



HAL
open science

Functional Genomics of Novel Secondary Metabolites from Diverse Cyanobacteria Using Untargeted Metabolomics

Richard Baran, Natalia Ivanova, Nick Jose, Ferran Garcia-Pichel, Nikos Kyrpides, Muriel Gugger, Trent Northen

► **To cite this version:**

Richard Baran, Natalia Ivanova, Nick Jose, Ferran Garcia-Pichel, Nikos Kyrpides, et al.. Functional Genomics of Novel Secondary Metabolites from Diverse Cyanobacteria Using Untargeted Metabolomics. *Marine drugs*, 2013, 11 (10), pp.3617 - 3631. 10.3390/md11103617. pasteur-01388177

HAL Id: pasteur-01388177

<https://pasteur.hal.science/pasteur-01388177>

Submitted on 26 Oct 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Article

Functional Genomics of Novel Secondary Metabolites from Diverse Cyanobacteria Using Untargeted Metabolomics

Richard Baran ¹, Natalia N. Ivanova ², Nick Jose ¹, Ferran Garcia-Pichel ³, Nikos C. Kyrpides ², Muriel Gugger ⁴ and Trent R. Northen ^{1,*}

¹ Life Sciences Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Rd, MS977R0181A, Berkeley, CA 94720, USA; E-Mails: RBaran@lbl.gov (R.B.); njose40707@gmail.com (N.J.)

² DOE Joint Genome Institute, Walnut Creek, CA 94598, USA; E-Mails: nnivanova@lbl.gov (N.N.I.); nckyrpides@lbl.gov (N.C.K.)

³ School of Life Sciences, Arizona State University, Tempe, AZ 85287, USA; E-Mail: ferran@asu.edu

⁴ Institute Pasteur, Collection of Cyanobacteria, Paris Cedex 15 75724, France; E-Mail: muriel.gugger@pasteur.fr

* Author to whom correspondence should be addressed; E-Mail: TRNorthen@lbl.gov; Tel.: +1-510-486-5240; Fax: +1-510-486-4545.

Received: 15 August 2013; in revised form: 21 August 2013 / Accepted: 9 September 2013 /

Published: 30 September 2013

Abstract: Mass spectrometry-based metabolomics has become a powerful tool for the detection of metabolites in complex biological systems and for the identification of novel metabolites. We previously identified a number of unexpected metabolites in the cyanobacterium *Synechococcus* sp. PCC 7002, such as histidine betaine, its derivatives and several unusual oligosaccharides. To test for the presence of these compounds and to assess the diversity of small polar metabolites in other cyanobacteria, we profiled cell extracts of nine strains representing much of the morphological and evolutionary diversification of this phylum. Spectral features in raw metabolite profiles obtained by normal phase liquid chromatography coupled to mass spectrometry (MS) were manually curated so that chemical formulae of metabolites could be assigned. For putative identification, retention times and MS/MS spectra were cross-referenced with those of standards or available spectral library records. Overall, we detected 264 distinct metabolites. These included indeed different betaines, oligosaccharides as well as additional unidentified metabolites with chemical formulae not present in databases of metabolism. Some of these metabolites were detected only in a single strain, but some were present in more than one. Genomic

interrogation of the strains revealed that generally, presence of a given metabolite corresponded well with the presence of its biosynthetic genes, if known. Our results show the potential of combining metabolite profiling and genomics for the identification of novel biosynthetic genes.

Keywords: cyanobacteria; metabolomics; mass spectrometry; MS/MS; betaines; oligosaccharides

1. Introduction

Cyanobacteria are photoautotrophic bacteria capable of oxygenic photosynthesis. Members of the cyanobacteria phylum are important primary producers of organic matter in diverse ecosystems ranging from temperate terrestrial and marine to extreme environments [1–4]. Their broad ecological presence makes cyanobacteria important determinants of global geochemical cycles of carbon and nitrogen [5]. Cyanobacteria also commonly produce diverse secondary metabolites and bioactive compounds [6,7] and the ability of cyanobacteria to utilize solar energy and to fix carbon dioxide has drawn interest for biotechnological applications [8–12].

Improved understanding of cyanobacterial metabolites and their utilization have great potential for drug development. Unfortunately, functional characterization of microbial metabolism in general has lagged behind the pace of sequencing [13] and significant proportions of microbial genes have no assigned function as well as numerous biochemical activities are not assigned to any specific gene [14]. Microbial functional genomics has benefited from advances in comparative genomics [15], availability of bacterial expression [16] and mutant libraries [17], large scale phenotyping [18,19] and genome-scale metabolic modeling [20], all synergistically enabling advances in assigning functions to specific genes.

Mass spectrometry-based metabolomics is an established platform in microbial functional genomics [21,22]. Examples of successful assignments of gene function include incubations of complex mixtures of metabolites with purified proteins of unknown function to discover enzymatic activities [23,24], and screening of libraries of bacterial mutants to identify genes of enzymes and transport proteins required for the utilization of specific metabolites [25]. Metabolite profiling often helps gene functional assignment but it more often points to novel metabolic capabilities. Uncharacterized biosynthetic capabilities are manifested by the frequent detection of novel metabolites or metabolites, which currently cannot be identified using mass spectrometry alone [26–28]. Additionally, utilization of uncharacterized metabolites from complex media or metabolite utilizations, which were not predicted based on available genome annotations, were also observed in bacteria [28,29].

Here we present untargeted metabolite profiling of cell extracts of nine additional cyanobacteria (Table 1) with available genome sequences [30–32] to test for the presence of unexpected metabolites previously detected in *Synechococcus sp.* PCC 7002 and to explore the diversity of small polar metabolites in this phylum. This exploratory study is intended to provide leads for downstream detailed structural and functional characterization of novel natural compounds.

Table 1. List of cyanobacteria used for metabolite profiling.

Taxonomic Subsection	Species	Strain	Abbreviation	Medium
1 (Chroococcales)	<i>Halotheca</i> sp.	PCC 7418	7418	ASNIII/Tu4X
1 (Chroococcales)	<i>Synechococcus elongatus</i>	PCC 6301	6301	BG11
1 (Chroococcales)	<i>Synechococcus</i> sp.	PCC 7002	7002	ASNIII/BG11 + vit. B12
2 (Pleurocapsales)	<i>Chroococcidiopsis</i> sp.	PCC 6712	6712	ASNIII/BG11
2 (Pleurocapsales)	<i>Pleurocapsa</i> sp.	PCC 7327	7327	BG11
3 (Oscillatoriales)	<i>Geitlerinema</i> sp.	PCC 7407	7407	BG11
3 (Oscillatoriales)	<i>Leptolyngbya</i> sp.	PCC 7376	7376	ASNIII + vit. B12
3 (Oscillatoriales)	<i>Microcoleus vaginatus</i>	PCC 9802	9802	BG11
4 (Nostocales)	<i>Calothrix</i> sp.	PCC 7507	7507	BG11o
4 (Nostocales)	<i>Nostoc</i> sp.	PCC 7107	7107	BG11o

BG11 and ASNIII as in ref. [33]; BG11o, BG11 without nitrate; vit. B12, Vitamin B12 at 10 µg/L final concentration; Tu4X, Turks Island salts 4× concentration.

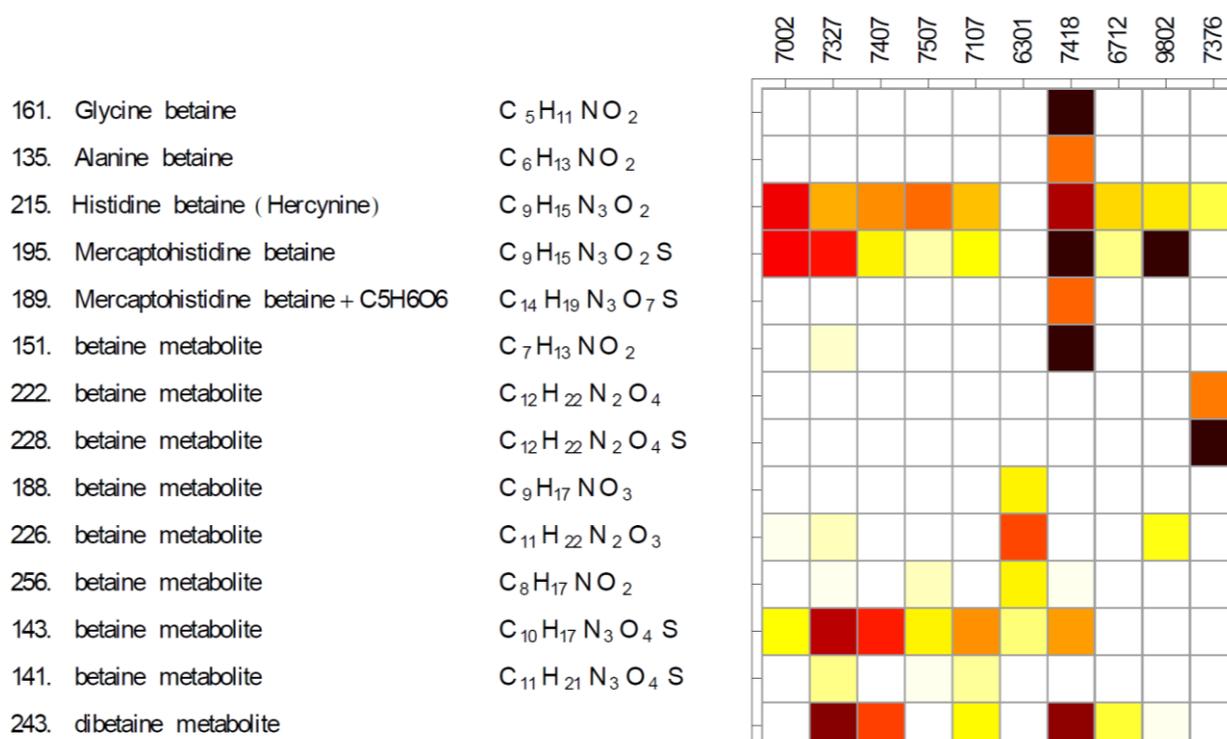
2. Results and Discussion

Analysis of raw metabolite profile data of selected cyanobacteria (Table 1) led to the curation of 264 metabolites across the ten analyzed cyanobacteria (Supplementary Table S1). The data analysis was performed combining the MathDAMP package [34] with iterative manual curation as described in the Experimental section. Annotations of metabolites were also visualized on density plots (Figure 1) to find any unannotated metabolites. Our aim was to annotate polar metabolites with at least one of its spectral features above 10,000 ion counts.

Tandem mass spectrometry (MS/MS) was performed on characteristic ions (primarily $[M + H]^+$ in positive mode and $[M - H]^-$ in negative mode) of metabolites following spectral feature annotation (Figure 1, Supplementary Table S1). Putative identifications of metabolites were based on our in-house database, built with identifications from our previous studies [26,29,30] and analysis of MS/MS spectra using public spectral libraries [35,36]. MS/MS spectra do not always provide sufficient information for full structural characterization of metabolites without corresponding true chemical standards, which limits the characterization power and overall degree of certainty of untargeted metabolomics [37]. Out of 264 metabolites annotated in this study, we could assign chemical formulae to 157 and putatively identify only 105 metabolites. This further underscores the relative ease of detecting metabolites which are not included in databases of metabolism or spectral libraries using untargeted metabolite profiling.

confirmed [25,38]. The biosynthesis of hercynine and ergothioneine has been characterized in mycobacteria. Homologs of mycobacterial genes for some steps of the biosynthesis have been reported for cyanobacteria [39]. In fact the biosynthesis of hercynine and ergothioneine has been confirmed experimentally in five cyanobacterial strains [40]. In this study, we detected hercynine in nine out of the ten strains, and a thiol of hercynine in eight of these nine (Figure 2). The genomes of these nine cyanobacteria contain homologs of some mycobacterial genes corresponding to ergothioneine biosynthesis, while the genome of *Synechococcus elongatus* PCC 6301 lacks homologs of these genes (Supplementary Table S2). Glycine betaine was detected only in *Halotheca* sp. PCC 7418 (Figure 2) what is consistent with the presence of the glycine methylation pathway [41]. Interestingly, this strain, which is a representative of the most halotolerant cyanobacteria known [42], contains the most diverse set of putative betaines of all strains tested.

Figure 2. Diversity of betaines in analyzed cyanobacteria.



Eleven additional metabolites were classified as betaines (Figure 2). Assignments to betaines were based on characteristic weak signals of their corresponding $[M - H]^-$ ions in negative mode profile (MS1) mass spectra and the presence of trimethylamine-related fragments (m/z 60.0808, 59.0730, 58.0651) and trimethylamine neutral loss (59.0735) in positive mode MS/MS spectra (Supplementary Figure S1 and Table S1). Chemical formulae of some unidentified betaines correspond to a few candidate compounds (e.g., $C_7H_{13}NO_2$ could correspond to crotonobetaine or proline betaine) but the analyses of MS/MS spectra were inconclusive. An interesting pair of betaines was detected exclusively in *Leptolyngbya* sp. PCC 7376. The chemical formulae of these two betaines differ by a single sulfur ($C_{11}H_{22}N_2O_4$, $C_{11}H_{22}N_2O_4S$)—an analogous pattern to hercynine and ergothioneine (Figure 2), pointing to a similar role. Accurate mass and isotopic profile of the betaine with metabolite number 243 is consistent with an oxidized form of ergothioneine ($C_{18}H_{28}N_6O_4S_2$). This metabolite is also

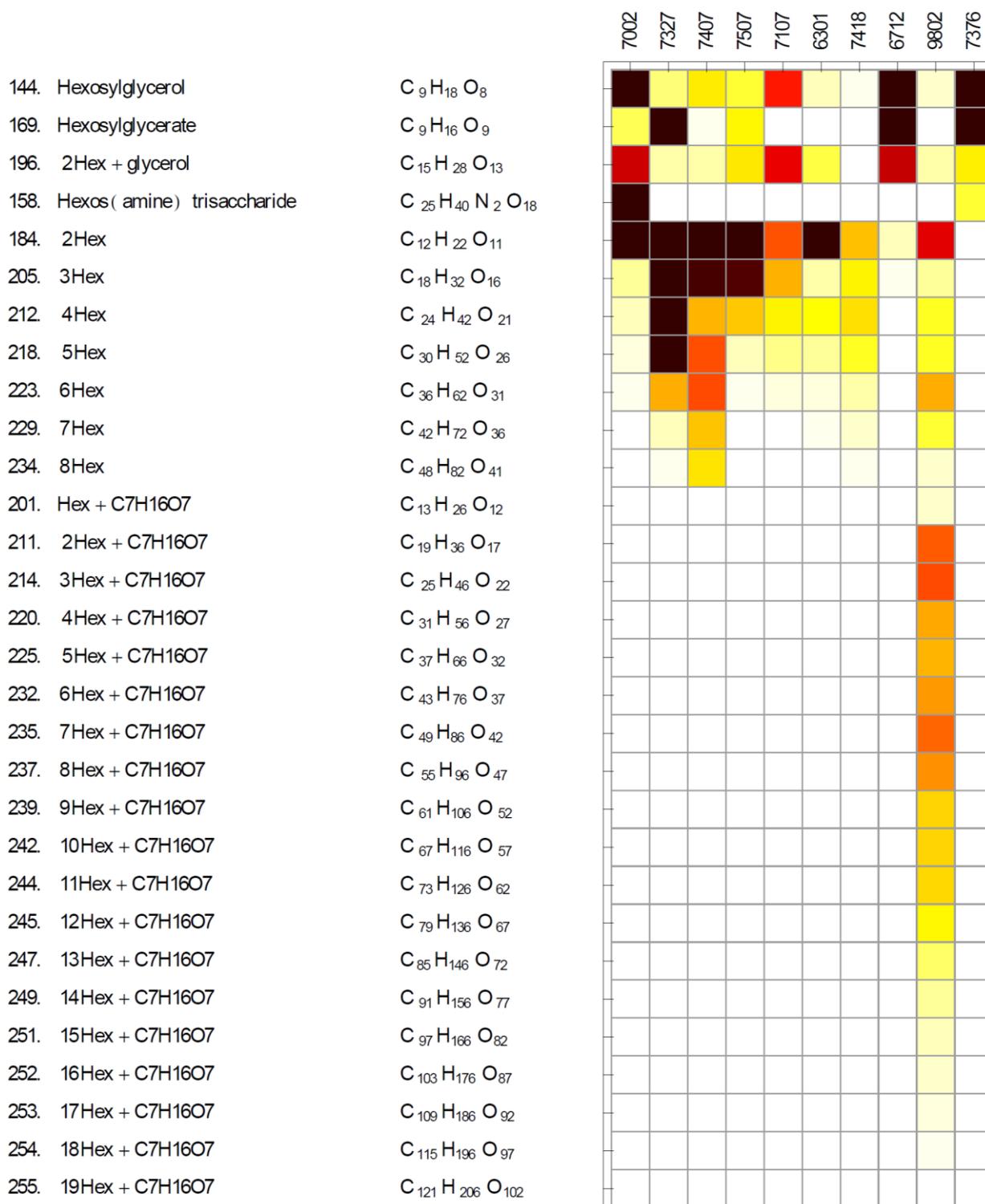
present in extracts of multiple strains in which the thiol of histidine betaine was detected (Figure 2). This coincidence and the mere presence of an oxidized form, would be consistent with an antioxidant role of ergothioneine [43].

2.2. Diversity of Glycosides and Oligosaccharides

Cyanobacteria are known to synthesize glycosides and oligosaccharides as compatible solutes [44], glycogen as a storage polysaccharide, and a variety of exopolysaccharides [45]. While glucosylglycerol and glucosylglycerate are common glycoside compatible solutes [44], other hexoses such as mannose and galactose may substitute for glucose [46]. Since we did not have authentic standards of these isomers to test if they have different chromatographic properties using our LC-MS method, we use combined names hexosylglycerol and hexosylglycerate for these glycerides (Figure 3). Homologs of glucosylglycerol-phosphate synthase and glucosylglycerol 3-phosphatase were identified by genomic annotation only in five of the strains analyzed (Supplementary Table S3), while hexosylglycerol was detected in all ten cyanobacteria (Figure 3, Supplementary Table S1). These findings suggest the presence of an alternative pathway for the biosynthesis of these glycerides. One simple possibility is transglycosylation reactions catalyzed by alpha-glucosidases, which have been shown to produce glucosylglycerol *in vitro* [47]. We detected hexosylglycerate in five of our strains (Figure 3, Supplementary Table S1). Homologs of glucosylglycerate biosynthetic genes were identified in four of these 5 cyanobacteria (Supplementary Table S4) showing overall good consistency between metabolite profiles and gene content while pointing to unusual hexosylglycerate biosynthesis in *Calothrix* sp. PCC 7507.

In addition to the detection of a hexose disaccharide, which could correspond to sucrose [44] or trehalose [31], a series of higher hexose oligomers was also detected (Figure 3). These could be storage glycogen-related maltooligosaccharides or structurally heterogeneous oligohexoses reported in different cyanobacteria [48,49]. Peak areas of ions of different metabolites cannot be used directly for absolute quantitative comparisons due to different ionization efficiencies of different metabolites [50]. However, results in Figure 3 clearly show different distribution of glycosides and oligosaccharides among the cyanobacteria studied here. Such comparisons may prove useful for identifying suitable starting points for engineering heterologous pathways or using cyanobacterial biomass as a feedstock for biotechnological applications.

We have previously identified an unusual trisaccharide of a hexose, *N*-acetylhexosamine and an oxidized version of *N*-acetylmuramic acid in *Synechococcus* sp. PCC 7002 [28]. This trisaccharide was detected as two distinct chromatographic peaks possibly caused by two anomers or mutarotation suggesting a reducing nature of this trisaccharide [28]. During this study, we detected this oligosaccharide in three additional cyanobacteria, also in two separate chromatographic peaks, yet at significantly lower signal intensities than in *Synechococcus* sp. PCC 7002 (Figure 3, Supplementary Table S1). Peak areas of the two peaks were combined and reported as a single metabolite number 158 (Supplementary Table S1). The role of this trisaccharide remains unclear; no other “decorated” hexosamine-based oligomers were identified in this study.

Figure 3. Diversity of glycosides and oligosaccharides in analyzed cyanobacteria.

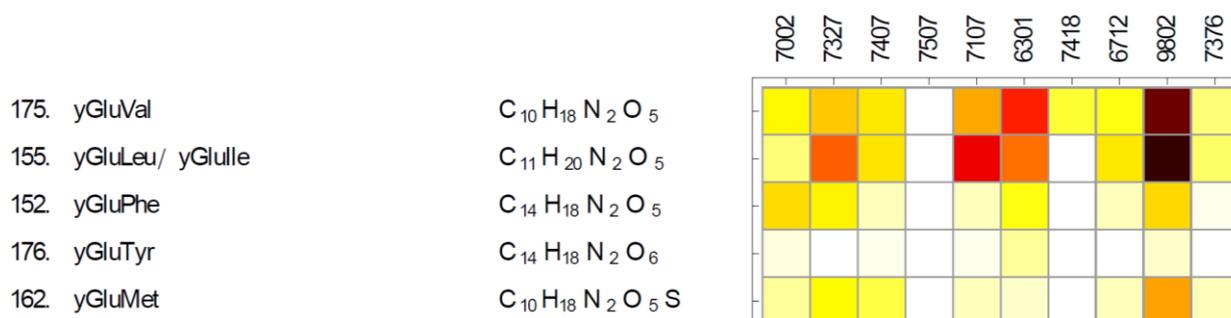
A range of unusual oligosaccharides was detected exclusively in *Microcoleus vaginatus* PCC 9802 (Figure 3). These oligosaccharides are condensation products of oligohexoses with $C_7H_{16}O_7$. This chemical formula is consistent with a chemical formula of seven-carbon sugar alcohols. Release of polysaccharides is considered to play a key role in cyanobacterial gliding [51] and oligohexoses with $C_7H_{16}O_7$ could be related to motility in *Microcoleus vaginatus*, a vertical migrant in desert soils [52]. Another possibility is that these glycans are related to exopolysaccharides which are known to play an

important ecological role in binding soils to form biological soil crusts [53]. Certainly, the large comparative diversity of these compounds in *Microcoleus vaginatus* PCC 9802 suggests that they may play a differential physiological role related to adaptations to life in desert soils.

2.3. Gamma-Glutamyl Dipeptides and Gamma-Glutamyltransferase (*ggt*)

A series of gamma-glutamyl dipeptides was among the unexpected findings in the profile of *Synechococcus* sp. PCC 7002 [28]. Gamma-glutamylation increases the solubility of non-polar amino acids [54] and may represent a strategy to prevent the loss of these amino acids via leakage through cell membranes [29]. Gamma-glutamyl dipeptides were detected in nine analyzed cyanobacteria and not detected in *Calothrix* sp. PCC 7507 (Figure 4). Homologs of gamma-glutamyltransferase gene (*ggt*) were identified in genomes of nine strains, interestingly; the strain missing the homolog is *Leptolyngbya* sp. PCC 7376 (Supplementary Table S5). *Calothrix* sp. PCC 7507 and two additional strains possess *ggt* sequences that are missing the catalytic threonine dyad shown to be important for autoprocessing in *Helicobacter pylori ggt* [55] (Supplementary Figure S2). This may indicate that *ggt* in *Calothrix* sp. PCC 7507 is not active or that it has some other (non-*ggt*) activity. *Pleurocapsa* sp. PCC 7327 has two additional *ggt* genes, at least one of them has perfectly conserved threonine dyad. And it is possible that both *Halotheca* sp. PCC 7418, which has the same possibly inactive *ggt* as *Calothrix* sp. PCC 7507 and *Leptolyngbya* sp. PCC 7376, which has no *ggt* at all, have an alternative form of gamma-glutamyltransferase, which is not similar to the known *ggt*.

Figure 4. Distribution of gamma-glutamyl dipeptides in analyzed cyanobacteria.



2.4. Other Metabolites

Similarly to the differences in the distribution of oligosaccharides, we also observed differences in the profiles of fatty acids, especially unsaturated ones (Supplementary Figure S3). A large proportion of detected metabolites remained unidentified and their chemical formulae are not present in databases of metabolism KEGG [56] or MetaCyc [57]. These metabolites represent a valuable resource for the characterization of novel metabolic capabilities. Having metabolite profile data of only ten cyanobacterial strains proved insufficient to link specific genes to individual metabolites as the number of candidate genes correlating with the presence or absence of specific metabolites was too large. Nevertheless, we believe that scaling up untargeted metabolite profiling to a larger number of strains would enable to both detect novel metabolites as well as to zoom in on a small number of potential corresponding biosynthetic genes.

3. Experimental Section

3.1. Strains and Culture Conditions

All strains are available at Pasteur Culture collection of Cyanobacteria [58]. Biomass of the nine PCC strains for chemical analyses was obtained by growing in 1.25 L of media suitable for the strain. The cultures were grown at 25 °C under continuous light provided by Osram Universal White fluorescent tubes ($20 \mu\text{mol quanta m}^{-2} \text{s}^{-1}$) with agitation and constant bubbling of 1% CO₂. Culture media are specified in Table 1. *Calothrix* sp. PCC 7507 and *Nostoc* sp. PCC 7107 are heterocystous cyanobacteria and these two strains were grown in nitrate free media to promote the development of heterocysts. All chemicals for growth media were purchased from Sigma.

3.2. Metabolite Extraction

Two milliliters of methanol were added to lyophilized biomass originating from approximately 2 mL of packed cell volume for each cyanobacterium (Table 1). In the case of *Microcoleus vaginatus* PCC 9802, 2 mL of methanol were added to cell pellet of approximately 1 mL of packed cell volume. The suspensions were sonicated for 15 min in sonic bath (VWR symphony) and then transferred to 2 mL microcentrifuge tubes and centrifuged for 10 min at $2348 \times g$ using an Eppendorf 5424 centrifuge. Supernatants were transferred to 1.8 mL glass vials and dried down using Savant Speedvac Plus (SC210A). 100 μL of methanol were added to each vial and the vials were stored at $-20 \text{ }^\circ\text{C}$. Prior to analysis by LC-MS, the samples were filtered using 0.22 μm PVDF microcentrifuge filters (Millipore).

3.3. LC-MS Analysis

An Agilent capillary 1200 liquid chromatography system coupled to an Agilent 6520 ESI-Q-TOF mass spectrometer was used for LC-MS analysis. ZIC-HILIC column (3.5 μm , 100 Å , $150 \times 1 \text{ mm}$) for normal phase liquid chromatography using analytical conditions as described previously [28]. Profile mode data (MS1) were acquired using a fast polarity switching mode. One LC-MS run was performed for a single sample of each strain. Fragmentation (MS/MS) spectra were acquired as two separate positive and negative polarity runs for each sample using data-dependent selection of precursor ions (Auto MS/MS) using collision energy of 10 V. Sample injection volume was 2 μL .

3.4. Data Analysis

Raw datasets from profile mode (MS1) analysis were exported to mzdata format using Agilent MassHunter Qualitative analysis software (B.05.00) and preprocessed by MathDAMP package [34] to unit mass resolution for comparative analysis. Differences among preprocessed datasets of the ten analyzed cyanobacteria were identified using MathDAMP by direct comparisons identifying outliers using quartile analysis [34]. Redundant spectral features potentially corresponding to a single metabolite (adducts, multimers, fragments) were grouped by correlation of their peak shape along the chromatographic dimension as described previously [29]. Resulting groups of spectral features were manually curated using Agilent MassHunter Qualitative Analysis software for chemical formula

calculation. Putative identification of metabolites was based on our previous results [28,29] and analysis of MS/MS spectra against spectral libraries Metlin [35] and MassBank [36].

3.5. Genome Analysis

Candidate genes and synthesis pathways were identified using the data and tools in Integrated Microbial Genomes database (IMG) [59]. GenBank accession numbers of 10 cyanobacterial genomes are as follows: CP003943 (*Calothrix* sp. PCC 7507), (*Chroococcidiopsis* sp. PCC 6712), CP003591 (*Geitlerinema* sp. PCC 7407), CP003945 (*Halotheca* sp. PCC 7418), CP003946 (*Leptolyngbya* sp. PCC 7376), CP003548 (*Nostoc* sp. PCC 7107), CP003590 (*Pleurocapsa* sp. PCC 7327), AP008231 (*Synechococcus elongatus* PCC 6301), CP000951 (*Synechococcus* sp. PCC 7002). Putative orthologs of experimentally characterized proteins in cyanobacterial genomes have been identified as bi-directional best BLASTp [60] hits using e-value cutoffs of 1.0×10^{-5} . Experimentally characterized proteins included in the analysis were ergothioneine biosynthesis proteins EgtB-E from *Mycobacterium smegmatis* MC2 155 [39], ovothiol biosynthesis protein OvoA from *Erwinia tasmaniensis* [61], GSMT and SDMT from *Ectothiorhodospira halochloris* [41], ggpPS from *Synechocystis* sp. PCC 6803 [62] and ggt from *H. pylori* [55]. Additional candidate genes in ergothioneine biosynthesis and gamma-glutamyltranspeptidase family proteins were identified by the matches to the corresponding TIGRfam models [63] and COG position-specific scoring matrices obtained from the CDD database [64].

4. Conclusions

In this study, we have shown that it is possible to correlate the presence of metabolites with known biosynthetic genes to the gene content of ten analyzed cyanobacteria. Additionally, we detected a series of novel betaines in some cyanobacteria and unusual oligohexoses with a degree of polymerization up to 19 with a single $C_7H_{16}O_7$ moiety in *Microcoleus vaginatus* PCC 9802. The scale-up of such comparative metabolite profiling may enable the linking of genes of unknown function to the biosynthesis of novel natural compounds.

Acknowledgments

This work was supported by the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 and the Institut Pasteur.

Conflict of Interest

The authors declare no conflict of interest.

References

1. Seckbach, J.; Oren, A. Oxygenic Photosynthetic Microorganisms in Extreme Environments. In *Algae and Cyanobacteria in Extreme Environments*; Seckbach, J., Ed.; Springer: Dordrecht, The Netherlands, 2007; pp. 5–25.
2. Whitton, B.A.; Potts, M. Introduction to the Cyanobacteria. In *Ecology of Cyanobacteria II*; Whitton, B.A., Ed.; Springer: Dordrecht, The Netherlands, 2012; pp. 1–13.

3. Garcia-Pichel, F.; Loza, V.; Marusenko, Y.; Mateo, P.; Potrafka, R.M. Temperature drives the continental-scale distribution of key microbes in topsoil communities. *Science* **2013**, *340*, 1574–1577.
4. Garcia-Pichel, F.; Belnap, J.; Neuer, S.; Schanz, F. Estimates of global cyanobacterial biomass and its distribution. *Algol. Stud.* **2003**, *109*, 213–227.
5. Falkowski, P.G.; Fenchel, T.; Delong, E.F. The microbial engines that drive Earth's biogeochemical cycles. *Science* **2008**, *320*, 1034–1039.
6. Gerwick, W.H.; Moore, B.S. Lessons from the past and charting the future of marine natural products drug discovery and chemical biology. *Chem. Biol.* **2012**, *19*, 85–98.
7. Gao, Q.; Garcia-Pichel, F. Microbial ultraviolet sunscreens. *Nat. Rev. Microbiol.* **2011**, *9*, 791–802.
8. Atsumi, S.; Higashide, W.; Liao, J.C. Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde. *Nat. Biotechnol.* **2009**, *27*, 1177–1180.
9. McNeely, K.; Xu, Y.; Bennette, N.; Bryant, D.A.; Dismukes, G.C. Redirecting reductant flux into hydrogen production via metabolic engineering of fermentative carbon metabolism in a cyanobacterium. *Appl. Environ. Microbiol.* **2010**, *76*, 5032–5038.
10. Ducat, D.C.; Way, J.C.; Silver, P.A. Engineering cyanobacteria to generate high-value products. *Trends Biotechnol.* **2011**, *29*, 95–103.
11. Lan, E.I.; Liao, J.C. ATP drives direct photosynthetic production of 1-butanol in cyanobacteria. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 6018–6023.
12. Xu, Y.; Guerra, L.T.; Li, Z.; Ludwig, M.; Dismukes, G.C.; Bryant, D.A. Altered carbohydrate metabolism in glycogen synthase mutants of *Synechococcus* sp. strain PCC 7002: Cell factories for soluble sugars. *Metab. Eng.* **2013**, *16*, 56–67.
13. Galperin, M.Y.; Koonin, E.V. From complete genome sequence to “complete” understanding? *Trends Biotechnol.* **2010**, *28*, 398–406.
14. Hanson, A.D.; Pribat, A.; Waller, J.C.; de Crécy-Lagard, V. “Unknown” proteins and “orphan” enzymes: The missing half of the engineering parts list—and how to find it. *Biochem. J.* **2009**, *425*, 1–11.
15. Gerdes, S.; El Yacoubi, B.; Bailly, M.; Blaby, I.K.; Blaby-Haas, C.E.; Jeanguenin, L.; Lara-Núñez, A.; Pribat, A.; Waller, J.C.; Wilke, A.; *et al.* Synergistic use of plant-prokaryote comparative genomics for functional annotations. *BMC Genomics* **2011**, *12* (Suppl. 1), S2.
16. Kitagawa, M.; Ara, T.; Arifuzzaman, M.; Ioka-Nakamichi, T.; Inamoto, E.; Toyonaga, H.; Mori, H. Complete set of ORF clones of *Escherichia coli* ASKA library (a complete set of *E. coli* K-12 ORF archive): Unique resources for biological research. *DNA Res.* **2005**, *12*, 291–299.
17. Baba, T.; Ara, T.; Hasegawa, M.; Takai, Y.; Okumura, Y.; Baba, M.; Datsenko, K.A.; Tomita, M.; Wanner, B.L.; Mori, H. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: The Keio collection. *Mol. Syst. Biol.* **2006**, *2*, 2006.0008.
18. Nichols, R.J.; Sen, S.; Choo, Y.J.; Beltrao, P.; Zietek, M.; Chaba, R.; Lee, S.; Kazmierczak, K.M.; Lee, K.J.; Wong, A.; *et al.* Phenotypic landscape of a bacterial cell. *Cell* **2011**, *144*, 143–156.

19. Deutschbauer, A.; Price, M.N.; Wetmore, K.M.; Shao, W.; Baumohl, J.K.; Xu, Z.; Nguyen, M.; Tamse, R.; Davis, R.W.; Arkin, A.P. Evidence-based annotation of gene function in *Shewanella oneidensis* MR-1 using genome-wide fitness profiling across 121 conditions. *PLoS Genet.* **2011**, *7*, e1002385.
20. Reed, J.L.; Patel, T.R.; Chen, K.H.; Joyce, A.R.; Applebee, M.K.; Herring, C.D.; Bui, O.T.; Knight, E.M.; Fong, S.S.; Palsson, B.O. Systems approach to refining genome annotation. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 17480–17484.
21. Baran, R.; Reindl, W.; Northen, T.R. Mass spectrometry based metabolomics and enzymatic assays for functional genomics. *Curr. Opin. Microbiol.* **2009**, *12*, 547–552.
22. Saito, N.; Ohashi, Y.; Soga, T.; Tomita, M. Unveiling cellular biochemical reactions via metabolomics-driven approaches. *Curr. Opin. Microbiol.* **2010**, *13*, 358–362.
23. Saito, N.; Robert, M.; Kitamura, S.; Baran, R.; Soga, T.; Mori, H.; Nishioka, T.; Tomita, M. Metabolomics approach for enzyme discovery. *J. Proteome Res.* **2006**, *5*, 1979–1987.
24. Saito, N.; Robert, M.; Kochi, H.; Matsuo, G.; Kakazu, Y.; Soga, T.; Tomita, M. Metabolite profiling reveals YihU as a novel hydroxybutyrate dehydrogenase for alternative succinic semialdehyde metabolism in *Escherichia coli*. *J. Biol. Chem.* **2009**, *284*, 16442–16451.
25. Baran, R.; Bowen, B.P.; Price, M.N.; Arkin, A.P.; Deutschbauer, A.M.; Northen, T.R. Metabolic footprinting of mutant libraries to map metabolite utilization to genotype. *ACS Chem. Biol.* **2013**, *8*, 189–199.
26. Kind, T.; Fiehn, O. Advances in structure elucidation of small molecules using mass spectrometry. *Bioanal. Rev.* **2010**, *2*, 23–60.
27. Bowen, B.P.; Northen, T.R. Dealing with the unknown: metabolomics and metabolite atlases. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 1471–1476.
28. Baran, R.; Bowen, B.P.; Bouskill, N.J.; Brodie, E.L.; Yannone, S.M.; Northen, T.R. Metabolite identification in *Synechococcus* sp. PCC 7002 using untargeted stable isotope assisted metabolite profiling. *Anal. Chem.* **2010**, *82*, 9034–9042.
29. Baran, R.; Bowen, B.P.; Northen, T.R. Untargeted metabolic footprinting reveals a surprising breadth of metabolite uptake and release by *Synechococcus* sp. PCC 7002. *Mol. Biosyst.* **2011**, *7*, 3200–3206.
30. Sugita, C.; Ogata, K.; Shikata, M.; Jikuya, H.; Takano, J.; Furumichi, M.; Kanehisa, M.; Omata, T.; Sugiura, M.; Sugita, M. Complete nucleotide sequence of the freshwater unicellular cyanobacterium *Synechococcus elongatus* PCC 6301 chromosome: gene content and organization. *Photosynth. Res.* **2007**, *93*, 55–67.
31. Starkenburg, S.R.; Reitenga, K.G.; Freitas, T.; Johnson, S.; Chain, P.S.; Garcia-Pichel, F.; Kuske, C.R. Genome of the cyanobacterium *Microcoleus vaginatus* FGP-2, a photosynthetic ecosystem engineer of arid land soil biocrusts worldwide. *J. Bacteriol.* **2011**, *193*, 4569–4570.
32. Shih, P.M.; Wu, D.; Latifi, A.; Axen, S.D.; Fewer, D.P.; Talla, E.; Calteau, A.; Cai, F.; Tandeau de Marsac, N.; Rippka, R.; *et al.* Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 1053–1058.

33. Rippka, R.; Deruelles, J.; Waterbury, J.B.; Herdman, M.; Stainer, R.Y. Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *J. Gen. Microbiol.* **1979**, *111*, 1–61.
34. Baran, R.; Kochi, H.; Saito, N.; Suematsu, M.; Soga, T.; Nishioka, T.; Robert, M.; Tomita, M. MathDAMP: A package for differential analysis of metabolite profiles. *BMC Bioinforma.* **2006**, *7*, 530.
35. Smith, C.A.; O’Maille, G.; Want, E.J.; Qin, C.; Trauger, S.A.; Brandon, T.R.; Custodio, D.E.; Abagyan, R.; Siuzdak, G. METLIN: A metabolite mass spectral database. *Ther. Drug Monit.* **2005**, *27*, 747–751.
36. Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K.; *et al.* MassBank: A public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.* **2010**, *45*, 703–714.
37. Scheubert, K.; Hufsky, F.; Böcker, S. Computational mass spectrometry for small molecules. *J. Cheminforma.* **2013**, *5*, 12.
38. Warren, C.R. Quaternary ammonium compounds can be abundant in some soils and are taken up as intact molecules by plants. *New Phytol.* **2013**, *198*, 476–485.
39. Seebeck, F.P. *In vitro* reconstitution of mycobacterial ergothioneine biosynthesis. *J. Am. Chem. Soc.* **2010**, *132*, 6632–6633.
40. Pfeiffer, C.; Bauer, T.; Surek, B.; Schomig, E.; Grundemann, D. Cyanobacteria produce high levels of ergothioneine. *Food Chem.* **2011**, *129*, 4.
41. Nyssola, A.; Kerovuoto, J.; Kaukinen, P.; von Weymarn, N.; Reinikainen, T. Extreme halophiles synthesize betaine from glycine by methylation. *J. Biol. Chem.* **2000**, *275*, 22196–22201.
42. Garcia-Pichel, F.; Nübel, U.; Muyzer, G. The phylogeny of unicellular, extremely halotolerant cyanobacteria. *Arch. Microbiol.* **1998**, *169*, 469–482.
43. Bello, M.H.; Barrera-Perez, V.; Morin, D.; Epstein, L. The *Neurospora crassa* mutant NcΔEgt-1 identifies an ergothioneine biosynthetic gene and demonstrates that ergothioneine enhances conidial survival and protects against peroxide toxicity during conidial germination. *Fungal Genet. Biol.* **2012**, *49*, 160–172.
44. Hagemann, M. Molecular biology of cyanobacterial salt acclimation. *FEMS Microbiol. Rev.* **2011**, *35*, 87–123.
45. Pereira, S.; Zille, A.; Micheletti, E.; Moradas-Ferreira, P.; De Philippis, R.; Tamagnini, P. Complexity of cyanobacterial exopolysaccharides: composition, structures, inducing factors and putative genes involved in their biosynthesis and assembly. *FEMS Microbiol. Rev.* **2009**, *33*, 917–941.
46. Luley-Goedl, C.; Nidetzky, B. Glycosides as compatible solutes: Biosynthesis and applications. *Nat. Prod. Rep.* **2011**, *28*, 875–896.
47. Ojima, T.; Saburi, W.; Yamamoto, T.; Kudo, T. Characterization of *Halomonas* sp. strain H11 α -glucosidase activated by monovalent cations and its application for efficient synthesis of α -D-glucosylglycerol. *Appl. Environ. Microbiol.* **2012**, *78*, 1836–1845.

48. Wieneke, R.; Klein, S.; Geyer, A.; Loos, E. Structural and functional characterization of galactooligosaccharides in *Nostoc commune*: β -D-galactofuranosyl-(1 \rightarrow 6)-[β -D-galactofuranosyl-(1 \rightarrow 6)]₂- β -D-1,4-anhydrogalactitol and β -(1 \rightarrow 6)-galactofuranosylated homologues. *Carbohydr. Res.* **2007**, *342*, 2757–2765.
49. Pontis, H.G.; Vargas, W.A.; Salerno, G.L. Structural characterization of the members of a polymer series, compatible solutes in *Anabaena* cells exposed to salt stress. *Plant Sci.* **2007**, *172*, 29–35.
50. Oss, M.; Krueve, A.; Herodes, K.; Leito, I. Electrospray ionization efficiency scale of organic compounds. *Anal. Chem.* **2010**, *82*, 2865–2872.
51. Hoiczky, E.; Baumeister, W. The junctional pore complex, a prokaryotic secretion organelle, is the molecular motor underlying gliding motility in cyanobacteria. *Curr. Biol.* **1998**, *8*, 1161–1168.
52. Garcia-Pichel, F.; Pringault, O. Microbiology: Cyanobacteria track water in desert soils. *Nature* **2001**, *413*, 380–381.
53. Rajeev, L.; da Rocha, U.N.; Klitgord, N.; Luning, E.G.; Fortney, J.; Axen, S.D.; Shih, P.M.; Bouskill, N.J.; Bowen, B.P.; Kerfeld, C.A.; *et al.* Dynamic cyanobacterial response to hydration and dehydration in a desert biological soil crust. *ISME J.* **2013**, doi:10.1038/ismej.2013.83.
54. Suzuki, H.; Yamada, C.; Kato, K. Gamma-glutamyl compounds and their enzymatic production using bacterial gamma-glutamyltranspeptidase. *Amino Acids* **2007**, *32*, 333–340.
55. Boanca, G.; Sand, A.; Okada, T.; Suzuki, H.; Kumagai, H.; Fukuyama, K.; Barycki, J.J. Autoprocessing of *Helicobacter pylori* gamma-glutamyltranspeptidase leads to the formation of a threonine-threonine catalytic dyad. *J. Biol. Chem.* **2007**, *282*, 534–541.
56. Kanehisa, M.; Goto, S.; Sato, Y.; Furumichi, M.; Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **2012**, *40*, D109–D114.
57. Caspi, R.; Altman, T.; Dreher, K.; Fulcher, C.A.; Subhraveti, P.; Keseler, I.M.; Kothari, A.; Krummenacker, M.; Latendresse, M.; Mueller, L.A.; *et al.* The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* **2012**, *40*, D742–D753.
58. The Pasteur Culture Collection of Cyanobacteria (PCC). Available online: http://www.pasteur.fr/pcc_cyanobacteria (accessed on 22 September 2013).
59. Markowitz, V.M.; Chen, I.-M.A.; Palaniappan, K.; Chu, K.; Szeto, E.; Grechkin, Y.; Ratner, A.; Jacob, B.; Huang, J.; Williams, P.; *et al.* IMG: The integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res.* **2012**, *40*, D115–D122.
60. Altschul, S.F.; Madden, T.L.; Schaffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402.
61. Braunshausen, A.; Seebeck, F.P. Identification and characterization of the first ovothiol biosynthetic enzyme. *J. Am. Chem. Soc.* **2011**, *133*, 1757–1759.
62. Schoor, A.; Hagemann, M.; Erdmann, N. Glucosylglycerol-phosphate synthase: target for ion-mediated regulation of osmolyte synthesis in the cyanobacterium *Synechocystis* sp. strain PCC 6803. *Arch. Microbiol.* **1999**, *171*, 101–106.
63. Haft, D.H.; Selengut, J.D.; White, O. The TIGRFAMs database of protein families. *Nucleic Acids Res.* **2003**, *31*, 371–373.

64. Marchler-Bauer, A.; Anderson, J.B.; Derbyshire, M.K.; DeWeese-Scott, C.; Gonzales, N.R.; Gwadz, M.; Hao, L.; He, S.; Hurwitz, D.I.; Jackson, J.D.; *et al.* CDD: A conserved domain database for interactive domain family analysis. *Nucleic Acids Res.* **2007**, *35*, D237–D240.

© 2013 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).