



European Union  
European Social Fund



MINISTRY OF EDUCATION & RELIGIOUS AFFAIRS  
MANAGING AUTHORITY

Co-financed by Greece and the European Union



## ΠΡΟΓΡΑΜΜΑ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗΣ ΑΕΙ ΓΙΑ ΤΗΝ ΕΠΙΚΑΙΡΟΠΟΙΗΣΗ ΓΝΩΣΕΩΝ ΑΠΟΦΟΙΤΩΝ ΑΕΙ (ΠΕΓΑ)

*«Οι σύγχρονες τεχνικές βιο-ανάλυσης στην υγεία, τη γεωργία, το περιβάλλον και τη διατροφή»*

# Genome sequence of the bioplastic-producing “Knallgas” bacterium *Ralstonia eutropha* H16

Anne Pohlmann<sup>1,5</sup>, Wolfgang Florian Fricke<sup>2,4,5</sup>, Frank Reinecke<sup>3,5</sup>, Bernhard Kusian<sup>4</sup>, Heiko Liesegang<sup>2</sup>, Rainer Cramm<sup>1</sup>, Thomas Eitinger<sup>1</sup>, Christian Ewering<sup>3</sup>, Markus Pötter<sup>3</sup>, Edward Schwartz<sup>1</sup>, Axel Strittmatter<sup>2</sup>, Ingo Voß<sup>3</sup>, Gerhard Gottschalk<sup>2</sup>, Alexander Steinbüchel<sup>3</sup>, Bärbel Friedrich<sup>1</sup> & Botho Bowien<sup>4</sup>

The H<sub>2</sub>-oxidizing lithoautotrophic bacterium *Ralstonia eutropha* H16 is a metabolically versatile organism capable of subsisting, in the absence of organic growth substrates, on H<sub>2</sub> and CO<sub>2</sub> as its sole sources of energy and carbon. *R. eutropha* H16 first attracted biotechnological interest nearly 50 years ago with the realization that the organism's ability to produce and store large amounts of poly[*R*-(-)-3-hydroxybutyrate] and other polyesters could be harnessed to make biodegradable plastics. Here we report the complete genome sequence of the two chromosomes of *R. eutropha* H16. Together, chromosome 1 (4,052,032 base pairs (bp)) and chromosome 2 (2,912,490 bp) encode 6,116 putative genes. Analysis of the genome sequence offers the genetic basis for exploiting the biotechnological potential of this organism and provides insights into its remarkable metabolic versatility.

*Ralstonia eutropha* H16 is a Gram-negative lithoautotrophic bacterium belonging to the  $\beta$ -subclass of the *Proteobacteria*. It is a ubiquitous inhabitant of soil and freshwater biotopes and is well adapted to life in habitats subject to transient anoxia<sup>1</sup>. One of the keys to the organism's lifestyle is its ability to use—alternatively or concomitantly—both organic compounds and molecular H<sub>2</sub> as sources of energy (Fig. 1). *R. eutropha* H16 can make use of the transiently available supplies of H<sub>2</sub> arising, for example, from the metabolic activity of N<sub>2</sub>-fixing microbes, because it is equipped with two energy-conserving hydrogenases<sup>2</sup>. These NiFe metalloproteins catalyze the oxidation of H<sub>2</sub>, providing the organism with energy and reductant. In the absence of environmental O<sub>2</sub>, *R. eutropha* H16 can switch to anaerobic respiration; a complete denitrification pathway allows the organism to exploit alternative electron acceptors such as NO<sub>3</sub><sup>-</sup> and NO<sub>2</sub><sup>-</sup>. Correlated with the flexible bioenergetics of the organism is its capacity to shift between heterotrophic and autotrophic growth modes. It fixes CO<sub>2</sub> via the Calvin-Benson-Bassham (CBB) cycle<sup>3</sup>. In addition, the bacterium can stockpile organic carbon in the form of poly[*R*-(-)-3-hydroxybutyrate] (PHB) in specialized storage granules<sup>4</sup>. This represents an adaptation to fluctuating O<sub>2</sub> levels, because PHB granules are formed whenever an abundance of carbon is available, but other factors such as O<sub>2</sub>, bound nitrogen or phosphate are growth-limiting.

Aside from its biological significance, this versatile, nonpathogenic organism is also of biotechnological interest. Perhaps the best known application of *R. eutropha* strains is the commercial production of the biodegradable thermoplastic Biopol<sup>5–7</sup> (<http://www.metabolix.com>). Production of biomolecules labeled with stable isotopes has been

carried out via lithoautotrophic fermentation of *R. eutropha* H16 (ref. 8). Whether cultivated under lithoautotrophic or heterotrophic conditions, *R. eutropha* H16 reaches high cell densities (~200 g dry weight per liter)<sup>6,9</sup>. Very recently an investigation demonstrated the potential application of the O<sub>2</sub>-tolerant, CO-resistant, membrane-bound hydrogenase of *R. eutropha* H16 for the construction of biological fuel cells<sup>10,11</sup> and for designing light-driven H<sub>2</sub> production complexes<sup>12</sup>. Another promising experimental approach documented the usefulness of *R. eutropha* H16 hydrogenase for the construction of an H<sub>2</sub>-sensing device<sup>13</sup>.

As a basis for future studies we undertook the sequencing and annotation of the tripartite *R. eutropha* H16 genome. The sequence of the smallest of the three replicons, the megaplasmid pHG1, has been reported<sup>14</sup>. Here we present the sequence analysis of the two main chromosomes.

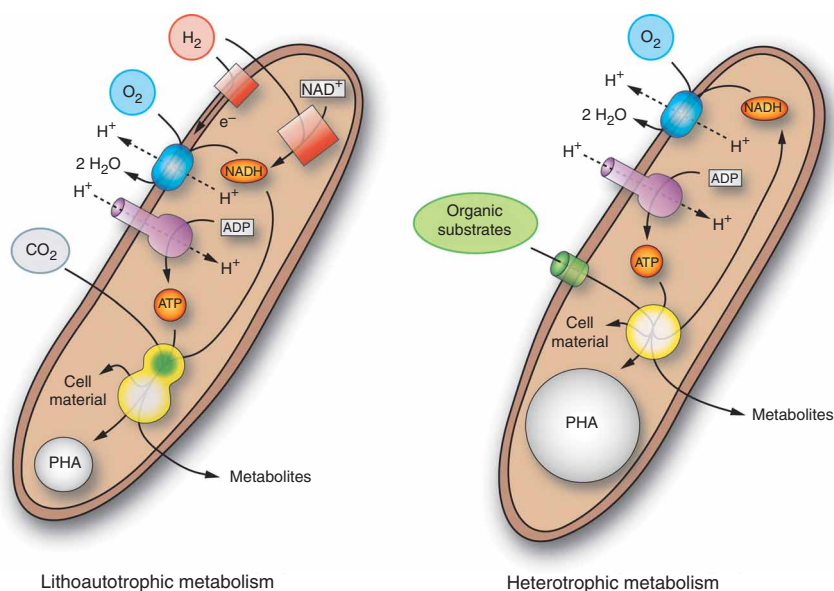
## RESULTS

### Organization and general features of the genome

The genome of *R. eutropha* H16 consists of three circular replicons: chromosome 1 (4,052,032 bp), chromosome 2 (2,912,490 bp) and megaplasmid pHG1 (452,156 bp), adding up to a total size of 7,416,678 bp (Supplementary Fig. 1 online). The number and size of the genomic replicons is in agreement with the physical mapping data obtained by pulsed-field electrophoresis<sup>15</sup>. General features of the three replicons are listed in Table 1. The two chromosomes have an almost identical G+C content and nearly the same proportion of coding regions. Both parameters differ from the corresponding values

<sup>1</sup>Humboldt-Universität zu Berlin, Institut für Biologie/Mikrobiologie, Chausseestraße 117, 10115 Berlin, Germany. <sup>2</sup>Georg-August-Universität Göttingen, Laboratorium für Genomanalyse, Grisebachstraße 8, 37077 Göttingen, Germany. <sup>3</sup>Westfälische Wilhelms-Universität Münster, Institut für Molekulare Mikrobiologie und Biotechnologie, Corrensstraße 3, 48149 Münster, Germany. <sup>4</sup>Georg-August-Universität Göttingen, Institut für Mikrobiologie und Genetik, Grisebachstraße 8, 37077 Göttingen, Germany. <sup>5</sup>These authors contributed equally to this work. Correspondence should be addressed to B.F. ([baerbel.friedrich@rz.hu-berlin.de](mailto:baerbel.friedrich@rz.hu-berlin.de)).

Received 13 April; accepted 3 August; published online 10 September 2006; corrected after print 30 March 2007; doi:10.1038/nbt1244



**Figure 1** The main growth modes of *R. eutropha* H16. Schematic representation illustrating the key aspects of lithoautotrophic and heterotrophic metabolism. The yellow circles represent the processes of central metabolism, whereas the yellow/green circle is the Calvin-Benson-Bassham cycle. The red squares symbolize the two energy-conserving hydrogenases. The gray circles indicate polyhydroxyalkanoate (PHA) storage granules.

for pHG1. A cumulative GC skew analysis of chromosome 1 pointed out a region with a typical 'backbone' configuration suggestive of an origin of replication: a *dnaA* ortholog with an adjacent consensus DnaA-binding site is flanked by the genes *dnaN* and *gyrB*. A cumulative GC skew analysis of chromosome 2 indicated an inflection point in the vicinity of a *repA* gene. Eight direct repeats are present in the 5'-flanking region of *repA*, representing typical iteron-like repeats that may be involved in RepA binding. Immediately adjacent to *repA* are orthologs of *parA* and *parB* for a plasmidlike partitioning system. Thus, the putative chromosome 2 origin has plasmidlike characteristics (for details see **Supplementary Fig. 1**). The *R. eutropha* H16 genome contains 59 transfer RNA (tRNA) genes. Of these 59 genes, 51, a complete set representing all possible codons, are carried on chromosome 1. Seven duplicates are located on chromosome 2 and one on pHG1. Five complete ribosomal RNA (rRNA) operons were identified in the sequence. Three of these operons are on chromosome 1 and two on chromosome 2 (**Table 1**).

A minimum of 32 coding sequences (CDSs) on the two chromosomes resemble transposase genes. Some of these are contained in complete insertion sequence elements or in remnants thereof. The genome of *R. eutropha* H16 contains 14 superficially intact insertion elements. Seven of these are located on megaplasmid pHG1. The latter also carries an extensive 'junkyard' region encompassing 39 remnants of transposases and phage-type integrases<sup>14</sup>. No comparable region was found on the chromosomes. Three copies of the insertion sequence ISAE1 are located on chromosome 1 and the other two on pHG1 (ref. 16). One of the plasmid-borne copies of ISAE1 and one chromosomal copy create in-frame fusions of ORF2 of the mobile element and a CDS at the target site. In the case of the chromosome 1 insertion, this is a leucyl-tRNA synthetase gene (H16\_A3139). Since additional leucyl-tRNA synthetase genes are not found in the genome, a putative fusion protein (with > 400 extra amino acid residues at the N terminus) must have leucyl-tRNA-charging activity.

Analysis of the distribution of genes representing major functional categories (**Supplementary Fig. 1**) revealed that chromosome 1 encodes most key functions of DNA replication, transcription and translation, including the ribosomal proteins. Chromosome 2 harbors genes for central steps of the 2-keto-3-deoxy-6-phosphogluconate (KDPG) sugar and sugar acid catabolism path, the decomposition of aromatic compounds and the utilization of alternative nitrogen sources.

Genes for the synthesis of cell appendages and for chemotaxis proteins are located on all three replicons of *R. eutropha* H16. Biosynthesis of flagella and the corresponding functions for chemotaxis are encoded by 72 genes arranged in four clusters on chromosome 2. Genes for 10 methyl-accepting sensor proteins were found on chromosome 2, including one gene for a putative aerotaxis sensor. Three additional putative sensor proteins are encoded on chromosome 1. Synthesis and secretion of type IV pili, necessary for twitching motility, are encoded by 41 genes on chromosome 1 and pHG1. Genes for Fli-like pili, which are responsible for tight nonspecific adherence to surfaces, were identified in two clusters on chromosome 1 and one cluster on chromosome 2 (**Supplementary Table 1** online).

#### Lithoautotrophic and organoautotrophic metabolism

Autotrophic fixation of CO<sub>2</sub> in *R. eutropha* H16 proceeds via the CBB cycle. Ribulose-1,5-bisphosphate carboxylase/oxygenase and the other enzymes of this pathway are encoded in duplicate *cbb* operons<sup>3</sup>. One set of *cbb* genes (PHG416-428, *cbb<sub>p</sub>*) maps on the megaplasmid pHG1, and the second (H16\_B1384-1396, *cbb<sub>c</sub>*) is located on chromosome 2. A regulatory gene *cbbR*, encoding a transcriptional activator, belongs to the chromosomal locus. A plasmid-borne copy of *cbbR* is present but defective, so the chromosomal copy controls the coordinate expression of both operons<sup>17</sup>. Two substrates supply energy for autotrophic CO<sub>2</sub> fixation. H<sub>2</sub> is oxidized by two NiFe hydrogenases that are encoded together with accessory and regulatory functions on pHG1 (ref. 14). Alternatively, formate is used as an energy source for organoautotrophic growth<sup>18</sup>. The *R. eutropha* H16 genome encodes at least four different formate dehydrogenases (**Supplementary Fig. 2** online). A soluble, NAD<sup>+</sup>-reducing, molybdenum-containing enzyme has been characterized<sup>19</sup>. This molybdoenzyme is encoded by the *fds* cluster on chromosome 1 (H16\_A0639-0644). The genomic sequence revealed determinants for three other membrane-bound formate dehydrogenases. Indeed, three distinct membrane-bound activities have been detected in *R. eutropha* H16 (ref. 20). The three additional formate dehydrogenases are exported across the cytoplasmic membrane by the twin-arginine-translocator (TAT) system and are bound to the periplasmic site of the membrane. They are encoded by the *fdh* genes (H16\_A2932-2937) on chromosome 1 and the *fdo* (H16\_B1452-1455) and *fdw* genes (H16\_B1700-1701), respectively, on chromosome 2. The latter enzyme is likely to be a tungsten-containing formate dehydrogenase.

**Table 1** General features of the *Ralstonia eutropha* H16 genome

Feature	Chromosome 1	Chromosome 2	Megaplasmid pHG1
Size (bp)	4,052,032	2,912,490	452,156
G+C ratio (mol%)	66.4	66.7	62.3
Percentage coding	88.1	88.6	79.7
tRNA	51	7	1
rRNA operons	3	2	0
IS elements	4	3	7
Total number of CDSs	3,651	2,555	420
No. of CDSs with assigned function	2,382	1,618	225
CDSs with unknown function	841	680	157
CDSs with general function prediction	428	257	38

### Heterotrophic carbon metabolism

As a facultative lithoautotroph *R. eutropha* is able to use various organic carbon and energy sources for heterotrophic growth. Typical substrates permitting high specific growth rates include tricarboxylic acid cycle intermediates, sugar acids like gluconic acid, fatty acids or other organic acids and amino acids; various alcohols and polyols also support growth<sup>21</sup>. The genes coding for organic acid metabolism are located on chromosome 1. The ability of strain H16 to metabolize sugars is restricted to fructose and *N*-acetylglucosamine<sup>21</sup>. Fructose is probably imported by an ABC-type transporter (H16\_B1498-1500), and catabolized via the Entner-Doudoroff (KDPG) pathway. The corresponding genes are located on chromosome 2 within three gene clusters coding for enzymes for the degradation of fructose, glucose, 2-ketogluconate and glucosaminic acid (see **Supplementary Fig. 2** for details). Previous biochemical analysis have failed to detect activities of the key enzymes of the Embden-Meyerhoff-Parnas and the oxidative pentose phosphate pathways, phosphofructokinase and 6-phosphogluconate dehydrogenase, respectively, in strain H16 (ref. 22). In agreement with these findings the respective genes were not identified in the *R. eutropha* H16 genome. Nevertheless genes for an anabolically operating Embden-Meyerhoff-Parnas pathway (gluconeogenesis) are present in *R. eutropha* H16 and scattered on chromosome 1 (**Supplementary Fig. 2**). A large set of genes located on chromosome 1 determines the utilization of the amino sugar *N*-acetylglucosamine, whose uptake is probably mediated by a phosphotransferase-type transport system (H16\_A0311-0316).

### Aerobic and anaerobic respiration

As expected for a strictly respiratory organism dwelling in an environment with fluctuating O<sub>2</sub> levels, *R. eutropha* H16 can draw on an extensive inventory of genes for respiratory chain components. Key determinants of aerobic energy metabolism include a cluster of 14 genes located on chromosome 1 (H16\_A1050-1063) encoding a typical NADH dehydrogenase. All of these genes are highly conserved and share similarities to the closest relatives of *R. eutropha* H16 (*R. solanacearum* GMI1000<sup>23</sup>, *Burkholderia mallei*<sup>24</sup>, *Burkholderia pseudomallei*<sup>25</sup>, *R. eutropha* JMP134 and *R. metallidurans* CH34; for the latter two genomic sequences see <http://www.jgi.doe.gov>). An alternative type 2 NADH dehydrogenase (H16\_A2740) might play a role in supplying electrons to the respiratory chain during growth of the cells on highly reduced substrates under aerobic conditions. The genes for succinate dehydrogenase (H16\_A2629-2632) and the cytochrome *bc1* complex (H16\_A3396-3398) were identified on the same replicon. Strain H16 is equipped with genes coding for eight distinct terminal oxidases (details given in **Supplementary**

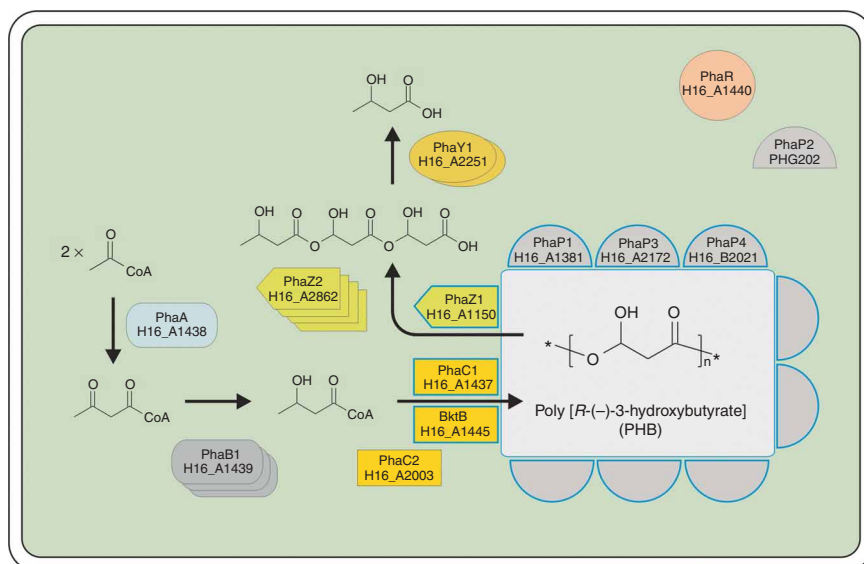
**Fig. 2**). Two gene clusters for cytochrome *c* oxidases of the heme-copper oxidase superfamily are located on chromosome 1 and chromosome 2, respectively. The gene cluster on chromosome 1 includes genes for heme *o* and heme *a* synthases, suggesting that an *aa*<sub>3</sub>-type oxidase is formed. Chromosome 1 also carries genes for a *cbb*<sub>3</sub>-type cytochrome oxidase (H16\_A2315-2320) and two *bo*<sub>3</sub>-type quinol oxidases (H16\_A1071-1074, H16\_A1640-1643). An additional *bo*<sub>3</sub> quinol oxidase as well as two alternative *bd*-type quinol oxidases are encoded on chromosome 2.

Under anoxic conditions *R. eutropha* H16 forms a complete denitrification pathway. Context analyses of the relevant gene regions revealed a unique distribution of genes for key enzymes and transcriptional regulators of the denitrification pathway and alternative enzymes involved in heme and deoxynucleotide synthesis under anaerobic conditions (**Supplementary Fig. 2**). Nearly identical copies of nitrate reductase and nitric oxide reductase genes are present on chromosome 2 and pHG1 (refs. 26, 27). Three putative transcriptional regulators (Fnr, DnrD, NarL), which may be involved in the control of anaerobic metabolism, are encoded in the pHG1-borne nitrate reductase cluster. These genes are missing in the chromosomal nitrate reductase gene cluster. On the other hand, cytochrome *cd*<sub>1</sub> nitrite reductase and nitrous oxide reductase are encoded only on chromosome 2 and megaplasmid pHG1, respectively. In addition to the earlier reported flavohemoglobin gene located on pHG1 (ref. 14), a second flavohemoglobin gene (*hmp*, H16A\_3533) was identified in the *R. eutropha* H16 genome on chromosome 1. The latter is probably regulated by a colocalized NsrR-like transcriptional regulator (**Supplementary Fig. 2**).

In the context of anaerobic metabolism, the genes for ribonucleotide reductase (RNR) deserve mention. RNRs provide deoxyribonucleotides by reduction of ribonucleotides and thus are essential for DNA synthesis. Three classes of RNRs are known, roughly classifiable as 'aerobic (NrdAB or NrdEF)', 'coenzyme B<sub>12</sub>-dependent (NrdJ)' and 'anaerobic (NrdDG)' enzymes. Comparison with the ribonucleotide reductase database (<http://nrndb.molbio.su.se>) revealed that strain H16 is one of only a few prokaryotes that encodes copies of all three classes of RNR: One aerobic class I ribonucleotide reductase (*nrdAB*) on chromosome 1, one anaerobic class III enzyme (*nrdG* and *nrdD*) on pHG1 (ref. 28) and one coenzyme B<sub>12</sub>-dependent class II enzyme (*nrdJ*) on chromosome 1. Three other coenzyme B<sub>12</sub>-dependent enzymes are encoded in the H16 genome: ethanolamine ammonium lyase (H16\_B0096-0097) on chromosome 2, methylmalonyl CoA mutase (H16\_B1841-1842) on chromosome 2 and methionine synthase (H16\_A0151) on chromosome 1. *R. eutropha* H16 is unable to synthesize coenzyme B<sub>12</sub> *de novo*. Nevertheless, the cells are able to assimilate cobalamin and its precursors. The 12 genes necessary for uptake and assimilation of cobalamin are clustered on chromosome 1 (H16\_A2961-H16\_A2972; **Supplementary Fig. 2**). *R. eutropha* H16 is able to bypass coenzyme B<sub>12</sub>-dependent reactions. It catabolizes, for example, propionyl-CoA via the methylcitric acid cycle instead of the methylmalonyl-CoA pathway<sup>29</sup>.

Aside from coenzyme B<sub>12</sub>, *R. eutropha* H16 contains genes for the *de novo* synthesis of all essential cofactors. The reconstructed pathways for the biosynthesis of biotin, thiamine, riboflavin, NAD<sup>+</sup> and genes for the assembly of Fe-S clusters are presented in the **Supplementary Figure 2**.





**Figure 2** Overview of polyhydroxybutyrate (PHB) metabolism in *R. eutropha* H16. The light gray field represents a PHB storage granule inside a bacterial cell. Arrows indicate the major enzymatic reactions of PHB synthesis and degradation. The boxes, circles and other symbols symbolize the enzymes and proteins involved directly in PHB metabolism, whereas stacked symbols indicate the existence of multiple alleles in the genome. Symbols outlined in blue represent proteins known to be associated with the PHB granule. See text for details.

### Transport

The *R. eutropha* H16 genome encodes a multitude of different transport systems (**Supplementary Table 2** online). In this context the numerous representatives of the extracytoplasmic solute-binding proteins deserves special mention. Extracytoplasmic solute-binding proteins were first described as periplasmic binding proteins of a new family of transporters called tripartite tricarboxylate transporter (TC 2.A.80). Subsequently, numerous orthologs, known as ‘Bug receptors’, have been identified in the genomes of various  $\beta$ -subclass proteobacteria, for example, *Bordetella pertussis*<sup>30</sup>. Notably, the majority of these genes are not linked to transporter genes.

A total of 154 representatives of this family were identified in the *R. eutropha* H16 genome (See **Supplementary Table 3** online). With two exceptions, all members of this group appear to be periplasmic proteins. About one-fourth (21.4%) have a putative TAT leader sequence predicted by the TatP algorithm (<http://www.cbs.dtu.dk>). Context analyses showed that a majority (100 CDSs; 64.1%) of the tripartite tricarboxylate transporter extracytoplasmic solute-binding proteins in the H16 genome are located immediately adjacent to (42.3%) or one to two CDSs away from (21.8%) the genes for transcriptional regulators. This suggests a regulatory role for the extracytoplasmic solute-binding proteins. In the case of the *B. pertussis* Bug protein BctC, it has been shown that citrate-liganded BctC regulates expression of citrate uptake genes by interacting with a two-component regulatory system<sup>31</sup>.

### Polyester synthesis

Like many bacteria, *R. eutropha* H16 can accumulate polyhydroxyalkanoates (PHAs) such as PHB in intracellular storage granules. Since the initial publication by Schlegel and co-workers describing enzymes involved in the synthesis of PHB in *R. eutropha* H16 (ref. 32), a tremendous amount of work has been done on this subject, making *R. eutropha* the model organism for the investigation of microbial

polyoxoester production. Many key components of PHA metabolism have been identified in previous studies. The analysis of the *R. eutropha* H16 genome has added to the list of known genes, providing a comprehensive view of PHA metabolism in *R. eutropha* (**Fig. 2**). The key players in PHB synthesis are (i) PHB synthase (*phaC1*), which catalyzes the polymerization of *R*-(-)-3-hydroxybutyryl-CoA, (ii)  $\beta$ -ketothiolase (*phaA*), which condenses two molecules of acetyl-CoA to acetoacetyl-CoA and (iii) NADPH-dependent acetoacetyl-CoA reductase (*phaB1*, *phaB2* and *phaB3*), which reduces acetoacetyl-CoA to *R*-(-)-3-hydroxybutyryl-CoA. A surprising finding of the genome analysis was the existence of a second gene for a PHB synthase, *phaC2*. Isologs of *phaA* (for example, *bktB*), which encode alternative  $\beta$ -ketothiolases with differing substrate spectra, have been reported<sup>33</sup>. The genomic sequence of *R. eutropha* H16 revealed 37 additional *phaA* isologs. In addition, 15 *phaB* isologs were identified. The contribution of the respective gene products to PHB synthesis remains to be investigated. Four phasin genes have been identified: *phaP1*, *phaP2*, *phaP3* and *phaP4*, all of which are expressed<sup>34</sup>. The phasin proteins build up a layer at the surface of the PHB granules, inhibiting their coalescence and thus regulating their number and size. Phasins may also play a role in the mobilization of the storage polymer.

Mobilization of PHB during carbon starvation requires specific enzymes for the depolymerization of PHB<sup>35</sup>. Seven genes in the *R. eutropha* H16 genome encode PHB depolymerase isoenzymes. The genes *phaZ1* and *phaZ2* are located on chromosome 1, *phaZ3*, *phaZ5*, *phaZ6* and *phaZ7* on chromosome 2 and *phaZ4* on pHG1. In addition, two PHB-oligomer hydrolases, which degrade the trimeric products of the PHB depolymerases, are encoded on chromosome 2 (*phaY1* and *phaY2*)<sup>36</sup>. The foregoing data and the results of systematic protein studies sketch a fascinating picture of the PHB granule as a complex and dynamic organelle<sup>34</sup>.

### Biodegradation

Traditionally, strains of the genus *Ralstonia* have played a key role in research on microbial degradation of aromatic compounds. Although *R. eutropha* strain JMP134 was used in several of these studies, it has been known for many years that *R. eutropha* H16 can grow on a similar spectrum of aromatic compounds<sup>37</sup>. It is therefore not surprising that the genomic sequence contains an impressive array of genes related to the metabolism of aromatics. Details are given in **Supplementary Figure 3** online.

### Toxin genes

A number of genes encoding potential toxins are encoded in the genome (**Supplementary Table 1**). H16\_B1353 might encode an insecticidal toxin similar to those found in *Phototrabdus luminescens*<sup>38</sup>. This putative toxin (3,406 amino acid residues, 374.2 kDa) is encoded by the largest CDS in the *R. eutropha* H16 genome. Two gene clusters for repeat-in-toxin type toxins are present on chromosome 2. The *tolC* gene, which is essential for toxin secretion, is present in both clusters but is defective in one. Two additional *tolC*-like genes are

located on chromosome 1 (H16\_A2296, H16\_A3731). A putative hemagglutinin/adhesin gene cluster (*fhaBC*; H16\_B0247-0248) was identified on chromosome 2. *R. eutropha* H16 has never been recognized as a human, animal or plant pathogen, and therefore has good potential for use in production processes.

## DISCUSSION

With a total size of 7,416,677 bp, *R. eutropha* H16 possesses one of the larger bacterial genomes sequenced to date. The genomes of the other sequenced members of the family Burkholderiaceae range between 5.8 and 8.6 Mbp<sup>23–25</sup>. Comparisons of the genomes of *R. eutropha* H16 and its close relatives reveal some important organizational similarities. A common feature of these organisms is a bipartite genome consisting of two chromosome-sized replicons. In addition, one or more plasmids may be present. Analysis of synteny for the large chromosomes showed a considerable degree of conservation of gene order (**Supplementary Fig. 4** online). In contrast, gene order was much less conserved among the small chromosomes. Reciprocal BLAST analyses revealed that chromosome 1 of *R. eutropha* H16 shares 76% and 54% orthologous genes with the corresponding replicons of *R. eutropha* JMP134 and *B. pseudomallei*, respectively. Chromosomes 2 of the two *R. eutropha* strains still have 58% orthologs in common, whereas the small chromosome of *B. pseudomallei* only shares 29% orthologs with that of *R. eutropha* H16.

Genes for H<sub>2</sub> oxidation and CO<sub>2</sub> fixation are present in *R. eutropha* H16 but not in *R. eutropha* JMP134 and *R. solanacearum* GM1000. The latter two organisms do not grow litho- or organoautotrophically<sup>39</sup>. In addition, genes for a nonribosomal polypeptide synthetase and putative repeat-in-toxin type toxin gene clusters are present in *R. eutropha* H16 but not in the other *Ralstonia* strains.

In several respects the *R. eutropha* H16 genome is typical of free-living bacteria. It contains 690 genes for potential regulatory and 83 genes for additional signaling proteins (**Supplementary Table 4** online). This represents 11.7% of its genes and is comparable to the statistics for *Pseudomonas putida* KT2440 and *Streptomyces coelicolor* A3(2). A large repertoire of regulatory genes is not surprising in an organism that must respond to the changing conditions of a highly variable habitat. Another characteristic of the soil biotope is its complex content of organic substances. A typical adaptation to this spectrum of diverse substrates is the large inventory of transport systems found in soil bacteria. A total of 832 CDS (12.3%) in the *R. eutropha* genome encode transport proteins (**Supplementary Table 2**). This is significantly higher than the average found in Eubacteria (9.2%) or Archaea (6.7%)<sup>40</sup>. An notable feature of the *R. eutropha* H16 genome is the scattering of genetic information. Genes for many metabolic processes (for example, denitrification) are distributed on all three replicons.

The enzymatic machinery for the synthesis of PHAs is robust and tunable; the composition and, hence, the physicochemical properties of the resulting polymer can be modified over a wide range by varying the carbon substrates in the feedstock<sup>7</sup>. The commercially produced Biopol, a 3-hydroxybutyrate/3-hydroxyvalerate copolymer, is one of >140 different PHAs described thus far. Recently it was discovered that *R. eutropha* H16 can, if fed with 3-mercaptopropionate or other organic thiochemicals, produce polythioesters, a novel class of bioplastics<sup>41,42</sup>. The variety of polymers that can be produced by this system is dependent on, among other things, the substrate spectra of the biosynthetic enzymes. The genome sequence revealed the presence of 38 candidate genes for  $\beta$ -ketothiolases and 15 candidate genes for acetoacetyl-CoA reductases, pointing to a potential for the production of many other polymer types.

This information may lead to novel PHAs containing substantial amounts of unusual building blocks such as 3-hydroxypropionic acid, 4-hydroxybutyric acid, 3-hydroxyvaleric acid, 4-hydroxyvaleric acid, 2-methyl-3-hydroxyvaleric acid, and 3-hydroxy-4-pentenoic acid. This is just one of the areas in which the biotechnological potential of *R. eutropha* remains to be further exploited. Moreover, the availability of the *R. eutropha* H16 genome sequence opens new perspectives for H<sub>2</sub>-based bioprocesses.

## METHODS

The *R. eutropha* H16 (DSM 428, ATCC 17699) genome sequence was determined by a whole-genome shotgun approach. Shotgun sequencing was done by Integrated Genomics. A plasmid library of 67,200 clones containing inserts of approximately 2.5 kilobases was generated in the vector pGEM-3Z. The vector Lorist6 was used to generate a cosmid library of 2,688 clones with insert sizes of 25–45 kilobases. All clones were sequenced using dye-terminator chemistry on ABI PRISM 3700 (Applied Biosystems) and MegaBACE 1000 (GE Healthcare) analyzers. Then, 101,060 reads were processed with Phred and assembled into contigs using the Phrap assembly tool (P. Green, University of Washington, Seattle; <http://www.phrap.org>). Assembly was guided by separate sequence data sets obtained from preparations of chromosome 1 and chromosome 2 isolated via pulsed-field gel electrophoresis as follows: Cassettes containing a homing site for the intron-encoded endonuclease I-SceI and an antibiotic resistance gene (kanamycin, gentamicin) were introduced into the structural genes for oxygen-independent coproporphyrinogen III oxidase (*hemN*) on chromosome 1 and nitric oxide reductase (*norB2*) on chromosome 2 as described<sup>15</sup>. Agarose-embedded genomic DNA was digested with I-SceI, subsequently separated by contour-clamped homogeneous field (CHEF; BioRad) electrophoresis and eluted from CHEF gels. The DNA was amplified and sequenced by GenomiPhi DNA Amplification Kit (GE Healthcare). Primer walking and PCR in conjunction with a cosmid library were used to close remaining gaps and to resolve assembly errors due to repetitive sequences. Sequence editing was done in GAP4 (ref. 43). Automatic CDS prediction was done using YACOP<sup>44</sup> and manually compared with a FrameD prediction<sup>45</sup>. The ERGO software suite<sup>46</sup> (Integrated Genomics) was used in the first phase of this study to obtain a preliminary automated annotation. All annotations were edited manually using the GeneSOAP annotation workbench<sup>14</sup> written by one of us (R.C.) and custom Perl scripts written by another of us (F.R.). GeneSOAP is a proprietary, Windows-based front-end software, which integrates a panel of established search and analysis tools including BLAST, FrameD, SMART and KEGG. The GenBank nonredundant protein database was the main database used for annotation. Annotations were checked by searches against curated databases including PFAM, PROSITE, PRODOM and COGS. Significant database hits were added to the note section of the GenBank entry of the corresponding CDS. Specialized databases such as the TCDB (<http://www.tcdb.org>) were used to classify functional groups of proteins. Context analysis and pathway reconstruction were done using the SEED<sup>47</sup> and KEGG resources. Putative frameshifts were checked and corrected manually. Genomic comparisons were carried out by bidirectional BLAST comparisons of whole-genome protein databases. Insertion sequences were located by computer-based and visual inspections. The flanking sequences of CDSs with similarity to transposases were analyzed for direct and inverted repeats. Global searches were done with the help of the online tool IS FINDER (<http://www-is.biotoul.fr>). Circular GC skew plots were created using the program GenomeViz<sup>48</sup>.

The complete genome sequence of *R. eutropha* H16 is available under EMBL accession numbers AM260479 (chromosome 1), AM260480 (chromosome 2) and AY305378 (megaplasmid pHG1).

Note: Supplementary information is available on the Nature Biotechnology website.

## ACKNOWLEDGMENTS

We are very grateful to Solveig Woitek and Christian Brämer for skillful support and detailed analysis in the initial phase of the project. We thank Dmitry Rodionov for helpful comments on genome annotation procedures. The *Ralstonia eutropha* genome project is part of the Competence Network Göttingen “Genome research on bacteria” (GenoMik) initiative financed by the Bundesministerium für Bildung und Forschung (BMBF). It was further

supported by a grant from the Niedersächsisches Ministerium für Wissenschaft und Kultur.

#### COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at <http://www.nature.com/naturebiotechnology/>

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>

- Aragno, M. & Schlegel, H.G. The mesophilic hydrogen-oxidizing (knallgas) bacteria. in *The Prokaryotes* (eds. Balows, A., Trüper, H.G., Dworkin, M., Harder, W. & Schleifer, K.-H.) 344–384, (Springer-Verlag, New York, NY, 1992).
- Schwartz, E. & Friedrich, B. The H<sub>2</sub>-metabolizing prokaryotes. in *The Prokaryotes: An Evolving Electronic Resource for the Microbiological Community* release 3.13, Edn. 3 (eds. Dworkin, M., Schleifer, K.H. & Stackebrandt, E.) (Springer-Verlag, New York, NY, 2003).
- Bowien, B. & Kusian, B. Genetics and control of CO<sub>2</sub> assimilation in the chemoautotroph *Ralstonia eutropha*. *Arch. Microbiol.* **178**, 85–93 (2002).
- Steinbüchel, A. & Fichtentbusch, B. Bacterial and other biological systems for polyester production. *Trends Biotechnol.* **16**, 419–427 (1998).
- Steinbüchel, A. Biodegradable plastics. *Curr. Opin. Biotechnol.* **3**, 291 (1992).
- Lee, S.Y. & Park, S.J. Fermentative production of short-chain-length PHAs. in *Biotechnology of Biopolymers—From Synthesis to Patents* Vol. 1 (eds. Steinbüchel, A. & Doi, Y.) 207–234, (Wiley-VCH, Weinheim, 2005).
- Asrar, J. & Gruys, K.J. Biodegradable polymer (Biopol®). in *Biopolymers* Vol. 4 (eds. Doi, Y. & Steinbüchel, A.) 53–90, (Wiley-VCH, Weinheim, 2002).
- Heumann, H. Method for labeling biopolymers using isotopes. International patent PCT Int. Appl. Publ. WO0012140 (2000).
- Schlegel, H.G. & Lafferty, R. Growth of Knallgas bacteria (*Hydrogenomonas*) using direct electrolysis of culture medium. *Nature* **205**, 308–309 (1965).
- Armstrong, F.A., Vincent, K.A., Friedrich, B. & Lenz, O. Biochemical fuel cell comprising oxygen-tolerant hydrogenases adsorbed on anode surfaces. *Int. patent PCT Int. Appl. Publ. WO2006030196* (2006).
- Vincent, K.A. *et al.* Electrocatalytic hydrogen oxidation by an enzyme at high carbon monoxide or oxygen levels. *Proc. Natl. Acad. Sci. USA* **102**, 16951–16954 (2005).
- Ihara, M. *et al.* Light-driven hydrogen production by a hybrid complex of a [NiFe]-hydrogenase and the cyanobacterial photosystem I. *Photochem. Photobiol.* **82**, 676–682 (2006).
- Lutz, B.J., Fan, Z.H., Burgdorf, T. & Friedrich, B. Hydrogen sensing by enzyme-catalyzed electrochemical detection. *Anal. Chem.* **77**, 4969–4975 (2005).
- Schwartz, E. *et al.* Complete nucleotide sequence of pHG1: a *Ralstonia eutropha* H16 megaplasmid encoding key enzymes of H<sub>2</sub>-based lithoautotrophy and anaerobiosis. *J. Mol. Biol.* **332**, 369–383 (2003).
- Schwartz, E. & Friedrich, B. A physical map of the megaplasmid pHG1, one of three genomic replicons in *Ralstonia eutropha* H16. *FEMS Microbiol. Lett.* **201**, 213–219 (2001).
- Kung, S.S., Chen, J. & Chow, W.Y. Molecular and genetic characterization of an *Alcaligenes eutrophus* insertion element. *J. Bacteriol.* **174**, 8023–8029 (1992).
- Windhövel, U. & Bowien, B. Identification of *cfxR*, an activator gene of autotrophic CO<sub>2</sub> fixation in *Alcaligenes eutrophus*. *Mol. Microbiol.* **5**, 2695–2705 (1991).
- Friedrich, C.G., Bowien, B. & Friedrich, B. Formate and oxalate metabolism in *Alcaligenes eutrophus*. *J. Gen. Microbiol.* **115**, 185–192 (1979).
- Oh, J.I. & Bowien, B. Structural analysis of the *fds* operon encoding the NAD<sup>+</sup>-linked formate dehydrogenase of *Ralstonia eutropha*. *J. Biol. Chem.* **273**, 26349–26360 (1998).
- Burgdorf, T., Bömmer, D. & Bowien, B. Involvement of an unusual *mol* operon in molybdopterin cofactor biosynthesis in *Ralstonia eutropha*. *J. Mol. Microbiol. Biotechnol.* **3**, 619–629 (2001).
- Kerstens, K. & De Ley, J. Genus *Alcaligenes* Castellani and Chalmers 1919. in *Bergey's Manual of Systematic Bacteriology* Vol. 1 (eds. Krieg, N.R. & Holt, J.G.) 381–373, (Williams & Wilkins, Baltimore, 1984).
- Gottschalk, G., Eberhardt, U. & Schlegel, H.G. Utilization of fructose by *Hydrogenomonas* H16 (I). *Arch. Mikrobiol.* **48**, 95–108 (1964).
- Salanoubat, M. *et al.* Genome sequence of the plant pathogen *Ralstonia solanacearum*. *Nature* **415**, 497–502 (2002).
- Nierman, W.C. *et al.* Structural flexibility in the *Burkholderia mallei* genome. *Proc. Natl. Acad. Sci. USA* **101**, 14246–14251 (2004).
- Holden, M.T. *et al.* Genomic plasticity of the causative agent of melioidosis, *Burkholderia pseudomallei*. *Proc. Natl. Acad. Sci. USA* **101**, 14240–14245 (2004).
- Cramm, R., Pohlmann, A. & Friedrich, B. Purification and characterization of the single-component nitric oxide reductase from *Ralstonia eutropha* H16. *FEBS Lett.* **460**, 6–10 (1999).
- Pohlmann, A., Cramm, R., Schmelz, K. & Friedrich, B. A novel NO-responding regulator controls the reduction of nitric oxide in *Ralstonia eutropha*. *Mol. Microbiol.* **38**, 626–638 (2000).
- Siedow, A., Cramm, R., Siddiqui, R.A. & Friedrich, B. A megaplasmid-borne anaerobic ribonucleotide reductase in *Alcaligenes eutrophus* H16. *J. Bacteriol.* **181**, 4919–4928 (1999).
- Brämer, C.O. & Steinbüchel, A. The methylcitric acid pathway in *Ralstonia eutropha*: new genes identified involved in propionate metabolism. *Microbiology* **147**, 2203–2214 (2001).
- Antoine, R. *et al.* Overrepresentation of a gene family encoding extracytoplasmic solute receptors in *Bordetella*. *J. Bacteriol.* **185**, 1470–1474 (2003).
- Antoine, R. *et al.* The periplasmic binding protein of a tripartite tricarboxylate transporter is involved in signal transduction. *J. Mol. Biol.* **351**, 799–809 (2005).
- Schlegel, H.G., Gottschalk, G. & Von Bartha, R. Formation and utilization of poly-β-hydroxybutyric acid by Knallgas bacteria (*Hydrogenomonas*). *Nature* **191**, 463–465 (1961).
- Slater, S. *et al.* Multiple β-ketothiolases mediate poly(β-hydroxyalkanoate) copolymer synthesis in *Ralstonia eutropha*. *J. Bacteriol.* **180**, 1979–1987 (1998).
- Pötter, M. *et al.* The complex structure of polyhydroxybutyrate (PHB) granules: four orthologous and paralogous phasins occur in *Ralstonia eutropha*. *Microbiology* **150**, 2301–2311 (2004).
- York, G.M. *et al.* *Ralstonia eutropha* H16 encodes two and possibly three intracellular Poly(D-)-3-hydroxybutyrate) depolymerase genes. *J. Bacteriol.* **185**, 3788–3794 (2003).
- Saegusa, H., Shiraki, M. & Saito, T. Cloning of an intracellular D(-)-3-hydroxybutyrate-oligomer hydrolase gene from *Ralstonia eutropha* H16 and identification of the active site serine residue by site-directed mutagenesis. *J. Biosci. Bioeng.* **94**, 106–112 (2002).
- Johnson, B.F. & Stanier, R.Y. Dissimilation of aromatic compounds by *Alcaligenes eutrophus*. *J. Bacteriol.* **107**, 468–475 (1971).
- Bowen, D. *et al.* Insecticidal toxins from the bacterium *Photorhabdus luminescens*. *Science* **280**, 2129–2132 (1998).
- Pemberton, J.M., Corney, B. & Don, R.H. Evolution and spread of pesticide-degrading ability among soil microorganisms. in *Plasmids of Medical, Environmental and Commercial Importance* (eds. Timmis, K.N. & Pühler, A.) 287–299 (Elsevier/North-Holland Biomedical Press, 1979).
- Ren, Q. & Paulsen, I.T. Comparative analyses of fundamental differences in membrane transport capabilities in prokaryotes and eukaryotes. *PLoS Comput. Biol.*, Published online 19 August 2005 (doi:10.1371/journal.pcbi.0010027).
- Lütke-Eversloh, T. *et al.* Biosynthesis of polythioesters by engineered *Escherichia coli* as novel thermoplastic materials. *Nat. Mater.* **1**, 236–240 (2002).
- Kim do, Y., Lütke-Eversloh, T., Elbanna, K., Thakor, N. & Steinbüchel, A. Poly(3-mercaptopropionate): a nonbiodegradable biopolymer? *Biomacromolecules* **6**, 897–901 (2005).
- Staden, R. The Staden sequence analysis package. *Mol. Biotechnol.* **5**, 233–241 (1996).
- Tech, M. & Merkl, R. YACOP: Enhanced gene prediction obtained by a combination of existing methods. *In Silico Biol.* **3**, 441–451 (2003).
- Schiex, T., Gouzy, J., Moisan, A. & de Oliveira, Y. Framed: A flexible program for quality check and gene prediction in prokaryotic genomes and noisy matured eukaryotic sequences. *Nucleic Acids Res.* **31**, 3738–3741 (2003).
- Overbeek, R. *et al.* The ERGO genome analysis and discovery system. *Nucleic Acids Res.* **31**, 164–171 (2003).
- Overbeek, R. *et al.* The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.* **33**, 5691–5702 (2005).
- Ghai, R., Hain, T. & Chakraborty, T. GenomeViz: visualizing microbial genomes. *BMC Bioinformatics* **5**, 198 (2004).

### Erratum: Live dynamic imaging of caveolae pumping targeted antibody rapidly and specifically across endothelium in the lung

Phil Oh, Per Borgström, Halina Witkiewicz, Yan Li, Bengt J Borgström, Adrian Chrastina, Koji Iwata, Kurt R Zinn, Richard Baldwin, Jacqueline E Testa & Jan E Schnitzer

*Nature Biotechnology* 25, 327–337 (2007); published online 4 March 2007; corrected after print 22 March 2007.

In the version of this article initially published, the phrase “through its caveolae” at the end of the second complete paragraph, on p. 335, right column, should have been deleted. The error has been corrected in the HTML and PDF versions of the article.

---

### Erratum: Incretin mimetics vie for slice of type 2 diabetes market

Catherine Shaffer

*Nature Biotechnology* 25, 263 (2007); published online 1 March 2007

In the version of this article initially published, the sentence on p. 263, column 2, at the start of the first full paragraph that begins, “Novo’s Galvus (vildagliptin)...” is incorrect. Galvus was developed by Novartis.

---

### Erratum: Editorial. Bioethanol needs biotech now

*Nature Biotechnology* 24, 725 (2006); published online 6 July 2006

In the version of this article originally published, on p. 725, column 2, paragraph 3, line 5, “atrazine” is not a fertilizer as stated, but an herbicide. The fertilizer cited should have been “anhydrous ammonia.”

---

### Corrigendum: Complete genome of the mutualistic, N<sub>2</sub>-fixing grass endophyte *Azoarcus* sp. strain BH72

Andrea Krause, Adarsh Ramakumar, Daniela Bartels, Federico Battistoni, Thomas Bekel, Jens Boch, Melanie Böhm, Frauke Friedrich, Thomas Hurek, Lutz Krause, Burkhard Linke, Alice C McHardy, Abhijit Sarkar, Susanne Schneiker, Arshad Ali Syed, Rudolf Thauer, Frank-Jörg Vorhölter, Stefan Weidner, Alfred Pühler, Barbara Reinhold-Hurek, Olaf Kaiser & Alexander Goesmann

*Nature Biotechnology* 24, 1385–1391 (2006); published online 22 October 2006; corrected after print 26 March 2007

In the version of this article initially published, the gene name *nosRZDYFLX* on p. 1388, left column, line 5, is incorrect; the correct name is *nosZRDFYL*. The error has been corrected in the PDF version of the article.

---

### Corrigendum: Genome sequence of the bioplastic-producing “Knallgas” bacterium *Ralstonia eutropha* H16

Anne Pohlmann, Wolfgang Florian Fricke, Frank Reinecke, Bernhard Kusian, Heiko Liesegang, Rainer Cramm, Thomas Eitingier, Christian Ewering, Markus Pötter, Edward Schwartz, Axel Strittmatter, Ingo Voß, Gerhard Gottschalk, Alexander Steinbüchel, Bärbel Friedrich & Botho Bowien

*Nature Biotechnology* 24, 1257–1262 (2006) Published online: 10 September 2006; corrected after print 30 March 2007

In the version of the article originally published, it was stated on p. 1259, column 2, line 14 of the first full paragraph, that “cytochrome *cd<sub>1</sub>* nitrite reductase and nitrous oxide reductase are encoded only on pHG1.” The sentence should have read, “cytochrome *cd<sub>1</sub>* nitrite reductase and nitrous oxide reductase are encoded only on chromosome 2 and megaplasmid pHG1, respectively.” In addition, there are mistakes in **Figure 2** in the structural formula for Poly[R-(–)-3-hydroxybutyrate] and the trimeric degradation product. In the former, the hydroxyl group should be a methyl group; in the latter, the hydroxyl groups at C-3 and C-6 should be methyl groups. The error has been corrected in the PDF version of the article.

---

### Corrigendum: Evolving symbiosis—venture capital and biotechnology

Terry C Bradford

*Nature Biotechnology* 21, 983–984 (2003)

In the version initially published, citations for the following reference should have made at the penultimate sentence in paragraph 3 and the last sentence of paragraph 5: Stephen P. Galante. An Overview of the Venture Capital Industry & Emerging Changes, 21<sup>st</sup> Annual Venture Capital Institute, September 17–21, 1995. <<http://www.vcinstitute.org/materials/galante.html>>

---