doi:10.1016/S0022-2836(03)00894-5 J. Mol. Biol. (2003) 332, 369–383

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Complete Nucleotide Sequence of pHG1: A *Ralstonia eutropha* H16 Megaplasmid Encoding Key Enzymes of H_2-based Lithoautotrophy and Anaerobiosis

Edward Schwartz[1]*, Anke Henne[2], Rainer Cramm[1], Thomas Eitinger[1] Bärbel Friedrich[1] and Gerhard Gottschalk[2]

[1]*Institut für Biologie Mikrobiologie Humboldt-Universität zu Berlin, Chausseestr. 117 10115 Berlin, Germany*

[2]*Laboratorium für Genomanalyse, Institut für Mikrobiologie und Genetik Georg-August-Universität Grisebachstr. 8 37077 Göttingen, Germany*

**Corresponding author*

The self-transmissible megaplasmid pHG1 carries essential genetic information for the facultatively lithoautotrophic and facultatively anaerobic lifestyles of its host, the Gram-negative soil bacterium *Ralstonia eutropha* H16. We have determined the complete nucleotide sequence of pHG1. This megaplasmid is 452,156 bp in size and carries 429 potential genes. Groups of functionally related genes form loose clusters flanked by mobile elements. The largest functional group consists of lithoautotrophy-related genes. These include a set of 41 genes for the biosynthesis of the three previously identified hydrogenases and of a fourth, novel hydrogenase. Another large cluster carries the genetic information for denitrification. In addition to a dissimilatory nitrate reductase, both specific and global regulators were identified. Also located in the denitrification region is a set of genes for cytochrome *c* biosynthesis. Determinants for several enzymes involved in the mineralization of aromatic compounds were found. The genes for conjugative plasmid transfer predict that *R. eutropha* forms two types of pili. One of them is related to the type IV pili of pathogenic enterobacteria. pHG1 also carries an extensive "junkyard" region encompassing 17 remnants of mobile elements and 22 partial or intact genes for phage-type integrase. Among the mobile elements is a novel member of the IS*5* family, in which the transposase gene is interrupted by a group II intron.

Keywords: Ralstonia eutropha strain H16; H_2 oxidation; denitrification; degradation of aromatics; group II intron

Introduction

Lithoautotrophy based on H_2 as a source of energy and CO_2 as a source of carbon is an important metabolic strategy in the microbial world†.[1] In recent years the discovery of lithoautotrophs in extreme habitats has attracted considerable attention. H_2-based lithoautotrophs thrive, for instance, in hydrothermal settings, where as primary producers, they support specialized ecosystems. However, the H_2-oxidizing, CO_2-fixing lifestyle is by no means confined to such exotic niches. Facultative and mixotrophic H_2 oxidizers are ubiquitous in terrestrial, marine and lacustrine environments.[1]

One of the best studied representatives of the H_2-/CO_2-based lithoautotrophs is the Gram-negative soil bacterium *Ralstonia eutropha* H16. This organism utilizes organic substances as sources of carbon and energy. In the absence of such compounds it can grow on CO_2, which is assimilated *via* the reductive pentose phosphate cycle (Calvin–Benson–Bassham cycle, CBB). The genetics and molecular biology of this pathway have been studied extensively.[2,3] Ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO)

Abbreviations used: CBB, Calvin–Benson–Bassham cycle; DR, direct repeat; EBS, exon-binding site; IBS, intron-binding site; IEP, intron-encoded protein; IR, inverted repeat; IS, insertion sequence; ORF, open reading frame; PHA, polyhydroxyalkanoate; PHB, polyhydroxybutyrate; RM, restriction/modification; Tn, transposon.

E-mail address of the corresponding author: edward.schwartz@rz.hu-berlin.de

† http://link.springer-ny.com/service/books/10125

and the other key genes of the pathway are encoded by dual operons. During autotrophic growth, *R. eutropha* can obtain energy from the oxidation of H_2. This biological "knallgas reaction" is catalyzed by two hydrogenases.

Early genetic studies revealed that in various strains of *R. eutropha* determinants for H_2 oxidation are carried on large plasmids.[4,5] The 450 kb megaplasmid of *R. eutropha* H16, designated pHG1, was shown to be self-transmissible and mobilized the capacity for H_2 oxidation into non-oxidizing strains. Furthermore, strains cured for pHG1 were not impaired in heterotrophic growth but were unable to grow lithoautotrophically on H_2.[6] Subsequent investigations showed that H_2 oxidation and CO_2 fixation were plasmid-linked in many phylogenetically diverse bacterial strains (reviewed by Friedrich & Schwartz[7]). It is now well known that naturally occurring plasmids frequently carry genetic information for facultative metabolic capabilities. Remarkably, a survey of independently isolated H_2-oxidizing strains of *Ralstonia* obtained from different geographical locations indicated that their megaplasmids are structurally related,[8] suggesting that a conserved plasmid "backbone" is widespread in this group of H_2 oxidizers.

R. eutropha H16 is not only versatile with respect to its range of carbon sources, but is also capable of utilizing alternative electron acceptors for growth under anoxic conditions. This ability is important for a soil organism that has to cope with transient periods of anoxia. The bacterium can grow on nitrate or nitrite, respiring them to N_2.[9] Several genes involved in anaerobic growth on nitrate and nitrite are located on pHG1.

Aside from its importance as a model organism for the analysis of H_2-/CO_2-based lithoautotrophy, *R. eutropha* H16 is attracting increased attention on account of its biotechnological potential, for instance, in the industrial production of isotope-labeled compounds for use in medical diagnostics and in the production of polyhydroxyalkanoates for the manufacture of biodegradable plastics.[10,11]

The different aspects of the versatile metabolism of *R. eutropha* H16 discussed above are the subjects of ongoing studies, which seek ultimately to understand the regulatory networks that coordinate metabolic activity in response to ambient conditions such as carbon source(s), energy source(s) and O_2 availability. Pursuant to this goal, we have initiated a genome sequencing project devoted to *R. eutropha* H16. The genome of *R. eutropha* H16 consists of three replicons: chromosome 1 (4.1 Mb), chromosome 2 (2.9 Mb) and the megaplasmid pHG1.[12] Here, we report the completion of the first stage of this project, the sequencing of megaplasmid pHG1.

Results and Discussion

General features

General features of the megaplasmid pHG1 are listed in Table 1. The 452,156 bp sequence of pHG1 has a mean GC content of 62.3%, which is slightly lower than that of the entire genome (64.8%). A total of 429 protein-coding regions were annotated. This represents a coding density of 79.7%. No genes for ribosomal RNAs were detected, corroborating the finding of previous hybridization studies.[12] A computerized search for tRNA genes predicted a gene for $tRNA^{met}$ comprising nucleotides 194,008–194,095. pHG1 carries seven apparently intact insertion (IS) elements representing different families (Table 2). Besides the previously published ISAE*1*,[13] a member of the ISL*3* family present in two copies, four new insertion sequences were identified. IS*882* is a relative of IS*1071* of *Comamonas testosteroni* BR60. The latter element is an intermediate between an insertion sequence and a class II transposon: It encodes a Tn*3*-like transposase flanked by Tn*3*-like inverted repeats but lacks a resolvase gene and non-transposition-related genes. IS*883* is related to the ISRm*14* of *Sinorhizobium meliloti*.[14] IS*884* and IS*885* belong to the IS*110* and IS*630* families, respectively.[15] IS*881* is a unique representative of the IS*5* family. The transposase gene of this element is interrupted by a group II intron. pHG1 also carries a gene for a Tn*21*-type transposase (open reading frame (ORF) PHG407) but it is not clear whether this gene belongs to an intact mobile element.

Table 1. General features of the *R. eutropha* H16 megaplasmid pHG1

Length (bp)	452,156
G + C content (%)	62.3
rRNA genes	0
tRNA genes	1
Percentage coding	79.7
Protein coding regions	429
ORFs with assigned functions	269
Plasmid maintenance and transfer	34
Hydrogen oxidation	41
CO_2 fixation	12
Anaerobic metabolism	34
Degradation of aromatics	13
Recombinases	27
Regulatory proteins	15
Transporters	20
ORFs with no database relatives	67
IS elements	7

Table 2. IS elements of pHG1

Element	Family	Length	Direct repeat
ISAE*1*	ISL*3*	1313	8
IS*881*	IS*5*	3033 (1195)[a]	4
IS*882*	Tn*3*/IS*1071*	3363	None
IS*883*	IS*66*	2461	8
IS*884*	IS*110*	1424	None
IS885	IS630	1100	2

[a] The size of the element without the group II intron is given in parentheses.

When the annotated ORFs of pHG1 are classified according to function, four major groups emerge. These contain genes for (i) H_2-/CO_2-based lithoautotrophy, (ii) denitrification and anaerobiosis, (iii) plasmid maintenance and transfer and (iv) degradation of aromatics. The genes assigned to these groups are not interleaved but rather form clusters bracketed by mobile elements (Figure 1). The lithoautotrophy cluster is the largest lump of genetic information (54,300 bp). Physically, this cluster is split into halves by a large conglomerate of partial and intact integrase genes, mobile elements and a few other odd genes. Aside from the functions mentioned above, pHG1 encodes some other metabolically important properties. Two genes related to the metabolism of polyhydroxyalkanoates (PHAs) are present: *phaP2* encodes a phasin, i.e. the major protein forming the coating of cytoplasmic PHA granules.[16,17] *phaZ* encodes a polyhydroxybutyrate (PHB) depolymerase required for the mobilization of the storage polyesters. The largest gene family on pHG1 consists of six helicase genes representing superfamilies I and II.[18]

Plasmid replication and maintenance

As in other naturally occurring plasmids, genes for replication and stable inheritance form a cluster on pHG1 (Figure 1). The arrangement of maintenance genes in a series, often referred to as the plasmid "backbone", is not uncommon.[19] Comparison of deduced gene products with database entries indicates that pHG1 is equipped with at least five different maintenance systems: (i) replication initiation, (ii) plasmid partitioning, (iii) multimer resolution, (iv) post-segregational killing and (v) restriction/modification. Many natural plasmids regulate their copy number at the level of the frequency of the initiation of DNA replication. The pHG1 replication/maintenance cluster encodes a RepA isolog. The proteins of this family act either alone or in conjunction with host DnaA protein to trigger replication by unwinding a stretch of DNA at the origin.[20] The pHG1 RepA protein is similar (24% identity) to RepA28, the initiation protein of pMOL28 of *Ralstonia metallidurans* CH34, a close relative of *R. eutropha*.[21] A gene for a RepB protein is situated immediately downstream of the *repA* gene on the opposite DNA strand. In many plasmids, the origin of replication (*oriV*) is located in the vicinity of the *repA* gene and is characterized by palindromic sequences and/or a set of tandem repeat elements.[19,22] Neither iteron-like repeats nor prominent palindromic elements are present in the neighborhood of *repA*. The most conspicuous hairpin-forming sequence is located in the region between *repA* and *repB*. Nevertheless, two distinctive sequence features suggest that an *oriV* may be nearby. First, an AT-rich island (47% AT) spanning 3 kb (nucleotides 389,001–392,000) overlaps *repB*. Furthermore, the cumulative GC-skew plot for pHG1 has an inflection point at nucleotide 384,550. The maxima and minima of GC-skew plots of bacterial replicons typically coincide with *ori* and *ter* regions.[23]

Adjacent to the *rep* genes are a pair of ORFs (*parA* and *parB1*) encoding partitioning proteins of the ParA and ParB families. Partitioning systems mediate the accurate distribution of plasmid copies to the daughter cells upon cell division.[24] pHG1 ParA contains the deviant Walker-type motifs A, A′ and B that are characteristic of partitioning ATPases.[25] Consistent with its size (398 amino acid residues), ParA clusters along with isologs found on *Escherichia coli* bacteriophage N15 (24% identity) and *Yersinia pestis* plasmid pCD1 (25% identity) in group Ia as defined by Gerdes and co-workers.[24] Database searches revealed several close relatives of pHG1 ParB1, including an isolog from *Pseudomonas fluorescens* plasmid pL6.5.

pHG1 also encodes an integrase/recombinase protein similar to plasmid-encoded resolvases of the XerC/XerD family.[26] Proteins of this family contribute to plasmid stability by resolving multimers arising *via* homologous recombination. Not far from *xerD* is a set of genes encoding PemI and PemK isologs, which are components of the plasmid addiction systems found on R100 and on other plasmids.[27] PemK proteins are growth inhibitors and are subject to inactivation by the cognate PemI proteins. Since the PemI proteins are short-lived compared to the corresponding PemK proteins, loss of plasmid leads to cell death at high frequencies.

Finally, a putative type II restriction/modification (RM) system is encoded several kilobases distant from the replication/maintenance region. ORF PHG325 encodes a protein similar to the *Salmonella enterica* serovar Paratyphi methyltransferase *Spt* AIM. The ORF PHG325 product harbors a typical N6-, N4-methyltransferase domain. RM systems are advantageous to the host cell because they pose a barrier against foreign DNA. Plasmid-borne RM systems also contribute to the stable inheritance of the plasmid by a post-segregational killing mechanism.[28]

Conjugative transfer and pilus biosynthesis

A 27 kb large region (nucleotides 357,521–384,589) of the *R. eutropha* megaplasmid contains genes related to conjugative plasmid transfer and pilus biosynthesis (Figure 2). This finding is not surprising, since pHG1 has been shown to be self-transmissable.[5] The ORFs of the pHG1 transfer region are organized in two divergent sets. Several ORFs of both sets predict proteins that are clearly isologous to gene products of the mating-pair formation and DNA transfer systems of well-studied plasmids. However, neither the overall arrangement of the genes nor the similarities at the level of amino acid sequences link the pHG1 transfer system to any one previously studied plasmid. Rather, the pHG1 transfer region seems to be a

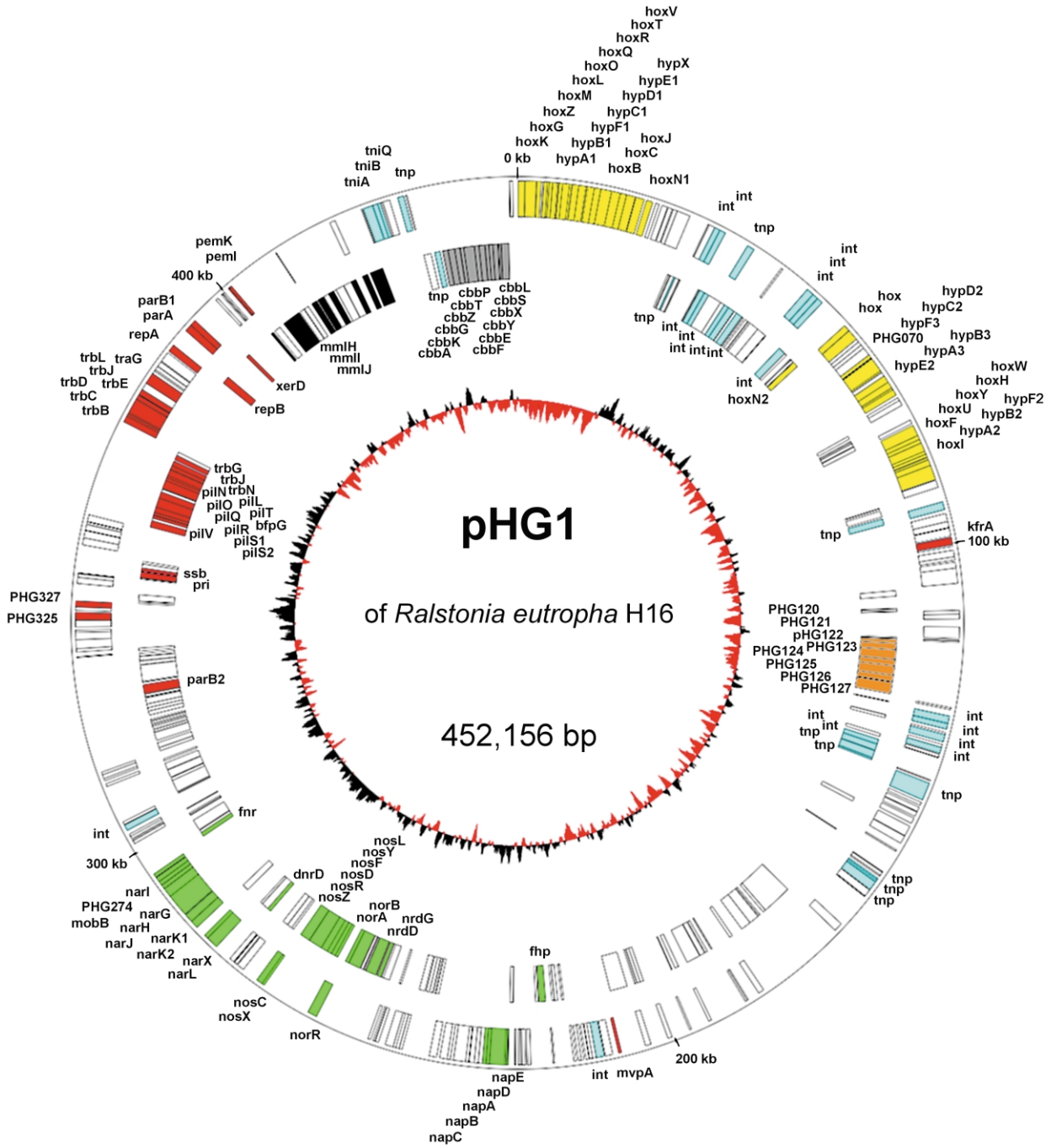

Figure 1. Genetic and physical map of megaplasmid pHG1. The outer and inner circle of boxes indicate genes transcribed in the clockwise and counterclockwise direction, respectively. ORFs are color-coded as follows: yellow, H_2 oxidation; grey, CO_2 fixation; green, denitrification; red, plasmid maintenance and transfer; black, degradation of aromatics; orange, iron uptake; blue, recombinases. ORFs with no functional assignment are left uncolored. A GC skew plot is given inside the ORF map. A G residue of a *Bgl* II site used for the physical mapping of pHG1,[12] was arbitrarily chosen as nucleotide 1 in the nucleotide sequence. The map is based on graphic output generated using the software package GenDB 2.0.[96]

pastiche of genetic information from diverse sources. Functions involved in pilus assembly and conjugative DNA processing intermingle in the divergent sets. On the basis of sequence comparisons with deduced pilin amino acid sequences, pHG1 appears to encode two different kinds of pili, one of them related to the type IV pili found in many pathogenic bacteria, the other to RP4-type sex pili. The major structural components of the former are pilins encoded by the pHG1 *pilS1* and *pilS2* genes (paralogs with 99% identity). PilS1/PilS2 and some of the other proteins encoded in the same set (oriented leftward in Figure 2) share significant similarity to gene

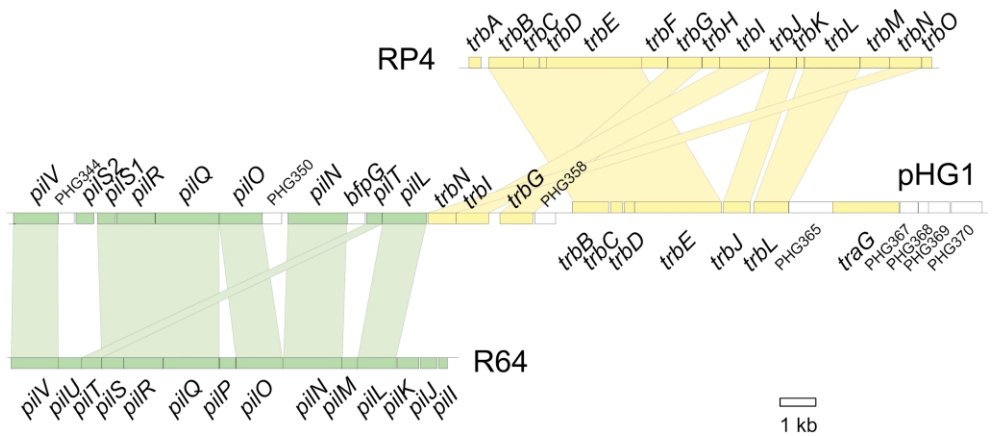

Figure 2. The pHG1 conjugative transfer region. Rectangles above and below the lines indicate genes transcribed to the right and left, respectively. For comparison, the *pil* operon of *Salmonella typhi* plasmid R64 and part of the Tra2 region of *Pseudomonas aeruginosa* plasmid RP4 are shown. Like coloring and shaded ribbons indicate sequence similarity. PHG1 ORFs with no similarity to R64 and RP4 genes are left uncolored.

products of the *pil* genes of plasmid R64. An adhesin, a minor pilus component, is encoded by *pilV*. The closest database relative of the pHG1 adhesin is PilV of R64. However, the characteristic invertable shufflon elements are absent from the pHG1 gene.

The assembly of type IV pili requires a specialized machinery constituting a type II secretion system. A key component of type II secretion systems is a hexameric ATPase called a "traffic warden protein".[29] This protein family includes the PilQ protein of plasmid R64, which is required for pilus assembly,[30] and *Agrobacterium tumefaciens* VirB11, which mediates the transfer of T-DNA to plant cells. The amino acid sequence of pHG1 PilQ contains typical signatures of the VirB11 family, including a Walker A box ($_{344}$GPT GSGKT$_{351}$), a Walker B box ($_{403}$RMDPDAV LLGEIRG$_{416}$) and a histidine box ($_{429}$TGHQ VLTTLHVTDP$_{442}$). An unusual feature of this sequence is the atypical aspartate box ($_{368}$KR LITIENPTE$_{378}$), which has asparagine in place of the aspartate residue.

Another essential part of the secretion pathway is a lipoprotein called a secretin. Secretins form annular aggregates in the bacterial outer membrane consisting of ten to 18 monomers.[31] It is thought that these secretin rings constitute pores through which type IV pili extend or through which protein factors are secreted. The product of pHG1 *pilN* is a lipoprotein with a typical leader peptide cleavage/modification site ($_{18}$LSGC$_{21}$)[32] and a characteristic secretin signature. The nearest database relative of pHG1 PilN is the isolog of *Yersinia pseudotuberculosis*. Significant similarity was found to the *Xanthomonas campestris* XpsD protein, which is involved in the hypersensitive response, and to Yss of *Yersinia enterocolitica*, which mediates export of Yops.

pHG1 PilO is related (20% identity) to the *S. typhi* PilO and to the R64 PilO.[33] The latter is an outer membrane protein of unknown function. pHG1 also encodes a PilR-like integral membrane protein. The sequences of ORFs PHG353 (*pilT*) and PHG355 (*trbN*) predict secreted or periplasmic proteins with N-terminal peptidase cleavage sites ($_{18}$AHA/DC$_{22}$ and $_{20}$AVA/SL$_{24}$, respectively). Both proteins contain motifs characteristic of lytic transglycosylases. Transglycosylases, such as the RP4 TrbN protein, are typical components of the ensemble of proteins involved in the biosynthesis of pili. Two other genes in the series (*trbI* and *trbG*) encode proteins that are apparently not related to Pil proteins but are similar to the *A. tumefaciens* Vir proteins VirB10 and VirB9 and their respective RP4 counterparts TrbI and TrbG. The pHG1 TrbI and TrbG may be part of the membrane-spanning channel required for the formation of type IV pili. TrbI seems to be an integral membrane protein and TrbG may be periplasmic.

The second set of transfer-related genes, oriented in the opposite direction (rightward in Figure 2), encodes a prepropilin of the RP4 type. While similarity to the RP4 pilin TrbC is not conspicuous at the level of amino acid sequence, pHG1 pilin carries the hallmarks of a pilin precursor:[34] The predicted protein (TrbC) is very hydrophobic and contains two putative transmembrane helices. Furthermore computer algorithms predict an N-terminal signal peptide with a leader peptidase cleavage site ($_{31}$TSA/VA$_{35}$). A C-terminal alanine motif ($_{96}$AAAA$_{99}$) may be the recognition site of a TraF-like endopeptidase. *trbB*, the first ORF of the set, encodes a hexameric ATPase. The other genes with clear database relatives seem to encode a membrane channel that may play a role in conjugative DNA transfer. TrbE is an integral membrane protein and contains a Walker A motif ($_{474}$GPIGA GKSV$_{482}$) indicative of NTP-binding proteins. These features suggest that it belongs to the TrbE/VirB4 family. Sequence comparisons support this classification. TrbJ is a secreted or periplasmic protein related to so-called "entry exclusion proteins" of other plasmid transfer systems. TrbL

is an integral membrane protein with six putative transmembrane helices. Finally, ORF PHG366 (*traG*) encodes a protein with an N-terminal signal peptide and three putative transmembrane helices. The derived amino acid sequence of TraG contains a Walker A box ($_{181}$GTTGSGKT$_{188}$) and a less conserved Walker B motif. This overall structure fits the picture of proteins of the TraG/VirD4 family.[35] Indeed, the pHG1 TraG is similar (31% identity) to a TraG isolog from *Vibrio cholerae*. According to a recent model, TraG/VirD4 proteins form an inner-membrane pore that binds the relaxosome complex, initiating conjugative DNA transfer.[35]

Two additional components of the conjugative transfer system are encoded a short distance away from the transfer region: a DNA primase and a single-stranded DNA-binding protein (SSB). Primases are multifunctional enzymes that bind DNA, unwind it and then synthesize short RNAs to prime DNA synthesis.[36] Primases encoded by plasmid transfer systems are transferred as a complex with the single-stranded DNA during conjugation.[37,38] The amino acid sequence of ORF PHG334 predicts a protein with typical primase features. The N-terminal part contains a Zn^{2+} binding motif ($_{43}$CPMCGG-14-WVCRKC$_{68}$). The corresponding domain could mediate sequence-specific interaction with the DNA. The amino acid pair glutamate–glycine embedded in a stretch of hydrophobic residues is a characteristic signature of bacterial primases.[36] The amino acid sequence of pHG1 primase contains such a motif ($_{248}$IGVAE GLESA$_{257}$). Sequence comparisons show, however, that the pHG1 primase is more closely related to a subgroup of phage-encoded primases than to the plasmid-encoded Sog/TraC proteins.

Metabolic capacity

Hydrogen metabolism

pHG1 contains structural and accessory genes for two energy-conserving [NiFe] hydrogenases, the key players in lithotrophic energy metabolism during growth on H_2: a cytoplasmic, NAD^+-reducing enzyme (SH),[39] and a membrane-bound enzyme (MBH) linked to the respiratory chain.[40] The MBH genes (*hoxKGZMLOQRTV*, nucleotides 115–8532) are part of a large operon comprising at least 18 genes. The distal gene, *hoxA*, encodes a regulator of the NtrC family that activates transcription at a remote promoter upstream of *hoxK*.[41] *hoxK*, *-G* and *-Z* encode the trimeric MBH consisting of a large subunit (HoxG) with the H_2-splitting [NiFe] center, a small subunit (HoxK) with three iron–sulfur clusters involved in electron transfer, and a membrane-integral *b*-type cytochrome (HoxZ) that anchors the complex on the outer surface of the cytoplasmic membrane.[42] The precise function of most of the conserved MBH accessory genes is not known. The so-called *hyp* genes (*hypA1*, *hypB1*, *hypF1*, *hypC1*, *hypD1*, *hypE1*, *hypX*, nucleotides 8525–15,474), located between the MBH genes and *hoxA*, are involved in metallocenter formation of both MBH and SH.[43–45] An H_2-sensing and signal-transduction system (reviewed by Lenz *et al.*[46]) that prevents transcription of hydrogenase genes in the absence of H_2, is encoded in the region (nucleotides 16,923–20,933) adjacent to *hoxA*. This system is composed of the H_2 sensor HoxBC and the histidine protein kinase HoxJ.

An intergenic region (402 bp) separates *hoxJ* from *hoxN1*. *hoxN1* (nucleotides 21,335–22,390) encodes a high-affinity nickel transporter of the NiCoT family (TC 2.A.52) that provides Ni^{2+} for incorporation into the hydrogenases.[47,48] Almost 58 kb separate *hoxN1* from the genes coding for the SH. The SH operon comprises the structural genes (*hoxFUYH*, nucleotides 79,712–84,331) of the heterotetrameric hydrogenase, two accessory genes (*hoxW*, *hoxI*, nucleotides 84,315–85,338),[39] and a partial set of *hyp* genes (*hypA2*, *hypB2*, *hypF2*, nucleotides 85,449–85,790).[49] This subset of *hyp* genes is not sufficient for synthesis of an active hydrogenase. Systematic mutagenesis experiments in *R. eutropha* H16 have revealed that *hypC1*, *hypD1*, and *hypE1* are essential for SH and MBH maturation, while only one intact copy each of *hypA*, *hypB*, and *hypF* is needed.[44,49] This result was surprising for the case of *hypF*. HypF catalyzes a key step in the biosynthesis of the two cyanide groups, located at the active site of [NiFe] hydrogenases, from carbamoyl phosphate.[50,51] HypF2, like HypF in *E. coli* and other organisms, contains three conserved signatures: an acylphosphate phosphatase domain, a zinc-finger-like motif, and an *O*-carbamoyl transferase domain. In contrast, HypF1 is a truncated version that lacks the phosphatase domain and the zinc finger. In the *E. coli* system, all three domains are essential for HypF activity[50] and thus, the finding that a $HypF1^+$ $HypF2^-$ *R. eutropha* strain produces functional hydrogenases was unexpected.

A major surprise came from analysis of the pHG1 sequence between nucleotides 58,061 and 74,032. This segment has the coding capacity for a fourth [NiFe] hydrogenase (Hyd4) in addition to the MBH, the SH, and the H_2 sensor. Since a triple mutant lacking both MBH and SH as well as the sensor is devoid of detectable hydrogenase activity,[52] the role of the Hyd4 genes remains obscure. The putative dimeric Hyd4 is a close relative of the putative [NiFe] hydrogenase of the Gram-positive *Streptomyces avermitilis*.[53] The small subunits of the two enzymes, like H_2-sensing hydrogenases, apparently lack an N-terminal leader, which suggests a cytoplasmic localization. The sequences of the large subunits contain C-terminal extensions (23 residues in Hyd4) as found in energy-conserving hydrogenases but not in H_2 sensors. These extensions are removed during metallocenter formation (reviewed by Blokesch *et al.*[54]). Certain cysteine residues involved in the ligation of [3Fe–4S] or [4Fe–4S] clusters in the small subunits of energy-conserving

and sensory [NiFe] hydrogenases are either replaced or displaced in Hyd4 and in the *S. avermitilis* sequence. The segment about 4 kb away from the Hyd4 genes encodes a putative protease (PHG070), which could be involved in C-terminal proteolysis of the Hyd4 large subunit, and a set of *hyp* genes in the order *hypF3*, *hypC2*, *hypD2*, *hypE2*, *hypA3*, *hypB3*. The physiological function of these extra *hyp* genes is unclear. A putative phosphoheptose isomerase that may convert D-sedoheptulose-7-phosphate into D-glycero-D-manno-heptose-7-phosphate is encoded by *gmhA*, localized between *hypC2* and *hypD2*. PHG063 (*hoxN2*), located upstream of the Hyd4 genes and oriented divergently, encodes a protein that is strikingly similar to the HoxN1 nickel permease. HoxN2 is 51% identical with HoxN1 on the level of the amino acid sequence and contains the conserved features of the NiCoT family.[47,48] The sequence identity between the *hoxN1* and *hoxN2* genes is only 67%, suggesting that the arrangement is not the result of a recent gene duplication. Preliminary experimental data suggest, however, that *hoxN2* does not encode a functional nickel permease (T.E. & S. Lorenz, unpublished results).

CO_2 fixation

Immediately adjacent to the MBH operon in divergent orientation is a 12 cistron operon encoding a red-type form I RubisCO and other key enzymes of the CBB cycle.[2] The first two genes of the operon (*cbbLp* and *cbbSp*) encode the large and small subunits of the RubisCO, respectively. Two downstream genes (*cbbXp* and *cbbYp*) code for products of unknown function. The other eight genes of the operon code for pentose-5-phosphate 3-epimerase (*cbbEp*), fructose-1,6-bisphosphatase/sedoheptulose-1,7-bisphosphatase (*cbbFp*), phosphoribulo-kinase (*cbbPp*), transketolase (*cbbTp*), 2-phosphoglycolate phosphatase (*cbbZp*), glyceraldehyde-3-phosphate dehydrogenase (*cbbGp*) and fructose-1,6-bisphosphate/ sedoheptulose-1,7-bisphosphate aldolase (*cbbAp*). A duplicate copy of the *cbb* operon is located on one of the chromosomes. The similarity of the paralogous genes is obvious, averaging 93% identity at the nucleotide sequence level.[2] Interestingly, both operons are under the control of a single LysR-type regulator encoded by the chromosomal gene *cbbR*.[3] A plasmid copy of this gene (*cbbR′*) is recognizable but is inactive due to deletions within the coding region.

Anaerobic metabolism

Genes for the anaerobic metabolism of *R. eutropha* have been mapped both on pHG1 and on chromosomal DNA.[12] This study revealed that more than 30 genes of pHG1 are related to denitrification or general anaerobic metabolism. These genes are clustered within a 76 kb region, bordered by the genes *fhp* (PHG200) and *fnr* (PHG276), respectively (Figure 3).

nar genes encoding a membrane-bound dissimilatory nitrate reductase occupy a 15 kb-large region of pHG1. The arrangement of the *nar* gene cluster is similar to the organization of nitrate reductase genes in other denitrifiers (reviewed by Philippot[55]). The three subunits of the nitrate reductase enzyme are encoded by *narG*, *narH* and *narI*. *narJ* may be a chaperone.[56] Two downstream genes with similarity to *nifM* and *mobB* of other denitrifiers may be involved in the formation of the molybdenum cofactor of NarGHI. *narK1* and *narK2*, encode members of the major facilitator superfamily. Since two *narK* copies are found frequently in *nar* gene regions, it has been speculated that one of them transports nitrate into the cytoplasm, whereas the other one is responsible for the extrusion of nitrite.[57] In most denitrifiers, gene regulation of the nitrate reductase depends on the NarXL two-component regulation system and an Fnr-like protein.[58] Genes encoding NarX/NarL and Fnr are located on pHG1 directly 5′ and 3′ of the *nar* genes, respectively. Furthermore, a gene encoding a Dnr/Nnr-like regulator is found 2 kb 5′ of *narX*. Dnr/Nnr proteins belong to the Fnr family of regulators but lack four conserved cysteine residues that coordinate the Fe–S center in Fnr. In contrast to the oxygen sensor Fnr, these proteins have been shown to be involved in NO-dependent activation of *nir* and *nor* gene transcription.[59]

pHG1 carries genes for a periplasmic nitrate reductase, Nap, that is supposed to facilitate the adaptation from aerobiosis to anaerobiosis. The genes *napA* and *napB* were cloned and sequenced previously.[60] The catalytic subunit NapA contains a molybdopterin guanidin (MGD) cofactor and a [4Fe–4S] cluster. NapB is a diheme cytochrome *c*. The *napC* gene product probably acts as a membrane anchor and electron donor for the NapAB dimer.[61] Two additional *nap* genes *napE* and *napD* are present 5′ of *napA* on pHG1. NapD may participate in the assembly of the MGD cofactor.[61] The function of NapE is unknown. Varying combinations of four other genes *napK*, *-F*, *-G* and *-H* are present in *nap* gene clusters of other bacteria,[55] which have no counterparts on pHG1. None of these proteins appears to be essential for nitrate reduction by the NapAB enzyme.[61]

The *napC* gene is followed by the eight genes *ccmA*, *-B*, *-C*, *-E*, *-F*, *-G*, *-L* and *-H* encoding a cytochrome *c* maturation machinery. CcmA, -B and -C are subunits of an ABC transporter that may participate in heme transport, while CcmE, -F, -G, -L, and -H may form a heme delivery and insertion pathway for apocytochrome *c*.[62] A similar arrangement of *nap* and *ccm* genes is present in the genomes of *E. coli* K12[63] and *R. metallidurans* CH34†. It is noticeable that the *nap/ccm* region of pHG1 is

† http://www.jgi.doe.gov

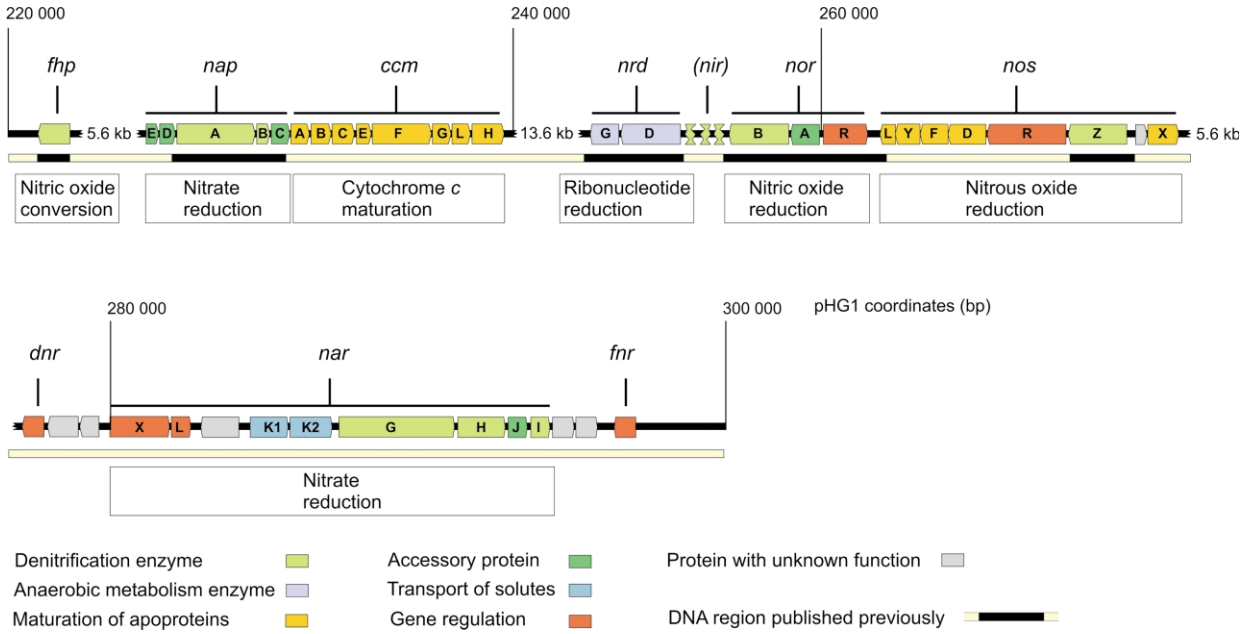

Figure 3. Organization of genes involved in anaerobic metabolism. An enlarged view of the region of pHG1 between nucleotides 220,000 and 300,000 is shown. Genes are depicted as arrows. Three fragments of a nitrite reductase gene (*nir*) are indicated. Previously sequenced regions are marked by filled black bars below the genetic map. Color codes are given at the bottom of the Figure.

flanked by truncated ORFs for transposases (PHG207, PHG208, PHG223), suggesting that this region was acquired by a transposition event.

About 6 kb 5′ of the *nap* genes we located the *fhp* gene encoding a flavohemoprotein. The *R. eutropha* flavohemoprotein was discovered more than 20 years ago,[64] and belongs to the flavohemoglobin family, which is distributed widely in bacteria and fungi. While several enzymatic activities have been attributed to flavohemoglobins, recent data indicate that most if not all flavohemoglobins are linked to stress response imposed by nitric oxide and nitrosation (reviewed by Poole & Hughes[65]). Although the precise role of Fhp in *R. eutropha* is unknown, the protein appears to be linked to nitric oxide conversion.

Nitric oxide reduction in *R. eutropha* is catalyzed by two qNor-type nitric oxide reductases.[66] In contrast to the heterodimeric cNor enzymes, which use cytochrome *c* as electron donor, qNor-type nitric oxide reductases are single-subunit quinol oxidases. The *norB* structural gene for the qNor is preceded by two genes, *norA* and *norR*. The *norR* gene encodes a sigma-54 dependent regulator that positively controls transcription of *norA* and *norB* in response to nitric oxide.[67] The physiological function of NorA is unknown. A second *norRAB* gene region (79% identity at the nucleotide level) is present on chromosome 2 of *R. eutropha* (GenBank accession AJ278372, AF002661). Either of the two gene regions is sufficient for denitrification.

Adjacent to *norB*, we identified remnants of a gene for a copper-containing dissimilatory nitrite reductase. This finding may explain previous hybridization results, which suggested the presence of a copper-type nitrite reductase in *R. eutropha*.[68] Subsequently, it was demonstrated unequivocally that the chromosomal *nirS* gene encodes the only functional nitrite reductase in *R. eutropha*.[69] Since NirS belongs to the cytochrome cd_1-type enzymes, the presence of the copper nitrite reductase gene fragments suggests that at least some of the denitrification genes of *R. eutropha* have been acquired by horizontal gene transfer. To date, no bacterial strain containing both a cd_1-type and a copper-type nitrite reductase has been identified. Furthermore, it is interesting to note that nitrite reductase is the only key enzyme of denitrification missing on pHG1. Apparently, the presence of two copies of *nar* and *nor* genes is either beneficial or at least not obstructive to *R. eutropha*, whereas two copies of nitrite reductase genes might facilitate the accumulation of toxic nitric oxide.

Two pHG1 genes (*nrdD* and *nrdG*) encoding an anaerobic class III ribonucleotide reductase (RNR) and its activase were characterized in an earlier study.[70] Class III RNRs form 2′-deoxyribonucleotides under anaerobic conditions by a radical mechanism that depends on a cognate activase.[71] The presence of these genes on a plasmid was surprising, since the chromosomally encoded denitrification enzymes are useless in the absence of an anaerobic ribonucleotide reductase. Indeed, the absence of *nrdDG* in plasmid-free mutants of *R. eutropha* accounts primarily for the inability of such strains to grow under anaerobic conditions.[70] Moreover, the non-denitrifier *Alcaligenes hydrogenophilus* can be transformed into a denitrifier simply by introduction of *nrdDG*.[70]

The structural gene for nitrous oxide reductase, *nosZ*, was identified on pHG1 several years ago.[72]

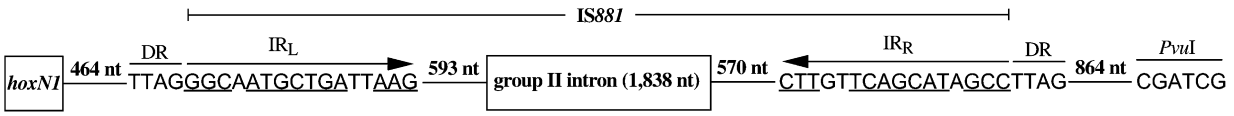

Figure 4. pHG1 region between nucleotides 21,335 and 26,767. Insertion of IS*881* has created a four nucleotide direct repeat (DR). Complementary nucleotides in the inverted repeats (IR_L, IR_R) are underlined. The transposase ORF in IS*881* is interrupted by a 1838 nucleotide large group II intron.

The present study revealed that there are at least six additional *nos* genes (*nosL*, *nosY*, *nosF*, *nosD*, *nosR* and *nosX*) on pHG1. Five of them may be involved in the maturation of the Cu metallocenters of NosZ.[73] The order of these accessory genes is conserved within the *nos* gene clusters of denitrifying bacteria.[55] The *norR* gene located 3′ of *nosZ* encodes a regulator protein that has been shown to be essential for the expression of nitrous oxide reductase.[74] In most of the *nos* sequences available so far, the order of *nosR* and *nosZ* is reversed with *norR* located 5′ of *norZ*.[55] A gene arrangement similar to *R. eutropha* is present only in the closely related species *Ralstonia solanacearum*[75] and *R. metallidurans* CH34†. *R. eutropha* cells carrying a mutation in the structural gene *nosZ* are unable to liberate dinitrogen (K. Schmelz, R.C. & B.F., unpublished results). Thus, in contrast to nitrate reduction and nitric oxide reduction, the ability to form dinitrogen is linked exclusively to pHG1.

Iron uptake

pHG1 carries a 10.5 kb large region that encodes seven putative proteins for the biosynthesis and transport of a siderophore. PHG126 contains a *tonB* box motif and shows considerable overall similarity to outer membrane ferric siderophore receptors from various organisms. PHG123 belongs to the protein family of multidrug-efflux transporters. The remaining ORFs may be involved in the formation of a hydroxymate siderophore. PHG122 and PHG124 are similar to *iucA* and *iucC*, respectively, of the aerobactin formation pathway of *Shigella flexneri*[76] and pathogenic *E. coli* strains.[77] Interestingly, the overall gene organization of PHG121 to PHG125 is highly similar to the *pvs* gene cluster of *Vibrio parahemolyticus*.[78] The *pvs* genes are involved in the formation of the hydroxamate siderophore vibrioferrin. PHG120 encodes a putative aldolase that has no counterpart within the *pvs* genes. However, a PHG120 isolog is encoded within a gene region of the *R. solanacearum* megaplasmid[75] that contains isologs of PHG120, PHG121, PHG122, and PHG124. Hence, ORFs PHG120 to PHG125 may take part in the formation of a vibrioferrin-like siderophore.

† http://www.jgi.doe.gov

Degradation of aromatic substances

Genes related to the catabolism of aromatic compounds occupy a 25 kb large segment of pHG1 between the transfer region and the *cbb* operon (Figure 1). Both the variety of the gene products predicted for the ORFs of this region and the heterogeneity of the DNA give the impression of a haphazard patchwork of genetic information derived from different organisms and specifying odd components of different catabolic pathways. Nevertheless, it is a fact that *R. eutropha* H16 can grow on benzoates as sole source of carbon and energy. It is not unlikely, therefore, that the pHG1 degradation genes complement or extend chromosomally encoded pathways for the catabolism of aromatics. In particular, ORFs PHG386, PHG385, and PHG384 encode a set of products with the potential for the utilization of methylated aromatics. The three ORFs predict proteins that are related (86%, 83% and 88% identity, respectively) to the products of the genes *mmlH*, *mmlI* and *mmlJ* found on plasmid pJP4 of *R. eutropha* JMP134. The latter genes encode a putative 4-methylmuconolactone transporter, a methylmuconolactone methyl isomerase and a methylmuconolactone isomerase, respectively.[79] 4-Methylmuconolactone is a "dead-end" product of the catabolism of methyl-substituted catechols *via ortho*-cleavage pathways. Hence, organisms equipped with the modified *ortho*-cleavage pathway specified by the *mml* genes can utilize 4-methylmuconolactone produced internally by "standard" *ortho* pathways. They can also take up and metabolize 4-methylmuconolactone secreted as a waste product by other degraders.

Mobile elements and a junkyard of transposases, integrases and a group II intron

The pHG1 region between nucleotides 22,390 and 94,364 that contains the Hyd4 and SH genes is crowded with ORFs encoding complete or partial transposases (17 ORFs) and phage-type integrases/recombinases (22 ORFs). It has long been known that this region is susceptible to spontaneous deletions. IS*881*, a novel insertion element of the IS*5* family, is located adjacent to the *hoxN1* nickel-permease gene. The segment is illustrated in Figure 4. IS*881* is flanked by direct repeats (four nucleotides) and contains imperfect inverted repeats at its ends. The presence of direct repeats is indicative of a transposition event. The ORF encoding the IS*881* transposase is interrupted by

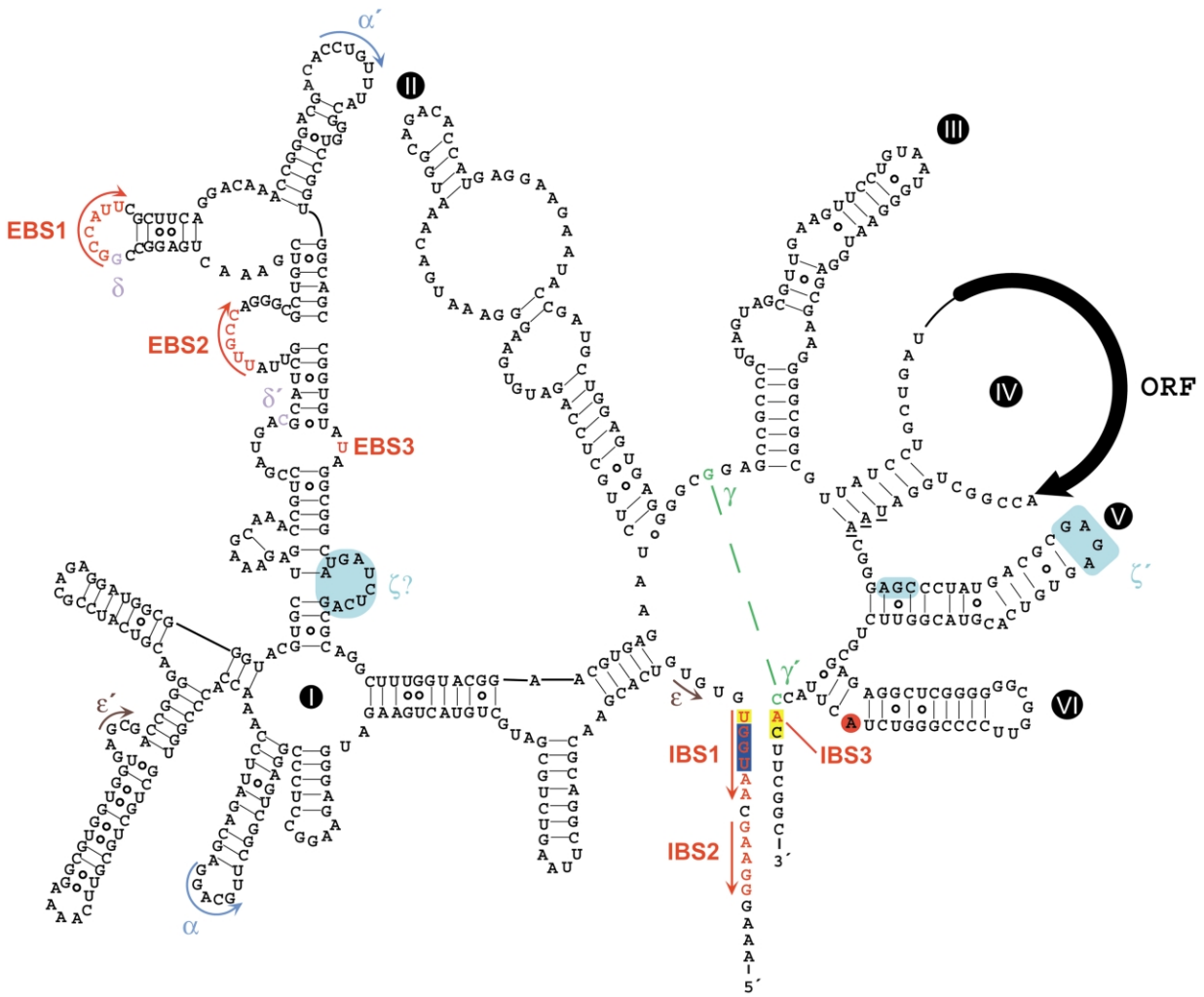

Figure 5. Secondary structure model of *R.e.*I1. Ribozyme domains dI–dVI are indicated by black circles. dI was modeled according to the consensus structure of class D bacterial group II introns (http://www.fp.ucalgary.ca/group2introns/B-likestructure.htm). Non-Watson–Crick GU pairs are shown by circles. Exon-binding sites (EBS) and intron-binding sites (IBS) are labeled in red. Additional sites of potential 3D interactions (α–α′, γ–γ′, δ–δ′, ε–ε′, ζ–ζ′) are marked. ζ and ζ′ are shown in blue. The ζ region is somewhat unusual. The sequence AGC in dV and the bulged A in dVI, essential for catalysis, are marked in blue and red, respectively. Underlined nucleotides at the 3′-end of dIV represent the stop codon downstream of the RNA maturase domain X. At the exon–intron borders, the boxed nucleotides indicate Trp (blue) and Tyr (yellow) codons of the IS*881* transposase gene.

1838 bp. Interestingly, the insertion disrupts a signature at the amino acid level (QW*YFGMK) that is conserved extensively in transposases of the IS*5* family and important for regulation of the transposition frequency. In IS*5* of *E. coli*, this motif (QWHFGMK) acts as a recognition site for proteolytic cleavage after the methionine residue resulting in a stable N-terminal product that lacks transposase activity.[80] The structure of the deduced RNA of the intervening sequence (*R.e.*I1, Figure 5) in IS*881* contains many features of a group II intron. Group II introns are found in organelles of fungi, plants and protists, in many bacteria, and in a few methanogenic archaea.[81–85] They are not related in sequence but share a six-domain architecture (ribozyme domains I–VI) and a conserved 5′ splice site (N*GUGYG).

Bacterial group II introns are located mostly in plasmids or mobile elements. Self-splicing of group II introns occurs *in vitro* under artificial conditions but splicing *in vivo* is facilitated by the host's machinery (in many organellar introns) or by an intron-encoded protein (IEP). IEPs are encoded in dIV of some eukaryotic and almost all known full-length bacterial group II introns, and behave as multi-functional enzymes. They contain reverse transcriptase, RNA maturase, and DNA endonuclease domains. The latter, however, is not present in more than half of the bacterial IEPs.

IEPs are required for group II intron mobility. These elements can move into homologous, intron-less sites ("homing") and, at a lower frequency, into ectopic sites. Since the two processes are dependent on reverse transcriptase activity, they are called retrohoming and retrotransposition, respectively. Splicing and intron mobility require

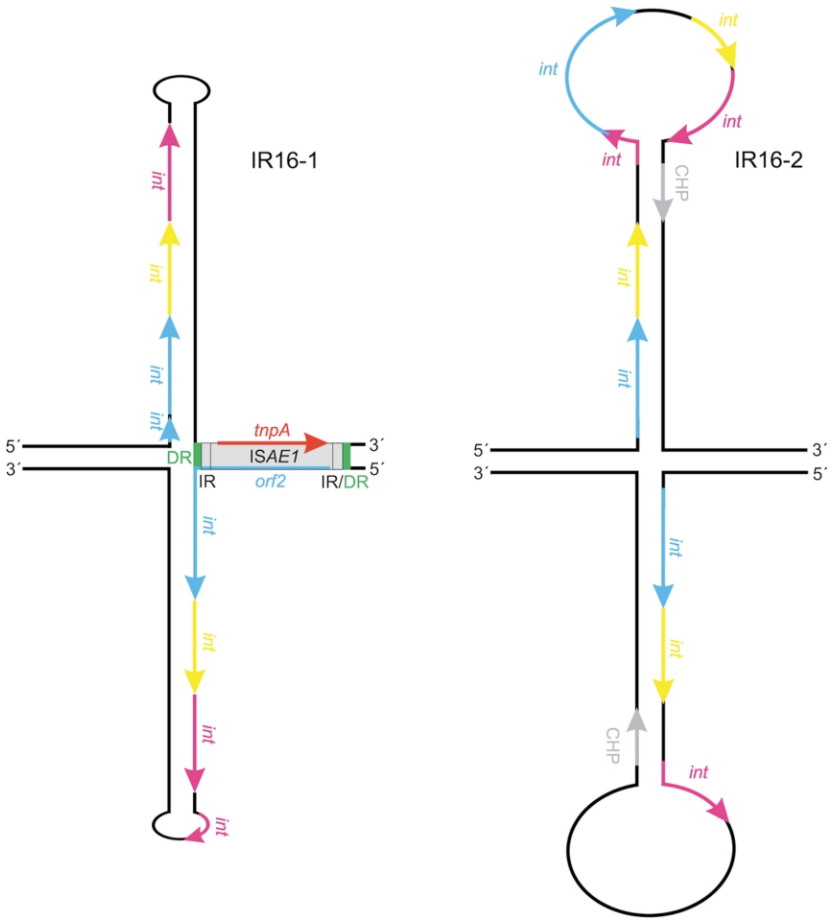

Figure 6. Models of pHG1-borne large inverted repeats. The models correlate with the snapback structures IR16-1 and IR16-2 observed by electron microscopy.[88] Stems in IR16-1 and IR16-2 are 96.3% and 100% complementary, respectively. The stem region in IR16-1 begins within the left direct repeat (DR) adjacent to ISAE*1*. IR, inverted repeats at the ends of ISAE*1*; *tnpA*, ISAE*1* transposase gene. The location and orientation of integrase genes (*int*) whose products are similar to the putative phage-type integrases/recombinases Y4RA (blue), Y4RB (magenta) and Y4RC (yellow) of *Rhizobium* sp. strain NGR234[89] are indicated. CHP, conserved hypothetical protein.

base-pairings between exon-binding sites (EBS) within the intron RNA and intron-binding sites (IBS) in the exons close to the exon–intron borders, as well as interactions among ribozyme domains.

According to the group II intron database†,[84] the *R.e.*I1 in IS*881* belongs to class D of bacterial group II introns. Based on the phylogenetic tree of the IEPs, its closest relatives are *B.j.*I1 from *Bradyrhizobium japonicum*, *M.a.*I5 from *Methanosarcina acetivorans*, *E.c.*I2 from *E. coli* and *R.m.*Int1 from *S. meliloti*.[84,85] The secondary structure model of *R.e.*I1 (Figure 5) shows the typical six-domain architecture and points to conserved elements involved in three-dimensional interactions. The structure of dI predicts a δ–δ′ interaction within this domain and an EBS3-IBS3 pairing, as has been demonstrated experimentally for the related *R.m.*Int1.[86] dIV contains the intronic ORF. Database searches and sequence alignments revealed reverse transcriptase domains 0–7 and the RNA maturase domain X. A DNA endonuclease domain is not obvious. Homing has been shown for the related *R.m.*Int1, the IEP of which, like the IEPS of most bacterial group II introns, lacks this domain.[87]. The intronic ORF in *R.e.*I1 harbors a frameshift between RT domains 6 and 7, resulting in a truncated IEP and suggesting that the intron is compromised. *R.e.*I1 is inserted within a TAC (Tyr) codon of the IS*881* transposase gene (Figure 5). Exon–exon ligation would lead to an mRNA in which the UGG (Trp) and UAC (Tyr) codons are fused in the correct frame, creating the full-length transposase ORF. The splicing reaction, however, requires a functional IEP and, therefore, it was not surprising that preliminary RT-PCR experiments failed to demonstrate *in vivo* splicing of *R.e.*I1 (T.E., unpublished results). This negative result is corroborated by additional observations. PCR experiments using total DNA of the pHG1-containing *R. eutropha* H16 (ATCC17699) and its relatives TF93 (ATCC17697), H1 (ATCC17698), H20 (ATCC17700), ATCC17701, ATCC17702, ATCC17704, ATCC17706, ATCC17707 and N9A (DSM518) as the template and primers directed against IS*881* resulted in a single product of the same size in each case that correlated with the intron-containing insertion element. A subset of these strains had been analyzed for the presence of extrachromosomal replicons and been shown to contain megaplasmids with estimated sizes ranging from 430 kb (strain H20) to 500 kb (strain ATCC17704).[8] Probing *Eco*RI-digested total DNA of the ten strains with labeled *R.e.*I1 DNA by Southern blotting identified a single band of 9.4 kb in each lane. In strain H16, this segment comprises

† http://www.fp.ucalgary.ca/group2introns/

a major part of the *hoxN1* nickel permease gene and its downstream region. The signal was absent, when the DNA of a plasmid-free derivative of strain H16 was probed (T.E., unpublished results). It is concluded that the sequence of this plasmid segment in the various strains is very similar or identical. The results predict that each copy of *R.e*.I1 is incapable of splicing. This would prevent IS*881* from transposition, because a full-length transposase cannot be produced, and explain the fact that the various *R. eutropha* strains contain only single copies of IS*881*. Another finding supports this assumption. The intron DNA was amplified from strain TF93 and a 328 bp-*Pst* I/*Apa* I fragment, comprising in *R.e*.I1 on pHG1 the frameshift region in the intronic ORF, was sequenced. The two sequences were 100% identical (T.E. & O. Degen, unpublished results).

As mentioned above, pHG1 harbors several complete insertion elements (IS*881*, IS*882*, IS*883*, IS*884*, IS*885* and ISAE*1*). Two copies of ISAE*1* are flanking the hydrogenase-4 and SH genes. They are located 17,938 bp upstream of *hoxN2* and 3798 bp downstream of *hypF2* and could form a 55 kb giant class I transposon. Kung and co-workers[13] have reported that both strands of ISAE*1* have coding capacity. ORF2 (on the opposite strand relative to the transposase gene) lacks a stop codon and in the case of PHG042, the insertion of ISAE*1* has fused ORF2 in-frame to a phage-type integrase/recombinase gene. This insertion site is illustrated in Figure 6. It is located immediately downstream of a repetitive element that could form stem-loop structures. The predicted stems consist of 3378 nucleotide (sense strands) and 3377 nucleotide (non-sense strands) and contain 93.6% complementary bases. The predicted loops are 603 nucleotide in length. Indeed, a snapback structure with a 3.4 kb stem and a 600 nucleotide loop (IR16-1) has been detected by electron microscopy of sheared, denatured and partially renatured pHG1 DNA.[88] Interestingly, the two stems have the coding capacity for partial or complete phage-type integrases/recombinases closely related to the Y4RA, Y4RB and Y4RC putative integrases encoded by plasmid pNGR234*a* in *Rhizobium* sp. strain NGR234.[89] It is tempting to speculate that the stem-loop structure of IR16-1 has guided ISAE*1* to this target, since the left direct repeat created at the insertion site is part of the stem. Electron microscopy of pHG1 DNA has predicted a second repetitive element (IR16-2) forming a 3.1 kb stem and a 3200 nucleotide loop.[88] This observation can be explained by the pHG1 sequence between nucleotides 128,214 and 137,807 that contains a perfect inverted repeat (Figure 6). Compared to IR16-1, the loops in IR16-2 (3390 nucleotide) are longer. The 3102 bp stems are 100% complementary. As in IR16-1 and in the same order, phage-type integrases/recombinases related closely to Y4RA, Y4RB and Y4RC are encoded in this element.

Materials and Methods

Library construction and genomic sequencing

The nucleotide sequence of pHG1 was determined using a shotgun approach. Megaplasmid DNA was enriched from cultures of *R. eutropha* H16 grown overnight in FN medium.[90] The DNA was sheared randomly and DNA fragments in the range of 1–3 kb were cloned in pTZ19R resulting in a shotgun library. The inserts of the recombinant plasmids were sequenced from both ends using ABI PRISM 377 (Applied Biosystems) or MegaBACE 1000 sequencers (Amersham Bioscience) with dye-terminator chemistry. In all, 5000 sequences (8.3-fold coverage) were processed with PHRED, assembled into contigs by using the PHRAP assembling tool[91] and edited with GAP4, which is part of the STADEN package software.[92] The resulting contigs of megaplasmid pHG1 were ordered according to the physical gene map and the restriction map.[12] In order to solve problems with misassembled regions caused by repetitive sequences and to close remaining sequence gaps, PCR-based techniques and primer walking on recombinant plasmids were applied.

ORF identification and annotation

Annotation was done using the workbench GENESOAP written by R. Cramm. This software is based on a database with encapsulated SQL and offers a graphical user interface for ORF viewing and editing. GENESOAP runs automated BLAST[93] searches and integrates the results of gene-finding resources (e.g. FRAMED and GENEMARK) and protein domain searches (SMART). ORFs with a minimum length of 33 codons were screened for similarity to the GenBank non-redundant protein database (release 2.2.4) with an *e* value of 10^{-4} as cutoff. ORFs with no significant similarity to a database entry were evaluated statistically using the program FRAMED† trained on a set of *R. solanacearum* genes. Global pairwise alignments were done using the program NEEDLE of the EMBOSS package. A subset of the hypothetical and conserved hypothetical proteins were investigated by searches of the PFAM database.[94] Searches using the WU-Blast2 program of W. Gish were done on the EBI server‡. TMHMM 2.0§ was used to predict membrane proteins. Signal peptides were predicted using the SignalP resource‖.

Other analyses

Screening for tRNAs was done on the TRNASCAN-SE server¶. The software MIROPEATS was used to localize large-scale sequence repetitions.[95] Analyses of GC skew and cumulative GC skew were done with programs written by P. Bucher.[23] ATPases of the VirB11 family were classified with the aid of the GeNomenclature system[a]. A computer-generated ORF map used for

† http://www.toulouse.inra.fr/FrameD/cgi-bin/FD
‡ http://www.ebi.ac.uk/blast2/index.html
§ http://www.cbs.dtu.dk/services/TMHMM-2.0
‖ http://www.cbs.dtu.dk/services/SignalP-2.0
¶ http://www.genetics.wustl.edu/eddy/tRNAscan-SE
[a] http://cpmcnet.columbia.edu/dept/figurski/genomenclature

preparing Figure 1 was kindly supplied by F. Meyer (Bielefeld).

Data Bank accession code

The complete nucleotide sequence of pHG1 has been deposited in the GenBank DNA sequence database under accession number AY305378.

Acknowledgements

We are grateful to E. Lanka for his advice on the analysis of plasmid transfer genes. We thank F. Meyer for providing a computer-generated ORF map. We thank M. Bömeke and A. Strack for expert technical assistance, and A. Wiezer for help with the graphical presentation of data. The project was carried out within the framework of the Competence Network Göttingen "Genome research on Bacteria" (GenoMik) financed by the German Federal Ministry of Education and Research (BMBF).

References

1. Schwartz, E. & Friedrich, B. (2003). The H_2-metabolizing prokaryotes. In *The Prokaryotes: an Evolving Electronic Resource for the Microbiological Community* (Dworkin, M., Schleifer, K. H. & Stackebrandt, E., eds), 3rd edit., release 3.14, Springer, New York.
2. Kusian, B. & Bowien, B. (1997). Organization and regulation of *cbb* CO_2 assimilation genes in autotrophic bacteria. *FEMS Microbiol. Rev.* **21**, 135–155.
3. Bowien, B. & Kusian, B. (2002). Genetics and control of CO_2 assimilation in the chemoautotroph *Ralstonia eutropha*. *Arch. Microbiol.* **178**, 85–93.
4. Andersen, K., Tait, R. C. & King, W. R. (1981). Plasmids required for utilization of molecular hydrogen by *Alcaligenes eutrophus*. *Arch. Microbiol.* **129**, 384–390.
5. Friedrich, B., Hogrefe, C. & Schlegel, H. G. (1981). Naturally occurring genetic transfer of hydrogen-oxidizing ability between strains of *Alcaligenes eutrophus*. *J. Bacteriol.* **147**, 198–205.
6. Hogrefe, C., Römermann, D. & Friedrich, B. (1984). *Alcaligenes eutrophus* hydrogenase genes (Hox). *J. Bacteriol.* **158**, 43–48.
7. Friedrich, B. & Schwartz, E. (1993). Molecular biology of hydrogen utilization in aerobic chemolithotrophs. *Annu. Rev. Microbiol.* **47**, 351–383.
8. Hogrefe, C. & Friedrich, B. (1984). Isolation and characterization of megaplasmid DNA from lithoautotrophic bacteria. *Plasmid*, **12**, 161–169.
9. Pfitzner, J. & Schlegel, H. G. (1973). Denitrifikation bei *Hydrogenomonas eutropha* strain H 16. *Arch. Mikrobiol.* **90**, 199–211.
10. Gorenflo, V., Schmack, G., Vogel, R. & Steinbüchel, A. (2001). Development of a process for the biotechnological large-scale production of 4-hydroxyvalerate-containing polyesters and characterization of their physical and mechanical properties. *Biomacromolecules*, **2**, 45–57.
11. Fukui, T., Abe, H. & Doi, Y. (2002). Engineering of *Ralstonia eutropha* for production of poly(3-hydroxybutyrate-*co*-3-hydroxyhexanoate) from fructose and solid-state properties of the copolymer. *Biomacromolecules*, **3**, 618–624.
12. Schwartz, E. & Friedrich, B. (2001). A physical map of the megaplasmid pHG1, one of three genomic replicons in *Ralstonia eutropha* H16. *FEMS Microbiol. Letters*, **201**, 213–219.
13. Kung, S. S., Chen, J. & Chow, W. Y. (1992). Molecular and genetic characterization of an *Alcaligenes eutrophus* insertion element. *J. Bacteriol.* **174**, 8023–8029.
14. Schneiker, S., Kosier, B., Pühler, A. & Selbitschka, W. (1999). The *Sinorhizobium meliloti* insertion sequence (IS) element ISRm*14* is related to a previously unrecognized IS element located adjacent to the *Escherichia coli* locus of enterocyte effacement (LEE) pathogenicity island. *Curr. Microbiol.* **39**, 274–281.
15. Mahillon, J. & Chandler, M. (1998). Insertion sequences. *Microbiol. Mol. Biol. Rev.* **62**, 725–774.
16. Stuart, E. S., Tehrani, A., Valentin, H. E., Dennis, D., Lenz, R. W. & Fuller, R. C. (1998). Protein organization on the PHA inclusion cytoplasmic boundary. *J. Biotechnol.* **64**, 137–144.
17. Pötter, M., Madkour, M. H., Mayer, F. & Steinbüchel, A. (2002). Regulation of phasin expression and polyhydroxyalkanoate (PHA) granule formation in *Ralstonia eutropha* H16. *Microbiology*, **148**, 2413–2426.
18. Hall, M. C. & Matson, S. W. (1999). Helicase motifs: the engine that powers DNA unwinding. *Mol. Microbiol.* **34**, 867–877.
19. Thomas, C. M. (2000). Paradigms of plasmid organization. *Mol. Microbiol.* **37**, 485–491.
20. del Solar, G. & Espinosa, M. (2000). Plasmid copy number control: an ever-growing story. *Mol. Microbiol.* **37**, 492–500.
21. Taghavi, S., Provoost, A., Mergeay, M. & van der Lelie, D. (1996). Identification of a partition and replication region in the *Alcaligenes eutrophus* megaplasmid pMOL28. *Mol. Gen. Genet.* **250**, 169–179.
22. Chattoraj, D. K. (2000). Control of plasmid DNA replication by iterons: no longer paradoxical. *Mol. Microbiol.* **37**, 467–476.
23. Grigoriev, A. (1998). Analyzing genomes with cumulative skew diagrams. *Nucl. Acids Res.* **26**, 2286–2290.
24. Gerdes, K., Møller-Jensen, J. & Bugge Jensen, R. (2000). Plasmid and chromosome partitioning: surprises from phylogeny. *Mol. Microbiol.* **37**, 455–466.
25. Koonin, E. V. (1993). A superfamily of ATPases with diverse functions containing either classical or deviant ATP-binding motif. *J. Mol. Biol.* **229**, 1165–1174.
26. Blakely, G., May, G., McCulloch, R., Arciszewska, L. K., Burke, M., Lovett, S. T. & Sherratt, D. J. (1993). Two related recombinases are required for site-specific recombination at *dif* and *cer* in *E. coli* K12. *Cell*, **75**, 351–361.
27. Jensen, R. B. & Gerdes, K. (1995). Programmed cell death in bacteria: proteic plasmid stabilization systems. *Mol. Microbiol.* **17**, 205–210.
28. Naito, T., Kusano, K. & Kobayashi, I. (1995). Selfish behavior of restriction-modification systems. *Science*, **267**, 897–899.
29. Salmond, G. P. & Reeves, P. J. (1993). Membrane traffic wardens and protein secretion in gram-negative bacteria. *Trends Biochem. Sci.* **18**, 7–12.
30. Sakai, D., Horiuchi, T. & Komano, T. (2001). ATPase activity and multimer formation of PilQ protein are

required for thin pilus biogenesis in plasmid R64. *J. Biol. Chem.* **276**, 17968–17975.
31. Bitter, W. (2003). Secretins of *Pseudomonas aeruginosa*: large holes in the outer membrane. *Arch. Microbiol.* **179**, 307–314.
32. Hayashi, S. & Wu, H. C. (1990). Lipoproteins in bacteria. *J. Bioenerg. Biomembr.* **22**, 451–471.
33. Sakai, D. & Komano, T. (2002). Genes required for plasmid R64 thin-pilus biogenesis: identification and localization of products of the *pilK*, *pilM*, *pilO*, *pilP*, *pilR*, and *pilT* genes. *J. Bacteriol.* **184**, 444–451.
34. Eisenbrandt, R., Kalkum, M., Lai, E. M., Lurz, R., Kado, C. I. & Lanka, E. (1999). Conjugative pili of IncP plasmids, and the Ti plasmid T pilus are composed of cyclic subunits. *J. Biol. Chem.* **274**, 22548–22555.
35. Schröder, G., Krause, S., Zechner, E. L., Traxler, B., Yeo, H. J., Lurz, R. *et al.* (2002). TraG-like proteins of DNA transfer systems and of the *Helicobacter pylori* type IV secretion system: inner membrane gate for exported substrates? *J. Bacteriol.* **184**, 2767–2779.
36. Ziegelin, G. & Lanka, E. (1995). Bacteriophage P4 DNA replication. *FEMS Microbiol. Rev.* **17**, 99–107.
37. Merryweather, A., Rees, C. E., Smith, N. M. & Wilkins, B. M. (1986). Role of Sog polypeptides specified by plasmid ColIb-P9 and their transfer between conjugating bacteria. *EMBO J.* **5**, 3007–3012.
38. Rees, C. E. & Wilkins, B. M. (1989). Transfer of Tra proteins into the recipient cell during bacterial conjugation mediated by plasmid ColIb-P9. *J. Bacteriol.* **171**, 3152–3157.
39. Tran-Betcke, A., Warnecke, U., Böcker, C., Zaborosch, C. & Friedrich, B. (1990). Cloning and nucleotide sequences of the genes for the subunits of NAD-reducing hydrogenase of *Alcaligenes eutrophus* H16. *J. Bacteriol.* **172**, 2920–2929.
40. Kortlüke, C., Horstmann, K., Schwartz, E., Rohde, M., Binsack, R. & Friedrich, B. (1992). A gene complex coding for the membrane-bound hydrogenase of *Alcaligenes eutrophus* H16. *J. Bacteriol.* **174**, 6277–6289.
41. Schwartz, E., Buhrke, T., Gerischer, U. & Friedrich, B. (1999). Positive transcriptional feedback controls hydrogenase expression in *Alcaligenes eutrophus* H16. *J. Bacteriol.* **181**, 5684–5692.
42. Bernhard, M., Benelli, B., Hochkoeppler, A., Zannoni, D. & Friedrich, B. (1997). Functional and structural role of the cytochrome *b* subunit of the membrane-bound hydrogenase complex of *Alcaligenes eutrophus* H16. *Eur. J. Biochem.* **248**, 179–186.
43. Dernedde, J., Eitinger, M. & Friedrich, B. (1993). Analysis of a pleiotropic gene region involved in formation of catalytically active hydrogenases in *Alcaligenes eutrophus* H16. *Arch. Microbiol.* **159**, 545–553.
44. Dernedde, J., Eitinger, T., Patenge, N. & Friedrich, B. (1996). *hyp* gene products in *Alcaligenes eutrophus* are part of a hydrogenase-maturation system. *Eur. J. Biochem.* **235**, 351–358.
45. Buhrke, T. & Friedrich, B. (1998). *hoxX* (*hypX*) is a functional member of the *Alcaligenes eutrophus hyp* gene cluster. *Arch. Microbiol.* **170**, 460–463.
46. Lenz, O., Bernhard, M., Buhrke, T., Schwartz, E. & Friedrich, B. (2002). The hydrogen-sensing apparatus in *Ralstonia eutropha*. *J. Mol. Microbiol. Biotechnol.* **4**, 255–562.
47. Eitinger, T. & Mandrand-Berthelot, M. A. (2000). Nickel transport systems in microorganisms. *Arch. Microbiol.* **173**, 1–9.
48. Eitinger, T. (2001). Microbial nickel transport. In *Microbial Transport Systems* (Winkelmann, G., ed.), pp. 397–417, Wiley, Weinheim.
49. Wolf, I., Buhrke, T., Dernedde, J., Pohlmann, A. & Friedrich, B. (1998). Duplication of *hyp* genes involved in maturation of [NiFe] hydrogenases in *Alcaligenes eutrophus* H16. *Arch. Microbiol.* **170**, 451–459.
50. Paschos, A., Bauer, A., Zimmermann, A., Zehelein, E. & Böck, A. (2002). HypF, a carbamoyl phosphate converting enzyme involved in [NiFe]-hydrogenase maturation. *J. Biol. Chem.* **277**, 49945–49951.
51. Reissmann, S., Hochleitner, E., Wang, H., Paschos, A., Lottspeich, F., Glass, R. S. & Böck, A. (2003). Taming of a poison: biosynthesis of the NiFe-hydrogenase cyanide ligands. *Science*, **299**, 1067–1070.
52. Kleihues, L., Lenz, O., Bernhard, M., Buhrke, T. & Friedrich, B. (2000). The H_2 sensor of *Ralstonia eutropha* is a member of the subclass of regulatory [NiFe] hydrogenases. *J. Bacteriol.* **182**, 2716–2724.
53. Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T. *et al.* (2003). Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nature Biotechnol.* **21**, 526–531.
54. Blokesch, M., Paschos, A., Theodoratou, E., Bauer, E., Hube, M., Huth, S. & Böck, A. (2002). Metal insertion into NiFe-hydrogenases. *Biochem. Soc. Trans.* **30**, 674–680.
55. Philippot, L. (2002). Denitrifying genes in bacterial and Archaeal genomes. *Biochim. Biophys. Acta*, **1577**, 355–376.
56. Liu, X. & DeMoss, J. A. (1997). Characterization of NarJ, a system-specific chaperone required for nitrate reductase biogenesis in *Escherichia coli*. *J. Biol. Chem.* **272**, 24266–24271.
57. Moir, J. W. & Wood, N. J. (2001). Nitrate and nitrite transport in bacteria. *Cell. Mol. Life Sci.* **58**, 215–224.
58. Stewart, V. (1994). Regulation of nitrate and nitrite reductase synthesis in enterobacteria. *Ant. v. Leeuwenhoek*, **66**, 37–45.
59. Zumft, W. G. (2002). Nitric oxide signaling and NO dependent transcriptional control in bacterial denitrification by members of the FNR-CRP regulator family. *J. Mol. Microbiol. Biotechnol.* **4**, 277–286.
60. Siddiqui, R. A., Warnecke-Eberz, U., Hengsberger, A., Schneider, B., Kostka, S. & Friedrich, B. (1993). Structure and function of a periplasmic nitrate reductase in *Alcaligenes eutrophus* H16. *J. Bacteriol.* **175**, 5867–5876.
61. Potter, L. C. & Cole, J. A. (1999). Essential roles for the products of the *napABCD* genes, but not *napFGH*, in periplasmic nitrate reduction by *Escherichia coli* K-12. *Biochem. J.* **344**, 69–76.
62. Thöny-Meyer, L. (2000). Haem-polypeptide interactions during cytochrome *c* maturation. *Biochim. Biophys. Acta*, **1459**, 316–324.
63. Blattner, F. R., Plunkett, G., III, Bloch, C. A., Perna, N. T., Burland, V., Riley, M. *et al.* (1997). The complete genome sequence of *Escherichia coli* K-12. *Science*, **277**, 1453–1474.
64. Probst, I., Wolf, G. & Schlegel, H. G. (1979). Oxygen-binding flavohemoprotein from *Alcaligenes eutrophus*. *Biochim. Biophys. Acta*, **576**, 471–478.
65. Poole, R. K. & Hughes, M. N. (2000). New functions for the ancient globin family: bacterial responses to nitric oxide and nitrosative stress. *Mol. Microbiol.* **36**, 775–783.
66. Cramm, R., Pohlmann, A. & Friedrich, B. (1999).

Purification and characterization of the single-component nitric oxide reductase from *Ralstonia eutropha* H16. *FEBS Letters*, **460**, 6–10.
67. Pohlmann, A., Cramm, R., Schmelz, K. & Friedrich, B. (2000). A novel NO-responding regulator controls the reduction of nitric oxide in *Ralstonia eutropha*. *Mol. Microbiol.* **38**, 626–638.
68. Ye, R. W., Fries, M. R., Bezborodnikov, S. G., Averill, B. A. & Tiedje, J. M. (1993). Characterization of the structural gene encoding a copper-containing nitrite reductase and homology of this gene to DNA of other denitrifiers. *Appl. Environ. Microbiol.* **59**, 250–254.
69. Rees, E., Siddiqui, R. A., Köster, F., Schneider, B. & Friedrich, B. (1997). Structural gene (*nirS*) for the cytochrome cd_1 nitrite reductase of *Alcaligenes eutrophus* H16. *Appl. Environ. Microbiol.* **63**, 800–802.
70. Siedow, A., Cramm, R., Siddiqui, R. A. & Friedrich, B. (1999). A megaplasmid-borne anaerobic ribonucleotide reductase in *Alcaligenes eutrophus* H16. *J. Bacteriol.* **181**, 4919–4928.
71. Jordan, A. & Reichard, P. (1998). Ribonucleotide reductases. *Annu. Rev. Biochem.* **67**, 71–98.
72. Zumft, W. G., Dreusch, A., Löchelt, S., Cuypers, H., Friedrich, B. & Schneider, B. (1992). Derived amino acid sequences of the *nosZ* gene (respiratory N_2O reductase) from *Alcaligenes eutrophus*, *Pseudomonas aeruginosa* and *Pseudomonas stutzeri* reveal potential copper-binding residues. Implications for the CuA site of N_2O reductase and cytochrome *c* oxidase. *Eur. J. Biochem.* **208**, 31–40.
73. Wunsch, P., Herb, M., Wieland, H., Schiek, U. M. & Zumft, W. G. (2003). Requirements for Cu(A) and Cu–S center assembly of nitrous oxide reductase deduced from complete periplasmic enzyme maturation in the nondenitrifier *Pseudomonas putida*. *J. Bacteriol.* **185**, 887–896.
74. Zumft, W. G. (1997). Cell biology and molecular basis of denitrification. *Microbiol. Mol. Biol. Rev.* **61**, 533–616.
75. Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S., Arlat, M. *et al.* (2002). Genome sequence of the plant pathogen *Ralstonia solanacearum*. *Nature*, **415**, 497–502.
76. Moss, J. E., Cardozo, T. J., Zychlinsky, A. & Groisman, E. A. (1999). The *selC*-associated SHI-2 pathogenicity island of *Shigella flexneri*. *Mol. Microbiol.* **33**, 74–83.
77. Martínez, J. L., Herrero, M. & de Lorenzo, V. (1994). The organization of intercistronic regions of the aerobactin operon of pColV-K30 may account for the differential expression of the *iucABCD iutA* genes. *J. Mol. Biol.* **238**, 288–293.
78. Funahashi, T., Moriya, K., Uemura, S., Miyoshi, S., Shinoda, S., Narimatsu, S. & Yamamoto, S. (2002). Identification and characterization of *pvuA*, a gene encoding the ferric vibrioferrin receptor protein in *Vibrio parahaemolyticus*. *J. Bacteriol.* **184**, 936–946.
79. Erb, R. W., Timmis, K. N. & Pieper, D. H. (1998). Characterization of a gene cluster from *Ralstonia eutropha* JMP134 encoding metabolism of 4-methylmuconolactone. *Gene*, **206**, 53–62.
80. Meinhof, C.-G. (1994). Studien zur IS*5*-Transposase und Entwicklung eines Informations-Managementsystems für rekombinante DNA. PhD thesis. Albert-Ludwigs-Universität Freiburg
81. Martínez-Abarca, F. & Toro, N. (2000). Group II introns in the bacterial world. *Mol. Microbiol.* **38**, 917–926.
82. Bonen, L. & Vogel, J. (2001). The ins and outs of group II introns. *Trends Genet.* **17**, 322–331.
83. Dai, L. & Zimmerly, S. (2002). Compilation and analysis of group II intron insertions in bacterial genomes: evidence for retroelement behavior. *Nucl. Acids Res.* **30**, 1091–1102.
84. Dai, L., Toor, N., Olson, R., Keeping, A. & Zimmerly, S. (2003). Database for mobile group II introns. *Nucl. Acids Res.* **31**, 424–426.
85. Toro, N. (2003). Bacteria and Archaea Group II introns: additional mobile genetic elements in the environment. *Environ. Microbiol.* **5**, 143–151.
86. Jiménez-Zurdo, J. I., Garcia-Rodriguez, F. M., Barrientos-Duran, A. & Toro, N. (2003). DNA target site requirements for homing *in vivo* of a bacterial group II intron encoding a protein lacking the DNA endonuclease domain. *J. Mol. Biol.* **326**, 413–423.
87. Martínez-Abarca, F., Garcia-Rodriguez, F. M. & Toro, N. (2000). Homing of a bacterial group II intron with an intron-encoded protein lacking a recognizable endonuclease domain. *Mol. Microbiol.* **35**, 1405–1412.
88. Rohde, C., Johannssen, W. & Mayer, F. (1986). Electron microscopic localization of hydrogenase genes on the megaplasmid pHG1 in *Alcaligenes eutrophus*. *Mol. Gen. Genet.* **202**, 476–480.
89. Freiberg, C., Fellay, R., Bairoch, A., Broughton, W. J., Rosenthal, A. & Perret, X. (1997). Molecular basis of symbiosis between *Rhizobium* and legumes. *Nature*, **387**, 394–401.
90. Nies, D., Mergeay, M., Friedrich, B. & Schlegel, H. G. (1987). Cloning of plasmid genes encoding resistance to cadmium, zinc, and cobalt in *Alcaligenes eutrophus* CH34. *J. Bacteriol.* **169**, 4865–4868.
91. Ewing, B., Hillier, L., Wendl, M. C. & Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**, 175–185.
92. Staden, R., Beal, K. F. & Bonfield, J. K. (2000). The Staden package, 1998. *Methods Mol. Biol.* **132**, 115–130.
93. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* **25**, 3389–3402.
94. Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L. & Sonnhammer, E. L. (2000). The Pfam protein families database. *Nucl. Acids Res.* **28**, 263–266.
95. Parsons, J. D. (1995). Miropeats: graphical DNA sequence comparisons. *Comput. Appl. Biosci.* **11**, 615–619.
96. Meyer, F., Goesmann, A., McHardy, A. C., Bartels, D., Bekel, T., Clausen, J. *et al.* (2003). GenDB—an open source genome annotation system for prokaryote genomes. *Nucl. Acids Res.* **31**, 2187–2195.

Edited by J. Karn

(Received 21 May 2003; received in revised form 9 July 2003; accepted 10 July 2003)