Fig. 3B). HPLC/MS analysis of preparations from cosmids 4DB5 (RVA5) and 1DF9 (RVA14) identified 265.4/468.4 and 586.4 as specific [M+H]+ masses for the respective molecules, which were not present in the *E. coli* control (Fig. S24). Two compounds were identified from cosmids 4DB5 (RVA5) encoding a gene cluster with high similarity to the yersiniabactin biosynthesis genes. The postulated sum formula C32H51O5N5, from 1DF9 cosmid (RVA14), fits well with the structure of a cyclic pentapeptide. Interestingly, an identical sum formula is found in sansalvamide A peptides (22–24), synthetic derivatives of the depsipeptide natural product sansalvamide A that were isolated from a marine fungus and showed potent anti-cancer activity (25). Isolation of all identified compounds for detailed structure elucidation is currently underway. Preliminary structural predictions of other secondary metabolites identified by RVA are presented in Table S1. One example is the predicted compound encoded by RVA 18 (Fig. S19) similar to a new antitumor antibiotic: glidobactin from *Polyangium brachysporum* (26). Two PKS clusters and one NRPS gene cluster were also encoded in mGOT RVA regions, suggesting that secondary metabolites produced from these operons could have potential roles in human pathogenicity. The rapid identification of new biological activities from these secondary metabolite synthetic genes suggests that RVA analysis can be a powerful method of drug discovery.

## Table 1. Regions of the *P. asymbiotica* ATCC43949 genome identified by RVA analysis

<table>
<thead>
<tr>
<th>RVA region</th>
<th>Accession number (locus tag range)</th>
<th>Assay</th>
<th>Gene clusters and predicted function</th>
</tr>
</thead>
<tbody>
<tr>
<td>RVA1</td>
<td>FM211043 (PA-RVA1–4467–4422)</td>
<td>iGOT nGOT</td>
<td>Type 3 secretion system genes and exoU effector</td>
</tr>
<tr>
<td>RVA2</td>
<td>FM211044 (PA-RVA2–4341–4267)</td>
<td>iGOT mGOT</td>
<td>Hemolysin and PKS modules</td>
</tr>
<tr>
<td>RVA3</td>
<td>FM211045 (PA-RVA3–4024–3964)</td>
<td>iGOT mGOT nGOT aGOT</td>
<td>Hemagglutinin-like gene and two NRPS clusters</td>
</tr>
<tr>
<td>RVA4</td>
<td>FM211046 (PA-RVA4–3953–3929)</td>
<td>mGOT</td>
<td>aaXAB homologues and hemagglutinin-like gene</td>
</tr>
<tr>
<td>RVA5</td>
<td>FM211047 (PA-RVAs–3244–3168)</td>
<td>iGOT mGOT nGOT aGOT</td>
<td>PVCumlpt, evp operon and hemagglutinin</td>
</tr>
<tr>
<td>RVA6</td>
<td>FM211048 (PA-RVA6–3081–3019)</td>
<td>iGOT nGOT</td>
<td>Fimbrial operon and Pdl-GI-1 island</td>
</tr>
<tr>
<td>RVA7</td>
<td>FM211049 (PA-RVA7–0626–0673)</td>
<td>iGOT mGOT nGOT aGOT</td>
<td>Invasin and unknown genes</td>
</tr>
<tr>
<td>RVA8</td>
<td>FM211050 (PA-RVA8–2849–2974)</td>
<td>nGOT</td>
<td>Unknown</td>
</tr>
<tr>
<td>RVA9</td>
<td>FM211051 (PA-RVA9–1890–1949)</td>
<td>iGOT mGOT aGOT</td>
<td>Prophage</td>
</tr>
<tr>
<td>RVA10</td>
<td>FM211052 (PA-RVA10–2214–2246)</td>
<td>mGOT aGOT</td>
<td>Fimbrial operon</td>
</tr>
<tr>
<td>RVA11</td>
<td>FM211053 (PA-RVA11–2280–2352)</td>
<td>iGOT mGOT nGOT aGOT</td>
<td>Hemagglutinin and type VI secretion system</td>
</tr>
<tr>
<td>RVA12</td>
<td>FM211054 (PA-RVA12–2456–2515)</td>
<td>nGOT mGOT</td>
<td>Unknown</td>
</tr>
<tr>
<td>RVA13</td>
<td>FM211055 (PA-RVA13–1315–1206)</td>
<td>iGOT nGOT mGOT</td>
<td>Pdl-Gl-4 and ast enterotoxin homologue</td>
</tr>
<tr>
<td>RVA14</td>
<td>FM211056 (PA-RVA14–1110–1068)</td>
<td>iGOT aGOT</td>
<td>Hemagglutinin-like and NRPS cluster</td>
</tr>
<tr>
<td>RVA15</td>
<td>FM211057 (PA-RVA15–17–0915–0951)</td>
<td>iGOT aGOT</td>
<td>rtXa homologue, NRPS and evp-like operon</td>
</tr>
<tr>
<td>RVA16</td>
<td>FM211057 (PA-RVA15–17–0951–0992)</td>
<td>iGOT mGOT nGOT aGOT</td>
<td>vgrG, rtXa homologues and NRPS</td>
</tr>
<tr>
<td>RVA17</td>
<td>FM211057 (PA-RVA15–17–1004–1042)</td>
<td>iGOT nGOT mGOT</td>
<td>kdp operon and a perforin-like gene</td>
</tr>
<tr>
<td>RVA18</td>
<td>FM211058 (PA-RVAs–1595–1632)</td>
<td>iGOT mGOT</td>
<td>NRPS and PKS-like gene clusters</td>
</tr>
<tr>
<td>RVA19</td>
<td>FM211059 (PA-RVA19–1780–1813)</td>
<td>iGOT</td>
<td>Pdl-Gl-2 and putative virulence factor mviN</td>
</tr>
<tr>
<td>RVA20</td>
<td>FM211060 (PA-RVA20–21–0210–0176)</td>
<td>iGOT mGOT nGOT aGOT</td>
<td>Two NRPS clusters and mcf1 cytotoxin</td>
</tr>
<tr>
<td>RVA21</td>
<td>FM211060 (PA-RVA20–21–0176–0072)</td>
<td>iGOT mGOT aGOT</td>
<td>PVCpfn and hemagglutinin-like gene</td>
</tr>
</tbody>
</table>

Illustrations of this information can be seen in Figs. S2–S22. Full annotation of these regions can be found in the EMBL entries indicated by accession numbers.
Interestingly, cloned pdl-flanking orf54 cosmid had negligible effect. (27) The genes all show reduced toxicity when injected to insects affected by pdl genes for fine-scale mapping (see below). Moreover, the toxicity islands of pdl-o54 tightly linked gene responsible for release of Tcd from the outer membrane of the bacteria. Interestingly, other Gram-negative pathogens, including homologues in the Pa pdl1-GI_2 island are also present in the genomes of many other pathogens. The RVA analysis of P. asymbiotica here identified the ORF responsible. In the case of the Pa pdl1-GL2 virulence island. The insertion mutants of the two pdl genes and the vgrG gene all show reduced toxicity when injected to G. mellonella, while the wild-type cosmid (4DE10) is fully toxic. Injection of control E. coli (pWEB cosmid) had negligible effect. (27) The vgrG toxin with or without the small flanking orf54 were cloned into the arabinose-inducible expression vector pBAD30. When induced, these clones were toxic by injection to G. mellonella. Interestingly, cloned orf54 reduced this effect suggesting its function is antagonistic to the toxic phenotype.

Identification of Virulence Gene Clusters. Our recent work on secretion of the orally active Toxin Complex d (Tcd) from P. luminescens indicates that a putative class III lipase pdl (27) is responsible for release of Tcd from the outer membrane of the bacterium and that this process is negatively regulated by the tightly linked gene orf54 (G. Yang and N.R.W., unpublished work). We detected several regions encoding pdl-o54 gene homologues in the P. asymbiotica genome that were toxic to either insects and/or nematodes. We therefore speculate that these represent virulence loci (Fig. 4A, and Figs. S7 and S20). We note that these regions have no effect on the mammalian macrophages again suggesting some specificity in their toxicity. Interestingly, pdl-islands are also present in the genomes of many other Gram-negative pathogens, including Vibrio cholerae, suggesting they also may be virulence determinants in other human pathogens. In characterizing RVA regions such as these, determination of the minimal region of overlap between “toxic” cosmids at any given locus may not provide sufficient clues to suggest the exact genes responsible. Thus, we carried out insertional mutagenesis to map the genes causing the phenotype by using an in vitro mutagenesis kit to construct insertion mutant libraries of toxic cosmids. Clones with insertions in different genes are then identified by sequencing out from the integrated transposon. The mutated cosmids are rescreened in the appropriate RVA assay to identify clones showing loss of function and therefore the ORF responsible. In the case of the Pa pdl1-GL2 insecticidal region, we selected the 4DE10 toxic cosmid for this fine-scale mapping (Fig. 4B). This identified vgrG and two of the pdl-homologues as necessary for the toxic effect on G. mellonella. Subcloning and heterologous expression of vgrG in E. coli (Fig. 4C) confirmed the toxicity of this gene product and serves to illustrate how RVA allows us to move from a whole genome to an individual toxin gene with ease.

Importantly, the RVA approach is also powerful in identifying genes with functions that could not be predicted using conventional genetic approaches. Three cosmids were identified in the iGOT screen, 3CD1, 3DB8, and 1AA3 (RVA17, Fig. S18), causing a moribund phenotype when injected to M. sexta. Fine-scale mapping traced the toxic effect to kdpEDCBAF encoding the high affinity potassium pump (28) of P. asymbiotica (Fig. S25). Over-expression of the Kdp pump appears to allow the bacteria to persist and even grow in the hemocytes (I. Vlisidou and N.R.W., unpublished work). This striking fundamental discovery is likely to be relevant to all immune cell-pathogen interactions, which rely upon high potassium gradients for maturation of the phagolysosome (29). We propose that the assumption-free RVA screening technology is ideal for making such discoveries and expect many more from a range of pathogens.

**Virulence Against Insect or Man?** It was anticipated that in the case of P. asymbiotica, a comparison between the invertebrate and macrophage (mGOT) RVA data could provide candidate genes responsible for facilitating human infection. Cosmids in RVA4 were identified by the mGOT screen but not in either of the invertebrate assays. This region does not encode genes unique to the human pathogenic P. asymbiotica and, therefore, has homologues in the insect-only pathogen P. luminescens TT01. RVA4 (Fig. S5) contains homologues of the proapoptotic xaxAB from Xenorhabdus (30) and a large hemolysin/hemagglutinin gene, either of which may represent potential virulence factors. Interestingly, the majority of mGOT regions are also shown to have effects in invertebrate hosts, supporting our hypothesis that genes evolved to combat invertebrates can be redeployed against mammalian hosts (3).

**Discussion**

Here, we have demonstrated the ability of RVA to identify virulence loci via the parallel screening of a recombinant DNA library in a non-pathogenic laboratory strain of E. coli against a range of different taxa. The RVA analysis of P. asymbiotica identified a range of different putative virulence factors including eight NRPS operons, two PKS operons, six hemolysin/hemagglutinin-like genes, three pilin/fimbrial operons, seven putative specialized secretion systems (PVCpnpf, T6SS, Evp-islands, T3SS and Pil-islands) and at least seven toxin-like genes (homologues of mcf1, xaxAB, vgrG, rtxA, xnt2, ast, and a VIP2 family gene). The identification of such a large number of virulence loci illustrates that RVA is useful as a means of generating preliminary biological annotation relating to pathogenic phenotypes. RVA annotation highlights operons or genes as potentially important in virulence that can then be investigated in greater detail either for accurate annotation or more specific applications (i.e., drug discovery or vaccine candidates). Therefore, fine-scale mapping of RVA regions, by transposon mutagenesis or cloning individual ORFs, is essential to confirm the exact genes responsible for toxicity in each cluster. The use of different invertebrate targets with less complex immune responses (insects, nematodes, and protozoa) allows the detection of less “potent” virulence factors that could otherwise be overlooked in whole animal mammalian models. It is interesting that most of the virulence factors identified in the mGOT assay were also toxic in the invertebrate screens confirming the general utility of invertebrates as model hosts for disease studies. Although it is likely that a small subset of host-specific virulence factors will not be detected by RVA, the screens are nevertheless sensitive enough to detect a very large number of more numerous general virulence factors.

The power of the RVA technique relies on several factors. First and highly important is the coverage and the depth of the
The genomic library used that covered >90% of the genome, with nearly 80% being covered by two or more clones. The analysis of the cosmid library distribution on the genome assembly confirmed that some regions are more represented than others and certain regions are absent. In future RVA studies it may be pertinent to "normalize" the library by selecting a subset of cosmids for the assays, reducing the level of redundancy in testing while still maximizing genome coverage. The second factor is the underlying genetic architecture of virulence factors. RVA's strength is in the identification of single-loci virulence factors and multilocus virulence factors that are tightly clustered within a bacterial genome. Molecular characterization of virulence factors to date suggests that the majority fall into these categories, further justifying our screening approach. However, there are several notable examples of virulence factors for which the underlying genetics basis is complex and involves multiple genes or regions of the genome, and these will naturally not be identified by RVA. This limitation can be overcome by using vectors that accommodate larger fragments such as BACs.

Frequent concerns regarding heterologous expression in E. coli include potential problems in differing G+C content between host and donor DNA and a failure of vectors that accommodate larger fragments such as BACs. Genomic library used that covered between host and donor DNA and a failure of vectors that accommodate larger fragments such as BACs. This limitation can be overcome by using vectors that accommodate larger fragments such as BACs.

The successful application of RVA to Photorhabdus, with an average G+C content of 42%, significantly lower than the 50% of E. coli, indicates that this is not a current limitation to these biological screens. One reason why E. coli-based expression is more successful than expected may relate to the fact that pathogenicity islands that have been recently horizontally acquired may be more prone to gene expression in differing host backgrounds, and indeed this may actually be a requirement of such elements if they are to be positively selected in nature. Furthermore, the RVA technique does not only rely on correct secretion of the heterologously expressed virulence factors as whole cultures are used in the screens. We note that cytoplasmically accumulated toxins can be detected in these assays, as demonstrated by the insect toxicity of E. coli expressing but not secreting the Mcf1 toxin (14). The application of RVA to more diverse bacterial pathogens will ultimately define the limits of its utility, although in principle the host strain used need not be E. coli and libraries from Gram-positive pathogens could be screened in different host species (such as Lactococcus).

We believe that RVA represents an excellent way to provide basic information about general virulence determinants in poorly understood pathogens and to add functional relevance to the ongoing annotation of pathogen genome sequences. It also provides an alternative strategy for virulence gene detection across whole genomes, important for current efforts toward the reduction and replacement of animal testing.

Materials and Methods

Cosmid Library Construction and Statistical Analysis. The P. asymbiotica ATCC43949 genomic library was constructed in the pWEB vector (Epicentre) in Waterfield's Cosmid Library Construction and Statistical Analysis. C. elegans were raised at 18 °C on NKM medium (32) using E. coli OP50 as feeding strain and transferred weekly to fresh plates. Adult worms were washed off plates with Phosphate Saline Buffer and the suspension adjusted to 10 nematodes per 1.0 μl drop. A. polyphaga was grown in PYG medium (33) at 25 °C, and subcultured weekly. Stationary phase (5 days) cultures were used for agot, adjusting the number of cells to 2 × 10^10 cells/ml using a haemocytometer.

Gain of Toxicity Assays (GOT). 100 μl of overnight cosmid cultures were injected into individual M. sexta larvae that were scored for death or severe delay in development daily over one week. Each cosmid showing an effect was retested on a further 10 insects allowing confirmation of the phenotype. For nGOTi/AgOTi, a 10 μl drop of overnight culture from each individual cosmid was grown on 25-well plates containing NGM agar with 50 μg/ml ampicillin. A. polyphaga was grown on 25-well plates containing PYG agar with 50 μg/ml ampicillin. Cosmid cultures were inoculated with 10 μl of the adjusted suspensions of C. elegans or A. polyphaga respectively and incubated at 22 °C for 7 days. Toxicity was assessed both by eye and under an inverted microscope and avoidance or delay in feeding was scored using an arbitrary scale of 0 to 5 (5 being no consumption of 1 half-consumed and 0 completely consumed). For mGOT, the mouse BALB/c monocyte macrophage cell line J774-2 was seeded at 1 × 10^5 cells/ml into 96-well plates, in Dulbecco's Modified Eagles Medium (DMEM) supplemented with 10% fetal bovine serum, 5% non-essential amino acids, and 100 μg/ml gentamicin and incubated under the same conditions for 2 h. Cell viability was ascertained using the XTT assay (34). Candidate cosmids clones were selected for their ability to reduce cell viability by 40% comparative to untreated cells.

HPLCMS Analyses and Secondary Metabolite Structure Prediction. Cosmids 4DB5 and 1DF9 were coexpressed in E. coli EC100 with pSUMtaA (35), which encodes a promiscuous phosphopantheteine transferase from Pseudomonas aeruginosa. HPLC/MS Analyses and Secondary Metabolite Structure Prediction. Cosmids 4DB5 and 1DF9 were coexpressed in E. coli EC100 with pSUMtaA (35), which encodes a promiscuous phosphopantheteine transferase from Pseudomonas aeruginosa. HPLC/MS Analyses and Secondary Metabolite Structure Prediction. Cosmids 4DB5 and 1DF9 were coexpressed in E. coli EC100 with pSUMtaA (35), which encodes a promiscuous phosphopantheteine transferase from Pseudomonas aeruginosa. HPLC/MS Analyses and Secondary Metabolite Structure Prediction. Cosmids 4DB5 and 1DF9 were coexpressed in E. coli EC100 with pSUMtaA (35), which encodes a promiscuous phosphopantheteine transferase from Pseudomonas aeruginosa. HPLC/MS Analyses and Secondary Metabolite Structure Prediction.


