Published in final edited form as:

Int J Pediatr Otorhinolaryngol. 2006 November ; 70(11): 1891–1900.

Construction and characterization of a highly redundant *Pseudomonas aeruginosa* genomic library prepared from 12 clinical isolates: application to studies of gene distribution among populations

Geza Erdos^{1,2,#}, Sameera Sayeed¹, Fen Ze Hu^{1,2}, Patricia T. Antalis¹, Kai Shen¹, Jay D. Hayes¹, Azad I. Ahmed¹, Sandra L. Johnson¹, J. Christopher Post^{1,3}, and Garth D. Ehrlich^{1,2,*}

¹ Center for Genomic Sciences, Allegheny-Singer Research Institute, 320 E. North Ave., Pittsburgh, PA 15212;

² Department of Microbiology and Immunology, Drexel University College of Medicine, Allegheny Campus, 320 E. North Ave., Pittsburgh, PA 15212 and

³Department of Otolaryngology, Drexel University College of Medicine, Allegheny Campus, 320 E. North Ave., Pittsburgh, PA 15212

Abstract

Objective—To create, array, and characterize a pooled, high-coverage, genomic library composed of multiple biofilm-forming clinical strains of the opportunistic pathogen, *Pseudomonas aeruginosa* (PA). Twelve strains were obtained from patients with otorrhea, otitis media, and cystic fibrosis as a resource for investigating: difference in the transcriptomes of planktonic and biofilm envirovars; the size of the PA supragenome and determining the number of virulence genes available at the population level; and for testing the distributed genome hypothesis.

Methods—High molecular weight genomic DNAs from twelve clinical PA strains were individually hydrodynamically sheared to produce mean fragment sizes of ~1.5Kb. Equimolar amounts of the 12 sheared genomic DNAs were then pooled and used in the construction of a genomic library with ~250,000 clones that was arrayed and subjected to quality control analyses.

Results—Restriction endonuclease and sequence analyses of 686 clones picked at random from the library demonstrated that >75% of the clones contained inserts larger than 0.5 Kb with the desired mean insert size of 1.4 Kb. Thus, this library provides better than 4.5x coverage for each of the genomes from the twelve component clinical PA isolates. Our sequencing effort (~1 million nucleotides to date) reveals that 13% of the clones present in this library are not represented in the genome of the reference *P. aeruginosa* strain PA01.

Conclusions—Our data suggests that reliance on a single laboratory strain, such as PA01, as being representative of a pathogenic bacterial species will fail to identify many important genes, and that to obtain a complete picture of complex phenomena, including bacterial pathogenesis and the genetics of biofilm development will require characterization of the *P. aeruginosa* population-based supragenome.

Address for Correspondence: Garth Ehrlich (gehrlich@wpahs.org), Center for Genomic Sciences, Allegheny-Singer Research Institute, 320 E. North Ave., Pittsburgh, PA 15212, Phone: 412-359-4228, Fax: 412-359-6995.

[#]present address, Department of Dermatology, University of Pittsburgh, Pittsburgh, PA 15212, email: gee2+@pitt.edu

Keywords

supragenome; distributed genome hypothesis; genomic plasticity; otitis media; otorrhea; bacterial evolution

Abbreviations

ppk = polyphosphate kinase gene; 16S = 16S ribosomal RNA gene

1. Introduction

Although the entire genomic sequence of the *Pseudomonas aeruginosa* (PA) laboratory strain PA01 strain has been completed [1], previous comparative studies among multiple strains have indicated that the genome size of individual strains within the species vary significantly [2]. Pulse field electrophoresis data of large restriction fragments of different *P. aeruginosa* strains suggested that the genome size could vary by more than 15% with respect to the PA01 genome [3]. Even more notably, other members of the *Pseudomonas* genus have genomes that display 30–40 % plasticity in size [4]. The differences measured for *P. aeruginosa* strains equate to a coding capacity for more than 500 proteins with an average size of 50 kilo Daltons (KD). Genomic plasticity via the acquisition of horizontally-transmitted genes has been associated with the evolution of epidemic strains for animals, plants, and humans among the *P. aeruginosa* [5–7]. Lewis et al identified and characterized a transmissible cystic fibrosis strain, designated as MA, which contains a unique pathogenicity island that includes 18 genes from two prophages and a cholera-toxin-like gene [8]. Similarly strain PA14 contains two pathogenicity islands harboring plant and animal virulence genes [9]. Dubois et al found that a nosocomial *P. aeruginosa* strain associated with a severe nursing home outbreak likely picked up and integrated an enterobacterial plasmid containing the *bla* resistance determinant [10]. Thus, it is highly probable that there exists a large uncharacterized pool of contingency genes (a supra-genome) distributed among the numerous *P. aeruginosa* strains that can be addressed via horizontal gene transfer mechanisms (the distributed genome hypothesis [8,9,10] for survival under diverse environmental conditions during polyclonal infections.

Genomic diversity in bacteria is of two types: 1) genetic heterogeneity wherein different strains have different alleles of the same gene; and 2) genomic plasticity wherein different strains have different genes, i.e. the actual genomic complement varies among strains of a species. The identification of extensive genomic plasticity among component strains comprising natural polyclonal infections would provide evidence that there exists ample genetic fodder to support recombination. This recombination could ultimately result in the generation of strains with increased fitness and enhanced survival characteristics for the prevailing environmental conditions, such as would be encountered in the face of host defenses.

Biofilm-associated growth of *P. aeruginosa* has been implicated in several chronic diseases. During progression of cystic fibrosis the decline of pulmonary function has been linked to biofilm-type infection by *P. aeruginosa* [11,12]. Moreover, the expression of *P. aeruginosa* genes involved in quorum sensing [13] and motility [14] has been documented both in biofilm formation *in vitro*, and during chronic lung infection in CF patients [12] *in vivo*. Biofilm-forming *P. aeruginosa* is also the primary pathogen of chronic suppurative otitis media, which is characterized by drainage of fluid (otorrhea) from the middle ear through a perforated tympanic membrane into the external auditory canal [15,16,17]. Bluestone and Klein [18] reported that otorrhea is the most common complication of childhood tympanostomy tube placement, developing in 12 to 32% of cases. Chronic infectious middle-ear disease is the most common cause of conductive hearing loss during childhood [19,20]. Teele *et al.* [21] noted

that the length of time with chronic middle-ear infection is negatively correlated with performance in school and on tests of cognitive ability, speech and language. Pediatric hearing loss associated with chronic otitis media can also result in antisocial behavior and withdrawal [22,23].

The clinical importance of *P. aeruginosa* and the suspected genomic diversity among clinical strains has prompted us to question the wisdom of defining the genome of one laboratory strain as the reference genome. Rather than using the genome of a single strain as a standard, we propose defining and characterizing a “supra-genome” that in its ideal form encompasses all of the nucleotide sequences that are present in the multitude of strains that make up the species. There are numerous biological and clinical ramifications that result from this paradigm shift. Thus, the establishment of a *P. aeruginosa* supra-genomic database would greatly aid diagnostics, typing, infection control, epidemiological studies, and rational drug design for this pathogen. Similar strategies could be adopted for other pathogenic microorganisms that display extensive genomic diversity.

To initiate the goal of developing a supra-genomic database, we report here the construction and characterization of a comprehensive, highly redundant addressable, pooled genomic library constructed from twelve low-passage clinical isolates of *P. aeruginosa*. The unique sequence data derived from this library will serve as the nucleus for the *P. aeruginosa* supra-genomic database.

2. Materials and methods

2.1 *Pseudomonas aeruginosa* clinical isolates

Six clinical strains of *P. aeruginosa* were obtained from otorrheic patients from the clinical microbiology laboratory of Children’s Hospital of Pittsburgh, and an additional five clinical strains isolated from various infections were obtained from Allegheny General Hospital as pure first plate isolates on LB agar. In addition, a clinical isolate (27853) was obtained from the ATCC. All strains were grown in LB broth (Difco Laboratories, Detroit, MI) at 37 °C for one passage and stored frozen in liquid nitrogen in 15% glycerol LB stock. Free-living planktonic cultures for RNA isolation were grown in LB broth to mid-log phase. Biofilm cultures were grown on the inner surface of glass roller-bottles at 37 °C, using 150 ml of 0.2× LB broth. The biofilm on the surface of the glass was rinsed with PBS every 12 h and the medium replaced with 150 ml of fresh broth. The biofilm was harvested after 4 days.

2.2 Design considerations for the construction of the *P. aeruginosa* genomic library

We used Seed *et al.*’s formula [24] to calculate that we needed ~ 230,000 1.5 Kb clones in a pooled genomic library in order to provide the desired 4.5X coverage for each of twelve, 6.2 megabase (Mb) *P. aeruginosa* genomes.

2.3 Preparation of *P. aeruginosa* genomic DNA

Total DNA was isolated from the twelve strains of *P. aeruginosa* using the chloroform/Cetyltrimethylammonium bromide (CTAB) extraction method [25] as previously described [8]. The purified genomic DNAs were tested for quantity and quality by UV absorbance at 260/280 nm and agarose gel electrophoresis, respectively.

2.4 Preparation of 1–3 Kb DNA fragments for ligation

The high molecular weight *P. aeruginosa* genomic DNA was used as starting material for a hydrodynamic shearing fragmentation protocol. Cellular DNA from each of the clinical isolates was independently fragmented to an average size of ~1.5 Kb (with a very small standard deviation) using the computer controlled HydroShear (GeneMachines, San Carlos, CA) nucleic

acid fragmentation device. The resulting hydrodynamically fragmented genomic DNAs were used for ligation into plasmid vectors.

2.5 Construction of the pooled genomic DNA library using fragmented DNA from *P. aeruginosa* clinical isolates

The pooled genomic DNA library was constructed according to the manufacturer's instructions in the highly efficient linearized pCR4Blunt-TOPO *amp^r kan^r* cloning vector which is contained in the TOPO Shotgun Subcloning Kit (Invitrogen, Carlsbad, CA, Cat#: K7060-01). Ten µg of fragmented DNA from each of the twelve clinical strains was pooled and mixed for use in the ligation assays. The pooled DNA fragments were end-repaired using T4 and Klenow DNA polymerases, dephosphorylated with calf intestinal phosphatase and used in ligation reactions with the pCR4Blunt-TOPO vector. The vector to insert ratio was adjusted to 10 ng vector: 200 ng genomic DNA fragments which yielded ~400 recombinant colonies per ng of vector DNA. After ligation the DNA was ethanol precipitated and used to transform *E. coli* TOP10 by electroporation. Transformants were plated on 22 cm² LB agar plates (Genetix Limited, UK, Cat# X6025) containing 30 µg/ml kanamycin (Sigma, St Louis, MO, Cat# 25389-94-0) at a colony density of ~2,000–3,000 per plate. The plates were incubated at 37 °C for 12–16 hours and colonies were picked into LB media containing 10 % glycerol and 100 µg/ml ampicillin (Sigma, St Louis, MO, Cat# 69-52-3) in 384 well plates (Genetix Limited, UK X6001) using the Q Bot, a multifunctional 3-axis robot (Genetix Limited, UK). The collected colonies were grown overnight at 37 °C and then the library was replica plated using the Q Bot system and stored at –80 °C.

2.6 Primary characterization of the *P. aeruginosa* pooled genomic library

To test the characteristics of the newly constructed pooled genomic library, four clones were picked randomly from one copy of each of the replica-plates into 1.5 ml LB/ampicillin media, grown overnight at 37 °C and the recombinant plasmids were isolated using NucleoSpin Robot-96 Plasmid purification system (Macherey-Nagel Inc., Easton, PA) on a Beckman/Coulter Biomek FX liquid handling robot, according to the manufacturer's instruction. The purified plasmid DNA was tested for quantity and quality by UV absorbance at 260/280 nm. 500 ng of each plasmid was subjected to *Eco* RI analysis, which releases the cloned insert, to estimate the insert size by agarose gel electrophoresis. Cloning efficiency was estimated by determining the percentage of clones with a detectable insert.

2.7 Macroarray printing and processing

The pooled genomic library of the *P. aeruginosa* clinical strains was gridded on duplicate 22 cm² nylon membranes (NEN, GeneScreenPlus) using the Q Bot (Genetix Limited, UK) with a 3×3 subarray. With this format one membrane can accommodate approximately ~ 3 complete genomes i.e. 18,400 duplicate colonies harboring ~ 2×10⁷ nucleotides of *P. aeruginosa* genomic DNA. The printed nylon membranes were placed on the surface of LB agar containing 100 µg/ml ampicillin with colony side up and incubated for 10–12 h at 37 °C. Once the proper sized colonies developed (0.4–0.5 mm), they were lysed with 3M filter paper soaked with a 0.5 n NaOH/1.5 M NaCl solution and then neutralized with 1 M Tris-HCl, pH7.5/1.5 M NaCl. The cellular debris was digested away using 50 µg/ml Proteinase K in the presence of 0.5% SDS, at 37 °C for one hour. The membranes were then dried overnight at room temperature. The plasmid DNA was UV-cross-linked to the nylon membranes using a Stratagene UV crosslinker.

2.8 Colony hybridization

The complexity of the *P. aeruginosa* genomic library was determined using colony hybridization of printed colony macroarrays to end-labeled oligonucleotide probes

complementary to the 16S rRNA gene (GGTGAAGTCGTAACAAGGTAG) of *P. aeruginosa*. The end-labeling reaction was carried out in 50 µl, contained 50 µCi ^{32}P - γ -ATP, 3 units of T4 polynucleotide kinase, 1× kinase buffer and 500 ng oligonucleotide. The reaction mix was incubated at 37 °C for 30 min and stopped by heat inactivation of the T4 kinase at 70 °C for 10 min, finally purified on a Sephadex G-25 (Amersham Biosciences, Piscataway, NJ) spin-column. The colony macroarrays were prehybridized at 55 °C for 4 h in prehybridization buffer (1% SDS, 10% dextran sulfate, 50 mg/ml denatures and sheared salmon sperm DNA and 1M NaCl). After prehybridization, the end-labeled oligonucleotide probe was added at a final concentration of 10^6 dpm/ml. Hybridization was for 16 h at 55 °C. The hybridized macroarrays were washed with 2× SSC for 5 min at room temperature followed by 2× SSC containing 1% SDS at 50 °C for 30 min and finally with 0.1× SSC for 30 min at room temperature. The hybridized macroarrays were air dried and exposed to X-ray film (BIOMAX MR, Kodak).

2.9 Differential screening of colony macroarrays

Radioactive cDNA probes for the detection of differentially expressed genes between planktonic and biofilm grown cultures were prepared in RT reactions using 50 µCi ^{32}P -dCTP and 5 µg total RNA as template, isolated from the biofilm forming strain 27853. The 50 µl reaction contained 10 ng/µl random hexamers (Promega), 10 mM DDT (Invitrogen), 0.5 mM dATP dGTP and dTTP, 1 unit/µl RNAsin (Promega) and one unit of MMLV RT (Invitrogen), was incubated at 42 °C for one hour then another unit of enzyme was added and incubated was continued for an additional hour at 42 °C. After the addition of 5 µl 0.5 M EDTA and 1N NaOH the template RNA was removed by incubation at 65 °C for 30 min. With the addition of 15 µl 1 M Tris-HCl, pH 7.5, the reaction mix was neutralized and finally purified on a Sephadex G-25 spin column. The hybridization conditions were as described above. The colony arrays first were hybridized to probe derived from biofilm grown culture. The hybridized membranes were not dried to enable us to strip the first probe off the membrane and reuse it for the hybridization with a second probe prepared from planktonic cultures.

2.10 DNA sequencing

Plasmid DNA templates were prepared for sequencing using the NucleoSpin Robot-96 plasmid purification kit (Macherey-Nagel, Easton PA). Only those constructs containing insert DNA that appeared to be larger than 0.5 kb were used as sequencing templates. The LiCor IR² Gene ReadIR™ and the Beckman Coulter CEQ 2000 XL automated fluorescence sequencing systems were used to sequence the clones as described [8].

2.11 Distribution of novel sequences in the clinical strains of the genomic library

Primer pairs were designed for eight *P. aeruginosa* clones for PCR-based distribution studies in which genomic DNA isolated from each clinical strain were used as templates to determine the number of clinical strains that encodes the tested non-PA01 sequence. Primer pairs used: 75_I15: GAA GGC AGG CAA GCA TCA AG and CGA AGC CGT TTG TTT TCT TG; 112_C5: GCGTGCTGTACCCTTCCTC and GCTCAAGTACCTGCTGCATC; 118_C17: GATCGTCTCGGCGGATTG and TCATCGAGGGCATCGTCAA; 211_O3: CTCTGTAGCCACACCCTTTC and CTTTGGTCCTTCCACTGTTC; 223_C22: TTCCCCACCCGAGCCACTAC and GCGCAGGTAGTGACGGATGA; 234_B2: TCGTAGTGCCGAGTACCAG and TACCTACCTGCCCCAGTTC; 338_G1: CCAGCCGAGCAGGTGATTGA and ATCAGTGCTGGTCCCACCTC; 419_M18: CCGCATACCATCGACATTCA and GCCAAGACGCATACATAGAA. A positive control reaction for each genomic DNA template was also performed using primers specific for the *P. aeruginosa* polyphosphate kinase gene (ppk): TCACAGTGCGGTAAGCACCGG and CTGGAAATGATCAACCGCGAGG. PCR was performed using the Eppendorf MasterTaq

Kit (Brinkman Instruments, Inc., Westbury, NY) in a 25 μ l reaction mixture. Standard reactions conditions included 0.6 units of *Taq* DNA polymerase, 100 ng of template DNA, 10 pmol of each primer, 1.5 mM MgCl₂, and 0.2 mM dNTPs. Reactions were carried out in a Perkin Elmer 9600 thermal cycler and included an initial 10-minute denaturation step at 95 °C. The amplification cycle consisted of 30 sec at 94 °C, 1 min at 55 °C, and 1 min at 72 °C for 35 cycles. This was followed by a final extension step of 7 min at 72 °C and then a 4 °C hold. Reactions were analyzed on 1.5% agarose gels in 1X TBE, stained with ethidium bromide, and imaged using a Kodak IS440 image station.

3. Results

3.1 Library construction

We chose to establish our library in a plasmid vector using a targeted average insert size of 1.5 Kb. This insert size facilitates the sequence analysis of the genomic clones, as in most cases it is possible to complete the sequence of individual inserts in a single forward and reverse sequencing step using standard vector-based M13 and T7 sequencing primers. Since the average genome size of *P. aeruginosa* is ~ 6.2 Mb to achieve the desired redundancy of 4–5x coverage for each of the component genomes, required the construction of a library comprised of 250,000 clones. This level of redundancy will produce a library with a $p > 0.99$ that any given nucleotide sequence will be included.

The design of the highly redundant pooled genomic library of *P. aeruginosa* required the fragmentation of the *P. aeruginosa* DNAs into a narrow fragment range. To achieve this goal the Hydroshear DNA fragmentation method was used which yielded very narrowly ranged fragment size distributions (1–3 Kb) for all 12 of the clinical strains eliminating any concerns of fragment size bias. To ascertain the size distribution of the cloned genomic DNA fragments we analyzed by restriction enzyme digestion two recombinant clones from each of the 650 plates into which we picked the *P. aeruginosa* genomic library. Figure 1 represents a typical result from the agarose gel electrophoretic separation of the fragments released by *Eco* RI digestion of the genomic clones. The linearized cloning vector migrates at ~ 4 Kb whereas the released genomic DNA fragments migrate at a higher mobility corresponding to a range of a few hundred bp to ~3 Kb in size. The apparent average molecular weight of the genomic inserts in the pooled *P. aeruginosa* genomic library is 1.4 Kb.

3.3 Technical characterization of the library

The complexity of the *P. aeruginosa* genomic library was tested by determining the percentage of clones in a subset of the library that contained 16S rDNA. We printed 27,648 clones, which represents $27 \pm 5 \times 10^6$ nucleotides yielding a ~ 4.5–5.2x genomic coverage, in a duplicate pattern on a 22 cm² membrane. Using a ³²P-labeled oligonucleotide probe complementary to *P. aeruginosa* 16S rRNA, a total of 20 positive colonies were detected. Figure 2 shows 1/6 of the membrane with 4 hybridizing clones in duplicate. The complexity of the library was further analyzed by random DNA sequence analysis of 1300 clones from the 249,600 clones comprising the entire pooled *P. aeruginosa* library. From this sequencing effort it was revealed that a single nucleotide sequence was over-represented. Sixteen of the 1300 (1.2%) clones examined possessed sequences that shared homology to the *P. aeruginosa* D3 phage. PCR-based gene distribution studies for the sequences uniquely identified in this study (not present in the reference PA genome of PA01) identified only the PITT E strain as possessing these D3-related sequences. The total number of genomic clones carrying sequences homologous to this phage is estimated at ~3,600. Thus, the presence of this phage does not significantly degrade the complexity of the library.

3.3 Identification of novel sequences

One hundred sixty nine of the 1300 clones (13%) contained unique sequences having no significant nucleotide homology to any DNA sequences contained in PA01 or GenBank. One hundred and thirty six of these (10.5%) analyzed were identified as containing no known sequence anywhere within the clone indicating that the clone was part of a chromosomal insert of a size larger than the clone, or that the clone was episomal in origin. Sequence data for the remaining 33 (2.5%) clones with unique sequences revealed that the nucleotide sequence at one end of the cloned insert was unique (with respect to the *P. aeruginosa* PA01 genome and GenBank), but also included genomic sequences homologous to those of the PA01 strain at the other end of the cloned genomic DNA. Thus, for these later clones it can be determined that the unique genes are integrated within the bacterial chromosome.

To validate that the clones with no known homology to the GenBank database originated from the genomes of the clinical strains of *P. aeruginosa* under study and were not contaminants of some kind, we tested a subset of these unique clones using genomic PCR. From the novel DNA sequences data we designed PCR primers specific for eight clones which showed no significant homology to the PA01 genome. These primer pairs were then used to support amplification from each of 11 of the original *P. aeruginosa* DNAs in a matrix format. The products of the PCR were then analyzed on ethidium bromide-stained agarose gels (Figure 3.). The presence of a band at the predicted molecular weight was taken as evidence that the novel sequence was present in the genome of that clinical isolate. All unique gene products were identified in one or more of the component library strains indicating that their origin was from within the *P. aeruginosa* supragenome.

3.4 Gene expression studies

To test the suitability of the library for the detection of differentially expressed genes, radioactive cDNA probes were prepared by RT-PCR using total RNA extracted from *P. aeruginosa* strain 27853 grown as biofilm on glass roller bottles, and under planktonic conditions. This probe was used to hybridize colony-microarrays prepared from the library that had been printed and grown on nylon membranes. These microarrays were printed using a 3×3 pattern resulting in 18,400 colonies per membrane. The processed membranes were hybridized first with a cDNA probe prepared from biofilm grown bacteria, exposed to x-ray film, stripped and checked for residual signal and then re-hybridized to labeled cDNA prepared from planktonic cells and exposed to X-ray film again. A small segment (<1%) of the 22 cm² autoradiograph is shown in Figure 4. The boxed area contains images of colonies that harbor a gene that is up-regulated in biofilms. This gene, CbrA, has been putatively annotated as part of a two-component regulator.

4. Discussion

In this study we report on our efforts to begin to delineate the supra-genome for *P. aeruginosa* and to explore the differential expression of planktonic and biofilm envirovars. Towards the latter goal, we have identified the CbrA gene as being upregulated in early biofilm development. This gene has been posited as being part of a two-component regulator [26], thus it is possible that this gene is part of the molecular signaling process associated with the transition from a free-swimming form to a sessile form. Infectious populations of *P. aeruginosa* can readily adapt and survive in changing environments. One of the underlying factors hypothesized to support this adaptive behavior is a high degree of genomic plasticity among strains coupled with high rates of horizontal gene transfer. In the current study we have documented significant heterogeneity among clinical strains and have identified numerous novel gene sequences, moreover, we have demonstrated that these unique genes originated within the strains under study. The genomic plasticity we observed among the PA01 and the

clinical strains comprising our pooled genomic library support the concept that there exists a species-level supra-genome that is much larger than the genome of any single organism.

It is not possible to address questions of genomic plasticity using extant libraries or microarrays [27,28,29] of *P. aeruginosa* genes prepared from a single reference strain. Moreover, the use of suppressive subtractive hybridization (SSH) techniques, while economical, does not provide for nearly as comprehensive or systematic an analysis of differences as can be achieved using a whole genome approach [30,31]. The most favorable way to address these problems is through the development of highly redundant addressable genomic libraries that are constructed from a large number of low-passage clinical isolates of the target organism and then use these libraries in conjunction with high throughput DNA sequencing, and high-density array-based screening studies.

We have established a high through-put automated laboratory system wherein it is possible using parallel processing to rapidly identify, isolate and study in depth those genes that are not present in the standard genome of laboratory strains of bacteria. The central element of this system is the Q Bot, a 3-axis multi-tasking robot (Genetix, England) supported by an Biomek FX liquid handling system (Beckman Coulter, Fullerton, CA) at the front end and by a ScanArray 4000 microarray imager (Packard BioScience, Biochip Technologies, Billerica, MA) and an automated DNA sequencing and analysis faculty at the back end. Using these core facilities we have constructed and characterized a highly redundant addressable genomic library of *P. aeruginosa* prepared from 12 clinical isolates. The DNA sequence analyses and the differential hybridization experiments using probes derived from biofilm and planktonic cultures of the same clinical strain demonstrate that this genomic library is an immense resource for differential gene expression studies.

The complexity of the library was demonstrated to be precisely what we had calculated (4–5x/genome) as the expected number of 16S-positive clones was 18–21 (due to the fact that there are four copies of the rRNA genes in the *P. aeruginosa* genome) and we identified 20 positive clones from among 27,000 colonies tested. After factoring in the redundancy of the library (4.5x), the size of the genome of the D3 phage (~56,000 bp), the average insert size of the genomic clones (~1.5 kb), and that the D3 phage sequence was detected only in one clinical strain (PITT E) we would predict ~100 copies of this phage per bacterial cell. Sequence analysis data of randomly picked genomic clones revealed that more than 13% of the sequences are not present in the PA01 reference genome. On the basis of our preliminary sequencing data of 1300 clones (~1 million nucleotides) it can be predicted that our pooled library contains DNA sequences equivalent at least to 4–6 million bp that are not represented in the reference PA01 genome. The roles that these unique distributed genes play in the pathophysiological behavior of this bacterium need to further analyzed. Apparently the novel nucleotide sequences that we detected are widely distributed among the constituent clinical strains of the library and no clustering within a strain (or strains) was observed for the eight unique clones tested. These findings of interstrain genomic plasticity for *P. aeruginosa* are similar to those for the human pathogens *Haemophilus influenzae* and *Streptococcus pneumoniae* [32,33] indicating that high degrees of genomic plasticity may be a common feature of bacterial pathogens.

Determining the origin and the role that genomic plasticity plays in bacterial pathogenesis is of paramount importance. If bacteria are able to re-assort distributed contingency genes from a supra-genome during polyclonal chronic infectious processes this would provide a major survival advantage to pathogens at the population level [9,10,32,33,34]. Recent observations using random amplification of polymorphic DNA (RAPD) [35] and amplified fragment length polymorphism (AFLP) [36] have established that a substantial number of *P. aeruginosa* infections (27 %), [30] are polyclonal. These studies combined with our findings underscore

the need for more extensive studies to determine the supra-genomic complement of *P. aeruginosa* and its role at the species level.

Acknowledgements

This work was supported by Allegheny Singer Research Foundation and by NIDCD grant nos 2R01DC04173 (GDE) and by grants from the Health Services and Resources Administration of DHHS (GDE).

References

1. Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrenner P, Hickey MJ, et al. Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. *Nature* 2000;406(6799):959–64. [PubMed: 10984043]
2. Romling U, Tummeler B. The impact of two-dimensional pulsed-field gel electrophoresis techniques for the consistent and complete mapping of bacterial genomes: refined physical map of *Pseudomonas aeruginosa* PAO1. *Nucleic Acids Res* 1991;19(12):3199–3206. [PubMed: 1905802]
3. Romling U, Schmidt KD, Tummeler B. Large genome rearrangements discovered by the detailed analysis of 21 *Pseudomonas aeruginosa* clone C isolates found in environment and disease habitats. *J Mol Biol* 1997;271(3):386–404. [PubMed: 9268667]
4. Claus H, Rotlich H, Filip Z. DNA fingerprints of *Pseudomonas* spp. using rotating field electrophoresis. *Microb Releases* 1992;1(1):11–16. [PubMed: 1364137]
5. Lewis DA, Jones A, Parkhill J, Speert DP, Govan JR, Lipuma JJ, Lory S, Webb AK, Mahenthiralingam E. Identification of DNA markers for a transmissible *Pseudomonas aeruginosa* cystic fibrosis strain. *American Journal of Respiratory Cell & Molecular Biology* 2005;33(1):56–64. [PubMed: 15834046]
6. He J, Baldini RL, Deziel E, Saucier M, Zhang Q, Liberati NT, Lee D, Urbach J, Goodman HM, Rahme LG. The broad host range pathogen *Pseudomonas aeruginosa* strain PA14 carries two pathogenicity islands harboring plant and animal virulence genes. *Proceedings of the National Academy of Sciences of the United States of America* 2004;101(8):2530–5. [PubMed: 14983043]
7. Dubois V, Arpin C, Noury P, Andre C, Coulangue L, Quentin C. Prolonged outbreak of infection due to TEM-21-producing strains of *Pseudomonas aeruginosa* and enterobacteria in a nursing home. *J Clin Microbiol* 2005;43(8):4129–38. [PubMed: 16081960]
8. Erdos G, Sayeed S, Antalis P, Hu FZ, Hayes J, Goodwin J, et al. Development and characterization of a pooled *Haemophilus influenzae* genomic library for the evaluation of gene expression changes associated with mucosal biofilm formation in otitis media. *Int J Pediatr Otorhinolaryngol* 2003;67(7):749–755. [PubMed: 12791450]
9. Shen, K.; Wang, X.; Post, JC.; Ehrlich, GD. Molecular and Translational Research. Chapter 7. In *Evidence-based Otitis Media*. 2nd Edition.. Rosenfeld, R.; Bluestone, CD.; Hamilton, BC., editors. Decker Inc; 2003.
10. Ehrlich, GD.; Hu, FZ.; Post, JC. Role for Biofilms in Infectious Disease. In: Ghannoum, M.; O'Toole, GA., editors. *Microbial Biofilms*. Washington, DC: ASM Press; 2004. p. 332–358.
11. Pedersen SS, Hoiby N, Espersen F, Koch C. Role of alginate in infection with mucoid *Pseudomonas aeruginosa* in cystic fibrosis. *Thorax* 1992;47(1):6–13. [PubMed: 1539148]
12. Singh PK, Schaefer AL, Parsek MR, Moninger TO, Welsh MJ, Greenberg EP. Quorum-sensing signals indicate that cystic fibrosis lungs are infected with bacterial biofilms. *Nature* 2000;407(6805):762–764. [PubMed: 11048725]
13. Davies DG, Parsek MR, Pearson JP, Iglewski BH, Costerton JW, Greenberg EP. The involvement of cell-to-cell signals in the development of a bacterial biofilm. *Science* 1998;280(5361):295–298. [PubMed: 9535661]
14. O'Toole GA, Kolter R. Flagellar and twitching motility are necessary for *Pseudomonas aeruginosa* biofilm development. *Mol Microbiol* 1998;30(2):295–304. [PubMed: 9791175]
15. Nakagawa T, Yadohisa O, Komune S, Uemura T. A 16-year survey of changes in bacterial isolates associated with chronic suppurative otitis media. *Eur Arch Otorhinolaryngol* 1994;251 (Suppl 1):S27–32. [PubMed: 11894770]

16. Kenna MA, White GW, Wadowsky RM, Post JC, Ehrlich GD. Bacteriology of otorrhea: polymerase chain reaction versus cultures. *Proceedings of the 6th International Symposium on Recent Advances in Otitis Media* 1996;428–430.
17. Melaku A, Lulseged S. Chronic suppurative otitis media in a children's hospital in Addis Ababa, Ethiopia. *Ethiop Med J* 1999;37(4):237–246. [PubMed: 11961874]
18. Bluestone, CD.; Klein, JO. Intratemporal complications and sequelae of otitis media. In: Bluestone, CD.; Stool, SE.; Kenna, MA., editors. *Pediatric Otolaryngology*. 3rd ed.. Philadelphia: W.B. Saunders; 1996. p. 583–647.
19. Eimas PD, Kavanagh JF. Otitis media, hearing loss, and child development: a NICHD conference summary. *Public Health Rep* 1986;101(3):289–293. [PubMed: 3086922]
20. Daly KA, Hunter LL, Giebink GS. Chronic otitis media with effusion. *Pediatr Rev* 1999;20(3):85–93. [PubMed: 10073070]quiz 94.
21. Teele DW, Klein JO, Chase C, Menyuk P, Rosner BA. Otitis media in infancy and intellectual ability, school achievement, speech, and language at age 7 years. Greater Boston Otitis Media Study Group. *J Infect Dis* 1990;162(3):685–694. [PubMed: 2387994]
22. Bennett KE, Haggard MP, Silva PA, Stewart IA. Behaviour and developmental effects of otitis media with effusion into the teens. *Arch Dis Child* 2001;85(2):91–95. [PubMed: 11466180]
23. Haggard MP, Birkin JA, Browning GG, Gatehouse S, Lewis S. Behavior problems in otitis media. *Pediatr Infect Dis J* 1994;13(1 Suppl 1):S43–50. [PubMed: 8159515]discussion S50–44.
24. Seed B, Parker RC, Davidson N. Representation of DNA sequences in recombinant DNA libraries prepared by restriction enzyme partial digestion. *Gene* 1982;19(2):201–209. [PubMed: 6293929]
25. Woo TH, Cheng AF, Ling JM. An application of a simple method for the preparation of bacterial DNA. *Biotechniques* 1992;13(5):696–698. [PubMed: 1358130]
26. Nishijyo T, Haas D, Itoh Y. The CbrA-CbrB two-component regulatory system controls the utilization of multiple carbon and nitrogen sources in *Pseudomonas aeruginosa*. *Mol Microbiol* 200;40(4):917–31. [PubMed: 11401699]
27. Whiteley M, Bangera MG, Bumgarner RE, Parsek MR, Teitzel GM, Lory S, et al. Gene expression in *Pseudomonas aeruginosa* biofilms. *Nature* 2001;413(6858):860–864. [PubMed: 11677611]
28. Wagner VE, Bushnell D, Passador L, Brooks AI, Iglewski BH. Microarray analysis of *Pseudomonas aeruginosa* quorum-sensing regulons: effects of growth phase and environment. *J Bacteriol* 2003;185(7):2080–2095. [PubMed: 12644477]
29. Firoved AM, Deretic V. Microarray analysis of global gene expression in mucoid *Pseudomonas aeruginosa*. *J Bacteriol* 2003;185(3):1071–1081. [PubMed: 12533483]
30. Malloff CA, Fernandez RC, Lam WL. Bacterial comparative genomic hybridization: a method for directly identifying lateral gene transfer. *J Mol Biol* 2001;312(1):1–5. [PubMed: 11545580]
31. Nesbo CL, Nelson KE, Doolittle WF. Suppressive subtractive hybridization detects extensive genomic diversity in *Thermotoga maritima*. *J Bacteriol* 2002;184(16):4475–4488. [PubMed: 12142418]
32. Shen K, Antalis P, Gladitz J, Sayeed S, Ahmed A, Yu S, Hayes J, Johnson S, Dice B, Dopico R, Keefe R, Janto B, Chong W, Goodwin J, Wadowsky RM, Erdos G, Post JC, Ehrlich GD, Hu FZ. Identification, Distribution, and Expression of Novel (nonRd) Genes in Ten Clinical Isolates of Nontypeable *Haemophilus influenzae*. *Infection and Immunity* 2005;73:3479–3491. [PubMed: 15908377]
33. Shen K, Antalis P, Gladitz J, Dice B, Janto B, Keefe R, Hayes J, Ahmed A, Dopico R, Ehrlich N, Jocz J, Kropp J, Yu S, Nistico L, Greenberg DP, Barbadora K, Post JC, Ehrlich GD, Hu FZ. Characterization, Distribution and Expression of Novel Genes Among Eight Clinical Isolates of *Streptococcus pneumoniae*. *Infection and Immunity* 2006;74(1):321–30.
34. Ehrlich GD, Hu FZ, Shen K, Stoodley P, Post JC. Bacterial Plurality as a General Mechanism Driving Persistence in Chronic Infections. *Clin Orthop Relat Res* 2005;437:20–24. [PubMed: 16056021]
35. Bonten MJ, Bergmans DC, Speijer H, Stobberingh EE. Characteristics of polyclonal endemicity of *Pseudomonas aeruginosa* colonization in intensive care units. Implications for infection control. *Am J Respir Crit Care Med* 1999;160(4):1212–1219. [PubMed: 10508809]

36. Pirnay JP, De Vos D, Cochez C, Bilocq F, Pirson J, Struelens M, et al. Molecular epidemiology of *Pseudomonas aeruginosa* colonization in a burn unit: persistence of a multidrug-resistant clone and a silver sulfadiazine-resistant clone. *J Clin Microbiol* 2003;41(3):1192–1202. [PubMed: 12624051]

Figure 1.

Restriction analysis of *P. aeruginosa* genomic DNA clones demonstrating that the average insert size is 1.4 Kb, with more than 80% of the clones carrying an insert. The genomic clones of the library were picked into 650 384 well plates using QBot, a multi-functional robotic system. Plasmid DNA of two randomly selected clones from each plate were digested with *EcoRI* restriction enzyme and analyzed on EtBr stained agarose gel.

Figure 2.

Hybridization of ^{32}P labeled 16S RNA probe to a colony array printed on a nylon membrane, demonstrate that the predicted complexity of the library is preserved. Approximately 27,500 clones (10 % of the total number of clones in the library) of the *P. aeruginosa* genomic library were printed on a 22×22 cm nylon membrane and screened for clones positive for the 16S rRNA genes. The area in this figure represents one sixth of the full membrane. The duplicate patterns of the four positive clones in the area are numbered.

Figure 3.

The distribution of novel DNA sequences in clinical isolates of *P. aeruginosa*. Genomic PCR reactions were performed to confirm the presence or absence of unique nucleotide sequences in the genomes of the clinical strains of *P. aeruginosa*. The resulting PCR products were separated on 1.5% agarose gels and stained with ethidium bromide. Lanes 1–11 are PittA, PittC, PittD, PittE, PittF, 27853, M18851, M18858, W27912, W27931 and W28869 strains respectively, lane 12 is the no-template control, lane MW is the 123 bp ladder (Invitrogen/BRL). Clone IDs: Panel A: ppk (control), Panel B: 118_C17, Panel C: 112_C5, Panel D: 223_C22, Panel E: 234_B2, Panel F: 338_G1, Panel G: 211_O3, Panel H: 75_I15, Panel I: 419_M18.

Figure 4.

Comparative hybridization of printed colony macroarrays to labeled cDNA prepared from planktonic- and biofilm-grown *P. aeruginosa* (strain 27853). Panel A and B are representing the same areas of the same membrane hybridized to ^{32}P labeled total RNA prepared from biofilm grown (Panel A) or planktonic (Panel B) *P. aeruginosa* cultures. In Panel A colonies in the boxed areas represent cloned genes up-regulated in biofilms compared to the free-living planktonic cells (Panel B). The area shown in this figure represents less than 1% of the full surface of the macroarray.