renewing daughter cell and a sister cell committed to differentiation as a gonialblast. Our data suggest that a somatic signal influences this decision by limiting the self-renewing potential. We cannot yet say how this potential is encoded. Our hypothesis is that in *raf*-deficient testes, where cyst progenitor cell identity is disturbed, the signal is lost, and excess stem cell potential is produced. Upon division, both daughters of the germline stem cell inherit some stem cell character. At steady state, this increases the number of cells that become gonialblasts, and somehow prolongs the active state of the stem cell population. Thus, a somatic cell defect leads to a tumour in the germline stem cell lineage, which suggests that some tumours of progenitor cell populations could be initiated by genetic lesions in support cells, rather than in the tumorous cells themselves. In the testis, the Raf-dependent signal may be delivered by cyst progenitor cells, or their cyst-cell daughters. Our mosaic analysis shows that depletion of Raf from just one of the two cyst progenitor cells surrounding a germline stem cell causes a defect. This may indicate a dose effect, where one heterozygous somatic cell is not sufficient to allow normal signalling. It is likely that the signal transducer Raf is engaged, owing to activation of the Epidermal growth factor receptor pathway in somatic cells[16]. We note that in *raf*-deficient testes, differentiation of 2° spermatogonia is blocked as they do not transit to the spermatocyte stage. We believe this is a secondary effect owing to the defective cyst lineage, as we previously showed that a cyst cell signal governs this later transition from 2° spermatogonia to spermatocytes[5].

Somatic signals have been postulated to affect germline stem cell behaviour in the *Drosophila* ovary[17,18]. However, the characterized signals are necessary to maintain germline stem cells, rather than restrict their self-renewing potential, as we find here. Additionally, *raf*-deficient ovaries exhibit no increases in early stage germ cells[19]. Thus, despite the superficial similarities of early germ cell development in ovary and testis, oogenesis and spermatogenesis are emerging as complementary systems from which different principles of stem cell regulation will emerge. □

Methods

The $raf^{11\text{-}29}$ or raf^{400B6}; HS-Raf/+ larvae were heat shocked for 1 hour at 38 °C, daily. At eclosion (day 1), heat shocking was halted, and flies were aged with wild-type FM6/Y siblings. Stains were carried out as in refs 3,11 and 15. Figure 3e, f shows wild-type and mutant cohorts that were aged and processed using Anti-Anillin for cell-cycle stage[20], Anti-Fasc III and Anti-1B1. Cells adjacent to the hub containing dot fusomes and exhibiting nuclear anillin accumulation were scored as late interphase germline stem cells. Wild-type and *raf*-deficient testes differed significantly at day 19 ($P < 0.003$). Figure 4a, b shows clones induced by a 1-hour heat shock at 38 °C in day 1 raf^{400B6}Act> DRaf+>nuc-*lacZ* / Y; HS-Flp/+ or raf^{400B6}Act5C>DRaf+>nuc-*lacZ* / Y; HS-Raf /+; HS-Flp/+ males that were fixed and stained on day 10. The DRaf+ flip-out is described in ref. 21. The continued production of persistent marked clones demonstrates the marking of a stem cell[6].

Received 2 June; accepted 28 July 2000.

1. Hall, P. A. & Watt, F. M. Stem cells:the generation and maintenance of cellular diversity. *Development* **106,** 619–633 (1989).
2. Horvitz, R. H. & Herskowitz, I. Mechanisms of asymmetric cell division: two B's or not two B's, that is the question. *Cell* **68,** 237–255 (1992).
3. Gönczy, P. & DiNardo, S. The germ line regulates somatic cyst cell proliferation and fate during *Drosophila* spermatogenesis. *Development* **122,** 2437–2447 (1996).
4. Fuller, M. T. in *The Development of Drosophila melanogaster* (eds Bate, M. & Martinez-Arias, A.) 71–147 (Cold Spring Harbor Press, Cold Spring Harbor, New York, 1993).
5. Matunis, E. *et al. punt* and *schnurri* regulate a somatically derived signal that restricts proliferation of committed progenitors in the germline. *Development* **124,** 4383–4391 (1997).
6. Gönczy, P., Matunis, E. & DiNardo, S. *bag-of-marbles* and *benign gonial cell neoplasm* act in the germline to restrict proliferation during *Drosophila* spermatogenesis. *Development* **124,** 4361–4371 (1997).
7. Ambrosio, L., Mahowald, A. P. & Perrimon, N. Requirement of the *Drosophila raf* homologue for *torso* function. *Nature* **342,** 288–291 (1989).
8. Dickson, B. *et al.* Raf functions downstream of Ras1 in the Sevenless signal transduction pathway. *Nature* **360,** 600–603 (1992).
9. Lee, T., Feig, L. & Montell, D. J. Two distinct roles for *Ras* in a developmentally regulated cell migration. *Development* **122,** 409–418 (1996).
10. Lin, H., Yue, L. & Spradling, A. C. The *Drosophila* fusome, a germline-specific organelle, contains membrane skeletal proteins and functions in cyst formation. *Development* **120,** 947–956 (1994).
11. de Cuevas, M. & Spradling, A. C. Morphogenesis of the *Drosophila* fusome and its implications for oocyte specification. *Development* **125,** 2781–2789 (1998).
12. Hime, G. R., Brill, J. A. & Fuller, M. T. Assembly of ring canals in the male germ line from structural components of the contractile ring. *J. Cell Sci.* **109,** 2779–88 (1996).
13. McKearin, D. & Ohlstein, B. A role for the *Drosophila* bag-of-marbles protein in the differentiation of cystoblasts from germline stem cells. *Development* **121,** 2937– 2947 (1995).
14. Hardy, R. W. *et al.* The germinal proliferation center in the testis of *Drosophila* melanogaster. *J. Ultrastruct. Res.* **69,** 180–190 (1979).
15. Gönczy, P., Viswanathan, S. & DiNardo, S. Probing spermatogenesis in *Drosophila* with P-element enhancer detectors. *Development* **114,** 89–98 (1992).
16. Kiger, A. A., White-Cooper, H. & Fuller, M. T. Somatic support cells restrict germline stem cell self-renewal and promote differentiation. *Nature* **407,** 750–754 (2000).
17. Cox, D. N. *et al.* A novel class of evolutionarily conserved genes defined by *piwi* are essential for stem cell self-renewal. *Genes Dev.* **12,** 3715–3727 (1998).
18. Xie, T. & Spradling, A. C. *decapentaplegic* is essential for the maintenance and division of germline stem cells in the *Drosophila* ovary. *Cell* **94,** 251–260 (1998).
19. Tran, T. *Genetic Analysis of Germ Line Stem Cell Regulation in* Drosophila 122 (Rockefeller Univ., New York, 1998).
20. Field, C. M. & Alberts, B. M. Anillin, a contractile ring protein that cycles from the nucleus to the cell cortex. *J. Cell Biol.* **131,** 165–178 (1995).
21. Struhl, G. & Basler, K. Organizing activity of wingless protein in *Drosophila*. *Cell* **72,** 527–540 (1993).

Acknowledgements

We thank our laboratory staff, B. Calvi, A. Kiger, M. Fuller, E. Matunis and S. Wasserman for comments and suggestions. Fly stocks and reagents were contributed by the Bonini, Lipschitz, McKearin, Struhl, Wasserman and Xu labs, as well as the Bloomington Stock center and Iowa Hybridoma bank. The NSF and the NIH supports our work.

Correspondence and requests for materials should be addressed to S.D. (e-mail: sdinardo@mail.med.upenn.edu).

The complete sequence of the mucosal pathogen *Ureaplasma urealyticum*

John I. Glass*†, Elliot J. Lefkowitz*, Jennifer S. Glass*†, Cheryl R. Heiner‡§, Ellson Y. Chen‡§ & Gail H. Cassell* †

* *Department of Microbiology, University of Alabama at Birmingham, Birmingham, Alabama 35294, USA*
† *Infectious Diseases Research and Clinical Investigation, Eli Lilly and Company, Indianapolis, Indiana 46285, USA*
‡ *Advanced Center for Genomic Technology, Perkin-Elmer Corporation, Foster City, California, 94404 USA*

The comparison of the genomes of two very closely related human mucosal pathogens, *Mycoplasma genitalium* and *Mycoplasma pneumoniae*, has helped define the essential functions of a self-replicating minimal cell, as well as what constitutes a mycoplasma. Here we report the complete sequence of a more distant phylogenetic relative of those bacteria, *Ureaplasma urealyticum* (*parvum* biovar), which is also a mucosal pathogen of humans. It is the third mycoplasma to be sequenced, and has the smallest sequenced prokaryotic genome except for *M. genitalium*. Although the *U. urealyticum* genome is similar to the two sequenced mycoplasma genomes[1,2], features make this organism unique among mycoplasmas and all bacteria. Almost all ATP synthesis is the result of urea hydrolysis, which generates an energy-producing electrochemical gradient. Some highly conserved eubacterial enzymes appear not to be encoded by *U. urealyticum*, including the cell-division protein FtsZ, chaperonins GroES and GroEL, and ribonucleoside-diphosphate reductase. *U. urealyticum* has six closely related iron transporters, which apparently arose through gene duplication, suggesting that it has a kind of respiration system not present in other small genome bacteria The genome is

§ Present address: Celera Genomics, PE Corporation, Foster City, California 94404, USA

only 25.5% G+C in nucleotide content, and the G+C content of individual genes may predict how essential those genes are to ureaplasma survival.

Ureaplasma urealyticum, a common commensal of the urogenital tract of humans, is gaining recognition as an important opportunistic pathogen during pregnancy. It is the most common organism isolated from the chorioamnion, even in the presence of intact fetal membranes. It is associated with inflammation, premature spontaneous delivery, septicaemia, meningitis and pneumonia in newborn infants[3]. This bacterium is a member of the class Mollicutes, commonly referred to as mycoplasmas. The Mollicutes are the smallest-known free-living microorganisms. Although they evolved from Gram-positive ancestors, mycoplasmas lack a cell wall, which is their most distinguishing feature. The species that parasitizes humans, *U. urealyticum*, is further subtyped into 14 serovars; however, recent molecular characterization of the *U. urealyticum* serovars has led to an effort to reclassify them into two species[4]. *U. urealyticum* would include the ten large genome serovars (0.88–1.2 megabase pairs (Mbp)); *U. parvum* would include the small genome serovars 1, 3, 6 and 14 (0.75–0.76 Mbp). We have sequenced the genome of an isolate of *U. urealyticum* serovar 3 (or *U. parvum* in the proposed revised classification), the serovar most commonly isolated from humans[3].

The genome of *U. urealyticum* serovar 3 is a circular chromosome comprising 751,719 bp. This is smaller than any other sequenced microbial genome except *M. genitalium*[1]. The G+C content is 25.5%. The genome contains 613 predicted protein-coding genes and 39 genes that code for RNAs. The genes comprise 93% of the genome. We assigned biological roles to 53% of the protein-coding genes (see Supplementary Information); 19% of the genes are

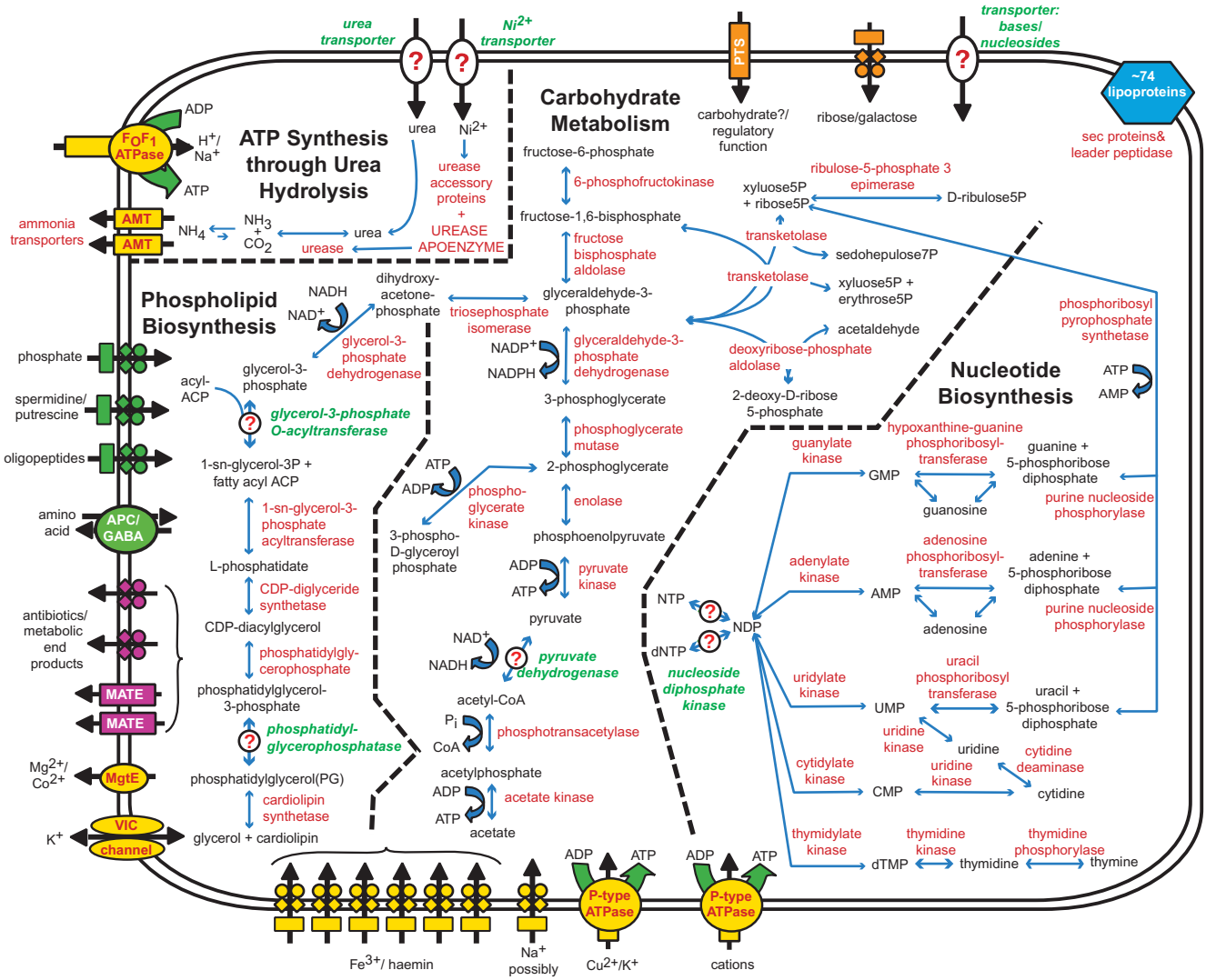

Figure 1 Metabolic pathways and substrate transport in *U. urealyticum*. An integrated view is shown of the transporters and main metabolism elements of an *U. urealyticum* cell, deduced from the set of genes for which we can predict functions. Metabolic products are shown in black and ureaplasma proteins in red. Question marks denote enzymes or transporters not identified that would be necessary to complete pathways, and missing enzyme and transporter names are shown in green italics. Transporters are coloured according to their substrates: yellow, cations; green, anions and amino acids; orange, carbohydrates; purple, multidrug and metabolic end product efflux. Arrows indicate the direction of substrate transport. Transporters are (1) an F-type ATPase; (2) two Amt ammonium transporters; (3) CopA, a copper-importing P-type ATPase, and PacL, a cation transport P-type ATPase; (4) a K^+ channel from the voltage-gated ion channel (VIC) superfamily; (5) a Mg^{2+} MgtE transporter; (6) XasA a glutamate:GABA antiporter from the amino-acid-polyamine-organocation (APC) transporter superfamily; (7) two multidrug antimicrobial extrusion family transporters (MATE); (8) a *ptsH* element of a phosphoenolpyruvate-dependent sugar phosphotransferase transport system (PTS), which probably has a regulatory instead of transport function as the genome lacks the PTS EI component needed for sugar transport; and (9) a broad array of ABC transporters. The ABC-type transporters are shown with a rectangle for the substrate-binding protein, diamonds for the membrane-spanning permeases and circles for the ATP-binding subunits. In some cases we cannot identify all the components of these transporters. ABC efflux transporters have no substrate-binding domains.

similar to hypothetical genes of unknown function, and 28% are unique hypothetical genes with no significant similarities to putative or demonstrated genes in other organisms. On the basis of the gene distribution on the two strands and on transitions in GC skew, we project that the *U. urealyticum* origin of replication is upstream of *dnaA*, which is gene UU001 in our nomenclature.

U. urealyticum generates 95% of its ATP through the hydrolysis of urea by urease[5]. Since the discovery in 1966 that ureaplasmas metabolize urea, the physiology of this energetic process has been studied to gain an understanding of what appears to be a unique system[6]. The hypothesis is that hydrolysis of urea generates an electrochemical gradient through accumulation of intracellular ammonia/ammonium. The gradient fosters a chemiosmotic potential that generates ATP (Fig. 1).

Studies of *U. urealyticum* energy production predict three enzymatic components in this process: urease; an ammonia/ammonium transporter; and an F_OF_1-ATPase[5]. The role in this process of the highly potent *U. urealyticum* urease, which is 30–180-fold more efficient than that reported for other bacterial ureases[7], has been shown[8,9]. Urease is a three-subunit Ni^{2+}-binding enzyme. The other four *ure* genes code for accessory proteins that supply the functional metallocentre to the apoenzyme. Although a nickel transporter could not be identified, candidates for genes encoding the Ni^{2+} transporter or elements of it include the Mg^{2+} transporter *mgtE* (ref. 10) and two hypothetical genes, UU101 and UU555, which had polyhistidine regions characteristic of Ni^{2+}-binding proteins. There were no proteins similar to known urea transporters in either prokaryotes or eukaryotes.

The intracellular concentration of ammonia generated through urea hydrolysis in *U. urealyticum* has been measured at 21 times the extracellular concentration. This excess suggests efflux of ammonium/ammonia through a saturable uniporter rather than by passive diffusion[5]. We identified a pair of genes that code for ammonia transporters of the Amt family. Although the physiology of Amt family transporters is controversial[11–13], the probable mechanism in ureaplasma is energy-independent-facilitated diffusion. The electrochemical gradient created by the export of ammonium presumably results in an ATP molecule generating proton influx through the *U. urealyticum* F-type ATPase.

As predicted by the inhibition of its ATP synthesis by *N*,*N*′-dicyclohexylcarbodiimide, *U. urealyticum* contains an F_OF_1-ATPase[5]. This family of enzymes comprises two oligomeric complexes: the F_O transmembrane proton channel, and the F_1 catalytic complex. The ureaplasma δ subunit of the F_1 complex, which in other organisms is the interface between the F_O and F_1 subunits and causes the permeability of the transmembrane proton channel F_O, is present as two distinct peptides. UU134 corresponds to the amino-terminal two-thirds of the consensus eubacterial δ, and UU133 corresponds to the normally more conserved carboxyl terminus. Speculatively, these atypical features may be linked to the unique ATP production pathways of ureaplasmas. The sources of the 5% of ureaplasma ATP production not resulting from urea hydrolysis are most probably substrate phosphorylation[5] (Fig 1). Enzymatic studies show that *U. urealyticum* has a limited capacity to metabolize carbohydrates.

We did not find genes for the *de novo* biosynthesis of purines or pyrimidines in *U. urealyticum*. Pathways for synthesis of RNA and DNA precursors are largely complete except for several notable missing enzymes (Fig. 1). There is no recognizable ribonucleoside-diphosphate reductase for conversion of ribonucleosides to deoxyribonucleosides, unlike in *M. genitalium* and *M. pneumoniae*. So *U. urealyticum* must import all its deoxyribonucleosides and/or deoxyribonucleoside precursors, or have a mechanism for converting ribonucleosides to deoxyribonucleosides.

As a result of its very limited biosynthetic capacity, ureaplasmas must import more of the nutrients needed for growth than do most other bacteria. We can identify 28 different *U. urealyticum* transporters, which represent 9 transporter families (Fig. 1). Despite this diversity of transport systems, there are a number of surprising absences, such as the previously mentioned lack of transporters for bases, nucleotides, nickel and urea. Missing transporters may be found among the 33 hypothetical *U. urealyticum* proteins that have 5 or more predicted transmembrane regions.

Unexpected among the 19 ABC transporters are six different Fe^{3+} and/or haemin transporters. Iron is essential for the function of many enzymes; however, the low solubility of Fe^{3+} at physiological pH forces most organisms to develop special mechanisms to acquire necessary amounts. Phylogenetic analysis of the ten membrane-spanning protein subunits of the *U. urealyticum* iron transporters indicated that they all cluster together with respect to other known bacterial proteins and form two related, but distinct paralogous groups within this *U. urealyticum* gene family. Thus, we think that this paralogous family arose from gene duplication events during ureaplasma evolution. In the apparent absence of the complement of transcriptional regulators found in most other bacteria, we speculate that *U. urealyticum* increased its capacity to import iron by increasing the number of iron transporter genes. This indicates that the availability of iron may be a limiting factor of ureaplasma growth; but this is puzzling. Because mycoplasmas have flavin-terminated respiratory chains, they lack the redox enzymes that normally use iron in the form of cofactors. This is especially true for *U. urealyticum* because none of its predicted enzymes requires iron. The most likely explanation is that it uses iron in respiration.

Iron is associated with the membranes of *Mycoplasma capricolum*[14] and *M. pneumoniae*[15]. In these mycoplasmas, the iron may act as a reversible electron carrier in concert with NADH oxidase, or as a component of an iron-containing electron storage protein[15]. Unlike the other mycoplasmas, *U. urealyticum* lacks both NADH oxidase activity[16] and an NADH oxidase gene. Even if ureaplasmas use iron in their respiration, the iron use is probably different from that in other mycoplasmas.

Five ureaplasma proteins, urease[17], immunoglobulin-α (IgA) protease[18], phospholipases A and C[20], and MBA[20], along with the ureaplasma enzymatic machinery for generating hydrogen peroxide have been proposed as virulence factors. The pathogenic effect of urease is caused by its generation of ammonia[17]. IgA protease may give ureaplasmas the capacity to invade the upper urogenital tract by degrading a principal component of the mucosal immune system. Ureaplasma phospholipases may be responsible for premature labour by altering prostaglandin biosynthesis[19]. Neither an IgA protease nor phospholipase A or C can be identified. These *U. urealyticum* enzymes may have diverged so far from orthologues in other bacteria that they are unrecognizable, or they may have convergently evolved an IgA protease and phospholipases with no recognizable sequence similarity to known enzymes. For instance, UU292 is a conserved hypothetical lipoprotein with a GDSL motif characteristic of some lipases[21], and UU441 is a unique hypothetical gene that contains a motif found in the subtilase family of serine proteases (the ureaplasma IgA protease is a serine protease[18]).

Hydrogen peroxide is the haemolysin of *M. pneumoniae* and other mycoplasmas, which led to the hypothesis that mycoplasma virulence is in part the result of reactive oxygen being secreted in close proximity to host cell membranes[22]. Unlike *M. pneumoniae*, *U. urealyticum* haemolysin activity is not inhibited by catalase[23]. Thus, ureaplasma haemolytic activity may be enzymatic. *U. urealyticum* has two haemolysins. The *hlyC* gene has a corresponding orthologue in *M. pneumoniae*; however, because *M. pneumoniae* haemolysis is attributed to H_2O_2, it is likely that the principal haemolysin in *U. urealyticum* is *hlyA*. Therefore *hlyA* may be a new ureaplasma virulence factor. Orthologues of the haemolysin *hlyA* have both haemolysin and cytotoxic activity in *Serpulina hyodysenteriae* and several mycobacteria. The haemolysis is thought to be due to pore formation. Serpulina and mycobacteria species lacking this gene are non-pathogenic[24].

Comparisons of two very closely related bacterial genomes, *M. pneumoniae* and *M. genitalium*, have helped to define the essential functions of a self-replicating minimal cell, as well as what constitutes a mycoplasma[25,26]. Determining what is a minimal cell was aided by the fact that the gene set of *M. genitalium* was essentially a subset of the gene set of *M. pneumoniae*; however, because the evolutionary divergence of those two mycoplasmas is so recent, the comparison has limited resolution. Phylogenetic analysis shows that *U. urealyticum* is a member of the same clade of the class Mollicutes as *M. pneumoniae* and *M. genitalium*; but it is different enough to make the comparison more informative. Of the 613 *U. urealyticum* protein-coding genes, only 324 are homologous to *M. genitalium* genes and *M. pneumoniae* genes (all *M. genitalium* genes have orthologues in *M. pneumoniae*[25]). No function can be predicted for 77 of the genes shared by the 3 mycoplasmas. Ten of those conserved hypothetical genes are found only in mycoplasmas (see Supplementary Information or http://genome.microbio.uab.edu/uu).

The evolutionary divergence of *U. urealyticum* from the other mycoplasmas is shown through analysis of the relative gene order of orthologous genes among the three species. The relative gene order in *M. genitalium* and *M. pneumoniae* is largely conserved, although additional genes are interspersed in the larger genome of *M. pneumoniae* (Fig. 2). Similar comparisons of the relative genome locations of *M. genitalium* and *U. urealyticum* orthologues, and *M. pneumoniae* and *U. urealyticum* orthologues show that there is very little conservation of gene order across large fractions of the genomes. The regions of synteny between *U. urealyticum* and *M. genitalium* or *M. pneumoniae* consist, in most cases, of genes that are functionally related, such as the ribosomal proteins or the oligopeptide transport genes.

We can predict a function or cellular location for 76 proteins coded for by *U. urealyticum* that are not in the *M. genitalium* and *M. pneumoniae* genomes. Many of those genes are involved in ATP production by urea hydrolysis, which is unique to ureaplasmas, and in iron acquisition. These differences may be why ureaplasmas grow more robustly *in vitro* than other minimal bacteria, and can replicate in both the lower and upper urogenital tracts. The ureaplasma phenotype may largely be the result of its unique set of energy production and respiration genes.

Because the *M. genitalium* genome is so far the smallest discovered for any free-living organism, this mycoplasma has become the benchmark for approximation of the minimal gene set necessary for life. With that view of *M. genitalium* in mind, it is more interesting to examine the *M. genitalium* genes that have no orthologues in *U. urealyticum*. There are 74 *M. genitalium* proteins with some functional annotation that do not have orthologues in *U. urealyticum*. Ten of the *M. genitalium* genes missing in *U. urealyticum* are involved in energy metabolism. In terms of the

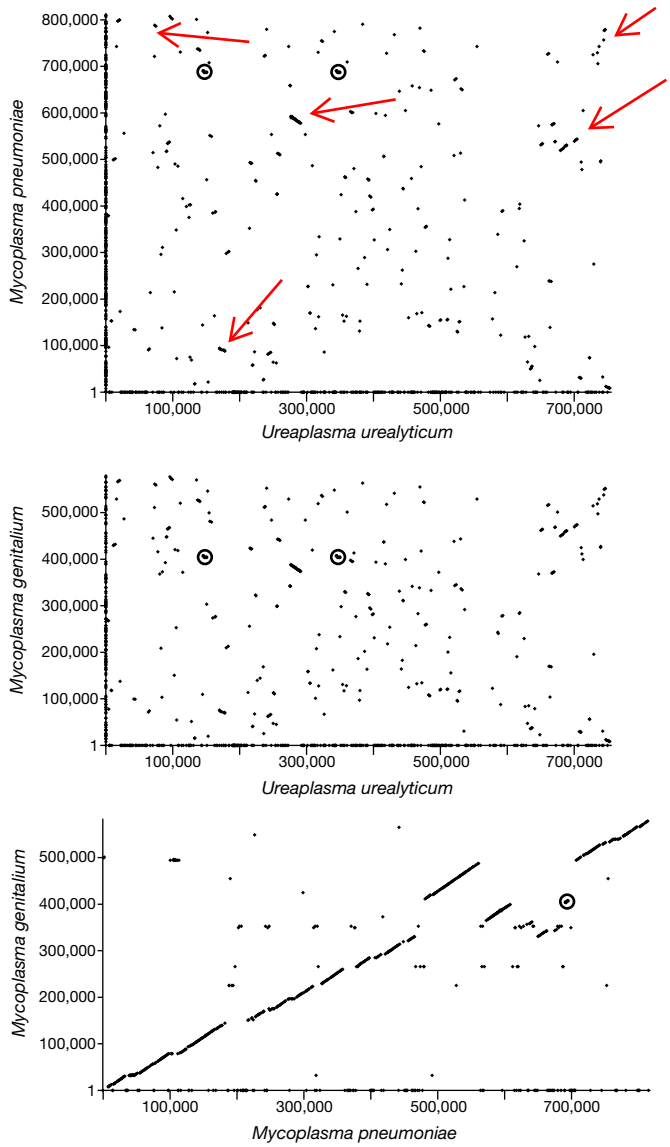

Figure 2 Conservation of gene order among *U. urealyticum*, *M. pneumoniae* and *M. genitalium*. Genes with corresponding orthologues in *U. urealyticum*, *M. genitalium* and *M. pneumoniae* were plotted with respect to their genome location. Base 1 for each organism corresponds to the starting base of the *dnaA* gene. The circled points represent the location(s) of the ribosomal RNA operon(s). There are two small syntenic regions between *U. urealyticum* and the other two organisms formed by the ribosomal proteins (275–291 kb) and the oligopeptide transporter genes (699–704 kb).

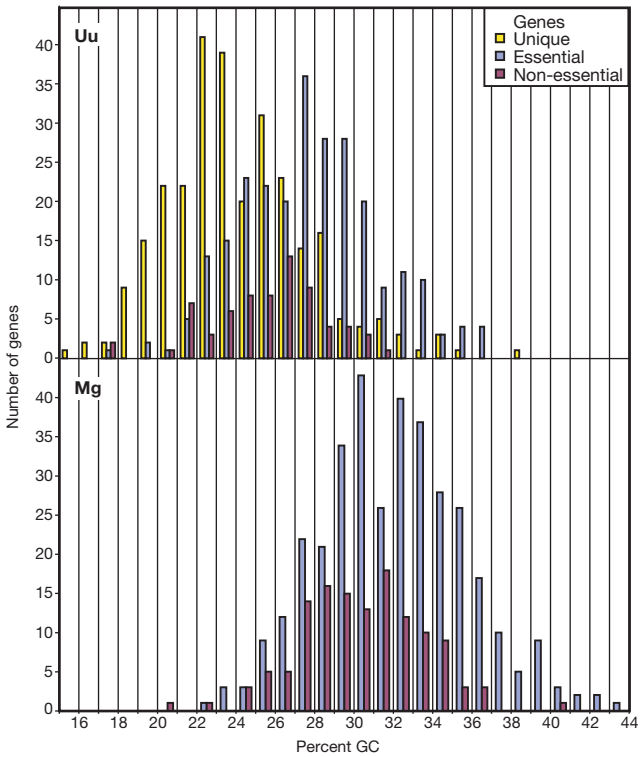

Figure 3 Distribution of *U. urealyticum* and *M. genitalium* genes according to percentage G+C. The base composition of protein-coding genes classified as unique, non-essential and potentially essential are plotted for *U. urealyticum* (Uu, top panel) and *M. genitalium* (Mg, bottom panel) (see text). The number of genes (indicated on the vertical axis) falling within a particular G+C frequency bin (indicated on the horizontal axis) is plotted for each of the gene classifications.

number of genes involved, ATP production by urea hydrolysis is a simpler process than is carbohydrate metabolism. The absence of other genes is more surprising. *U. urealyticum* lacks the heat shock protein/chaperonins GroEL and GroES. These double-ring prokaryotic chaperonins, which mediate protein folding within the cell, are found in all other sequenced microbial genomes, although they are not essential genes *in vitro*[26]. Perhaps the most intriguing absence in ureaplasma is the cell-division protein FtsZ. All bacterial genomes sequenced to date, except for the chlamydias and the archeaon *Aeropyrum pernix* K1, contain *ftsZ*. Because chlamydia cell division takes place in eukaryotic cell vacuoles, FtsZ protein is not necessary. Among other eubacteria, FtsZ is presumed to be an essential component of the cell-division mechanism.

The mean base composition of the *U. urealyticum* genome is 25.5% G+C. The *U. urealyticum* genome is much more A+T-rich than any of the other prokaryotic genomes sequenced to date. Low G+C content is a general characteristic of the Mollicutes. The enzymatic cause of the mutation pressure leading to A+T enrichment in *U. urealyticum* is possibly a decreased capacity to remove uracil from DNA. Sources of uracil in DNA are mis-incorporation of dUTP by DNA polymerase and spontaneous deamination of deoxycytidine residues. Repair of the resulting G·U mismatches by DNA polymerase creates an A/T-biased mutation pressure. Two enzymes that could affect this process are dUTPase, which prevents dUTP from being incorporated into DNA, and uracil-DNA glycosylase, which removes uracil residues from DNA. We could not identify a dUTPase, nor could any dUTPase activity be detected in *U. urealyticum*[16]. *U. urealyticum* does have a uracil-DNA glycosylase; however, enzymatic analysis of ureaplasma extracts did not detect any uracil-DNA glycosylase activity[27]. The uracil-DNA glycosylase of *U. urealyticum* and other low G+C mycoplasmas may remove uracil from DNA inefficiently. This is suggested by analysis of the uracil-DNA glycosylase of *Mycoplasma lactucae* (30% G+C), which has a K_m for uracil-containing DNA that is 40–1,000 times greater than values reported for other organisms[28].

Although the biased mutational pressure acts on the whole genome, the genome does not respond uniformly. Presumably, essential genes may have fewer sites where mutations are viable. We examined the G+C contents of sets of *U. urealyticum* genes that we believed were likely to be essential or non-essential to determine whether the high and low G+C contents were predictive of essentiality. We designated a subset of ureaplasma genes on the basis of an *M. genitalium* study that used transposon mutagenesis to classify genes as either non-essential or possibly essential. A set of 129 of the 480 protein-coding genes of *M. genitalium* was deleted using transposon mutagenesis, and thus shown to be non-essential. Of the remaining 351, at least 265 are probably essential[26]. There are *U. urealyticum* homologues for 69 of the *M. genitalium* non-essential genes and for 255 of the possibly essential genes. That leaves 289 *U. urealyticum* genes that have no homologues in *M. genitalium* (referred to as 'unique' in this analysis). The three sets had different but overlapping G+C percentage distributions (Fig. 3). The histogram suggests that genes with lower G+C contents are likely to be expendable, at least under *in vitro* growth conditions. □

Methods

U. urealyticum serovar 3 isolate

We obtained the *U. urealyticum* serovar 3 isolate from E. A. Freundt (Institute of Medical Microbiology, University of Aarhus, Aarhus, Denmark). The sequenced isolate of *U. urealyticum* serovar 3 is available from the American Type Culture Collection as ATCC number 700970. We grew bacteria in 10B media supplemented with 20% fetal calf serum[20].

Genome sequencing

We used a hybrid sequencing strategy that we refer to as CROSS (for complete random and ordered shotgun sequencing). In the initial phase of CROSS, we generated a series of libraries containing *U. urealyticum* DNA that had been sheared or partially digested with restriction endonucleases and then size fractionated before insertion into vectors. We generated dye-primer sequences from lambda libraries containing 7–11-kilobase pair (kbp) inserts, and plasmid libraries containing 1.5–2.5-kbp inserts. All sequencing was done using ABI 373 or 377 sequencers. When sequence coverage was at ~5×, we shifted from random shotgun to ordered shotgun sequencing[29]. We identified those lambda clones that had poor sequence coverage between their two end sequences, and generated sequences from 1.5–2.5-kbp insert pUC18 libraries made from sheared polymerase chain reaction (PCR) amplicons of the lambda clone inserts. In this phase, we used a mixture of dye-primer and dye-terminator sequencing. Next, small mapped gaps (< 3.0 kbp) in the assembled sequence were closed by sequencing PCR products that covered those gaps. Unmapped gaps were mapped either by performing PCRs using all combinations of primers that annealed to the ends of all the contigs, or by sequencing directly from genomic DNA templates[30].

We used AutoAssembler 1.4 (PE Biosystems) for sequence assembly. Because AutoAssembler could not assemble all 11,291 sequences used in the project at once, we assembled the genome in ten segments by grouping those sequences from defined regions of the genome into subassemblies. Each DNA sequence was manually examined and edited. Every base determined from cloned *U. urealyticum* genomic DNA was sequenced using at least two different templates. Regions sequenced only from PCR products or genomic DNA templates were sequenced on both strands. We confirmed the sequence assembly by a combination of Southern blotting, PCR, and verifying that the map locations of the end sequences from lambda and plasmid inserts agreed with insert sizes. A description of the computational tools and databases that we used to annotate the individual genes can be found at http://genome.microbio.uab.edu/uu (see also Supplementary Information).

Received 13 March; accepted 26 July 2000.

1. Fraser, C. M. *et al.* The minimal gene complement of *Mycoplasma genitalium*. *Science* **270,** 397–403 (1995).
2. Himmelreich, R. *et al.* Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Res.* **24,** 4420–4429 (1996).
3. Cassell, G. H., Waites, K. B., Watson, H. L., Crouse, D. T. & Harasawa, R. *Ureaplasma urealyticum* intrauterine infection: role in prematurity and disease in newborns. *Clin. Microbiol. Rev.* **6,** 69–87 (1993).
4. Bradbury, J. M. International Committee on Systematic Bacteriology Subcommittee on Mollicutes: Minutes of the Interim Meeting 12 and 18 July 1996. Orlando, Florida, USA. *Int. J. System. Bacteriol.* **47,** 911–914 (1997).
5. Smith, D. G., Russell, W. C., Ingledew, W. J. & Thirkell, D. Hydrolysis of urea by *Ureaplasma urealyticum* generates a transmembrane potential with resultant ATP synthesis. *J. Bacteriol.* **175,** 3253–3258 (1993).
6. Shepard, M. C. Human mycoplasma infections. *Health Laboratory Science* **3,** 163–169 (1966).
7. Mobley, H. L., Island, M. D. & Hausinger, R. P. Molecular biology of microbial ureases. *Microbiol. Rev.* **59,** 451–480 (1985).
8. Romano, N., La Licata, R. & Russo Alesi, D. Energy production in *Ureaplasma urealyticum*. *Pediatr. Infect. Dis.* **5,** 308–312 (1986).
9. Neyrolles, O., Ferris, S., Behbahani, N., Montagnier, L. & Blanchard, A. Organization of *Ureaplasma urealyticum* urease gene cluster and expression in a suppressor strain of *Escherichia coli*. *J. Bacteriol.* **178,** 647–655 (1996).
10. Smith, R. L., Thompson, L. J. & Maguire, M. E. Cloning and characterization of MgtE, a putative new class of Mg^{2+} transporter from *Bacillus firmus* OF4. *J. Bacteriol.* **177,** 1233–1238 (1995).
11. Siewe, R. M. *et al.* Functional and genetic characterization of the (methyl)ammonium uptake carrier of Corynebacterium glutamicum. *J. Biol. Chem.* **271,** 5398–5403 (1996).
12. Jayakumar, A., Epstein, W. & Barnes, E. M. J. Characterization of ammonium (methylammonium)/potassium antiport in *Escherichia coli*. *J. Biol. Chem.* **260,** 7528–7532 (1985).
13. Soupene, E., He, L. D. Y. & Kustu, S. Ammonia acquisition in enteric bacteria: physiological role of the ammonium/methylammonium transport B (AmtB) protein. *Proc. Natl Acad. Sci. USA* **95,** 7030–7034 (1998).
14. Bauminger, E. R. *et al.* Iron storage in *Mycoplasma capricolum*. *J. Bacteriol.* **141,** 378–381 (1980).
15. Pollack, J. D., Merola, A. J., Platz, M. & Booth, R. L. J. Respiration-associated components of *Mollicutes*. *J. Bacteriol.* **146,** 907–913 (1981).
16. Pollack, J. D., Williams, M. V. & McElhaney, R. N. The comparative metabolism of the mollicutes (Mycoplasmas): the utility for taxonomic classification and the relationship of putative gene annotation and phylogeny to enzymatic function in the smallest free-living cells. *Crit. Rev. Microbiol.* **23,** 269–354 (1997).
17. Ligon, J. V. & Kenny, G. E. Virulence of ureaplasmal urease for mice. *Infect. Immun.* **59,** 1170–1171 (1991).
18. Kilian, M., Brown, M. B., Brown, T. A., Freundt, E. A. & Cassell, G. H. Immunoglobulin A1 protease activity in strains of *Ureaplasma urealyticum*. *Acta Pathol. Microbiol. Scand. B* **92,** 61–64 (1984).
19. De Silva, N. S. & Quinn, P. A. Localization of endogenous activity of phospholipases A and C in *Ureaplasma urealyticum*. *J. Clin. Microbiol.* **29,** 1498–1503 (1991).
20. Zheng, X. *et al.* Small repeating units within the *Ureaplasma urealyticum* MB antigen gene encode serovar specificity and are associated with antigen size variation. *Infect. Immun.* **63,** 891–898 (1995).
21. Upton, C. & Buckley, J. T. A new family of lipolytic enzymes? *Trends Biochem. Sci.* **20,** 178–179 (1995).
22. Somerson, N. L., Walls, B. E. & Chanock, R. M. Hemolysin of *Mycoplasma pneumoniae*: tentative identification as a peroxide. *Science* **150,** 226–228 (1965).
23. Shepard, M. C. & Masover, M. C. in *The Mycoplasmas* (eds Barile, M. F. & Razin, S.) 451–494 (Academic, New York, 1979).
24. Wren, B. W. *et al.* Characterization of a haemolysin from *Mycobacterium tuberculosis* with homology to a virulence factor of *Serpulina hyodysenteriae*. *Microbiology* **144,** 1205–1211 (1998).
25. Himmelreich, R., Plagen, H., Hilbert, H., Reiner, B. & Herrmann, R. Comparative analysis of the genomes of the bacteria *Mycoplasma pneumoniae* and *Mycoplasma genitalium*. *Nucleic Acids Res.* **25,** 701–712 (1997).
26. Hutchison, C. A. *et al.* Global transposon mutagenesis and a minimal mycoplasma genome. *Science* **288,** 2165–2169 (1999).

27. Williams, M. V. & Pollack, J. D. in *DNA Replication and Mutagenesis* (eds Moses, R. C. & Summers, W. C.) 440–444 (American Society for Microbiology, Washington DC, 1988).
28. Williams, M. V. & Pollack, J. D. A mollicute (mycoplasma) DNA repair enzyme: purification and characterization of uracil-DNA glycosylase. *J. Bacteriol.* **172,** 2979–2985 (1990).
29. Chen, E. Y., Schlessinger, D. & Kere, J. Ordered shotgun sequencing, a strategy for integrated mapping and sequencing of YAC clones. *Genomics* **17,** 651–656 (1993).
30. Heiner, C. R., Hunkapiller, K. L., Chen, S. M., Glass, J. I. & Chen, E. Y. Sequencing multimegabase-template DNA with BigDye terminator chemistry. *Genome Res.* **8,** 557–561 (1998).

Supplementary information is available on *Nature*'s World-Wide Web site (http://www.nature.com) or as paper copy from the London editorial office of *Nature*.

Acknowledgements

The authors thank D. Schlessinger for establishing the collaboration that led to this project, Y. Hale for growing the *U. urealyticum*; R. Belo, P. Babayan, K. Hunkapiller and T. Nguyen for assistance with DNA sequencing; and C.-N. Chen, S. Peterson, K. Ketchum and S. Payne for helpful disscussions. This work was supported by PE Biosystems, the National Institute Allergy and Infectious Diseases of the National Institutes of Health, Eli Lilly and Company, and the University of Alabama at Birmingham, Department of Microbiology.

Correspondence and requests for materials should be addressed to J.I.G. (e-mail: Glass_John_I@Lilly.com). The annotated genome is available on ⟨http://genome.microbio.uab.edu/uu/uugen.htm⟩ and ⟨http://www.stdgen.lanl.gov/⟩. The sequence has been deposited with GenBank with accession number AF222894.

Quorum-sensing signals indicate that cystic fibrosis lungs are infected with bacterial biofilms

Pradeep K. Singh*†, Amy L. Schaefer†‡, Matthew R. Parsek†§, Thomas O. Moninger‖, Michael J. Welsh* & E. P. Greenberg‡

** Howard Hughes Medical Institute & Department of Internal Medicine, University of Iowa College of Medicine, Iowa City, Iowa 52242, USA*
‡ Department of Microbiology, University of Iowa College of Medicine, Iowa City, Iowa 52242, USA
§ Department of Civil Engineering, Northwestern University, Evanston, Illinois 60208, USA
‖ Central Microscopy Research Facility, University of Iowa College of Medicine, Iowa City, Iowa 52242, USA
† These authors contributed equally to this work

The bacterium *Pseudomonas aeruginosa* permanently colonizes cystic fibrosis lungs despite aggressive antibiotic treatment[1–3]. This suggests that *P. aeruginosa* might exist as biofilms—structured communities of bacteria encased in a self-produced polymeric matrix—in the cystic fibrosis lung[1,4]. Consistent with this hypothesis, microscopy of cystic fibrosis sputum shows that *P. aeruginosa* are in biofilm-like structures. *P. aeruginosa* uses extracellular quorum-sensing signals (extracellular chemical signals that cue cell-density-dependent gene expression) to coordinate biofilm formation[5]. Here we found that cystic fibrosis sputum produces the two principal *P. aeruginosa* quorum-sensing signals; however, the relative abundance of these signals was opposite to that of the standard *P. aeruginosa* strain PAO1 in laboratory broth culture. When *P. aeruginosa* sputum isolates were grown in broth, some showed quorum-sensing signal ratios like those of the laboratory strain. When we grew these isolates and PAO1 in a laboratory biofilm model, the signal ratios were like those in cystic fibrosis sputum. Our data support the hypothesis that *P. aeruginosa* are in a biofilm in cystic fibrosis sputum. Moreover, quorum-sensing signal profiling of specific *P. aeruginosa* strains may serve as a biomarker in screens to identify agents that interfere with biofilm development.

Pseudomonas aeruginosa is the principal pathogen in the lungs of patients with cystic fibrosis (CF). Chronic colonization by this bacterium leads to progressive lung damage, and eventually respiratory failure and death in most CF patients[2,3,6]. Once the CF lung has been colonized, even long-term antibiotic therapy does not eradicate *P. aeruginosa*[2,3,7]. One hypothesis for the antibiotic resistance of *P. aeruginosa* in the CF lung is that this bacterium exists there as a biofilm, as bacterial biofilms are innately resistant to antimicrobial treatment[1].

Because little is known about the physiology of *P. aeruginosa* cells in biofilms, the hypothesis that *P. aeruginosa* in the CF lung is a biofilm has not been rigorously tested[1], and a marker for the biofilm-state has not been identified. Here, we first used transmission electron microscopy to examine CF sputum from patients infected with only *P. aeruginosa*. We prepared samples using a perflurocarbon-based protocol to avoid the extraction of water-soluble material inherent in aqueous-based methods. We observed clusters of *P. aeruginosa* encased in a densely stained matrix (Fig. 1). Bacterial clusters were easily distinguished from the surrounding heterogeneous material containing host cells. We examined sputum samples from three different patients, and we did not observe bacteria outside these structures. This appearance is consistent with previous microscopic observations and consistent with the hypothesis that in CF sputum, *P. aeruginosa* exists in biofilms. As microscopic observations are open to interpretation, however, we sought physiological evidence of a biofilm lifestyle.

Biofilm development involves several steps. After attachment and multiplication on a surface, bacterial microcolonies form and then differentiate into characteristic antibiotic-resistant structures with cells encased in an extracellular polysaccharide matrix. In *P. aeruginosa*, differentiation is cued by one of two quorum-sensing signals that control the expression of dozens of *P. aeruginosa* genes in a cell-density-dependent fashion[8,9]. The signals are *N*-(3-oxododecanoyl)-L-homoserine lactone (3OC12-HSL) and *N*-butyryl-L-HSL (C4-HSL). The signal required for microcolony differentiation in biofilms is 3OC12-HSL[5]). Although studies with mice indicate that quorum sensing is important for virulence of *P. aeruginosa*[10,11],

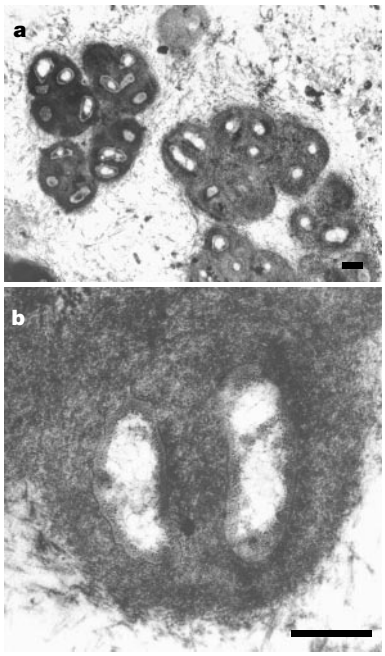

Figure 1 Transmission electron microscope images of *P. aeruginosa* in sputum. **a**, Low magnification; **b**, high magnification. *P. aeruginosa* are organized in clusters encased in a densely staining matrix. Scale bars, 1 μm.