Computational identification of BioR, a transcriptional regulator of biotin metabolism in *Alphaproteobacteria*, and of its binding signal

Dmitry A. Rodionov[1] & Mikhail S. Gelfand[1,2,3]

[1]Institute for Information Transmission Problems RAS, Moscow, Russia; [2]State Scientific Center GosNIIGenetika, Moscow, Russia; and [3]Department of Bioengineering and Bioinformatics, Moscow State University, Moscow, Russia

Correspondence: Dmitry A. Rodionov, The Burnham Institute, 10901 N. Torrey Pines Rd., La Jolla, CA 92037. Tel.: +1 858 646 3100 ext. 3082; fax: +1 858 713 9949; e-mail: rodionov@burnham.org

Received 3 September 2005; revised 14 November 2005; accepted 15 November 2005.

doi:10.1111/j.1574-6968.2005.00070.x

Editor: Michael Galperin

Keywords
biotin regulation; *Alphaproteobacteria*; BioR.

Abstract

Comparative genomic analysis was applied to identify the biotin transcriptional regulator, BioR, in most *Alphaproteobacteria*, and to identify its recognition signal TTATMKATAA. BioR belongs to the GntR family of transcriptional repressors. The functional assignment is supported by three lines of evidence: (1) *bioR* is positionally clustered with various *bio* genes, both for biotin biosynthesis and transport; (2) in most cases, candidate BioR-binding sites (BIOR boxes) are observed upstream of the *bioR* genes, suggesting autoregulation; (3) the phyletic distribution of the BIOR boxes coincides exactly with the phyletic distribution of the *bioR* genes, as the genomes lacking BIOR boxes do not have orthologs of *bioR*. Thus, in *Alphaproteobacteria*, BioR seems to have assumed the role of the biotin regulator that in most other bacteria is fulfilled by the dual function biotin–protein ligase BirA having the DNA-binding helix-turn-helix domain.

Introduction

Biotin (vitamin H) is an essential cofactor for a class of metabolic enzymes catalyzing various carboxylation reactions in fatty acid biosynthesis and the central intermediate metabolic pathways (Samols *et al*., 1998; Perkins & Pero, 2001; Dunn *et al*., 2002). The biotin biosynthesis pathway is widespread among microorganisms. Enzymes forming the pathway from pimeloyl-CoA to biotin are encoded by universal genes *bioF, bioA, bioD* and *bioB*, whereas the pimeloyl-CoA synthesis is more diverse and involves both experimentally determined and computationally predicted genes *bioW, bioI, bioC, bioG, bioH, bioK* and *bioZ* in a variety of bacterial species (Ifuku *et al*., 1994; Lemoine *et al*., 1996; Stok & De Voss, 2000; Rodionov *et al*., 2002). Although active biotin uptake was demonstrated experimentally (Eisenberg, 1985; Piffeteau & Gaudry, 1985), for about 20 years genes encoding biotin transporters remained unknown (Heinz *et al*., 1999; Koonin & Galperin, 2003). Recently, computational identification of the *bioY* gene family predicted to encode biotin transporters has been achieved (Rodionov *et al*., 2002). At the same time, the *bioY* locus has been found to influence biotin uptake in *Sinorhizobium meliloti* (Entcheva *et al*., 2002). These findings were followed by experimental validation of a role of BioY in biotin transport in *Rhizobium etli* (Guillen-Navarro *et al*., 2005a).

In many bacteria, the biotin metabolism and transport genes are regulated by the bifunctional protein, BirA (Cronan, 1989; Bower *et al*., 1995). The enzymatic domain of BirA constitutes the biotin–protein ligase mediating biotinylation of an apo-biotin-dependent carboxylase. The BirA proteins having an additional N-terminal DNA-binding helix-turn-helix domain act as transcriptional repressors of the *bio* genes in biotin-replete conditions (Kwon *et al*., 2000). The BirA-binding signal is conserved, with minor variations, in bacteria belonging to the *Bacillus/Clostridium* group, many proteobacteria, *Thermus thermophilus*, *Chlorobium tepidum*, as well as some archaea (Rodionov *et al*., 2002; Rodionov *et al*., 2004).

BirA from all *Alphaproteobacteria* with sequenced genomes lack the DNA-binding domain (Rodionov *et al*., 2002; Guillen-Navarro *et al*., 2005b), and the mode of biotin-dependent regulation in most of these species is not known. However, it is known that biotin influences expression of several genes in *Sinorhizobium meliloti* (Heinz & Streit, 2003; Guillen-Navarro *et al*., 2005b).

This gap is partially filled in this study, where we have applied comparative genomic analysis to identify the transcriptional regulator of the *bio* genes and its binding signal in eight out of 19 species of *Alphaproteobacteria* with available genomic sequences.

Materials and methods

Complete genomes of *Agrobacterium tumefaciens* (AT), *Bartonella quintana* (BQ), *Bradyrhizobium japonicum* (BJ), *Brucella melitensis* (BME), *Caulobacter crescentus CB15* (CC),

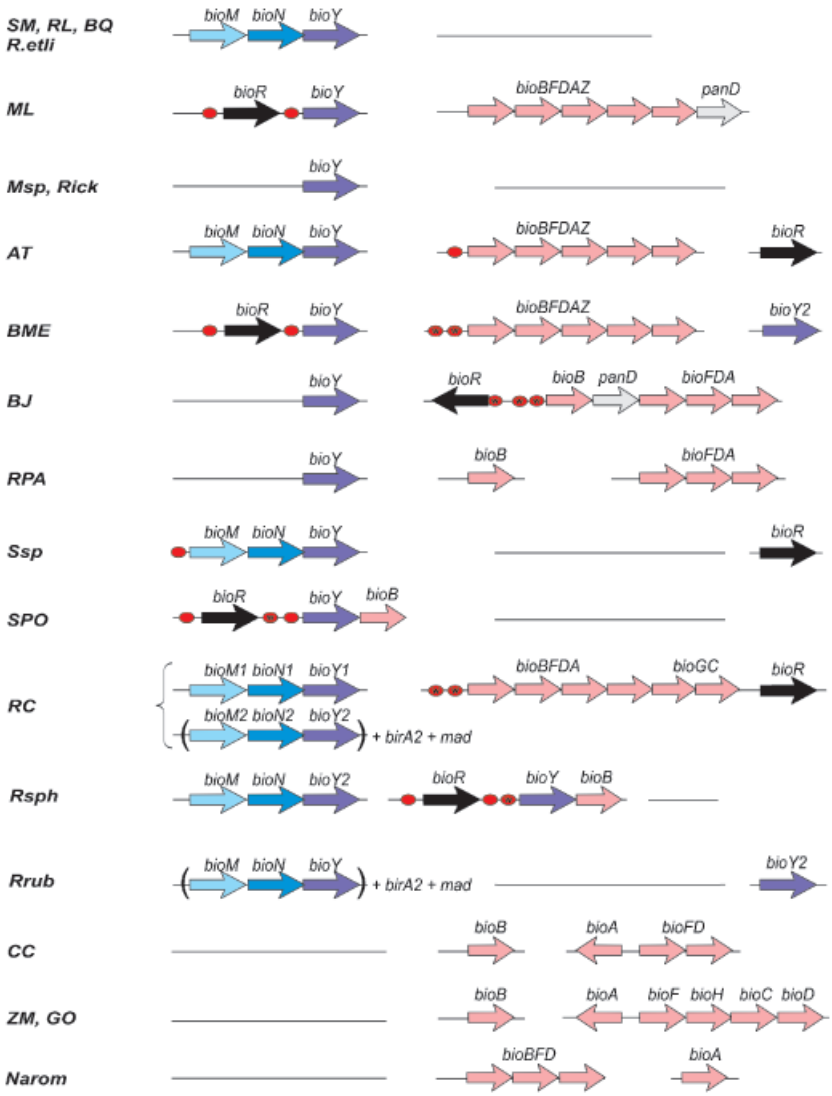

Fig. 1. Organization of bio loci in *Alphaproteobacteria*. Pink arrows: biotin biosynthesis genes. Blue arrows: biotin transport genes. Black arrows: *bioR* genes. Red circles: candidate BIOR boxes (BioR-binding sites). Red circles with 'w' inside: weaker sites with minor deviations from the consensus. For genome abbreviations, see Materials and methods.

Gluconobacter oxydans (GO), *Mesorhizobium loti* (ML), *Rhodopseudomonas palustris* (RPA), four *Rickettsia* species (Rick), *Silicibacter pomeroyi* (SPO), *Sinorhizobium meliloti* (SM) and *Zymomonas mobilis* (ZM), as well as unfinished annotated genomic sequences of *Mesorhizobium* sp. (Msp), *Novosphingobium aromaticivorans* (Narom), *Rhodobacter sphaeroides* (Rsph), *Rhodospirillum rubrum* (Rrub) and *Silicibacter* sp. TM1040 (Ssp) were downloaded from GenBank (Benson *et al.*, 2005). Incomplete genomes of *Rhizobium leguminosaurum* (RL) and *Rhodobacter capsulatus* (RC) were downloaded from the websites of the Wellcome Trust Sanger Institute (http://www.sanger.ac.uk) and Integrated Genomics Inc. (http://www.integratedgenomics.com), respectively.

Similarity search was performed using basic local alignment search tool (McGinnis & Madden, 2004). The phylogenetic tree was constructed by the maximum likelihood method implemented in PHYLIP (Felsenstein, 1981) using multiple sequence alignment of proteins produced by ClustalX (Thompson *et al.*, 1997). Comparative genomic analysis was performed using GenomeExplorer (Mironov *et al.*, 2000) and SEED (Overbeek *et al.*, 2005) (http://theseed.uchicago.edu/FIG/index.cgi, see the Biotin biosynthesis subsystem). Identification of DNA signals and construction of positional weight matrix for identification of candidate sites was achieved using SignalX (Gelfand *et al.*, 2000). Helix-turn-helix DNA-binding motifs were analyzed using the weight matrix method (Dodd & Egan, 1990) (http://npsa-pbil.ibcp.fr). The PFAM database was used to verify protein functional annotations (Bateman *et al.*, 2002).

Results and discussion

Orthologs of known biotin biosynthesis (*bioA*, *bioB*, *bioD*, *bioF*) and transport (*bioY*) genes in available genomes of

(a)
```
ML:mlr7428          MAKGT--------DDTIAVRISKVLADRIISGAIEPGARLRQDHVAEEFNTSHVPVREAFR
BME:BMEI0320        MNQNVPASRPAPRSETIAARISRILADRIIAGEIEPGTKLRQDHIAEEFQTSHVPVREAFR
AT:AGR_C_1447       MPKPEKTTP----EDTIAARISRTLAERIVRGELSPGERLRQDHIAAEFGASHVPVREAFR
Rsph:Rsph03000991   MQHSP--------EETQSARIARILADRIISGEIPPGARLRQDHVAIEFGASHVPVREAFR
BJ:bll2094          MITAD--------NMTMVARIADSIAERIISGALAPGAPLRQDHVAREFNSSHVPVREAFR
Ssp:STM1w01002633   MGGFAPYRT--FMSETAVDKIFRALMERIVTGDLPAGEKLRQDHIARDYDVSHVPVREAFL
SPO:SPO3340         ------------MTATIADTIYARLSQQIITGALSAGEKLRQDHIARAFDTSHVPVREALL
RC:008_0866         MPVSP-----------TPQLIADEIARRIIRGELSAGERLRQDYVAAEFGVSHVPVREALL
                                     *   :  :*: * : .*  ****::*  :  ********:

ML:mlr7428          RLEAQGLAVSEPRRGVRVAAFDLGEVKEVAEMRAALEVLALRHAAPHLTSAILDLAEEATK
BME:BMEI0320        RLEAQGLAVSEPRRGVRVASFDIGEIREVAEMRAALEVLALRHAAPHITRAVLDAAEQATL
AT:AGR_C_1447       RLEAQGLAASIPRRGVRVADFGLDDVREVAEMRAALEVLALRHAIPNMTPAILDQAEAATL
Rsph:Rsph03000991   ELQTRGLAESEPRRGFRVTEFDVAELREVAEMRSSLESLALRHAAPNLTRAILLQAEEITR
BJ:bll2094          HLQAQHLVVAVPRRGVRVAPLDTRSVKEIAEMRAALEVVALRHSGPRLTTAHLAQIELALI
Ssp:STM1w01002633   RLEAKGLAKSIPHKGMRVTALEPTEIREVIEMRVALEVLALTHALPRHDADTLAQVEAARQ
SPO:SPO3340         RLEAHGLAQSEPRRGMRVTALDPAEIREVIEMRVALEPLALQHAVQNATPGDLQAIDVARL
RC:008_0866         KLAARGLVVSQHNKGVQVAPLDPLAQRELKLMRLALEPLTLLHSVPHLTPAQVREADRLRE
                    .* :: *. :  .:*.:*: :     :*:  ** :** ::* *:  .     :   :

ML:mlr7428          AGDKSRDVRSWEEANRAFHRLILAPCGMPRLLATIDDLHAASARFLFAAWR-SEWEARTDQ
BME:BMEI0320        EGDKSRDVRSWEDANRRFHRLILTPCKMPRLLAAIDDLHAASARFLFATWR-SAWEARTDH
AT:AGR_C_1447       EGDSAVDVQHWEEANRRFHRLITAPCNMPRLMEAIDGLHAASARFLFSAWR-AGWERRTDT
Rsph:Rsph03000991   RGDSASTVREWEEANRHFHRLILSPCRMPRLLRTIDDLQAASARFLFAAWR-RDWEARTDH
BJ:bll2094          EGDNAKTIEEFEMANRAFHQALVAPCGMPRLLASLDGLQLANSRLVFAMARSVGWRPRSNQ
Ssp:STM1w01002633   ACDAAQDMVLWDRCNRQFHTALLQPCEMPRLLSAIEDLHITAARHLFVHWR-DRWTSRIDS
SPO:SPO3340         ACDGAEDMPTWEARNRAFHRAILTPCAMPRLLSAIDDLHIASSRHLHAHWR-QRWRQRTDT
RC:008_0866         ACDQAPDIYTWEEMNRAFHMQIYEACEMPRLLETVENMQNLAARYILVHYR-SRWRPRIDT
                      * :  :  ::  ** **  :  .* ****: :::.::   :* :    *   *  * :

ML:mlr7428          DHRAILTALRQGNTESAAATLGRHVQWIGRKPVRTASGATREAFAIVG
BME:BMEI0320        DHRAILAALRQNDIESAATILARHVQWIGHRPVKTASGKVRDSFAIVG
AT:AGR_C_1447       DHRAILDALRGGRAEEAVRILDGHVQWIGRKAIRTSSGSVRDAFAIVG
Rsph:Rsph03000991   DHRAILDALRKGQTDLACSTLARHVGWVGRRRPSANNGDRPGTYDIHD
BJ:bll2094          DHRLILQALRTRNVDHACNLLTRHIQTIERLALPAS------------
Ssp:STM1w01002633   DHDALLEAVTAQDLPRACDILRRHLRRVS-------------------
SPO:SPO3340         DHAAILMAMKRRDAETACDVLRRHLRRAL-------------------
RC:008_0866         DHHGIIAAIRRRDAKGAAAILERHLQRLSQVFDPR-------------
                    **  :: *:       *   *  *:
```

Fig. 2. (a) Multiple alignment of the BioR proteins. The predicted helix-turn-helix DNA-binding motifs are underlined. (b) Maximum likelihood phylogenetic tree of the GntR family of regulators. The BioR branch for predicted biotin regulators in *Alphaproteobacteria* is marked by background gray. All GntR-type proteins from Swiss-Prot (http://www.expasy.ch/sprot/) were used to construct the tree. For genome abbreviations, see Materials and methods.

Alphaproteobacteria were identified by a similarity search (Fig. 1). An initial attempt to identify a common pattern in upstream regions of all *bio* genes of *Alphaproteobacteria* failed. However, we noticed that in most genomes the biotin biosynthesis genes are colocalized with a gene encoding a GntR-type transcription factor, tentatively named *bioR* (Fig. 1). The alignment of the BioR proteins demonstrated high conservation of the helix-turn-helix domain (Fig. 2a), making it likely that the binding signal is also conserved. This DNA-binding domain is representative of the GntR family of DNA-binding domains (PFAM database accession number PF00392), a family that includes transcription factors regulating biosynthesis of important bacterial metabolites (Haydon & Guest, 1991). The BioR proteins identified in *Alphaproteobacteria* form a monophyletic branch on the phylogenetic tree of the GntR family (Fig. 2b).

Upstream regions of the biotin biosynthesis genes from these genomes were collected and found to contain sites conforming to the palindromic consensus (BIOR box) TTATMKATAA, where M is C or A and K is T or G (Table 1). A recognition profile (positional weight matrix) was constructed and used to scan all genomes, resulting in identification of additional candidate sites upstream of the *bioR* genes (Table 1, Fig. 1). In some cases, two BIOR boxes occur in tandem, and then one or both sites may deviate slightly from the consensus. No candidate sites were observed upstream of genes unrelated to biotin metabolism or in genomes lacking *bioR* (Table 2).

Thus, three lines of evidence support identification of BioR as the transcriptional regulator of the *bio* genes in *Alphaproteobacteria*. Firstly, *bioR* is positionally clustered with various *bio* genes, both for biotin biosynthesis and transport, forming divergons (in *Bradyrhizobium japonicum*) or candidate operons (in *M. loti, Brucella melitensis, Silicibacter pomeroyi, Rhodobacter capsulatus, Rhodobacter sphaeroides*); such clustering with transcription factors and

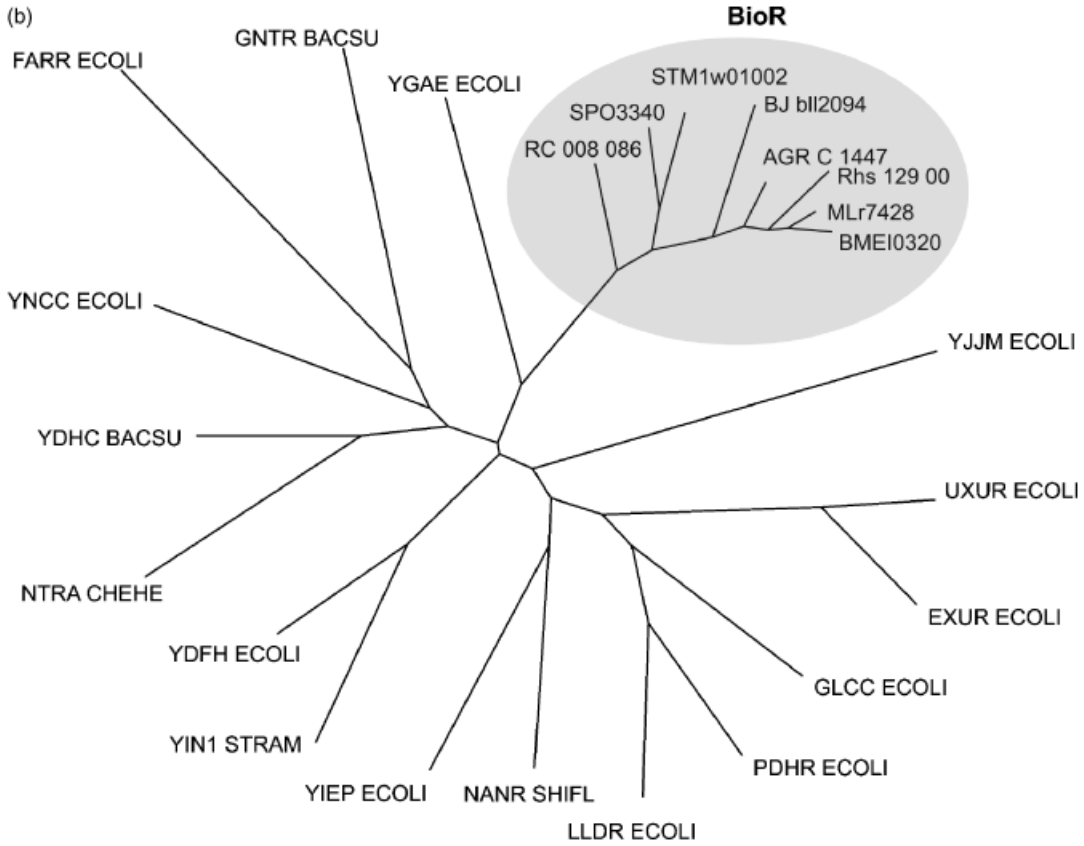

Fig. 2. Continued

Table 1. Candidate BIOR boxes (BioR-binding sites)

Genome	Gene or operon	Position	Score	Site
Ssp	*STM1w01002019 (bioMNY)*	– 93	4.66	**TTATAGATAA**
SPO	*SPO3339 (bioYB)*	– 22	4.66	**TTATCTATAA**
		– 63	4.30	TTATAGATAg
SPO	*SPO3340 (bioR)*	– 12	4.66	TTATCTATAA
Rsph	*Rsph03000990 (bioYB)*	– 125	4.66	**TTATCTATAA**
		– 81	4.30	TTATAGATAg
Rsph	*Rsph03000991 (bioR)*	– 11	4.66	TTATCTATAA
RC	*008_0857 (bioBFDA)*	– 181	3.72	TcATATATtA
		– 74	3.83	**TcATAGATAg**
BME	*BMEI0319 (bioY)*	– 21	4.66	**TTATCTATAA**
BME	*BMEI0320 (bioR)*	– 13	4.66	TTATCTATAA
BME	*BMEII0775 (bioBFDAZ)*	– 93	4.19	TTATCTATtA
		– 23	4.18	**TTATCTAcAA**
AT	*AGR_L_1708 (bioBFDAZ)*	86	4.66	**TTATCTATAA**
ML	*mlr7428 (bioR)*	– 14	4.66	TTATCTATAA
ML	*mlr7429 (bioY)*	135	4.66	**TTATCTATAA**
BJ	*bll2094 (bioR)*	– 119	3.72	TcATAGATtA
BJ	*bioB-panD-bioFDA*	– 47	4.30	**cTATAGATAA**
		– 28	4.18	TTATCTAcAA

Position is given relative to the start codon. Score is computed using the recognition profile (position weight matrix). The sites used to construct the profile are shown in bold. For genome abbreviations see Materials and methods.

the genes they regulate has been observed in many other cases (Doerks *et al*., 2004). Secondly, in most cases (*M. loti, Brucella melitensis, Bradyrhizobium japonicum, Silicibacter pomeroyi*, *Rhodobacter sphaeroides*) candidate BIOR boxes were observed upstream of the *bioR* genes, suggesting autoregulation. Thirdly, the phyletic distribution of the BIOR boxes coincides exactly with the phyletic distribution of the *bioR* genes, as the genomes lacking BIOR boxes do not have orthologs of *bioR*; cf. identification of NrdR as the repressor of ribonucleotide reductase genes based on the same type of reasoning (Borovok *et al*., 2004; Rodionov & Gelfand, 2005).

From the genomic data available at the moment, the occurrence of the BioR regulon is restricted to two lineages of *Alphaproteobacteria*, *Rhizobiales* and *Rhodobacterales*. The latter lineage is represented by two *Rhodobacter* and two *Silicibacter* species, all of which possess BioR. In contrast, among nine *Rhizobiales* genomes, only four appear to have the novel biotin regulon, whereas the mode of regulation of the biotin uptake operons in other species including *Sinorhizobium meliloti* and *Rhizobium leguminosaurum* remains unknown. The biotin repression of the *bioMNY* operon in *Rhizobium etli*, a close relative of *Sinorhizobium meliloti*, was observed recently (Guillen-Navarro, 2005a); however, the

Table 2. Distribution of predicted BioR sites in bacterial genomes

	Strong BioR sites, cut-off = 4.65				Pairs of weak BioR sites, cut-off = 3.70		
Genome	Total number of sites	Candidate sites in intergenic regions	False sites in intergenic regions	False sites in the coding regions	Total number of site pairs	Candidate site pairs in intergenic regions	False site pairs in intergenic regions
BME	5	2	2	1	6	1	5
AT	6	1	3	2	3	0	3
ML	9	2	6	1	2	0	2
BJ	4	0	1	3	2	1	1
Ssp	1	1	0	0	0	0	0
SPO	2	2	0	0	1	1	0
Rsph	2	2	0	0	1	1	0
RC	1	0	0	1	1	1	0

To account for both single strong BioR sites and pairs of two weak BioR sites, we analyzed the distributions of sites in the genomes separately for two different cut-off values. Cut-off 4.66 selects strong BioR sites without deviations from the consensus sequence. For pairs of weak BioR sites, the maximal distance of 110 base pairs between the sites was used.

regulatory region of this operon lacks a BIOR box. Furthermore, the complete genome of *Rhizobium etli* encodes no BioR ortholog (http://www.cifn.unam.mx/retlidb/). The BioR regulon was not detected in other lineages of *Alphaproteobacteria* (*C. crescentus, Z. mobilis*, *N. aromaticivorans, G. oxydans*, *Rhodospirillum rubrum* and *Rickettsia*) that also lack a DNA-binding repressor domain in their BirA proteins. This suggests that BioR evolved to substitute for the loss of the regulatory function of BirA rather late in the history of *Alphaproteobacteria*, probably in the last common ancestor of *Rhizobiales* and *Rhodobacterales*.

One possibility is that the *bio* genes in the rhizobia genomes without BioR are not regulated at all. Indeed, even in the genomes with BirA- or BioR-dependent regulation, some *bio* genes encoding both enzymes and transporters remain nonregulated (Fig. 1) (Rodionov *et al*., 2002). Only four genomes in this study (*A. tumefaciens*, *Bradyrhizobium japonicum, Brucella melitensis* and *Rhodobacter capsulatus*) have the entire biotin biosynthetic pathway under predicted control of BioR. This shows that biotin-dependent regulation may not be essential. Several bacteria with the BioR-regulated biotin transporter are most likely biotin auxotrophic species as they lack the entire biotin synthesis pathway. However, the presence of the biotin synthase gene *bioB* in a cluster with *bioY* in two of them (*Silicibacter pomeroyi* and *Rhodobacter sphaeroides*) is of special interest, and may indicate that these species could also utilize dethiobiotin, a biotin precursor.

Another possibility is the existence of a different regulatory mechanism for the control of the *bio* genes in some rhizobia. In *Sinorhizobium meliloti*, a regulatory role has been ascribed to the biotin-induced gene *bioS* based on claimed similarity to the LysR-family regulators and on observed positive autoregulation by interaction with its own promoter region (Heinz *et al*., 1999). Plant-derived biotin is an important factor that stimulates the growth of *Sinorhizobium meliloti* in the rhizosphere (Streit *et al*., 1996). Other target genes regulated by BioS are not yet known, although a mutation in *bioS* increases biotin uptake and extends the stationary phase in the presence of biotin, but has no influence on the survival under biotin limitation (Heinz *et al*., 1999; Heinz & Streit, 2003). On the other hand, *bioS* gene does not have orthologs in other *Alphaproteobacteria*, and thus its possible regulatory role is in any case confined to *S. inorhizobium meliloti* and its close relatives (such as *Sinorhizobium fredii*) (Heinz *et al*., 1999).

Acknowledgements

The authors are grateful to Thomas Eitinger for helpful discussions and to the *Rhizobium etli* database team for sharing genomic data prior to publication. This study was partially supported by grants from the Howard Hughes Medical Institute (55000309) and Russian Academy of Sciences (programs 'Molecular and Cellular Biology' and 'Origin and Evolution of the Biosphere') to M. G. and the Russian Fund of Basic Research (04–04–49361) to D. R., and M. G. was also supported by the Russian Science Support Fund.

References

Bateman A, Birney E, Cerruti L, *et al.* (2002) The Pfam protein families database. *Nucleic Acids Res* **30**: 276–280.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J & Wheeler DL (2005) GenBank. *Nucleic Acids Res* **33**: D34–D38.

Borovok I, Gorovitz B, Yanku M, *et al.* (2004) Alternative oxygen-dependent and oxygen-independent ribonucleotide reductases in *Streptomyces*: cross-regulation and physiological role in response to oxygen limitation. *Mol Microbiol* **54**: 1022–1035.

Bower S, Perkins J, Yocum RR, *et al*. (1995) Cloning and characterization of the *Bacillus subtilis birA* gene encoding a repressor of the biotin operon. *J Bacteriol* **177**: 2572–2575.

Cronan JE Jr (1989) The *E. coli bio* operon: transcriptional repression by an essential protein modification enzyme. *Cell* **58**: 427–429.

Dodd IB & Egan JB (1990) Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. *Nucleic Acids Res* **18**: 5019–5026.

Doerks T, Andrade MA, Lathe W III, vonMering C & Bork P (2004) Global analysis of bacterial transcription factors to predict cellular target processes. *Trends Genet* **20**: 126–131.

Dunn MF, Araíza G, Encarnación S, Finan TM & Mora J (2002) Characteristics and metabolic roles of biotin-dependent enzymes in rhizobia. *Nitrogen Fixation: Global Perspectives* (Finan T, O'Bran MR, Layzell DB, Vessey JK & Newton W, eds), pp. 158–162. CABI, Wallingford, Oxon, UK.

Eisenberg MA (1985) Regulation of the biotin operon in *E. coli*. *Ann N Y Acad Sci* **447**: 335–349.

Entcheva P, Phillips DA & Streit WR (2002) Functional analysis of *Sinorhizobium meliloti* genes involved in biotin synthesis and transport. *Appl Environ Microbiol* **68**: 2843–2848.

Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* **17**: 368–376.

Gelfand MS, Koonin EV & Mironov AA (2000) Prediction of transcription regulatory sites in Archaea by a comparative genomic approach. *Nucleic Acids Res* **28**: 695–705.

Guillen-Navarro K, Araiza G, Garcia-de Los Santos A, Mora Y & Dunn MF (2005a) The *Rhizobium etli bioMNY* operon is involved in biotin transport. *FEMS Microbiol Lett* **250**: 209–219.

Guillen-Navarro K, Encarnacion S & Dunn MF (2005b) Biotin biosynthesis, transport and utilization in rhizobia. *FEMS Microbiol Lett* **246**: 159–165.

Haydon DJ & Guest JR (1991) A new family of bacterial regulatory proteins. *FEMS Microbiol Lett* **79**: 291–296.

Heinz EB, Phillips DA & Streit WR (1999) BioS, a biotin-induced, stationary-phase, and possible LysR-type regulator in *Sinorhizobium meliloti*. *Mol Plant Microbe Interact* **12**: 803–812.

Heinz EB & Streit WR (2003) Biotin limitation in *Sinorhizobium meliloti* strain 1021 alters transcription and translation. *Appl Environ Microbiol* **69**: 1206–1213.

Ifuku O, Miyaoka H, Koga N, Kishimoto J, Haze S, Wachi Y & Kajiwara M (1994) Origin of carbon atoms of biotin. 13C-NMR studies on biotin biosynthesis in *Escherichia coli*. *Eur J Biochem* **220**: 585–591.

Koonin EV & Galperin MY (2003) *Sequence–Evolution–Function: Computational Approaches in Comparative Genomics*, Kluwer Academic Publishers, Dordrecht, the Netherlands.

Kwon K, Streaker ED, Ruparelia S & Beckett D (2000) Multiple disordered loops function in corepressor-induced dimerization of the biotin repressor. *J Mol Biol* **304**: 821–833.

Lemoine Y, Wach A & Jeltsch JM (1996) To be free or not: the fate of pimelate in *Bacillus sphaericus* and in *Escherichia coli*. *Mol Microbiol* **19**: 645–647.

McGinnis S & Madden TL (2004) BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res* **32**: W20–W25.

Mironov AA, Vinokurova NP & Gelfand MS (2000) GenomeExplorer: software for analysis of complete bacterial genomes. *Mol Biol (Moscow)* **34**: 222–231.

Overbeek R, Begley T, Butler RM, *et al*. (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* **33**: in press.

Perkins JB & Pero JG (2001) Vitamin biosynthesis. *Bacillus subtilis and its Relatives: From Genes to Cells* (Sonenshein AL, Hoch JA & Losick R, eds), pp. 279–293. American Society for Microbiology, Washington, DC.

Piffeteau A & Gaudry M (1985) Biotin uptake: influx, efflux and countertransport in *Escherichia coli* K12. *Biochim Biophys Acta* **816**: 77–82.

Rodionov DA & Gelfand MS (2005) Identification of a bacterial regulatory system for ribonucleotide reductases by phylogenetic profiling. *Trends Genet* **21**: 385–389.

Rodionov DA, Mironov AA & Gelfand MS (2002) Conservation of the biotin regulon and the BirA regulatory signal in Eubacteria and Archaea. *Genome Res* **12**: 1507–1516.

Rodionov DA, Dubchak I, Arkin A, Alm E & Gelfand MS (2004) Reconstruction of regulatory and metabolic pathways in metal-reducing δ-proteobacteria. *Genome Biol* **5**: R90.

Samols D, Thornton CG, Murtif VL, Kumar GK, Haase FC & Wood HG (1998) Evolutionary conservation among biotin enzymes. *J Biol Chem* **263**: 6461–6464.

Stok JE & De Voss J (2000) Expression, purification, and characterization of BioI: a carbon-carbon bond cleaving cytochrome P450 involved in biotin biosynthesis in *Bacillus subtilis*. *Arch Biochem Biophys* **384**: 351–360.

Streit WR, Joseph CM & Phillips DA (1996) Biotin and other water-soluble vitamins are key growth factors for alfalfa root colonization by *Rhizobium meliloti* 1021. *Mol Plant Microbe Interact* **9**: 330–338.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F & Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**: 4876–4882.