Simple and Inexpensive but Highly Discriminating Method for Computer-Assisted DNA Fingerprinting of Pseudomonas aeruginosa

Taha H. Al-Samarrai, Ningxin Zhang, Iain L. Lamont, Lois Martin, John Kolbe, Margaret Wilsher, Arthur J. Morris and Jan Schmid

Simple and Inexpensive but Highly Discriminating Method for Computer-Assisted DNA Fingerprinting of *Pseudomonas aeruginosa*

TAHA H. AL-SAMARRAI, NINGXIN ZHANG, IAIN L. LAMONT, LOIS MARTIN, JOHN KOLBE, MARGARET WILSHER, ARTHUR J. MORRIS, AND JAN SCHMID

Institute of Molecular BioSciences, Massey University, Palmerston North, Department of Biochemistry, University of Otago, Dunedin, and Departments of Respiratory Services and Clinical Microbiology, Green Lane Hospital, Auckland, New Zealand

Received 16 May 2000/Returned for modification 18 August 2000/Accepted 18 September 2000

We describe here a method for computer-assisted fingerprinting of *Pseudomonas aeruginosa*. In this method, DNA is digested with *SalI*, and bands with molecular sizes of ≥9.7 kb are visually scored after electrophoresis on agarose gels. Pattern scores are entered into a Microsoft Excel database. In scoring, the number of bands within each of a set of molecular size ranges is scored, rather than the absolute molecular size of each band, substantially enhancing the speed and reproducibility of the method, while eliminating the need for using expensive gel scanning equipment and software. Pattern scores are used to generate matrices of genetic distance values, which can be visualized in neighbor-joining trees. The method reliably distinguishes two epidemiologically unrelated isolates in 99.3% of all comparisons. The genetic relationships between isolates observed with the method were consistent with those obtained by analysis of two *P. aeruginosa* genes, indicating that it provides valid estimates of genetic divergence between isolates. Using the method, respiratory tract isolates from cystic fibrosis patients in Green Lane Hospital in Auckland, New Zealand, were shown to be genetically less diverse than epidemiologically unrelated isolates from other patients. This finding was not due to the existence of clusters of related strains specialized toward colonization of the respiratory tract and thus was indicative of transmission between patients. Analysis of multiple isolates from individual cystic fibrosis patients suggested that up to five separate clusters of genetically related strains may simultaneously be present in a patient. The method described should significantly enhance our ability to investigate the epidemiology of *P. aeruginosa*.

*Pseudomonas aeruginosa* is the fifth most frequent nosocomial pathogen, and infections with it are often difficult to treat due to antibiotic resistance (5, 7). A large number of typing approaches have been developed to learn more about the epidemiology of *P. aeruginosa* and to monitor transmission between high-risk patients (15). The ideal typing method should be highly discriminating, reproducible, quick, and fairly easy to perform, and it should not require expensive specialized equipment or software. It should also easily handle comparisons between large numbers of isolates, as necessary for large-scale epidemiological studies. Finally, the genetic relationships between isolates deduced with the method should be indicative of the overall similarity, or dissimilarity, of their genomes, i.e., the pattern of a given strain should remain stable, and convergent evolution of the same typing pattern in distant lineages should be rare.

To our knowledge, no single method, meeting all of the above criteria currently exists for *P. aeruginosa*. Two currently favored methods are ribotyping and pulsed-field gel electrophoresis (4, 14, 16). While these methods are fairly discriminating (4, 14, 16), they are rather slow and/or require sophisticated and expensive equipment. In 1993, Maher et al. (12) demonstrated that *SalI* digestion of *P. aeruginosa* DNA produced polymorphic high-molecular-weight bands which could be resolved on low-agarose-content gels, run in standard horizontal electrophoresis units; these researchers suggested that these restriction fragments could be used for typing. Nociari et al. (14) later demonstrated the high discriminatory power of these polymorphisms but suggested that convergent evolution of the same *SalI* types in distant lineages might limit the usefulness of the method. A greater, and to date unaddressed, obstacle preventing widespread use of the method as an epidemiological tool was the lack of a simple, reliable, and cost-effective method for comparing large numbers of *SalI* patterns of *P. aeruginosa* isolates.

We have therefore developed a computer-assisted approach which overcomes this significant obstacle. We have evaluated the discriminatory power of the resulting computer-assisted *P. aeruginosa* typing method, both in general and when applied to isolates from cystic fibrosis patients. To assess long-term stability and the possible convergence of the typing patterns in distant lineages, we have determined the correlation between the divergence of the typing patterns and divergence at two genomic loci of *P. aeruginosa*.

**MATERIALS AND METHODS**

Strains typed. Table 1 lists the isolates used. Isolates were obtained from routine clinical specimens sent for bacterial culture. Representative single colonies were subcultured, mixed with glycerol, and stored at −70°C (3). All isolates had been identified as *P. aeruginosa* on the basis of their typical colonial appearance or on the basis of a yellow-green pigmentation on Chromocult agar (Merck) and a positive reaction in the oxidase test (8). For isolates obtained in Dunedin, identification as *P. aeruginosa* was in addition confirmed on the basis of their ability to grow at 42°C, the production of the characteristic blue pigment pyo-
cytochrome on King's A agar, and the ability to use pyoverdine in cross-feeding assays (11).

DNA extraction. Ten milliliters of liquid medium containing 1% (wt/vol) tryptone (Difco) in a test tube was inoculated from glycerol stocks, and the cultures were incubated with slow shaking at 30°C until they reached an optical density of 0.6 to 1.0 at 650 nm. Then, 50 μl was transferred into 5 ml of fresh medium and incubated. When an optical density of 0.2 to 0.3 was reached, cells were harvested by centrifugation. DNA extraction was carried out in a modification of the method of Al-Samarrai and Schmid (1). Cells were suspended in 0.5 ml of a lysis buffer containing 40 mM Tris-acetate (pH 7.8), 20 mM sodium acetate, 1 mM EDTA, and 1% sodium dodecyl sulfate, and 165 μl of 5 M NaCl was added. The suspension was centrifuged at 13,000 rpm for 15 min in a microcentrifuge at 4°C. Next, 500 μl of supernatant was removed and extracted with chloroform. The aqueous phase was removed, mixed with 37.5 μl of lysis buffer and 12.5 μl of 5 M NaCl, and extracted once more with chloroform. DNA was precipitated with 2 volumes of cold 95% ethanol. The pellet was rinsed three times with cold 70% ethanol, dried, and subsequently dissolved in 25 μl of TE buffer (pH 7.8) (3). DNA concentration was measured fluorometrically using the Hoechst dye 33258 (3).

Digestion and electrophoresis. A total of 2 μg of DNA were digested with 20 U of SalI for 1 h in a volume of 20 μl. Then, 5 μl of a loading buffer, containing 40% sucrose and 0.075% (wt/vol) each of bromophenol blue and xylene cyanol, was added, and the entire sample was loaded onto a 0.5% agarose gel in TBE buffer (3). Gels were loaded so that each lane of Pseudomonas DNA was flanked by two lanes containing 0.3 μg of XV molecular weight standard (Roche Diagnostics). Electrophoresis was carried out at 30 V. After that gels were transferred into 5 ml of 5 M NaCl, and electrophoresis continued for another 20 h at 4°C. Gels were stained with ethidium bromide (1.7 μg/ml) for 30 min and then destained for at least 1 h.

Scoring of patterns and analysis of relationships between isolates. Patterns were visualized on a transilluminator, and an image was acquired on an IS 1000 Alpha Innotech Corporation gel analysis system. Printouts from the system's printer or printouts made from the image file, using a laser writer, were used for visual scoring. For scoring, the number of bands in the intervals between the bands of the molecular weight marker were counted (see Fig. 1). The score was entered into a computer database, either Dendron 1.0 (22) or Excel. The maximum score per interval used was three, even if more than three bands appeared to be present. A simple genetic distance, D (19), was then calculated between the electronically stored patterns using the equation:

$$D = \frac{1}{1 + \sum_{i=1}^{n} (a_i + b_i - |a_i - b_i|)}$$

where $a_i$ and $b_i$ are the number of bands in the molecular weight interval $i$ in patterns A and B, respectively, and $k$ is the number of bands. Neighbor-joining trees were generated from matrices of distance values by using PAUP* 4.0 (Sinauer Associates, Inc.). Genetic separation between groups of isolates was tested in a modification of the method of Schmid et al. (20), by determining how often members of one group had an isolate from the other group as their closest related counterpart and by comparing this with the number of instances expected if there was random mixing of the two groups. The number of times with which isolates from group A, containing $n_A$ isolates, should have an isolate from group B, containing $n_B$ isolates, as their closest related counterparts upon random mixing was calculated as follows: $N_{random} = n_A \times n_B \times (n_A - 1 + n_B)^{-1}$. The presence or absence of the fpvA gene in an isolate was scored on the basis of a PCR assay using the primers, 5'-CTCATGCAATTCACAAGTAGCACAA (TCACTATCATC-3') and reverse (5'-GCCGATCCAAGATTGTTGAC-3') primers in the presence of 0.2 mM deoxynucleoside triphosphates and 1.5 mM MgCl2; samples were incubated at 94°C for 5 min, followed by 30 cycles of 94°C (30 s), 48°C (45 s), and 72°C (55 s), with a final extension step at 72°C for 5 min. Sequencing was done with an Applied Biosystems 377 Automated Sequencer using the forward and reverse primers. Sequences were aligned using Auto Assembler 2.0 software (Applied Biosystems), and genetic distances between sequences were calculated using PAUP* (distance setting: uncorrected “p”). The presence or absence of the fvdS gene in an isolate was scored on the basis of a PCR assay using the primers, 5'-CCATAGCGGCGATCCAGC-3' and 5'-CCGTGCCTGTTGTCGCGTCG-3', designed from the sequence of the gene (17) that gave rise to a 756-bp product; the PCR was carried out as for fpvA except that amplification of primers was carried out at 55°C instead of at 48°C. The presence or absence of fvdS was confirmed for each strain by using purified type I pyoverdine in growth stimulation tests (13), with pyoverdine selectively stimulating the growth of fvdS-containing strains, and by Southern blotting (18) using the fpa PCR product from the P. aeruginosa type strain PAO (ATCC 15692) as a radiolabeled probe. Sequence analyses were carried out as described for the fpvA locus.

**RESULTS**

Development of computer-assisted SalI fingerprinting. Analysis of high-molecular-weight bands of SalI digests of P. aeruginosa DNA separates epidemiologically unrelated isolates into a large number of different types (14). To best exploit this diversity, it is necessary to compare the patterns of isolates quantitatively, generating genetic distance values between them. The number of comparisons for a set of $n$ isolates equals $0.5 \times n \times (n - 1)$; thus, already for only 30 isolates, for example, 435 comparisons need to be made to describe the relationships between them. It was therefore essential to convert the patterns into a form in which the comparisons could be carried out by a computer. Based on the diversity of patterns reported (12, 14), it seemed feasible to employ a rapid and very simple scoring and data entry method, one not necessitating expensive specialized equipment and yet retaining sufficient discriminatory power. The number of bands in molecular-weight intervals was counted, converting the patterns into short strings of numbers. The latter were entered into a computer and used to calculate genetic distances (Fig. 1 and Materials and Methods for details).

**Discriminatory power of the method.** To test the discriminatory power of the method, we carried out 85 repeat analyses of isolates' SalI patterns and determined the genetic distances between electronically stored scores obtained for the same isolates. Each distance was based on a comparison of two patterns run on two different gels, each derived from a separate DNA preparation made from a separate liquid culture of the isolate. Scoring was done without referring to previously scored patterns. A histogram of the distribution of the 85 distance values between repeat scores of the same isolates is shown in Fig. 2A. The average distance between repeat scores was 0.113, and the upper 95% confidence limit of the distance between repeat scores was 0.139. We then generated a matrix of 136 SalI pattern-based distance values between a set of 17 epidemiologically unrelated P. aeruginosa isolates and deter-
mined the frequency of values in this second set, which fell below the upper confidence limit of the distance between repeat scores (Fig. 2B). Only 1 of 136 values was below the upper confidence limit of the distance between repeat scores; this is equivalent to 0.7% of all values. Thus, the method will on average distinguish two unrelated isolates in at least 99.3% of comparisons, a result equivalent to a conservative estimate of its discriminatory power (9, 10) of 0.993.

Correlation between divergence of SalI patterns and divergence of two protein coding loci, pvdS and fpvA. The usefulness of a typing method depends on how well the degree of divergence between two isolates deduced with the typing method correlates with the overall divergence of their genomes. A given locus cannot be assumed to evolve at a constant rate in each lineage (25), and therefore one cannot expect a tight correlation between typing patterns and changes at each individual locus. Nevertheless, if a correlation between typing pattern divergence and divergence of the remainder of the genome exists, (i) the pattern of the same strain will be stable over prolonged periods of time, (ii) closely related strains will have similar patterns, and (iii) distantly related strains will have very different patterns.

We initially tested the first aspect directly by growing four strains for 2,000 generations and found no pattern alterations (data not shown). We then investigated the correlation between divergence as measured by SalI typing and divergence at two other loci, pvdS, which codes for a sigma factor (15), and fpvA, which codes for a ferrisiderophore receptor (17).

To compare divergence at the pvdS locus and the divergence of SalI patterns, we first generated for 22 isolates a distance matrix of 231 values, describing their relationships based on their SalI patterns. Next, we generated a second matrix containing the 231 distances between the same set of strains based on pvdS sequence comparisons. We then compared the matching distance values obtained with the two methods (Fig. 3). Because of the large number of datum points and their scatter, the figure shows the average distance values based on sequence comparisons for different categories of SalI-based distances. There was a loose correlation between the two types of distances: among isolates with SalI-based distances below the average SalI-based distance in the set (the average was 0.491), the average pvdS-based distance was 0.003. Among isolates with SalI-based distances above the average SalI distance, the average pvdS-based distance was 0.007. This difference was statistically significant (t test, P < 0.000015).

The second gene used for these comparisons, fpvA, codes for a siderophore uptake protein and is only present in approximately 42% of all P. aeruginosa strains (13). Figure 4 shows a tree of the 22 isolates used, based on SalI-based distances, in which isolates with the fpvA gene are marked. The tree suggests that isolates which appear closely related according to SalI type often have an identical status as far as the presence or
absence of \( fpvA \) is concerned; see, for instance, the cluster containing P6870, P6990, and P8615 and the cluster containing IAI25, IAI17, and P8265. A quantitative analysis showed that if two isolates had a \( SalI \)-based distance of \( \leq 0.2 \), they had a probability of 82% of having identical \( fpvA \) status; that was almost twice as often as isolates with \( SalI \)-based distances of >0.2 (46%). The difference was statistically significant (\( z \) test, \( P < 0.05 \)). For isolates with the \( fpvA \) gene, we compared sequence-based distances with \( SalI \)-based distances as described above for the \( pvdS \) gene. For pairs of isolates with low \( (\leq 0.2) \) \( SalI \)-based distances, the average \( fpvA \)-based distance was almost 10 times lower than for pairs of isolates with larger \( SalI \)-based distances between them (0.0004 versus 0.0031). The difference was statistically significant (\( t \) test, \( P < 0.006 \)). No apparent correlation existed between the \( SalI \)-based distances and the \( fpvA \)-based distances when \( SalI \)-based distances exceeded 0.2 (data not shown).

These results show that \( SalI \) pattern-based relationships are loosely correlated with relationships based on the divergence of other genetic loci. Note that this correlation is stronger than that observed when the same statistical methods were used to compare the divergence at the \( pvdS \) and \( fpvA \) loci with each other (data not shown).

**Evaluation of typing method in \( P. \) aeruginosa isolates from cystic fibrosis patients.** Our primary reason for developing the typing method was to use it in future investigations of the epidemiology of respiratory tract \( P. \) aeruginosa infection in cystic fibrosis patients.
cystic fibrosis patients. It was important to confirm that the
discriminatory power of the method, calculated above by using
isolates from a variety of patient types and body locations,
would also apply to respiratory tract isolates from cystic fibrosis
patients. This is necessary since it is conceivable that special-
ized groups of related strains may preferably colonize the re-
spiratory tract either in general or in cystic fibrosis patients in
particular. Such specialization would significantly reduce the
genetic diversity among respiratory tract cystic fibrosis isolates
and thereby diminish the discriminatory power of the method
in this category of isolates. Indeed, we found an indication that
genetic diversity may be reduced among cystic fibrosis isolates:
Distances of <0.25 were about twice as frequent among iso-
lates from the respiratory tract of different cystic fibrosis pa-
patients at Green Lane Hospital than among isolates from dif-
ferent body sites from epidemiologically unrelated patients
(5% [7 of 136] versus 11% [20 of 190]; P < 0.05 and P < 0.10,
in one-sided and two-sided z tests, respectively; because of the
low number of isolates from Dunedin and because in different
geographical areas different specialized clonal groups might
predominate, this analysis was restricted to isolates obtained in
Auckland). To determine whether this reduced diversity was a
result of strain specialization, we analyzed the relationships
between representative isolates (one per patient) from the
respiratory tracts of cystic fibrosis patients and other patients
(20 and 16 isolates, respectively) and 23 isolates from other
sites. The tree shown in Fig. 5 does not suggest a clear-cut
separation between respiratory or nonrespiratory isolates or
between respiratory tract isolates obtained from patients with
cystic fibrosis and isolates from patients who did not have cystic
fibrosis. A more stringent test for such separation was to de-
terminate whether representative isolates from sites other than
the respiratory tract have a respiratory tract isolate as their
closest related counterpart less often than expected if there is
no separation. When this analysis was carried out, there was no
significant difference (z test) between the frequency observed
(13 of 23) and the frequency expected upon random mixing of
the two sets of isolates (14 of 23; see Materials and Methods
for calculation of estimated frequency). Likewise representa-
tive respiratory tract isolates from non-cystic fibrosis patients
had respiratory tract isolates from cystic fibrosis patients as
their closest related counterpart approximately as often as
expected upon random mixing. In this analysis 16 representa-
tive non-cystic fibrosis isolates and 20 representative cystic
fibrosis isolates were compared. The observed frequency was 7
of 16 compared to 9 of 16 expected upon random mixing.
The difference was not significant in a z test. Thus, there was no
evidence of significant host specificity among the isolates
tested, and the discriminatory power of the method calculated
earlier should also apply to cystic fibrosis patients. Reduced
genetic diversity among cystic fibrosis isolates was therefore
likely to indicate transmission between patients (see Discus-

A last prerequisite for future applications of the method for
epidemiological studies was to obtain an initial estimate of the
diversity of strains on individual cystic fibrosis patients. The
study design must take the diversity of strains within a patient
into account when deciding on how many isolates from a pa-

tient need to be tested. For 13 of the cystic fibrosis patients we
had obtained and typed more than one isolate over periods of
up to 20 months. As illustrated by the pattern of isolates from
one of the patients shown in Fig. 6, the same P. aeruginosa
genotypes can be maintained for prolonged periods of time,
but different genotypes can coexist in the same individual. To
quantitate this diversity, we used a distance of <0.25 as a
threshold to divide a patient’s isolates into groups; since 95%
of distance values between epidemiologically unrelated isolates
are ≥0.25 (see above), a distance of <0.25 between isolates
from the same patient would indicate a 95% probability that
they are derived from a single clonal group of identical or
highly related cells, which had colonized the patient. We gen-
erated trees for the isolates of each patient (see Fig. 6B for an
example) and determined the number of groups in each tree
(three groups are labeled in Fig. 6B). We then plotted these
numbers against the inverse of the number of isolates typed
per patient (Fig. 7). The figure suggests that five or more
separate clusters of genetically closely related strains could be
present on a patient (intercept of trendline with y axis at x = 0,
i.e., when an infinite number of isolates per patient are typed).
However, the data in Fig. 7 seem to best fit a two-component
trendline, suggesting that one or two clonal groups of highly
similar strains may predominate since they are readily identi-
fied when small numbers of isolates from a patient are ana-
lyzed. We note that our typing method does not take into account band intensity and that some of the patterns in Fig. 6, although otherwise identical, differ in the intensity of one of the bands. Thus, some of the groups defined by use of our method could potentially be further subdivided, indicating a possibly even larger genetic diversity of the *P. aeruginosa* flora in a given patient.

No correlation was found between the length of the time interval in which the samples had been collected from a patient—these intervals ranged between 0 days and 20 months—
and the genetic diversity of his or her isolates (calculated by dividing the number of groups per patient by the number of isolates typed; data not shown).

**DISCUSSION**

A conservative estimate of discriminatory power of our method of computer-assisted SalI typing of *P. aeruginosa* is 0.993. This compares favorably with the discriminatory power reported for other DNA-based methods used to type the organism (0.956 to 1.00 [6, 14]). Only the reported discriminatory power of one of these other methods, pulsed-field gel electrophoresis, was determined by Grundmann et al. (6) to exceed the discriminatory power of our method. However, their estimate is based on visual side-by-side comparisons of patterns; in their own words, this “becomes virtually impossible” (6) when large collections of isolates are analyzed. In contrast, we determined the discriminatory power using digitally stored scores of patterns determined independently on different gels and using different DNA preparations. It is therefore probable that in “real life” the discriminatory power of our computer-assisted SalI typing exceeds that of pulsed field gel electrophoresis. Computer-assisted SalI fingerprinting might thus be the most discriminating typing method for *P. aeruginosa* currently available. The method also performs well with other DNA-based methods in terms of simplicity and speed. In addition, it requires only the most basic molecular biology and computing equipment and affordable, user-friendly, commercially available software packages, namely, Microsoft Excel and PAUP*. Indeed, it is possible to carry out both data entry and data analysis using PAUP* alone. However, we found data entry to be easier in Excel, and the genetic distance calculation favored by us was not supported by PAUP*, although future releases may include it as an option (D. L. Swoford, personal communication).

We have also assessed by serial transfer of strains that the patterns are stable (no change in four strains observed for 2,000 generations each), which is in accordance with earlier observations of Nociari et al. (14). In addition, we have conducted an extensive analysis confirming long-term pattern stability and showing that the degree of divergence of the SalI pattern scores is positively correlated with sequence divergence at two genomic loci. This correlation is not in disagreement with the findings of Nociari and coworkers (14), who observed occasional disparities between grouping of strains by SalI patterns and other DNA-based methods. Indeed, such disparities are to be expected, because genomic change is caused by individual events and loci evolve at different rates (25). Our more extensive and quantitative analysis does, however, contradict the assumptions made by Nociari et al. on the basis of these disparities (14) that the usefulness of SalI patterns for typing is compromised by frequent convergent evolution of the same SalI pattern in distant lineages.

It is known that for bacterial pathogens specialized clones or clonal groups can exist which are associated with particular diseases (23, 26). Such specialization can reduce the discriminatory power of a method on a given patient group. In preparation for future use of the method to investigate the epidemiology of *P. aeruginosa* in cystic fibrosis patients, we established that cystic fibrosis isolates are not a specialized group and that our method will have an undiminished power of discrimination between isolates from these patients. A reduced genetic diversity among isolates from cystic fibrosis patients from the same hospital, as observed by us, is therefore an indicator of an epidemiological connection between patients (21). We have now begun to use our initial observations as a starting point for pinpointing more precisely the routes of transmission between local cystic fibrosis patients, thus complementing our earlier work (27, 28) on the epidemiology of the related pathogen *Burkholderia cepacia*.

Lastly, we have made a rough estimate of the genetic diversity of the *P. aeruginosa* respiratory tract flora within an individual cystic fibrosis patient as assessed by our method. The estimate is broadly consistent with the results of previous studies (2, 12, 15, 24), which suggested that one or two genotypes predominate, often over prolonged periods of time, but that other genotypes can coexist with these. Our best preliminary estimate of the maximum number of clonal groups on a patient, which are so distinct from each other that they are most likely derived from separate ancestors, is five. This is a surprisingly high number, which needs to be verified in future studies. If it is correct, studies aimed at elucidating transmission will need to take this diversity into account and assure collection of a sufficient number of isolates from each patient. Otherwise, failure to detect minor strains in a patient may obscure transmission, especially if strains playing a minor role on one patient may act as the predominating etiological agent on another. By monitoring the complete flora of a patient, we may also hope to learn more about the etiology of *P. aeruginosa* if we can determine how multiple strains can coexist on a patient and what determines the balance between them.

In summary, computer-assisted SalI fingerprinting allows fast, reliable typing of *P. aeruginosa* with a minimum of effort and capital expenditure, and the relationships obtained are a valid indicator of genetic divergence at other loci.

**ACKNOWLEDGMENTS**

This work was supported by a grant from the New Zealand Health Research Council to J.K., M.W., A.J.M., and J.S. and by a grant from the New Zealand Lotteries Board (Health) to I.L.L.

**REFERENCES**


