

Chimeric 16S rDNA sequences of diverse origin are accumulating in the public databases

Philip Hugenholtz† and Thomas Huber

Correspondence
Philip Hugenholtz
philiph@nature.berkeley.edu

ComBinE group, Advanced Computational Modelling Centre, The University of Queensland, Brisbane 4072, Australia

A significant number of chimeric 16S rDNA sequences of diverse origin were identified in the public databases by partial treeing analysis. This suggests that chimeric sequences, representing phylogenetically novel non-existent organisms, are routinely being overlooked in molecular phylogenetic surveys despite a general awareness of PCR-generated artefacts amongst researchers.

INTRODUCTION

Culture-independent studies based on obtaining 16S rRNA genes directly from the environment by broad-specificity primer PCR and cloning have greatly improved our understanding of microbial diversity (Hugenholtz *et al.*, 1998; Pace, 1997). However, such PCR-based surveys have a number of recognized limitations (Hugenholtz & Goebel, 2001; von Wintzingerode *et al.*, 1997), perhaps the most insidious of which is the formation of recombinant or chimeric sequences during PCR amplification. Chimera formation is thought to occur when a prematurely terminated amplicon reanneals to a foreign DNA strand and is copied to completion in the following PCR cycles. This results in a sequence composed of two or more phylogenetically distinct parent sequences and, when comparatively analysed with other 16S rDNA sequences, suggests the presence of a non-existent organism. This problem was recognized early on in the application of PCR-clone library studies (Kopczynski *et al.*, 1994; Liesack *et al.*, 1991) and significant efforts have been made both to quantify (and hopefully reduce) chimera formation (Qiu *et al.*, 2001; Speksnijder *et al.*, 2001; Wang & Wang, 1996, 1997) and to improve their detection (Komatsoulis & Waterman, 1997; Liesack *et al.*, 1991; Maidak *et al.*, 2001; Robinson-Cox *et al.*, 1995). Despite these precautions, a surprising number of chimeric 16S rDNA sequences from molecular phylogenetic surveys were detected in the public databases during a recent collation (Hugenholtz, 2002).

METHODS

Phylogenetic analysis. 16S rDNA sequences from several PCR-clone library studies were obtained from the public databases and

imported into an ARB (<http://www.arb-home.de/>) database, where they were automatically aligned against existing sequences using FAST ALIGNER (version 1.03) followed by a manual refinement of the alignment. Studies were selected on the basis that at least one sequence in the study had been putatively identified as chimeric during routine database updating. Only almost-complete 16S rDNA sequences (>1300 nt) were included in the analysis, because the phylogenetic placement of shorter sequences can be unreliable, particularly if they lack close relatives in the database (Hugenholtz *et al.*, 1998). Datasets comprising all sequences (>1300 nt) from a single study and 341 or 200 reference sequences representing the bacterial or archaeal domains, respectively, were selected for phylogenetic inference (Hugenholtz, 2002). These datasets are available through the Ribosomal Database Project (RDP; Maidak *et al.*, 2001; <http://rdp.cme.msu.edu/html/alignments.html>).

Evolutionary distance trees were inferred independently from 5' and 3' halves of each dataset (partial treeing) applying the Lane mask (Lane, 1991) from absolute positions 0 to 4000 (635 nt for comparative analysis equivalent to *Escherichia coli* positions 28–762; 5' half) and 4000 to 0 (653 nt equivalent to *E. coli* positions 762–1512; 3' half) using the 'column selection' option in the filter selection menu. The environmental clone sequences in the dataset were then marked and tree topologies were compared for branching incongruencies indicative of chimeric sequences (Wang & Wang, 1997). The alignments of putatively identified chimeras were examined against their closest 5' and 3' matches (at least two of each) and inspected for nucleotide signature shifts characteristic of chimeric sequences (Wang & Wang, 1997). Breakpoints (also known as chimeric junctions or recombination sites) were estimated as being halfway between the change of nucleotide signatures characteristic of each parent group. Exact breakpoints are difficult to determine because the parent sequences are usually identical around the recombination site (Hugenholtz & Goebel, 2001). Positively identified chimeras were flagged in the database by appending '#' to the clone name and annotating the 'warning' field with the affiliations of the parent sequences and approximate breakpoint. This information is summarized in Table 1.

RESULTS AND DISCUSSION

A sampling of nine published and three unpublished studies with publicly available sequence data revealed 21 inter-phylum and 18 intra-phylum chimeras by partial treeing (Table 1). Numerous smaller local topological rearrangements of sequences in the partial trees were observed,

Published online ahead of print on 26 July 2002 as DOI 10.1099/ijs.0.02441-0.

†Present address: Environmental Science, Policy and Management, Division of Ecosystem Sciences, 151 Hilgard Hall, University of California Berkeley, Berkeley, CA 94720-3110, USA.

Table 1. Chimeric 16S rDNA sequences detected in the public databases

Putative chimeric sequence (accession no.)	5' Parent sequence		3' Parent sequence		Approx. breakpoint (<i>E. coli</i> numbering)	Reference study	Chimera detection‡
	Phylum (Class)*	Closest BLAST match†	Phylum (Class)*	Closest BLAST match†			
Inter-phylum							
Arctic96AD-3 (AF354607)	<i>Proteobacteria</i> (Gamma)	Arctic96AD-9 (AF354608) (98 %)	<i>Gemmatimonas</i>	BD7-2 (AB015578) (92 %)	935	Bano & Hollibaugh (2002)	C, P
Arctic95A-3 (AF355043)	<i>Proteobacteria</i> (Delta)	Arctic97A-12 (AF355044) (97 %)	Marine group A	Arctic96B-7 (AF355047) (91 %)	675	Bano & Hollibaugh (2002)	C, P
AT425_EubG1 (AY053495)	OP1	KTK 32 (AJ133617) (87 %)	<i>Proteobacteria</i> (Gamma)	<i>Pseudomonas</i> sp. G2 (AF326356) (98 %)	1215	Lanoil <i>et al.</i> (2001)	N
BD7-9 (AB015584)	<i>Proteobacteria</i> (Epsilon)	BD7-6 (AB015582) (98 %)	<i>Actinobacteria</i>	BD2-10 (AB015539) (94 %)	1085	Li <i>et al.</i> (1999)	S
BPC043 (AF154080)	<i>Firmicutes</i>	BPC060 (AF154081) (98 %)	OP9	GCA018 (AF154105) (98 %)	1070	Unpublished	U
BPC061 (AF154092)	<i>Planctomycetes</i>	<i>Pirellula</i> sp. Schlesner 516 (X81940) (90 %)	<i>Acidobacteria</i>	Sva0725 (AJ241003) (93 %)	990	Unpublished	U
BPC005 (AF154090)	<i>Proteobacteria</i> (Delta)	<i>Desulfocapsa sulfexigens</i> (Y13672) (94 %)	<i>Acidobacteria</i>	BPC015 (AF154085) (96 %)	930	Unpublished	U
BPC066 (AF154095)	<i>Acidobacteria</i>	wb1_A08 (AF317741) (93 %)	<i>Chloroflexi</i>	BPC110 (AF154084) (91 %)	875	Unpublished	U
d011 (AF422631)	TM6	Ebpr8 (AF255643) (92 %)	<i>Firmicutes</i>	d154 (AF422677) (99 %)	380	Unpublished	U
MNB2 (AF293011)	<i>Proteobacteria</i> (Alpha)	<i>Methylosinus</i> sp. LW4 (AY007293) (90 %)	<i>Nitrospira</i>	MNC2 (AF293010) (98 %)	350	Stein <i>et al.</i> (2001)	C
MNF8 (AF293012)	<i>Nitrospira</i>	MNC2 (AF293010) (97 %)	<i>Proteobacteria</i> (Gamma)	SM2E06 (AF445725) (89 %)	1005	Stein <i>et al.</i> (2001)	C
MNA3 (AF293013)	<i>Nitrospira</i>	MNC2 (AF293010) (98 %)	<i>Proteobacteria</i> (Gamma)	Iron-oxidizer m-1 (AF387301) (93 %)	930	Stein <i>et al.</i> (2001)	C
UASB_TL84 (AF254390)	<i>Chloroflexi</i>	BPC110 (AF154084) (90 %)	<i>Proteobacteria</i> (Delta)	UASB_TL15 (AF254406) (99 %)	900	Wu <i>et al.</i> (2001a)	C, S, P
UASB_TL94 (AF254401)	OP9	BA021 (AF323760) (94 %)	<i>Spirochaetes</i>	' <i>Leptospira illini</i> ' (M88719) (99 %)	1080	Wu <i>et al.</i> (2001a)	C, S, P
UASB_TL56 (AF254405)	<i>Proteobacteria</i> (Delta)	UASB_TL44 (AF254395) (99 %)	OP8	OPB95 (AF027060) (98 %)	620	Wu <i>et al.</i> (2001a)	C, S, P
UASB_TL15 (AF254406)	OP11	TA18 (AF229791) (99 %)	<i>Proteobacteria</i> (Delta)	UASB_TL11 (AF254397) (99 %)	930	Wu <i>et al.</i> (2001a)	C, S, P
VC2.1 Bac7 (AF068788)	<i>Aquificae</i>	VC2.1 Bac13 (AF068793) (99 %)	<i>Proteobacteria</i> (Epsilon)	VC2.1 Bac8 (AF068789) (99 %)	420	Reysenbach <i>et al.</i> (2000)	C, H
VC2.1 Bac9 (AF068790)	<i>Aquificae</i>	VC2.1 Bac13 (AF068793) (100 %)	<i>Proteobacteria</i> (Epsilon)	VC2.1 Bac8 (AF068789) (99 %)	420	Reysenbach <i>et al.</i> (2000)	C, H
VC2.1 Bac32 (AF068806)	<i>Proteobacteria</i> (Epsilon)	VC1.2-cl26 (AF367490) (98 %)	<i>Aquificae</i>	VC2.1 Bac10 (AF068791) (99 %)	1090	Reysenbach <i>et al.</i> (2000)	C, H

Table 1. cont.

Putative chimeric sequence (accession no.)	5' Parent sequence		3' Parent sequence		Approx. breakpoint (<i>E. coli</i> numbering)	Reference study	Chimera detection‡
	Phylum (Class)*	Closest BLAST match†	Phylum (Class)*	Closest BLAST match†			
VC2.1 Bac43 (AF068810)	<i>Ferribacter</i>	DO008 (AF385508) (< 94 %)	<i>Proteobacteria (Epsilon)</i>	VC2.1 Bac19 (AF068796) (99 %)	675	Reysenbach <i>et al.</i> (2000)	C, H
YNPFFP89 (AF391983)	<i>Acidobacteria</i>	SHA-18 (AJ249099) (93 %)	<i>Actinobacteria</i>	YNPFFP1 (AF391984) (96 %)	550	Unpublished	U
Intra-phylum							
Arctic96B-6 (AF353224)	<i>Proteobacteria (Alpha)</i>	Arctic97A-1 (AF353228) (98 %)	<i>Proteobacteria (Alpha)</i>	Arctic96B-10 (AF353211) (99 %)	530	Bano & Hollibaugh (2002)	C, P
Arctic96BD-1 (AF354605)	<i>Proteobacteria (Gamma)</i>	<i>Alcanivorax borkumensis</i> (AF062642) (99 %)	<i>Proteobacteria (Gamma)</i>	Arctic96B-16 (AF354595) (99 %)	615	Bano & Hollibaugh (2002)	C, P
AT425_EubD5 (AY053489)	<i>Proteobacteria (Beta)</i>	AP009 (AY005030) (99 %)	<i>Proteobacteria (Gamma)</i>	BD5-14 (AB015570) (98 %)	555	Lanoil <i>et al.</i> (2001)	N
BD3-1 (AB015547)	<i>Proteobacteria (Alpha)</i>	BD3-6 (AB015548) (99 %)	<i>Proteobacteria (Gamma)</i>	BD5-14 (AB015570) (100 %)	790	Li <i>et al.</i> (1999)	S
BD6-4 (AB015574)	<i>Proteobacteria (Gamma)</i>	str. 61716 (AF227866) (99 %)	<i>Proteobacteria (Gamma)</i>	<i>Shewanella violacea</i> (D21225) (100 %)	520	Li <i>et al.</i> (1999)	S
BPC023 (AF154087)	<i>Proteobacteria (Gamma)</i>	Ebpr13 (AF255638) (94 %)	<i>Proteobacteria (Gamma)</i>	BPC028 (AF154088) (100 %)	1070	Unpublished	U
d035 (AF422644)	<i>Proteobacteria (Alpha)</i>	d163 (AF422687) (99 %)	<i>Proteobacteria (Alpha)</i>	d041 (AF422650) (99 %)	283	Unpublished	U
MNG7 (AF292997)	<i>Proteobacteria (Alpha)</i>	H34 (AF234750) (93 %)	<i>Proteobacteria (Alpha)</i>	MND8 (AF292999) (99 %)	1030	Stein <i>et al.</i> (2001)	C
MNH4 (AF293002)	<i>Proteobacteria (Beta)</i>	Tui3-12 (AF353297) (90 %)	<i>Proteobacteria (Alpha)</i>	<i>Azospirillum</i> sp. B510 (AB049111) (90 %)	550	Stein <i>et al.</i> (2001)	C
MNA5 (AF293005)	<i>Proteobacteria (Beta)</i>	A15 (AF234683) (95 %)	<i>Proteobacteria (Beta)</i>	MNC9 (AF293007) (98 %)	1165	Stein <i>et al.</i> (2001)	C
pIVWA101 (AB019722)	<i>Crenarchaeota (C1)</i>	19H08 (AF393305) (99 %)	<i>Crenarchaeota (C1)</i>	pIVWA2 (AB019730) (99 %)	975	Takai & Horikoshi (1999)	C
SAGMA-C (AB050207)	<i>Crenarchaeota (C1)</i>	SAGMA-A (AB050205) (99 %)§	<i>Crenarchaeota (C1)</i>	SAGMA-2 (AB050233) (98 %)	520, 790	Takai <i>et al.</i> (2001)	C
SAGMA-Z (AB050231)	<i>Crenarchaeota (C1)</i>	SAGMA-8 (AB050238) (96 %)	<i>Crenarchaeota (C1)</i>	SAGMA-2 (AB050233) (97 %)	720	Takai <i>et al.</i> (2001)	C
SAGMA-3 (AB050234)	<i>Crenarchaeota (C1)</i>	SAGMA-D (AB050208) (99 %)	<i>Crenarchaeota (C1)</i>	SAGMA-8 (AB050238) (96 %)	520	Takai <i>et al.</i> (2001)	C
TA07 (AF229780)	<i>Proteobacteria (Delta)</i>	BA053 (AF323776) (99 %)	<i>Proteobacteria (Delta)</i>	TA14 (AF229787) (99 %)	790	Wu <i>et al.</i> (2001b)	N
TA09 (AF229782)	<i>Proteobacteria (Delta)</i>	TA11 (AF229784) (99 %)	<i>Proteobacteria (Delta)</i>	<i>Syntrophus gentianae</i> (X85132) (99 %)	970	Wu <i>et al.</i> (2001b)	N

Table 1. cont.

Putative chimeric sequence (accession no.)	5' Parent sequence		3' Parent sequence		Approx. breakpoint (<i>E. coli</i> numbering)	Reference study	Chimera detection†
	Phylum (Class)*	Closest BLAST match†	Phylum (Class)*	Closest BLAST match†			
TA15 (AF229788)	Proteobacteria (Delta)	<i>Syntrophus genitiana</i> (X85132) (96%)	Proteobacteria (Delta)	TA16 (AF229789) (96%)	350	Wu <i>et al.</i> (2001b)	N
VC2.1 Arc7 (AF068818)	Euryarchaeota (Thermococci)	<i>Thermococcus siculi</i> (AJ291808) (99%)	Euryarchaeota (Thermoplasmata)	vadinCA11 (U81778) (86%)	935	Reysenbach <i>et al.</i> (2000)	C, H

*Where possible, the nomenclature of the taxonomic outline for *Bergey's Manual of Systematic Bacteriology* (Garrity *et al.*, 2001) has been used. Candidate phyla are named as described previously (Hugenholtz *et al.*, 1998; Hugenholtz, 2002).

†Closest match to a sequence > 1300 nt long.

‡Method(s) used by researchers who generated the sequence: C, CHIMERA_CHECK; H, long-range helical base pairing; S, signature nucleotide shift; P, partial treeing; N, none stated; U, unknown.

§A second breakpoint reintroduces this parent sequence at the 3' end.

suggesting the presence of additional chimeras created from closely related parent sequences (data not shown). Inter-phylum chimeras are particularly problematic in phylogenetic inference, as they can result in novel lines of descent that may be misinterpreted as representing novel species, genera or even families of prokaryotes, although they should also be the easiest to detect (von Wintzingerode *et al.*, 1997). The most popular method for detecting chimeric 16S rDNA sequences is the CHIMERA_CHECK program available through the RDP (Maidak *et al.*, 2001). This is one of a number of nearest-neighbour methods that detect chimeric sequences by determining whether fragments of two independent database entries have a higher overall similarity to the query sequence than a single, full-length database entry (Komatsoulis & Waterman, 1997; Robinson-Cox *et al.*, 1995). Unfortunately, once a chimeric sequence is added to the RDP database it becomes invisible to CHIMERA_CHECK because it is simply compared against itself in the analysis. Currently, there is no way to use CHIMERA_CHECK against subsets of the database, such as 16S rDNA sequences from cultivated organisms, to bypass this problem.

One instance of a chimeric sequence with two breakpoints was detected, SAGMA-C (Table 1). PCR-mediated recombinant sequences with multiple recombination sites have been documented previously (Bradley & Hillis, 1997) and are more likely to occur between closely related sequences, as seen in this instance. The sequences presented in Table 1 reduce the quality of the public databases and should be removed, or divided at the breakpoint and resubmitted as separate entries designated A and B to distinguish the chimeric fragments.

This study is by no means an exhaustive search of the public databases and simply serves to illustrate that chimeric 16S rDNA sequences are being overlooked in molecular phylogenetic surveys, despite a general appreciation of PCR-generated chimeras amongst researchers. Nearest-neighbour chimera detection methods should be routinely supplemented with partial treeing analysis, as this method is less sensitive to the absence of closely related parent sequences in the databases and is relatively simple to implement using ARB. In addition, we have recently written a program, called BELLEROPHON, that detects chimeric sequences in aligned datasets based on partial treeing analysis; this program is available online (<http://cassandra.visac.uq.edu.au/perl/bellerophon.pl>).

NOTE ADDED IN PROOF

One of the three unpublished studies addressed in this paper has now been published (Lowe *et al.*, 2002; sequences d011 and d035 in Table 1).

REFERENCES

Bano, N. & Hollibaugh, J. T. (2002). Phylogenetic composition of bacterioplankton assemblages from the Arctic Ocean. *Appl Environ Microbiol* **68**, 505–518.

- Bradley, R. D. & Hillis, D. M. (1997).** Recombinant DNA sequences generated by PCR amplification. *Mol Biol Evol* **14**, 592–593.
- Garrity, G. M., Winters, M. & Searles, D. B. (2001).** *Taxonomic Outline of the Prokaryotes. Bergey's Manual of Systematic Bacteriology*, release 1.0, April 2001. New York: Springer-Verlag (<http://www.cme.msu.edu/bergeys/>).
- Hugenholtz, P. (2002).** Exploring prokaryotic diversity in the genomic era. *Genome Biol* **3**, REVIEWS0003.
- Hugenholtz, P. & Goebel, B. M. (2001).** The polymerase chain reaction as a tool to investigate microbial diversity in environmental samples. In *Environmental Molecular Microbiology: Protocols and Applications*, pp. 31–42. Edited by P. A. Rochelle. Wymondham, UK: Horizon Scientific Press.
- Hugenholtz, P., Goebel, B. M. & Pace, N. R. (1998).** Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J Bacteriol* **180**, 4765–4774; erratum, 6793.
- Komatsoulis, G. & Waterman, M. (1997).** A new computational method for detection of chimeric 16S rRNA artifacts generated by PCR amplification from mixed bacterial populations. *Appl Environ Microbiol* **63**, 2338–2346.
- Kopczynski, E. D., Bateson, M. M. & Ward, D. M. (1994).** Recognition of chimeric small-subunit ribosomal DNAs composed of genes from uncultivated microorganisms. *Appl Environ Microbiol* **60**, 746–748.
- Lane, D. J. (1991).** 16S/23S rRNA sequencing. In *Nucleic Acid Techniques in Bacterial Systematics*, pp. 115–175. Edited by E. Stackebrandt & M. Goodfellow. New York: Wiley.
- Lanoil, B. D., Sassen, R., La Duc, M. T., Sweet, S. T. & Neilson, K. H. (2001).** Bacteria and Archaea physically associated with Gulf of Mexico gas hydrates. *Appl Environ Microbiol* **67**, 5143–5153.
- Li, L., Kato, C. & Horikoshi, K. (1999).** Bacterial diversity in deep-sea sediments from different depths. *Biodivers Conserv* **8**, 659–677.
- Liesack, W., Weyland, H. & Stackebrandt, E. (1991).** Potential risks of gene amplification by PCR as determined by 16S rDNA analysis of a mixed-culture of strict barophilic bacteria. *Microb Ecol* **21**, 191–198.
- Lowe, M., Madsen, E. L., Schindler, K., Smith, C., Emrich, S., Robb, F. & Halden, R. U. (2002).** Geochemistry and microbial diversity of a trichloroethene-contaminated Superfund site undergoing intrinsic *in situ* reductive dechlorination. *FEMS Microbiol Ecol* **40**, 123–134.
- Maidak, B. L., Cole, J. R., Lilburn, T. G. & 7 other authors (2001).** The RDP-II (Ribosomal Database Project). *Nucleic Acids Res* **29**, 173–174.
- Pace, N. R. (1997).** A molecular view of microbial diversity and the biosphere. *Science* **276**, 734–740.
- Qiu, X. Y., Wu, L. Y., Huang, H. S., McDonel, P. E., Palumbo, A. V., Tiedje, J. M. & Zhou, J. Z. (2001).** Evaluation of PCR-generated chimeras: mutations, and heteroduplexes with 16S rRNA gene-based cloning. *Appl Environ Microbiol* **67**, 880–887.
- Reysenbach, A. L., Longnecker, K. & Kirshtein, J. (2000).** Novel bacterial and archaeal lineages from an *in situ* growth chamber deployed at a Mid-Atlantic Ridge hydrothermal vent. *Appl Environ Microbiol* **66**, 3798–3806.
- Robinson-Cox, J. F., Bateson, M. M. & Ward, D. M. (1995).** Evaluation of nearest-neighbor methods for detection of chimeric small-subunit rRNA sequences. *Appl Environ Microbiol* **61**, 1240–1245.
- Speksnijder, A., Kowalchuk, G. A., De Jong, S., Kline, E., Stephen, J. R. & Laanbroek, H. J. (2001).** Microvariation artifacts introduced by PCR and cloning of closely related 16S rRNA gene sequences. *Appl Environ Microbiol* **67**, 469–472.
- Stein, L. Y., La Duc, M. T., Grundl, T. J. & Neilson, K. H. (2001).** Bacterial and archaeal populations associated with freshwater ferromanganese micronodules and sediments. *Environ Microbiol* **3**, 10–18.
- Takai, K. & Horikoshi, K. (1999).** Genetic diversity of archaea in deep-sea hydrothermal vent environments. *Genetics* **152**, 1285–1297.
- Takai, K., Moser, D. P., DeFlaun, M., Onstott, T. C. & Fredrickson, J. K. (2001).** Archaeal diversity in waters from deep South African gold mines. *Appl Environ Microbiol* **67**, 5750–5760.
- von Wintzingerode, F., Gobel, U. B. & Stackebrandt, E. (1997).** Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol Rev* **21**, 213–229.
- Wang, G. C.-Y. & Wang, Y. (1996).** The frequency of chimeric molecules as a consequence of PCR co-amplification of 16S rRNA genes from different bacterial species. *Microbiology* **142**, 1107–1114.
- Wang, G. C.-Y. & Wang, Y. (1997).** Frequency of formation of chimeric molecules is a consequence of PCR coamplification of 16S rRNA genes from mixed bacterial genomes. *Appl Environ Microbiol* **63**, 4645–4650.
- Wu, J. H., Liu, W. T., Tseng, I. C. & Cheng, S. S. (2001a).** Characterization of a 4-methylbenzoate-degrading methanogenic consortium as determined by small-subunit rDNA sequence analysis. *J Biosci Bioeng* **91**, 449–455.
- Wu, J. H., Liu, W. T., Tseng, I. C. & Cheng, S. S. (2001b).** Characterization of microbial consortia in a terephthalate-degrading anaerobic granular sludge system. *Microbiology* **147**, 373–382.