Environmental Microbiology (2002) **4**(12), 819–823

The genome structure of *Pseudomonas putida*: high-resolution mapping and microarray analysis

Diana Stjepandić,[1]* Christian Weinel,[2] Helmut Hilbert,[3] Hean L. Koo,[4] Frank Diehl,[1] Karen E. Nelson,[4] Burkhard Tümmler[2] and Jörg D. Hoheisel[1]

[1]*Functional Genome Analysis, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 506, D-69120 Heidelberg, Germany.*

[2]*Klinische Forschergruppe OE6711, Medizinische Hochschule Hannover, D-30623 Hannover, Germany.*

[3]*QIAGEN GmbH, Max-Volmer-Str. 4, D-40724 Hilden, Germany.*

[4]*The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA.*

Summary

As part of a collaborative project aimed at sequencing and functionally analysing the entire genome of *Pseudomonas putida* strain KT2440, a physical clone map was produced as an initial resource. To this end, a high-coverage cosmid library was arrayed and ordered by clone hybridizations. Restriction fragments generated by rare-cutting enzymes and plasmids containing the *rrn* operon and 23S rDNA of *Pseudomonas aeruginosa* were used as probes and, parts of the cosmids were end-sequenced. This provided the information necessary for merging and comparing the macro-restriction map, cosmid clone order and sequence information, thereby assuring co-linearity of the eventual sequence assembly with the actual genome. A tiling path of clones was selected, from the shotgun clones used for sequencing, for the production of DNA microarrays that represent the entire genome including its non-coding portions.

Introduction

Pseudomonas putida is a non-pathogenic, Gram-negative bacterium from the group of fluorescent pseudomonads. Its genome size has been determined at about 6 Mb in length (Ramos-Diaz and Ramos, 1998), with a G/C content of about 60%. As a root colonizer, the bacterium stimulates plant growth and suppresses root pathogens (O´Sullivan and O´Gara, 1992). As a result of its ability to grow in different environments and to metabolize various natural and synthetic compounds, *P. putida* is an excellent model organism for the study of such phenomenons. Strain KT2440 also acts as a host for cloning in *Pseudomonas* (Bagdasarian and Timmis, 1982) and, represents the first host-vector biosafety system for gene cloning in Gram-negative soil bacteria. For these reasons, networks were established for sequencing the entire genome and the performance of global functional analyses. Sequencing operations are the combined efforts of a German consortium (www.mh-hannover.de/kliniken/kinderheilkunde/kfg) and The Institute for Genomic Research (TIGR; www.tigr.org/tdb), based on a global shotgun strategy. As part of this project, we prepared a scaffold-clone map to assure the correct alignment and assembly of the emerging sequence contigs. The clone order was established by hybridization-based mapping, an efficient method to such end (Hoheisel *et al*., 1993; Frohme *et al*., 2000; Aign *et al*., 2001). This mapping information was integrated with existing macro-restriction data, which was obtained by digests with *Swa*I and the intron-encoded endonuclease I-*Ceu*I, and subsequent two-dimensional pulsed-field gel electrophoresis (Weinel *et al*., 2001), extending the map of 160 kb resolution that had been reported previously (Ramos-Diaz and Ramos, 1998). In addition, two plasmids containing sequences that represent the repetitive *rrn* operon and part of the 23S rDNA from *P. aeruginosa* strain PAO1 were used in order to determine the position of these repeats. Eventually, a representative number of cosmids was end-sequenced, adding shotgun information to the global sequencing and also providing links between map and sequence. All this information was integrated in an effort to assure co-linearity of the genome and the final sequence compilation.

Results

Mapping procedures

In the initial phase of the mapping process, DNA from individual cosmid clones was isolated and hybridized to the entire clone library. Probes were picked at random from the ever-decreasing number of library clones that had not been not positive in any prior hybridization experiment and should, by definition, originate from still unmapped areas (Mott *et al*., 1993). Using this strategy, the probes

Received 16 February, 2002; accepted 30 April, 2002. *For correspondence. E-mail d.stjepandic@dkfz.de; Tel. +49 6221 42 4681; Fax +49 6221 42 4687.

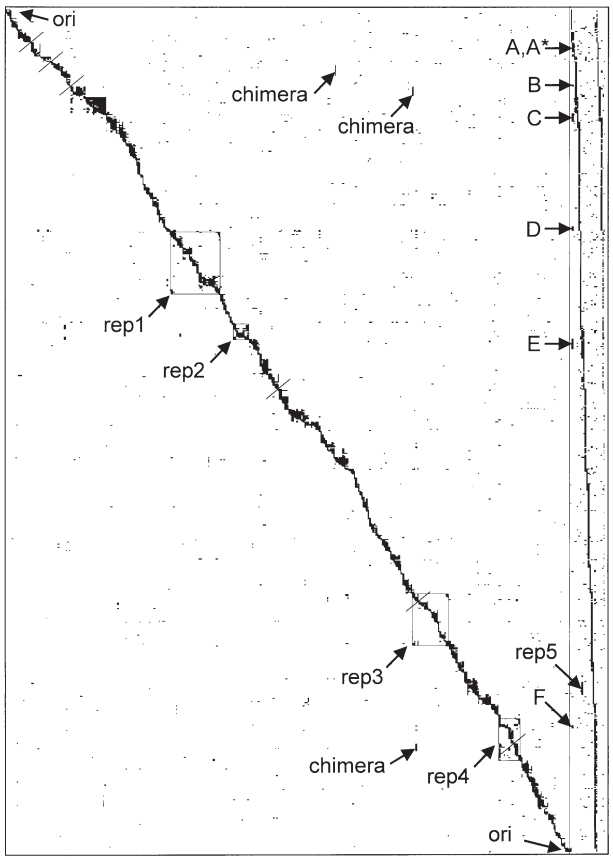

Fig. 1. Physical map of the *Pseudomonas putida* genome. In the two-dimensional matrix, 362 hybridization probes are arranged as columns, while 3620 clones are arranged in rows. A positive hybridization signal is represented by a black spot at the respective cross-section. The entire data set is presented, including all false-positive or negative results. The map starts and ends at the origin of replication (ori). Gaps in the contig coverage are indicated by lines across the diagonal. On the right margin, the results obtained from hybridizing genomic restriction fragments are shown; they are ordered left to right according to their position in the macro-restriction maps produced with *Swa*I and I-*Ceu*I, respectively. These data were not immediately taken into account for the actual clone-ordering process but served as an independent control of co-linearity. Hybridizing a short fragment resulting from the *Swa*I-digest of genomic DNA, the positions of the ribosomal operon were highlighted (rDNA: A*, A–F). The position of four other repetitive sequences (rep1 to rep4) and the cross-hybridization patterns produced by three chimeric clones (chimera) are indicated. The reason for a cross-hybridization of a genomic fragment with a specific but unrelated area (rep5) is unknown; no such effect could be observed in any of the relevant cosmid probe hybridizations. Upon request, a much more detailed presentation of the map can be obtained from the authors.

should also be spaced relatively evenly. After 264 hybridizations, 3540 clones – equivalent to 92% of the library – had been positive in at least one hybridization, assembling the clones into 85 relatively small contigs. At this stage, genomic restriction fragments produced by cleavage with either *Swa*I or I-*Ceu*I were hybridized to the cosmid library in order to help orienting and connecting contigs. Then, cosmid clones were chosen as probes for gap closure that were located at contig ends or had been positive in the hybridizations of two supposedly neighbouring restriction fragments.

Few regions were identified during the mapping procedure that had high potential for being of repetitive nature, indicated by a string of hybridization signals located outside the main diagonal of the map (Fig. 1). Also, the known repeat sequences of the *rrn* operon and 23S rDNA were localized by hybridizing the respective sequences from *Pseudomonas aeruginosa*. In addition, few apparently chimeric cosmid clones were identified by their specific hybridization pattern. By virtue of their insert-DNA, linking artificially DNA-fragments that originate from different regions of the genome, they disturb the clone order in a manner similar to repeat sequences. However, in all likelihood, any such clone is a unique artefact. Because of the high redundancy of the information acquired during the mapping process and, the fact that no apparently overlapping regions were accepted as such unless confirmed by more than two independent clones, it is unlikely that chimeras significantly disturbed the map.

Map analysis

The final clone order is shown in Fig. 1, starting and ending at the position of the origin of replication. Obvious chimeric clones were removed from the map but for three typical ones, which were left in and identified as such. The majority of the signals are located along the diagonal of the two-dimensional presentation of clones and probes, representing the best correlation between the ordered probes and the fitting clone order. The background of false positive hybridization results located outside of the diagonal was low, with less than 2.9% of the scored signals. On the right margin, the signals produced by the hybridizations of the restriction fragments are shown, the fragments being ordered according to the macro-restriction map (Weinel *et al*., 2001). The apparent overlap between the signal-strings of some adjacent fragments is caused by the fact that cosmid clones, which contain the relevant restriction site, hybridize to both fragments.

All seven rDNA repeats (A, A*, B-F) could be identified as re-occurring signals produced by hybridization of the *Swa*I-fragment M, which contains the entire ribosomal DNA, and by hybridizing the respective probe isolated from *P. aeruginosa*. Unsurprisingly, the adjacent copies A and A* were indistinguishable by means of hybridization. The relative distances between the repeat units in the map fit perfectly the positions predicted from the macro-restriction map. In addition to the ribosomal operon, four other repeat sequences were identified during the mapping analysis by obvious cross-hybridization events and are indicated in the map.

In the final map assembly there are six gaps left that cannot be covered by cosmid clones in spite of the high statistical coverage of the library. It is well known that the actual clone coverage fluctuates widely along a given chromosome, explained by the variation in cleavage activity of the restriction enzyme at different sites, differences in clonability of given sequences and pure statistical fluctuation. An uninterrupted representation of the genome could have been obtained by a recently published method for specific isolation of the missing stretches of DNA by a comparison of genomic DNA and the cosmid coverage (Frohme *et al.*, 2001). As simultaneous shotgun sequencing was ongoing, however, no such effort was pursued. It is likely that the few gaps are caused by sequences which inhibited *E. coli* growth entirely or were deleted by this host bacterium. The latter is most likely for two gaps located exactly at the position of repeat sequences, which are known to cause cloning problems. As shotgun clones have much shorter inserts, there should be plasmids that contain only part of the relevant toxic or repeated sequences, thus permitting gap closure by cloning and subsequent sequencing at this level.

Comparison of map and sequence

Sequences of cosmid-ends were determined along the analysis of the shotgun plasmid library. Besides adding equivalent sequence information, these results physically linked the sequence contigs assembled from the shotgun information to the clone map and ordered restriction fragments. Comparing this information immediately indicated inconsistencies and, thus, inaccuracies in the interpretation of either data set. During the duration of the project, the congruence increased steadily. Most discrepancies were minor, indicating false orientation of short sequence contigs, or could be assigned to the presence of repeats. In addition, some differences occurred as insufficient mapping information was available for certain clones, for example being positive with one probe only. Such a signal could be a false-positive resulting in the clone being misplaced. However, all clones that were assigned to their position by at least two hybridization probes – the actual backbone of the map – exhibited complete co-linearity with the sequence.

Production of DNA microarrays for functional analyses

The map formed a basis for the alignment of the emerging sequence contigs and permitted a check of the co-linearity between sequence assembly and the actual genome. In turn, the sequence was utilized for ordering the shotgun clones in a map, which provides a starting point for in-depth functional analyses. From this map, a set of some 4600 shotgun clones was selected, which represent a tiling path across the genome and thus – by definition – also a normalized gene repertoire. As a result of the size selection process prior to the shotgun cloning, the fragments were rather similar in length at around 2 kb. This is advantageous for microarray production, as the overall amplification yield is influenced by size differences. Also, variation in the efficiency of priming events is minimized, as a single primer pair could be used in all PCR amplifications. Of the set of clones, 96.7% could be PCR-amplified successfully, producing a single band when checked on agarose gels. In the remaining 3.3% of the reactions more than two bands, or no distinct fragment at all, were visible in the gels. The resulting DNA was spotted onto poly L-lysine glass slides and subjected to hybridizations with DNA-samples. Chip quality was examined with different samples of strain KT2440 obtained from different sources. Concomitantly, the genome structure of the strains was compared. In competitive hybridizations, one DNA sample was labelled with the fluorophor Cy3, the other with Cy5. Unsurprisingly, the analysed KT2440 strains gave rise to identical signal intensities (Fig. 2a). Any variance should have shown as

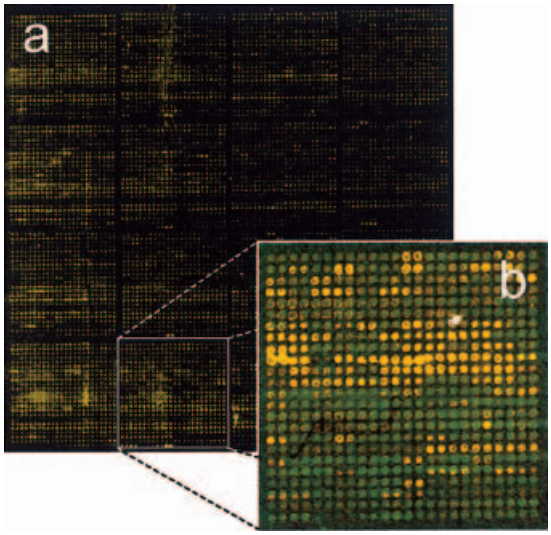

Fig. 2. Microarray analysis of genomic DNA.
a. Only one half of a microarray is shown. PCR-products were spotted in duplicate onto poly L-lysine slides. Cy5- and Cy3-labelled DNA-target was made from the genomic DNA of two samples of strain KT2440 kindly provided by Soeren Molin (BioCentrum-DTU, Lyngby, Denmark) and Kenneth Timmis and Edward Moore (GBF, Braunschweig, Germany). Any genomic difference between the two samples should show up as a red or green signal. As can be judged by the yellow colour of all spots, no discernable difference in genomic representation was identified. Overall signal intensity varied across the microarray due to differences in the amount of PCR-product present at the individual positions. However, on close scrutiny, all positions with PCR-product gave rise to a signal.
b. The difference in genome content can be seen for one square of the microarray comparing samples of strain *P. pudita* KT2440, labelled green, and *P. fluorescens*, labelled red. As the microarray represents the KT2440 genome, only the lack of such sequence in the *P. fluorescens* genome could be detected, indicated by a green signal.

a clear red or green signal from one fluorophor as a result of the lack of the relevant DNA-fragment in the respective sample. The mixed yellow colouring of all spots indicates the lack of genomic differences big enough to be detectable by the PCR-product probes. However, when one sample was replaced by DNA from a different strain, such as *Pseudomonas fluorescens* for example, clearly distinct signals could be observed (Fig. 2b). By design, only a sequence that is lacking in *P. fluorescens* or significantly different in base composition can be detected. Any extra sequence could not bind to the microarray made from the *P. putida* genome because of the lack of the respective PCR-product. Currently, the microarrays are used for comparisons of the genomic DNA content of different *Pseudomonas* strains and, for transcriptional profiling analyses with RNA-samples obtained under various growth conditions. Upon completion of the annotation of the genomic sequence, these results will be analysed further.

Discussion

A high-density cosmid map of the *Pseudomonas putida* KT2440 genome was produced. The number of clones in the library, representing a 24-fold coverage of the genome, was well above the 10-fold representation needed to cover the entire genome with very high probability. Nevertheless, six gaps were left in the final map that could not be bridged even by directed probe hybridization; only redundant information was produced in the end. This suggests that the relevant genomic regions were actually missing from the library. As has been done before for other genomes upon the availability and full annotation of the final sequence, the gene content of these regions will be analysed in an attempt to correlate potential gene function with the lack of clonability in large-insert vector systems of the *E. coli* host. This process will extend the list of incompatibility genes from Gram-negative bacteria for *E. coli* (Frohme *et al*., 2000), which should add to the definition of both the functions that the genes encode for, as well as the functioning of the host bacterium.

The current version of the genomic microarray represents some 96% of the entire genome. For the missing portion of less than 4% – DNA equivalent to a single microtitre dish worth of shotgun clones despite an intentionally high degree of redundancy in their selection – amplification will be performed at modified experimental conditions, e.g. PCR with proof-reading capable enzymes (Barnes, 1994) or the use of alternative clone templates. Any gaps that cannot be covered by such means will be amplified directly from genomic DNA using specific primers upon final sequence assembly. This ability to check the completeness of coverage is one advantage of using genomic as opposed to gene-specific arrays. Representation of the latter is much dependent on the quality of the sequence annotation, while preparation of the former could be started even before sequence analysis. In addition, all parts of the genome can be assayed, extending possible analysis procedures from mere transcriptional profiling to concomitant studies of promoter activity and such like. Comparisons of the genome structure of different *P. putida* strain derivatives and related organisms, such as *Pseudomonas aeruginosa,* are under way. Also, the analysis of transcript levels in isolates from marine water and soil and strains relevant to the investigation of biofilm formation is an emphasis of ongoing work. Simultaneously, various mutants will be used in protein binding assays for the elucidation of related and unrelated DNA–protein interactions.

Experimental procedures

Cosmid library preparation

Genomic DNA was isolated from *Pseudomonas putida* KT2440, which was kindly provided by Edward Moore and Kenneth Timmis. This strain is a plasmid-free derivative of the mt-2 strain, which carries the pWW0 TOL plasmid (Bagdasarian and Timmis, 1982). Construction of the cosmid library followed procedures described in detail elsewhere (Hanke and Hoheisel, 1999). Briefly, cosmid Lawrist-4 was linearized with *Sca*I and dephosphorylated, followed by cleavage of the *Bam*HI cloning site. The genomic DNA was partially cut with *Mbo*I and treated with phosphatase prior to ligation to the vector. The average insert size was 37 kb. Some 3800 clones, representing a 24-fold genome coverage, were picked into microtitre plates. The entire library was spotted on filters using a commercial robotic device (BioGrid I, BioRobotics, Cambridge, UK).

Hybridization and data analysis

Cosmid-DNA was isolated by a fast protocol (Frohme *et al*., 2000). *Swa*I and I-*Ceu*I restriction fragments were produced in agarose blocks and isolated from pulsed-field gels following standard procedures (Sambrook *et al*., 1989). The inserts of plasmids pPA and pHF360 were isolated in order to prepare a probe of the *rrn* operon and part of the 23S rDNA from *P. aeruginosa* respectively. DNA was labelled by random hexamer priming (Feinberg and Vogelstein, 1983) and hybridized to the filter arrays as described (Frohme *et al*., 2000; Aign *et al*., 2001). Results were detected by a Fuji FLA3000 phosphorimager and scored either manually or with the VisualGrid software (GPC-Biotech AG, Martinsried, Germany). Clone ordering and map construction were carried out using established analysis software (Mott *et al*., 1993).

Microarray production and hybridization

Purified DNA of some 4600 overlapping shotgun clones in vector pCR-BluntII-TOPO (Invitrogen, Groningen, The

Netherlands) was used as template in standard PCR amplification with the vector-specific primer pair d(TCGGATC CACTAGTAACG) and d(GGCCGCCAGTGTGATG). Shotgun cloning had been performed according to the procedures described by the vector's manufacturer. Each PCR reaction was individually checked by agarose gel electrophoresis. Prior to spotting onto microarrays, the DNA concentration of the fragments was adjusted to about 300 ng μl^{-1}. Spotting was done using an SDDC-2 DNA Micro-Arrayer from Engineering Services Inc. (Toronto, Canada) and SMP3 pins (TeleChem International Inc., Sunnyvale, USA) onto glass slides coated with poly L-lysine as described in detail earlier (Diehl *et al*., 2001).

For labelling DNA-samples, random priming (Feinberg and Vogelstein, 1983) was performed on 500 ng genomic *Pseudomonas* DNA in the presence of 15 μM each of dATP, dGTP and dTTP, and 25 μM Cy3- or Cy5-labelled dCTP (Amersham-Pharmacia, Amersham, UK). Hybridization of such samples to the microarrays was done as described (Diehl *et al*., 2001). Fluorescence signals were detected on a ScanArray5000 unit (Packard, Billerica, USA) and analysed with the GenePix software package (Axon Instruments, Union City, USA).

Acknowledgements

We are grateful to Marita Schrenk and Mareike Grees for their skilful technical assistance and to Edward Moore and Kenneth Timmis (GBF, Braunschweig, Germany) as well as Soeren Molin (BioCentrum-DTU, Lyngby, Denmark) for the provision of material. This work was funded by the German Ministry of Education and Research (BMBF) and the United States Department of Energy.

References

Aign, V., Schulte, U., and Hoheisel, J.D. (2001) Hybridization mapping of *Neurospora crassa* linkage groups II and V. *Genetics* **157:** 1015–1020.

Bagdasarian, M., and Timmis, K.N. (1982) Host: vector systems for gene cloning in *Pseudomonas*. *Curr Top Microbiol Immunol* **96:** 47–67.

Barnes, W.M. (1994) PCR amplification of up to 35 kb DNA with high fidelity and high yield from 1 bacteriophage templates. *Proc Natl Acad Sci USA* **91:** 2216–2220.

Diehl, F., Grahlmann, S., Beier, M., and Hoheisel, J.D. (2001) Manufacturing DNA-microarrays of high spot homogeneity and reduced background signal. *Nucleic Acids Res* **29:** e38.

Feinberg, A.P., and Vogelstein, B. (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem* **132:** 6–13.

Frohme, M., Camargo, A.A., Heber, S., Czink, C., Simpson, A.J.D., Hoheisel, J.D., and de Souza, A.P. (2000) Mapping analysis of the *Xylella fastidiosa* genome. *Nucleic Acids Res* **28:** 3100–3104.

Frohme, M., Camargo, A., Czink, C., Matsukuma, A.Y., Simpson, A.J.G., Hoheisel, J.D., and Verjovski-Almeida, S. (2001) Directed gap closure in large scale sequencing projects. *Genome Res* **11:** 901–903.

Hanke, J., and Hoheisel, J.D. (1999) Construction and screening of cosmid libraries. In *The Nucleic Acid Protocols Handbook*. Rapley, R.R. (ed.) Totowa, New Jersey: Humana Press, pp. 405–414.

Hoheisel, J.D., Maier, E., Mott, R., McCarthy, L., Grigoriev, A.V., Schalkwyk, L.C., *et al.* (1993) High-resolution cosmid and P1 maps spanning the 14-Mbp genome of the fission yeast *Schizosaccharomyces pombe. Cell* **73:** 109–120.

Mott, R., Grigoriev, A., Maier, E., Hoheisel, J.D., and Lehrach, H. (1993) Algorithms and software tools for ordering clone libraries; application to the mapping of the genome of *Schizosaccharomyces pombe. Nucleic Acids Res* **21:** 1965–1974.

O'Sullivan, D.J., and O'Gara, F. (1992) Traits of fluorescent *Pseudomonas* spp. involved in suppression of plant root pathogens. *Microbiol Rev* **56:** 662–676.

Ramos-Diaz, M.A., and Ramos, J.L. (1998) Combined physical and genetic map of the *Pseudomonas putida* KT2440 chromosome. *J Bacteriol* **180:** 6352–6363.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning, A Laboratory Manual.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Weinel, C., Tümmler, B., Hilbert, H., Nelson, K.E., and Kiewitz, C. (2001) General method of rapid Smith/Birnstiel mapping adds for gap closure in shotgun microbial genome sequencing projects: application to *Pseudomonas putida* KT2440. *Nucleic Acids Res* **29:** e110.