The less travelled road in microbial genetics

Bruce W. Holloway

Fax: +61 3 9857 8756. e-mail: hollowab@ozemail.com.au

Department of Biological Sciences, Monash University, Clayton, Victoria 3168, Australia

Keywords: Pseudomonas aeruginosa, microbial genetic analysis, whole genome analysis, physical and genetic maps, genome sequencing

Background

The use of genetics as a tool for the analysis of the biology of bacteria essentially started in 1946 with the monumental discovery of conjugation in Escherichia coli by Lederberg & Tatum (1946a, b). The key feature of this work was the ability to select for rare recombinant events. A parallel activity was the investigation of the biology of bacteriophages by M. Delbrück, S. E. Luria, A. Lwoff and others. This period in modern microbiological science has been elegantly documented in the book ‘Phage and the Origins of Molecular Biology’ (Cairns et al., 1966). An additional linkage of these two mainstream research activities occurred through the studies on lambda phage, lysogeny in E. coli and the discovery of transduction in Salmonella typhimurium (Lederberg et al., 1951).

The progression of studies using E. coli K-12 involving mutant isolation, mapping, bacteriophages, lysogeny and plasmids laid the foundation for recombinant DNA technology that has been universally applicable to all living organisms. These techniques were developed in the period from about 1980 onwards and have been applied to genomic analysis of a variety of microorganisms. Prior to these advances, there was only a limited application of the genetic techniques developed with E. coli K-12 to other bacteria. Difficulties were encountered in demonstrating that the phenomena of conjugation, transformation and transduction occurred over the spectrum of bacteria. In addition, there were few attempts to show that these genetic techniques could be used for the analysis of bacteria important for human and animal disease, agriculture and the biotechnology industry. Genetic modification of micro-organisms selected for special characteristics was largely restricted to mutational change.

The genetics of Pseudomonas aeruginosa

Selecting an organism in 1953 as a suitable new candidate for genetic analysis required certain parameters to be met. It should be:

- Prototrophic, enabling auxotrophic mutants to be isolated
- Safe in the laboratory, thus excluding many human pathogens
- Lysogenic, enabling the possibility of transduction as a gene transfer mechanism
- Freely available as different isolates.

An optional but desirable extra was that it had some medical importance as funds for medical research were easier to obtain.

P. aeruginosa met all these requirements and a genetic analysis was initiated using a collection of clinical strains obtained from various Australian and South African hospitals. In a few months, auxotrophs were isolated using a newly discovered mutagen, manganous chloride (Demerec, 1951). Using the techniques pioneered by Lederberg & Tatum (1946a, b), crosses were set up between auxotrophic mutants of different strains and from some of these, prototrophic recombinants were produced. The non-random segregation of some non-selected markers was an early indication of linkage (Holloway, 1955, 1956). This was followed by the demonstration of an infectious fertility factor, FP2 (Holloway & Jennings, 1958). These results showed that the system in P. aeruginosa was comparable to that in E. coli, which enabled crosses to be arranged in a more orderly way to detect linkage (Holloway & Fargie, 1960).

Given the importance that the geneticists working on E. coli placed on bacteriophage, it seemed appropriate to obtain some phages for these P. aeruginosa strains. There was no difficulty in identifying lysogenic strains; indeed one strain was shown to have five different prophages (Holloway et al., 1960). This enabled the demonstration of transduction (Holloway & Monk, 1959). As with phages from other genera, the host range and efficiency of infection with some of these phages was shown to be dependent upon the host strain on which they had been propagated, the phenomenon of host induced modification (Holloway & Rolfe, 1963). This is
a manifestation of restriction endonucleases in distinguishing between self and non-self DNA.

During this study on phages, a phenomenon was discovered, which while still unexplained, is of considerable use in matings between two different strains or genera. If P. aeruginosa PAO is grown at 43 °C for five generations or more, it loses the ability to restrict phage previously grown on other P. aeruginosa strains. This loss of restriction function persists when such 43 °C-grown strains are then grown at 37 °C, and this phenotypic loss remains for about 60 generations of growth at the lower temperature. After this period of growth, the wild-type restriction phenotype is regained. The persistence of this restriction-deficient phenotype depends on cell division at the lower temperature. The number of generations required for return to the original phenotype is unaffected by any changes of media, or storage in non-growing conditions (Holloway, 1965). Cultures grown at 43 °C become equally more receptive to the transfer of heterologous DNA of plasmids and bacterial chromosome in both transduction and conjugation (Rolfe & Holloway, 1966). The genetic basis of the persistence of this phenotype is intriguing. It is most unlikely to be a dilution effect of some cellular component produced during growth at 43 °C. The almost clock-like precision of the return to the wild-type phenotype is difficult to explain by some rearrangement of genetic elements in the bacterial genome. It would have been pleasing to think of an experiment that would provide an explanation of this so-called ‘43 °C effect’.

A major objective at this stage of the study of the genetics of P. aeruginosa was the establishment of a comprehensive map. This objective persisted for more than 30 years and new data and new techniques were used to achieve such a map as they became available. Two strains were used to achieve this objective, and they became known as PAO and PAT (P. aeruginosa One and P. aeruginosa Two). A significant feature of gene arrangement in P. aeruginosa had already been detected in data from early crosses and this concerned the arrangement of genes controlling enzymes in biosynthetic pathways. In enteric bacteria it had been shown that loci controlling related functions were contiguous or closely linked to each other, and this arrangement was shown to be important for the regulation of these genes. By contrast, in P. aeruginosa, there is almost no clustering of biosynthetic genes and as was shown subsequently the pattern of regulation is different (Fargie & Holloway, 1965; Isaac & Holloway, 1968).

It became clear that the first isolated sex factor, FP2, was not going to be effective in transferring chromosome at sufficient frequencies to map the entire chromosome of P. aeruginosa. A search was undertaken for other sex factors that would do this. Sex factors like FP2 were easy to find in clinical strains of this organism. However, none of those tested were more effective than FP2 as mapping tools. Maps of P. aeruginosa were prepared with an increasing number of genes mapped but the maps remained linear. The expected goal of circularity was not readily realized (Holloway et al., 1971).

The availability of plasmids carrying resistance genes to antibiotics, resistance transfer factors or more simply R factors, provided a solution to this problem. Most R factors studied then had restricted host ranges; for example, most R factors that could replicate in enteric bacteria could not transfer to or replicate in pseudomonads. Lowbury et al. (1969) isolated a group of R factors from burns patients that were able to move readily between different Gram-negative bacteria. It was subsequently shown that the host range of these so-called IncP1 R factors is most Gram-negative bacteria. When tested in P. aeruginosa some were found to mobilize chromosome, although at low frequency (Stanisch & Holloway, 1971). By using a variety of plasmids promoting chromosome transfer, including FP plasmids and R factors, the first circular genetic map of P. aeruginosa was constructed in strain PAT (Watson & Holloway, 1978).

In the continuing search for effective sex factors, variants of IncP1 plasmids were sought with more efficient properties of chromosome transfer. Such variants were isolated from R68 and one of these, R68.45, was studied in detail (Haas & Holloway, 1976, 1978). R68.45 is 10²-fold more effective in its chromosome-mobilizing ability than R68. Unlike FP2, it can transfer chromosome from a number of chromosome sites. Using a combination of different FP sex factors, including FP2, FP5, FP110 and R68.45, a circular chromosome map of strain PAO was eventually constructed (Royle et al., 1981).

The molecular structure of R68.45 was of particular interest. Jacob et al. (1977) showed that R68.45 had acquired an extra 2.1-2.4 kb fragment compared with R68. Riess et al. (1980) showed that this extra DNA segment involves a repetition of a 2.1 kb sequence already present nearby on R68. This repetitive sequence was studied by Willetts et al. (1981) and found to be a new insertion sequence, IS21. There is one copy of IS21 in R68 and two copies are arranged contiguously in tandem in R68.45. IS21 can transpose at high frequency and it was proposed that the chromosome-mobilizing ability of R68.45 is due to the formation of an R68.45–chromosome cointegrate during transposition of IS21. Schurter & Holloway (1986) showed that IS21 has a promoter site at the left-hand end. IS21 transposes at high frequency from the tandem configuration, because a transposition gene on the left-hand copy is transcribed from the copy of this promoter located in the right-hand copy. The fact that the tandem configuration of IS21 in R68.45 can promote chromosome transfer in a variety of Gram-negative bacteria (Holloway, 1986) suggests that this insertion sequence has a very low specificity of insert sites for bacterial DNA. R68.45 was used by a range of workers to establish effective conjugation systems in a range of organisms including Agrobacterium tumefaciens, Azospirillum brasilense, Erwinia carotovora, Erwinia chrysanthemi, Escherichia coli, Klebsiella pneumoniae, Rhizobium leguminosarum, Sinorhizobium (Rhizo-
bium) meliloti, Rhizobium trifolii, Rhodopseudomonas capsulata, Rhodopseudomonas sphaeroides and Zymomonas mobilis (for references, see Holloway, 1986).

The construction of a circular map for strain PA0 with a series of different sex factors meant that there was no certainty that the different plasmids mobilized chromosome at the same rate. Hence the measured distances between genes on the map were of questionable accuracy. What was needed was a circular chromosome map constructed using only one sex factor. This was achieved using two new tools. A mutant of the R factor R68 was isolated that was temperature-sensitive for replication (Holloway et al., 1982). This mutant was then loaded with a new transposon, Tn2521, isolated from clinical strains of P. aeruginosa. Tn2521 was mapped to the chromosome in those clinical isolates, demonstrating that the mapping techniques that had been developed were available for direct analysis of new strains (Sinclair & Holloway, 1982). The hybrid plasmid, designated pMO514, was then transferred to selected PAO strains with appropriate auxotrophic markers. These were grown at 43 °C and selection was made for streptomycin resistance, a marker encoded by Tn2521. It was found that pMO514 showed stable integration into the PAO chromosome at a range of different sites. When integrated, it mobilized chromosome at high frequency from each site, thus creating a donor strain comparable with the Hfr strains of E. coli.

Time of entry data obtained from a range of crosses enabled the recalibration of the P. aeruginosa PAO chromosome map to 75 min, compared with the former value of 95 min (O'Hoy & Krishnapillai, 1987).

**Genetic analysis of other species of Pseudomonas**

Considerable effort was expended into extending the findings of genetic analysis in P. aeruginosa to other species of the genus Pseudomonas. Pseudomonas putida, given its importance for the study of the unusual metabolic activities of this genus, was an obvious first choice. Lysogeny is very rare in this species, but a transduction system was developed with a phage iso-

New techniques for transformation have meant that uptake of DNA fragments by bacterial cells is no longer dependent upon very specific conditions relating to competency. The importance of these developments is that manipulation of DNA for any organism is of little biological importance unless DNA fragments can be reintroduced into a living cell, be it the one they came from or another organism.

The essential techniques that have enabled the construction of combined physical and genetic maps of bacteria are:

- An individual cosmid library
- PFGE
• Restriction enzyme site mapping
• Complementation.
• DNA probes and Southern hybridization
• DNA and protein sequence databases.

The essential feature of the techniques for general mapping of bacterial genomes is that all involve a fragmentation of the bacterial genome, followed by the identification of the natural order of the fragments as they exist in the chromosome from which they were derived. Cloning has been pre-eminent in this respect, particularly with the development of cloning vehicles that could carry fragments of DNA of increased size. The cosmid has been a major force in genomic mapping of bacteria because up to 40 kb of inserted DNA can be cloned in cosmid vectors. The development of cosmids with wide bacterial host ranges such as pLA2917 (Allen & Hanson, 1983) or pLAFR1 (Friedman et al., 1982) has been particularly fruitful for the application of these techniques to a wider range of organisms.

Methylotrophs provide a good example. These organisms, characterized by being able to use one-carbon compounds like methanol for growth and energy, do not appear to have any gene transfer mechanisms such as conjugation, transformation or transduction. Hence alternative means must be used for genetic analysis. Cosmid libraries using pLA2917 were constructed from Methylobacterium viscosum and Methylobacterium methylophilus AS1 and used to complement known auxotrophic mutants of P. aeruginosa PAO and E. coli K-12. As well, prime plasmids were constructed using a naturally occurring IncP1 plasmid, pMO172, with similar chromosome-mobilizing properties to those of R68.45. Genetic linkage groups were identified for each methylotroph by complementation of a wide range of P. aeruginosa and E. coli auxotrophs using these cosmids and primes. The two species of methylotrophs showed similarities in gene arrangement for the markers examined (Moore et al., 1983; Lyon et al., 1988).

Two key techniques that have enabled the construction of physical maps are PFGE and the identification and use of restriction endonucleases that cut infrequently. For bacteria, about 40 or less fragments should result from endonuclease digestion of the whole genome. The use of the appropriate rare-cutting restriction enzyme results in fragments that can be separated by PFGE. The fragments that can be separated by PFGE range from 5 Mb to 15 kb.

Using these combined techniques, a physical/genetic map of P. aeruginosa PAO has been constructed (Ratnakingsih et al., 1990; Romling & Tümmler, 1991; Holloway et al., 1994). The PAO genome has a 5.9 Mb circular chromosome, 38 SpeI sites and 15 DpnI sites. Further restriction sites have been mapped including DraI, HpaI, ScaI and PacI, but those sites have not been correlated to gene location (for references, see Holloway et al., 1994). More than 200 genes have been located on the SpeI fragments. This has been done by means of probes derived from chromosomal segments cloned in plasmids, bacteriophage or cosmids, individual cloned genes, oligonucleotides synthesized from sequence information from either a protein or DNA base, or PCR-generated sequences.

Physical/genetic maps are now being generated for many bacteria. This approach has resulted in genetic analysis being available to more bacterial genera than with any previous techniques. Organisms for which physical/genetic maps are available include Anabaena, Bacillus subtilis, Bordetella pertussis, Brucella melitensis, Campylobacter, Caulobacter crescentus, Clostridium perfringens, E. coli, Haemophilus influenzae, Haloferax volcanii, Helicobacter pylori, Lactococcus lactis, Listeria monocytogenes, Methanobacterium thermoautotrophicum, Methanococcus voltae, Mycoplasma pneumoniae, Myxococcus xanthus, Neisseria gonorrhoeae, Rhodobacter sphaeroides, Salmonella typhimurium, Streptococcus mutans, Streptomyces coelicolor and Thermococcus celer (see Holloway, 1993, for references).

**Whole genome sequencing**

The next step to the construction of physical/genetic maps of bacteria is the generation of nucleotide sequences of the entire genome. The logistics of this have undergone major changes in the last few years. When the third chromosome of Saccharomyces cerevisiae was sequenced in its entirety, the cost per base was more than US$10, which meant that the whole project had cost more than US$3 million. On this basis, the cost of sequencing the genome of any individual bacterial species would be in the vicinity of US$50 million, which would not be economically achievable. Improved and less expensive technology has totally changed that situation, and the cost of sequencing bacterial genomes is now US$10 cents or less per base and an entire bacterial genomic sequence can be obtained in about a month's work.

This has been largely due to the work of Craig Venter at The Institute for Genomic Research (TIGR) in Maryland, USA. Venter and his group questioned whether it was really necessary for bacterial genome sequencing to go through the labour intensive stage of first determining the order of cloned fragments. They predicted that if more powerful computational methods were available, it should be possible to sequence bacterial genomes using a shotgun sequencing strategy. A single random DNA fragment library could be used, and if the ends of a sufficient number of randomly selected fragments were sequenced and assembled, it should be possible to define the sequence of the whole genome. This approach has been an outstanding success.

The first organism chosen was Haemophilus influenzae Rd and this was the first free-living organism to be entirely sequenced (Fleischmann et al., 1995). Haemophilus influenzae has a circular chromosome of 1830137 bp. It was predicted that this genome should encode 1743 genes and it was found that 736 of these have no similarity with known genes or indeed have any
identifiable role. Of these ‘new’ genes, there was no database match for 389, while 347 matched so-called ‘hypothetical proteins’. Thus whole genomic sequencing is the method par excellence for the discovery of new genes, a procedure until now almost entirely dependent upon mutational strategies.

This technique of whole genome sequencing has now been applied to other bacteria including *Methanococcus jannaschii* (Bult et al., 1996), *Mycoplasma genitalium* (Fraser et al., 1995) and *Helicobacter pylori* (Tomb et al., 1997), to name but a few. Other bacteria have been sequenced and the data have not been published for commercial reasons. Increasingly, the sequence data of bacterial genomes are being made available on the Internet.

The sequencing of the *P. aeruginosa* PAO genome is in progress as a collaborative project between the US Cystic Fibrosis Foundation, the University of Washington Genome Center and the Pathogenesis Corporation, all located in the United States. The Centre for Molecular and Cellular Biology, University of Queensland, is the other collaborator. The data obtained are being progressively released on the Internet at http://www.pseudomonas.com and http://www.bit.uq.edu.au/pseudomonas/Pseudomonas.html.


Having the sequence of a bacterial genome is really the beginning, not the end of genetic analysis of any given bacteria. Genomic data provide an entirely new profile on basic questions of cellular structure, function and evolution. Using available genomic sequence data, a small sector of the genome can be examined in great detail and used to provide knowledge of a particular gene, its neighbouring genes and the region of the chromosome in which it is located. A search can be made for specific genes to identify unusual features of any metabolic feature.

Sequencing of whole bacterial genomes will continue to be limited to a few research centres with the appropriate infrastructure, human resources and financing. It is likely that most of this information will become generally available, so the challenge for microbiologists is how to take advantage of this opportunity. More complex questions and experiments to answer those questions are now possible than ever before. It is important that young microbiologists entering this new whole genome era are adequately and broadly trained in all the disciplines that will be needed to use this major genetic tool of the 21st century.

Acknowledgements

I would like to thank the Society for General Microbiology and the Institute for Biology for the honour of being selected to give the 1998 Kathleen Barton-Wright Lecture. My own work was carried out at the Australian National University, the University of Melbourne and Monash University. This work involved many collaborators over a 40-year period. I am much indebted to all of them and they share the honour of this award.

References


