

The Plasmid Genome Database

The Plasmid Genome Database (PGD) is a regularly updated collection of all fully sequenced plasmids (approaching 500 as of May 2003) with links to structural maps of each plasmid (<http://www.genomics.ceh.ac.uk/plasmiddb/>). The amount of whole genome and whole plasmid sequence data has been growing exponentially (see Fig. 1), generating enormous amounts of data that, if the information can be arranged in a comprehensive and structural way, represent a major resource for many researchers. To our knowledge, this is the first database that has collated all fully sequenced plasmids, including core features, their genetic composition and structural maps (Wackett, 2002).

By definition, plasmids are non-essential extra-chromosomal fragments of DNA that replicate with different degrees of autonomy from the hosts' replicative proteins. Plasmids are present in nearly all bacterial species (Amabile-Cuevas & Chicurel, 1992), range in size from a few to more than 1000 kbp and, as such, can represent a large proportion of the whole bacterial genome. In nature, plasmids appear to increase bacterial genetic diversity and to promote bacterial adaptation by horizontal gene spread (Bergstrom *et al.*, 2000; Gogarten *et al.*, 2002; Levin & Bergstrom, 2000).

The first plasmids were isolated and characterized in the 1950s and were associated with newly acquired antibiotic resistances. Plasmids have since been studied intensively for both their genetic and phenotypic properties, including antibiotic and toxic heavy metal resistance, degradation of xenobiotic compounds, symbiotic and virulence determinants, bacteriocin production, resistance to radiation and increased mutation frequency. These so-called 'accessory functions' (Levin & Bergstrom, 2000),

which facilitate rapid adaptation to new or transient environmental selection pressures, are typically located on mobile genetic elements (MGEs) such as genomic islands, conjugative transposons, mobilizable transposons as well as plasmids. Evidence from bacterial sequencing projects clearly indicates that bacteria adapt and genomes evolve by rearranging existing DNA and by acquiring new sequences (Gogarten *et al.*, 2002; Levin & Bergstrom, 2000). Thus, MGEs have contributed to the evolution of bacteria.

Due to their physical separation from the chromosome, plasmids constitute a substantial and easily identifiable component of this accessory gene pool, but one that was not represented comprehensively in any database. The PGD contains all the plasmid genomes listed in the Entrez Genome pages of the National Center for Biotechnology Information (NCBI) web site (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Genome>) under Archaea/Plasmids ($n=18$), Bacteria/Plasmids ($n=305$), Eukaryotes/Plasmids ($n=13$) and Plasmids ($n=36$ on 2 May 2003). In addition, manual searches identified additional sequenced plasmids elsewhere in GenBank ($n=88$). At the time of writing, the database contains 460 plasmids, including ones of eukaryotic and mitochondrial origin, and meta-data from the informational categories included in

NCBI submissions. These include plasmid name, host, NCBI genome number, accession number, genome size (bp), chromosome type (circular or linear) and date of submission/last update (when defined). Those genomes not included in the Refseq collection are listed by accession number only. Other features of the database include the ability to sort all data in the PGD by category and search locally held plasmid genomes using standard BLAST (Altschul *et al.*, 1990) tools.

One of the intriguing areas of biology that is being highlighted by sequencing large numbers of bacterial genomes is the blurred view of 'plasmids', mega-plasmids and secondary chromosomes. Comparative genomics is raising questions about how to differentiate between secondary chromosomes (apparently of plasmid origins in, for example, vibrio and rhizobium species) and mega-plasmids. Fig. 1 might suggest that there is a size cut-off between most bacterial plasmids and chromosomes at about 1×10^6 bp. This could be a consequence of a lack of basic knowledge; for example, whether a large replicon is truly a secondary chromosome or a mega-plasmid or *vice versa*, an artefact of sampling bias (of sequenced genomes) or simply a consequence of how we define what constitutes a plasmid or a chromosome. Certainly, the sequestration of plasmid

Microbiology Comment provides a platform for readers of *Microbiology* to communicate their personal observations and opinions in a more informal way than through the submission of papers.

Most of us feel, from time to time, that other authors have not acknowledged the work of our own or other groups or have omitted to interpret important aspects of their own data. Perhaps we have observations that, although not sufficient to merit a full paper, add a further dimension to one published by others, or we may have a useful piece of methodology that we would like to share.

Guidelines on how to submit a *Microbiology Comment* article can be found in the Instructions for Authors at <http://mic.sgmjournals.org>

It should be noted that the Editors of *Microbiology* do not necessarily agree with the views expressed in *Microbiology Comment*.

Chris Thomas, Editor-in-Chief

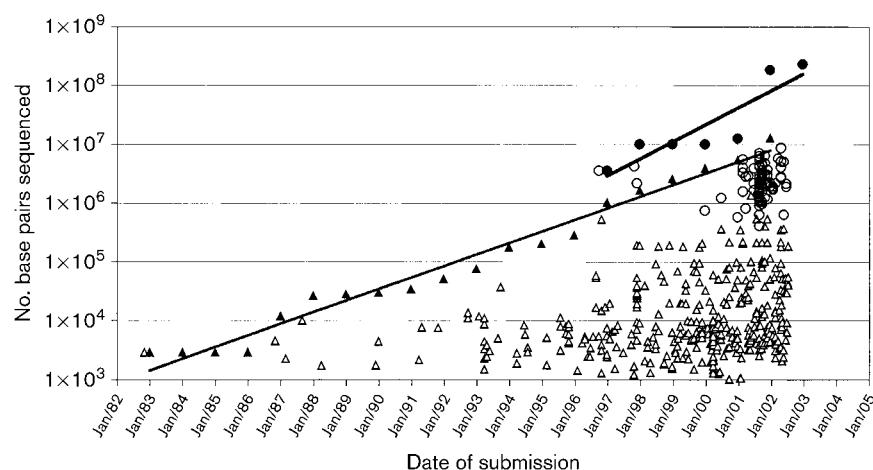


Fig. 1. Graph showing the size of all sequenced plasmids in the Plasmid Genome Database (Δ) and all sequenced chromosomes of bacteria and archaea from the NCBI web site (as of 6 January 2003) ($\circ$) according to date of submission. Additionally, the figure shows trend lines of the total number of base pairs for the sequenced plasmids ($\blacktriangle$) and chromosomes ($\bullet$).

genes to fulfil the role of secondary chromosomes has important general evolutionary implications.

Once all plasmid genomes have been collected into a single resource, analyses can be more easily applied across the entire data set. The PGD currently contains links to graphic structural maps for each plasmid in the database. These plots are constructed directly from the NCBI genome files by the Center for Biological Sequence Analysis (CBS) server in Denmark (<http://www.cbs.dtu.dk/services/GenomeAtlas>) (Pedersen *et al.*, 2000; Jensen *et al.*, 1999). The structural plasmid atlases provide an overview of plasmid structure, including features such as base composition, DNA flexibility, GC-skew, palindrome distributions, the presence of local and global repeats of various types, and gene content (when annotated). These plots highlight the mosaic structure of many plasmids, especially the larger ones. They clearly show that 'backbone' functions, responsible for self-maintenance, for example, genes encoding replication, copy number control, multimer resolution, partitioning, post-segregation killing and horizontal transfer, have similar physical characteristics. By contrast, adaptive genes, probably acquired relatively recently as a consequence of recent environmental selection, can be associated with blocks of

DNA with distinct composition. These blocks often include gene cassettes including ones carried on smaller MGEs (transposons and IS elements) nested between the 'backbone' operons. The observation that recent horizontal gene acquisition gives rise to (or is associated with) atypical nucleotide signatures, relative to the rest of the genome, first proposed by Lawrence & Ochman (1997), has been highlighted in numerous genome sequencing projects since. That this phenomenon is observable among, at least, the larger plasmid replicons has clear implications for plasmid biology. In the context of bacterial adaptation, it perhaps indicates a hierarchy of horizontal gene spread. For example, self-mobilizing plasmids may act more as accidental mediators of intra- and inter-species spread of hitch-hiking adaptive traits associated with the smaller MGEs which are otherwise 'locked' within a host cell/clonal population. This contrasts with the concept of plasmids as drivers of adaptation *per se*, or with them existing as parasites within their host (for discussion, see Bergstrom *et al.*, 2000). Systematic interrogation of the PGD's comprehensive collection of plasmid genomes and structures should reveal patterns that improve our understanding of the roles that different types of plasmid contribute to the biology of their hosts in addition to plasmid biology.

A final point to make about plasmids is to emphasize their biological diversity and the resulting fact that plasmids currently lack a naming convention with real biological meaning. A number of sequenced plasmids lack any name. Plasmids do not share a single phylogenetic history and therefore can not be assigned a classic taxonomy, but they can move through bacterial populations in an independent manner acquiring and losing genes over time. The continued development of the PGD, including the collection of a large amount of meta-data describing each plasmid, should allow the selection and analysis of plasmids based on their phenotypic and genomic characteristics. Therefore, the PGD should improve the effective interrogation of these diverse but important genomic components.

Lars Mølbak,^{1,2} Adrian Tett,²
David W. Ussery,³ Kerr Wall,⁴
Sarah Turner,² Mark Bailey²
and Dawn Field²

¹Danish Veterinary Institute, Bülowsvej 27, DK-1790 Copenhagen V, Denmark

²CEH-Oxford, Mansfield Road, Oxford, OX1 3SR, UK

³Center for Biological Sequence Analysis, Institute of BioZentrum-DTU, Technical University of Denmark, Building 208, DK-2800, Lyngby, Denmark

⁴Pennsylvania State University, Department of Biology, 208 Mueller Lab., University Park, PA 16802, USA

Correspondence: Adrian Tett
(adet@ceh.ac.uk)

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403–410.

Amabile-Cuevas, C. F. & Chicurel, M. E. (1992). Bacterial plasmids and gene flux. *Cell* **70**, 189–199.

Bergstrom, C. T., Lipsitch, M. & Levin, B. R. (2000). Natural selection, infectious transfer and the existence conditions for bacterial plasmids. *Genetics* **155**, 1505–1519.

Gogarten, J. P., Doolittle, W. F. & Lawrence, J. G. (2002). Prokaryotic evolution in light of gene transfer. *Mol Biol Evol* **19**, 2226–2238.

Jensen, L. J., Friis, C. & Ussery, D. W. (1999). Three views of microbial genomes. *Res Microbiol* **150**, 773–777.

Lawrence, J. G. & Ochman, H. (1997).

Amelioration of bacterial genomes: rates of change and exchange. *J Mol Evol* **44**, 383–397.

Levin, B. R. & Bergstrom, C. T. (2000). Bacteria

are different: observations, interpretations, speculations, and opinions about the mechanisms of adaptive evolution in prokaryotes.

Proc Natl Acad Sci U S A **97**, 6981–6985.

Pedersen, A. G., Jensen, L. J., Brunak, S.,

Staerfeldt, H. H. & Ussery, D. W. (2000).

A DNA structural atlas for *Escherichia coli*.

J Mol Biol **299**, 907–930.

Wackett, L. P. (2002). Plasmid and insertion

sequence databases. An annotated selection of

World Wide Web sites relevant to the topics in

Environmental Microbiology. *Environ Microbiol*

4, 916–917.

DOI 10.1099/mic.0.C0123-0