

# What can genome-scale metabolic network reconstructions do for prokaryotic systematics?

Francisco Barona-Gómez · Pablo Cruz-Morales · Lianet Noda-García

Received: 6 September 2011 / Accepted: 7 October 2011 / Published online: 21 October 2011  
© Springer Science+Business Media B.V. 2011

**Abstract** It has recently been proposed that in addition to Nomenclature, Classification and Identification, Comprehending Microbial Diversity may be considered as the fourth tenet of microbial systematics [Staley JT (2010) The Bulletin of BISMis, 1(1): 1–5]. As this fourth goal implies a fundamental understanding of microbial speciation, this perspective article argues that translation of bacterial genome sequences into metabolic features may contribute to the development of modern polyphasic taxonomic approaches. Genome-scale metabolic network reconstructions (GSMRs), which are the result of computationally predicted and experimentally confirmed stoichiometric matrices incorporating all enzyme and metabolite components encoded by a genome sequence, provide a platform that can illustrate bacterial speciation. As the topology and the composition of GSMRs are expected to be the result of adaptive evolution, the features of these networks may provide the prokaryotic taxonomist with novel tools for reaching the fourth tenet of

microbial systematics. Through selected examples from the *Actinobacteria*, which have been inferred from GSMRs and experimentally confirmed after phenotypic characterisation, it will be shown that this level of information can be incorporated into modern polyphasic taxonomic approaches. In conclusion, three specific examples are illustrated to show how GSMRs will revolutionize prokaryotic systematics, as has previously occurred in many other fields of microbiology.

**Keywords** Genome-scale metabolic network reconstruction · Chemotaxons · Cardiolipin · Menaquinones · Natural products · *Streptomyces* · Actinobacteria

## Evolutionary implications of GSMRs

With the advent of the genomic era, which has prompted a major revolution in microbiology as a whole, it has been anticipated that genomics will also play a major role in the field of prokaryotic systematics (Sutcliffe et al. this issue). Nevertheless, how genomic data can be embraced by modern polyphasic taxonomy is still an open question. As chemotaxons are the result of metabolism, one obvious possibility is the reconstruction of metabolism from genome sequences, which can provide novel *in silico* chemotaxons useful as potential evolutionary markers. Indeed, classical studies of bacterial metabolism have been re-evaluated

---

**Electronic supplementary material** The online version of this article (doi:10.1007/s10482-011-9655-1) contains supplementary material, which is available to authorized users.

---

F. Barona-Gómez (✉) · P. Cruz-Morales · L. Noda-García  
Evolution of Metabolic Diversity Laboratory, Laboratorio Nacional de Genómica para la Biodiversidad (Langebio), CINVESTAV-IPN, Km 9.6 Libramiento Norte, Carretera Irapuato - León, CP 36822 Irapuato, Mexico  
e-mail: fbarona@langebio.cinvestav.mx

after the availability of whole genome sequences, as this data provides novel metabolic insights that could not be discovered otherwise. Genomic insights support the general conclusion that bacterial central or primary metabolism is less universally conserved than as presented in textbooks, and many versions of central metabolic pathways have been discovered (Osterman and Overbeek 2003). Furthermore, beyond these well-integrated central metabolic networks, genomics has also prompted the discovery of novel peripheral metabolic pathways, revolutionizing the field of natural products biosynthesis (Nett et al. 2009).

As a genome sequence represents a catalogue of putative genes, many encoding enzymes, it may seem trivial to start reaching metabolic conclusions at the sequence level simply by functional annotation. However, prediction of an enzyme function from sequence alone does not provide metabolically relevant information: metabolism is not the result of enzymes acting independently, but the integration of these enzymes and their substrates/products within a complex network. Here is where Genome-Scale Metabolic-network Reconstruction (GSMRs) platforms provide a fundamental approach for a holistic and realistic interpretation of metabolism (Feist et al. 2009). Moreover, metabolic features cannot be interpreted from an evolutionary standpoint until the components of a gene catalogue are placed into the correct metabolic context through an *in silico* stoichiometric matrix prediction and experimental confirmation (Wagner 2009; Conrad et al. 2011).

Arguably, metabolism provides bacteria with the means to adapt to their ecological niches, implying a role for metabolism within bacterial speciation. Although many of the typical examples used to support this statement (such as the antibiotic activity of certain metabolites as a means to mediate bacterial interactions [Yim et al. 2007]) still need to be confirmed in real environmental conditions, this hypothesis finds support in evolutionary theory (Conrad et al. 2011). Therefore, the topology and composition of any given bacterial metabolic network must provide hints into the evolutionary history and biochemical adaptation of bacterial species, implying that this information may find an application within modern polyphasic taxonomic approaches. Furthermore, as phenotypes are the result of these evolutionary metabolic processes, this is in agreement with the strong emphasis on phenotypic characterisation that

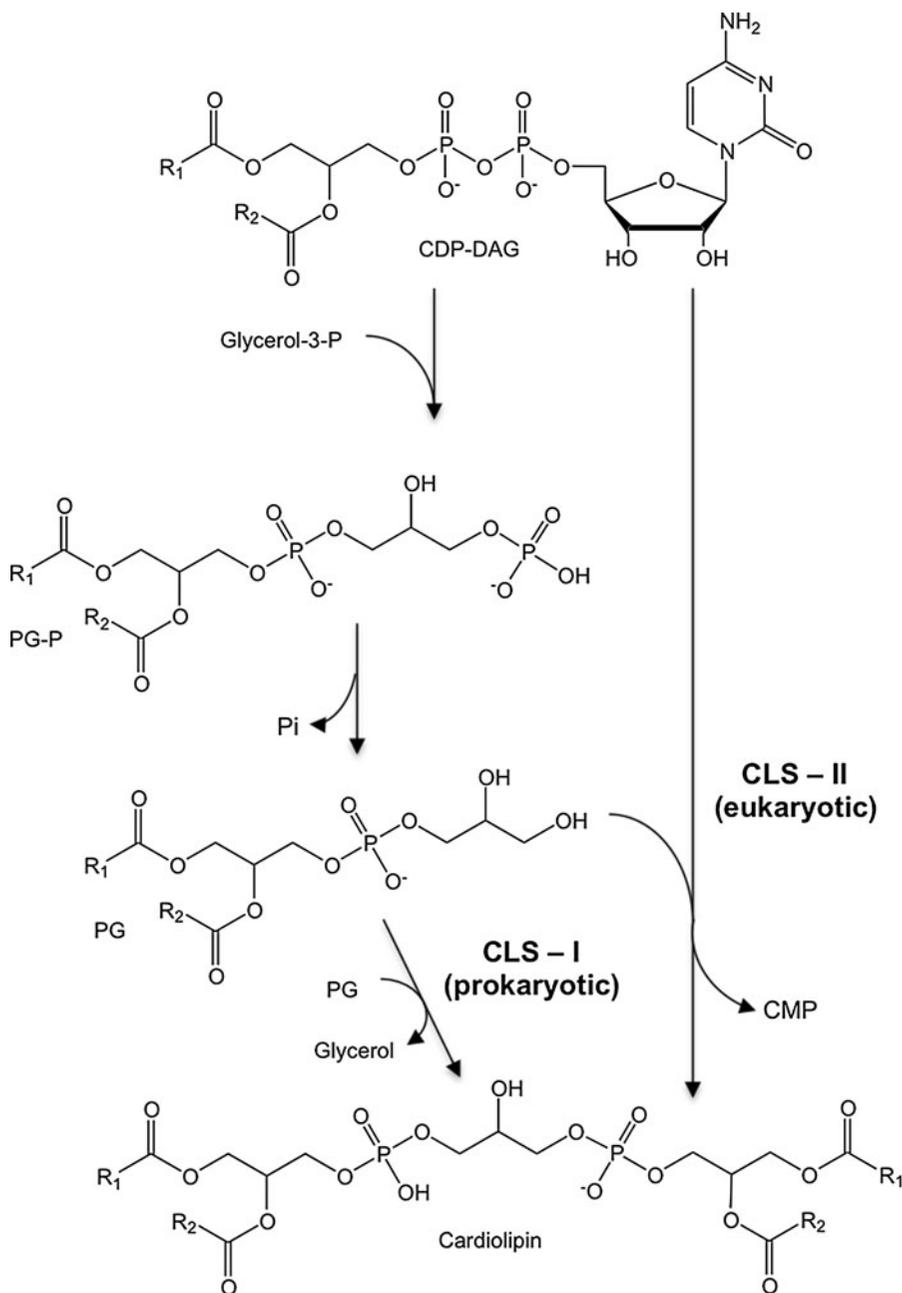
polyphasic taxonomy should embrace (Tindall et al. 2010).

Two metabolic scenarios with evolutionary and phenotypic consequences are developed in this perspective article. Both were revealed after actinobacterial GSMRs and argue in favour of the possibility of incorporating genome-scale views of metabolism into polyphasic taxonomy. The first example relates to pathway gaps and how these may be used as (chemo)taxons, whereas the second example uses enzymatic expansions with taxonomic resolution.

### Beyond the final product: pathways as (chemo)taxons

GSMRs have been valuable in identifying pathway gaps, where a gene expected to be present in a complete genome sequence cannot be found (Osterman and Overbeek 2003; Feist et al. 2009). For instance, some of the actinobacterial metabolic missing genes, identified by the GSMR of the model actinomycete *Streptomyces coelicolor* (Borodina et al. 2005), have been subsequently solved through experimentation (Hiratsuka et al. 2008; Marineo et al. 2008; Sandoval-Calderon et al. 2009). In these cases, the importance of novel enzymes filling pathway gaps is not limited to the final metabolite of the pathway, but also to the way in which the newly discovered analogous enzyme catalyses the synthesis of intermediary metabolites. Analogous enzymes may not only be different in sequence and structure, but also in the way they exert their enzymatic activities, to the point that different substrates and co-factors may be used. This is expected to have a bearing on the topology and function of the metabolic network.

An example of this scenario is given by the synthesis of cardiolipin in *Actinobacteria* (Fig. 1), where two classes of cardiolipin synthases (CL-S) have been reported: the prokaryotic (Nampoothiri et al. 2002) and eukaryotic (Sandoval-Calderon et al. 2009) versions. Comparative genome analyses with emphasis on these two versions of the *cls* genes (Supplementary Information) revealed that the prokaryotic version of CL-S is present only in *Corynebacterium* (the exception being *C. kroppenstedtii*, a lipophilic bacteria that lacks mycolic acids [Tauch et al. 2008]) and in certain deep-rooted

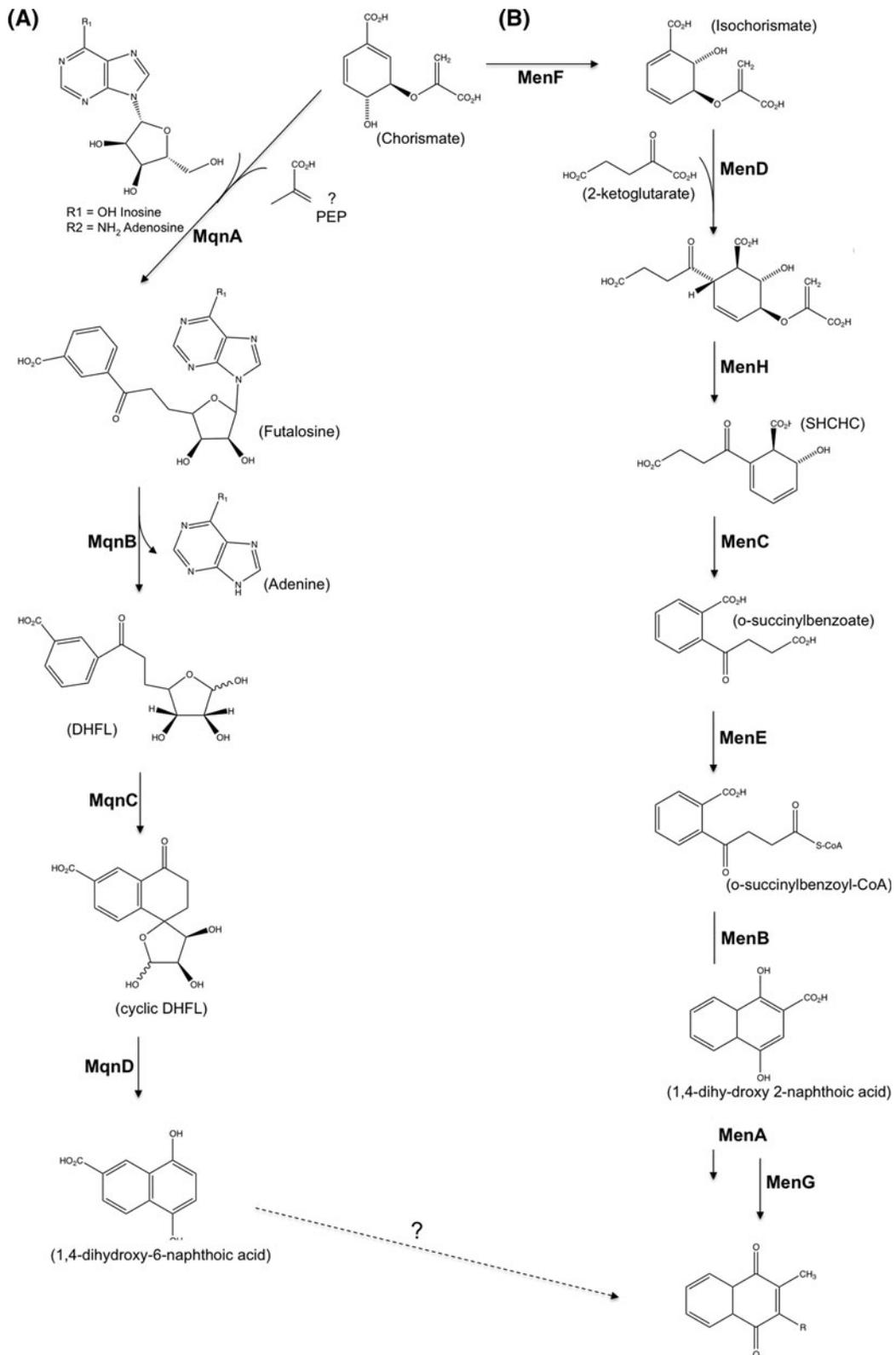


**Fig. 1** Analogous cardiolipin synthases. The prokaryotic and eukaryotic versions of the enzyme cardiolipin synthase are shown (CLS-I and CLS-II, respectively). Abbreviations to

actinobacteria). Enzymes of this class use as a substrate two molecules of phosphatidylglycerol (PG) to synthesise cardiolipin in a reversible fashion. However, in most actinobacteria, including *S. coelicolor*, where it was originally shown to be operative (Sandoval-Calderon et al. 2009), the eukaryotic-like

intermediaries are as follows: CDP-DAG cytidine diphosphate-diacylglycerol; PI phosphatidylinositol; PG phosphatidylglycerol; PG-P phosphatidylglycerol 3-phosphate

CL-S seems to be the rule. Interestingly, the eukaryotic-like CL-S synthesizes cardiolipin from one molecule of PG and one molecule of cytidine diphosphate-diacylglycerol (CDP-DAG) as the donor of the phosphatidyl group, implying not only different substrates but also an irreversible reaction.



◀ **Fig. 2** Convergent biosynthesis of menaquinones. Two alternative menaquinone biosynthetic pathways have been shown to be present in actinobacteria. **a** The futasolose biosynthetic pathway (*mqn* genes) and **b** the chorismate biosynthetic pathway (*men* genes). Abbreviations to intermediaries and enzymes are as follows: PEP, phosphoenolpyruvate; DHFL, dehydropyruvyl futasolose; SEPHCHC, 2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylic acid; SHCHC, 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate; MenA, 1,4-dihydroxy-2-naphthoate octaprenyltransferase; MenB, naphthoate synthase; MenC, O-succinylbenzoate synthase; MenD, 2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylate synthase; MenE, O-succinylbenzoic acid-CoA ligase; MenF, Isochorismate synthase; MenG, S-adenosylmethionine:2-demethylmenaquinone methyltransferase; MenH, Menaquinone biosynthesis methyltransferase; MqnABCD, enzyme names to be defined

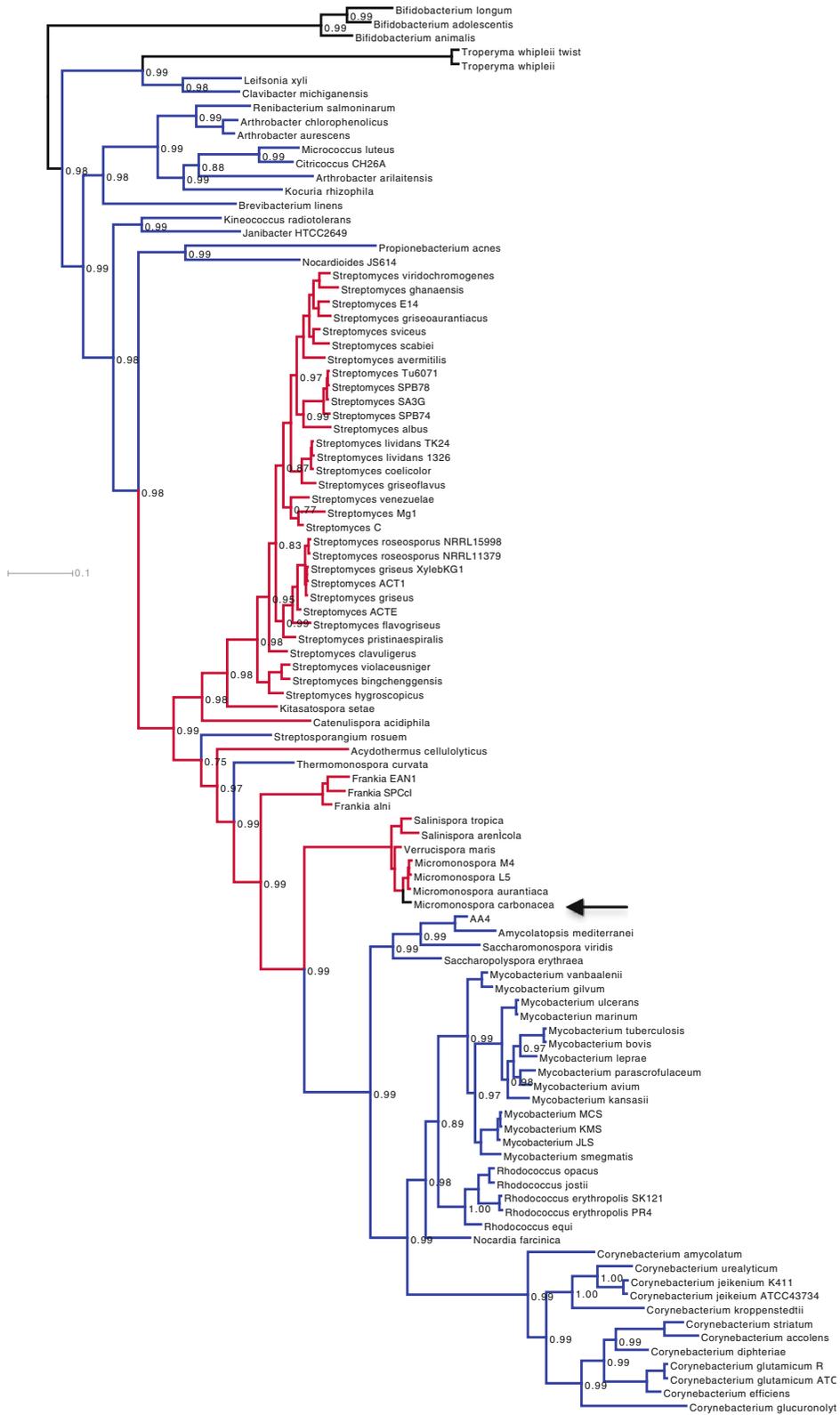
In the light of these differences it is reasonable to expect that the profiles of the polar phospholipids synthesized by these convergent biosynthetic pathways (Fig. 1), whose products and intermediary metabolites are commonly used as chemotaxons in *Actinobacteria*, will be determined by which of the analogous CL-S is present. This is a very interesting hypothesis that warrants reinvestigation of type strains using modern analytical techniques to accurately determine their quantitative phospholipids profiles. This would imply that a polyphasic taxonomic approach of determining what type of CL-S is present in any given newly characterised actinobacteria may be more relevant and informative than quantifying the final products of these pathways. Once more genome sequences from key organisms diverging at the nodes with alternative CL-S become available, this hypothesis could be experimentally tested. Nevertheless, it already is clear from this example that genome-scale notions of metabolism can inform on the bacterial speciation processes, and thus genomes interpreted in such a way have a bearing on the fourth tenet of microbial systematics.

Remarkably, these pathway gaps are not limited to single enzymes. Several complete convergent biosynthetic pathways, which account for many missing genes, have been reported (reviewed by Dairi et al. 2011). From a chemotaxonomic point-of-view, the convergent biosynthesis of the menaquinones (MK) stands-out and therefore it is analysed in further detail. The classic MK biosynthetic pathway (*men* genes; Fig. 2) starts from chorismate and it is present in most bacterial biological models, e.g. *Escherichia coli* and *Bacillus subtilis*. Comparative genome analysis centred on the *men* genes (Hiratsuka et al. 2008) showed

that this pathway is also present in some actinobacteria. However, this was not the case for many actinobacteria, including *Streptomyces*, which lack the *men* genes all together. This situation was at odds with the fact that MKs had been extensively used as chemotaxons in *Streptomyces* and related organisms (Collins et al. 1985). Again, a convergent pathway was first inferred from the GSMRs of *S. coelicolor* (Borodina et al. 2005) and this discrepancy was experimentally solved by the discovery of a completely novel MK biosynthetic pathway (Hiratsuka et al. 2008; Seto et al. 2008; Arakawa et al. 2011) that starts from futasolose or aminodeoxyfutasolose (*mqn* genes; Fig. 2).

Comparative genome analysis of a hundred actinobacteria, with an emphasis on the *men* and *mqn* genes, is used here to reveal the actinobacterial occurrence and taxonomic distribution of the alternative MK biosynthetic pathways (Fig. 3; see also Supplementary Information). The chorismate biosynthetic pathway is present in most of the deep-rooted organisms, such as *Arthrobacter*, but also in some of the most recently diverged genera, such as *Corynebacterium* and *Mycobacterium* (blue branches, Fig. 3). This observation suggests a differential gene gain or loss that is lineage-specific, something that can be indicative of speciation processes. Indeed, the occurrence of the *men* and *mqn* genes may help to clarify phylogenetic relationships. For instance, two organisms which possess the *men* genes, *Streptosporangium roseum* and *Thermomonospora curvata*, are located within branches populated by organisms that have the *mqn* genes. This incongruence in our RpoB species-tree suggests that the taxonomic coverage of organisms whose genomes have been sequenced is still limited. However, the proposed approach of using pathways as (chemo)taxons does solve the taxonomic relationships of these organisms.

Given that experimental determination of MKs for many actinobacteria are available (Collins et al. 1977; Collins et al. 1985) it is feasible to test the hypothesis that a correlation may exist between the type of pathway occurring and the type of MK that is being synthesized. In agreement with this, most actinobacteria synthesizing MKs through the futasolose pathway (*mqn* genes) have been reported to primarily synthesize MKs with nine isoprenoid units. This contrasts with those organisms using the chorismate pathway (*men* genes), which synthesize MKs with eight



◀ **Fig. 3** RpoB-based actinobacterial species-tree and taxonomic distribution of menaquinones and their biosynthetic genes. Coloured blue branches indicate organisms with *men* genes, and coloured red branches indicate organisms with *mqn* genes. Black branches denote organisms (e.g. *Bifidobacterium*) that lack MKs and their biosynthetic genes (Kindberg et al. 1987). *Micromonospora carbonacea*, whose genome encode for both the *men* and *mqn* genes, is marked with an arrow. Details on the parameters used for construction of the phylogenetic tree are provided as [Supplementary Information](#). The tree was constructed with MrBayes and posterior probabilities are provided at the nodes (Huelsenbeck and Ronquist 2001). (Color figure online)

isoprenoid units (Fig. 3; see also Supplementary Information). Despite the fact that evidence of co-occurrence of both pathways could not be found in previous analyses (Dairi et al. 2011), our comparative analysis suggests that the genome of *Micromonospora carbonacea* may encode for both the *men* and *mqn* genes. This observation makes this organism a suitable biological model for testing the hypothesised pathway—MK relationship, supporting the idea that GSMRs may become part of polyphasic taxonomy. Indeed, early data suggests that the genus *Micromonospora* is more heterogeneous with respect to MK composition than other actinomycete genera that exhibit similar high levels of intrageneric relatedness (Koch et al. 1996; Trujillo et al. 2005).

### A link between enzyme evolution and taxonomy

Once GSMRs have been established for model organisms, this information can be compiled into the ‘pan-metabolism’ of any given bacterial group. For *Actinobacteria*, the GSMRs of *S. coelicolor* (Borodina et al. 2005), *M. tuberculosis* (Jamshidi and Palsson 2007) and *C. glutamicum* (Kjeldsen and Nielsen 2009) have been reported. This data provides a universe of protein sequences that can be used to mine genomic databases in search of distinctive metabolic features, with taxonomic resolution. Inspired by the use of microbial natural products as chemotaxons (Frisvad et al. 2008), we propose herein the use of conserved enzymatic expansions (EEs), as a distinctive metabolic feature for the resolution of taxonomic relationships in natural products-producing organisms. This proposal finds its foundation in evolutionary theory, which postulates gene duplication leading to enzymatic expansion and divergence as the main source of

novel enzyme activities. Furthermore, looking for conserved expansions may add metabolic markers with real physiological implications to those already identified as lineage-specific actinobacterial conserved genes (Gao & Gupta, this issue).

Our comparative phylogenomic analysis within the context of the biosynthesis of natural products, which will be reported elsewhere in more detail (FBG & PCM, manuscript in preparation), used as a proof-of-concept nine selected central metabolic pathways, plus about one hundred actinobacterial genome sequences and the GSMRs of *S. coelicolor*, *M. tuberculosis* and *C. glutamicum*. The nine central metabolic pathways used were selected due to the fact that these pathways are known to provide precursors for the synthesis of natural products. Moreover, functional annotation of the enzymes taking part in these pathways, using only sequence data, is straightforward with little or no ambiguity. As can be seen in Figure 4, after a statistical treatment of the output sequences using as background the phylogeny of the *Actinobacteria*, the suborders *Micromonosporineae*, *Streptomycineae* and *Pseudonocardineae*, as well as the *Nocardiaceae* family, are clearly enriched for EEs within these pathways. Given that these actinobacterial families include species that are renowned for synthesizing a broad range of natural products, it is tempting to speculate that natural product biosynthetic diversity and the occurrence of EEs may be related traits.

From a systematics point-of-view, it is proposed that these traits can be used to solve taxonomic relationships. Indeed, strain AA4, which was originally thought to be a streptomycete but groups in the same clade as the *Pseudonocardineae* within our RpoB-based species tree (Fig. 3), shows a different EEs profile than the rest of the streptomycetes (Fig. 4). In agreement with this, it has recently been reported that strain AA4 actually belongs to the genus *Amycolatopsis* (Seyedsayamdost et al. 2011). EE traits may also be used for understanding actinobacterial speciation involving metabolic adaptation, which may have a strong bearing on the fourth tenet of microbial systematics. For instance, as *Streptomyces* species are probably the most important natural products-producing genus, it is interesting to note that certain EEs, not found in the context of natural products biosynthesis (FBG & PCM, unpublished results), are highly conserved in these organisms. These EEs include the glycolytic enzymes phosphoglycerate mutase (Pgm)



novel taxa, getting involved in GSMRs represents an area of opportunity. This in turn may not only help to revalue prokaryotic systematics within the broader microbiological community, but also may provide the bacterial taxonomist with more and better (chemo)taxons related to pathway gaps, analogous enzymes and complete novel pathways. Next generation genomics-driven taxons can inform studies of bacterial speciation, and will be fundamentally important to reach Comprehending Microbial Diversity, the fourth tenet of microbial systematics.

**Acknowledgments** This work was supported by Conacyt, Mexico (grant No. 82319).

## References

- Arakawa C, Kuratsu M, Furihata K, Hiratsuka T, Itoh N, Seto H, Dairi T (2011) Diversity of the early step of the futasoline pathway. *Antimicrob Agents Chemother* 55(2):913–916
- Borodina I, Krabben P, Nielsen J (2005) Genome-scale analysis of *Streptomyces coelicolor* A3(2) metabolism. *Genome Res* 15(6):820–829
- Collins MD, Pirouz T, Goodfellow M, Minnikin DE (1977) Distribution of menaquinones in actinomycetes and corynebacteria. *J Gen Microbiol* 100(2):221–230
- Collins MD, Goodfellow M, Minnikin DE, Alderson G (1985) Menaquinone composition of mycolic acid-containing actinomycetes and some sporoactinomycetes. *J Appl Bacteriol* 58(1):77–86
- Conrad TM, Lewis NE, Palsson BO (2011) Microbial laboratory evolution in the era of genome-scale science. *Mol Syst Biol* 7:509
- Dairi T, Kuzuyama T, Nishiyama M, Fujii I (2011) Convergent strategies in biosynthesis. *Nat Prod Rep* 28(6):1054–1086
- Feist AM, Herrgard MJ, Thiele I, Reed JL, Palsson BO (2009) Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol* 7(2):129–143
- Frisvad JC, Andersen B, Thrane U (2008) The use of secondary metabolite profiling in chemotaxonomy of filamentous fungi. *Mycol Res* 112(Pt 2):231–240
- Hiratsuka T, Furihata K, Ishikawa J, Yamashita H, Itoh N, Seto H, Dairi T (2008) An alternative menaquinone biosynthetic pathway operating in microorganisms. *Science* 321(5896):1670–1673
- Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17(8):754–755
- Jamshidi N, Palsson BO (2007) Investigating the metabolic capabilities of *Mycobacterium tuberculosis* H37Rv using the in silico strain iNJ661 and proposing alternative drug targets. *BMC Syst Biol* 1:26
- Kindberg C, Suttie JW, Uchida K, Hirauchi K, Nakao H (1987) Menaquinone production and utilization in germ-free rats after inoculation with specific organisms. *J Nutr* 117(6):1032–1035
- Kjeldsen KR, Nielsen J (2009) *In silico* genome-scale reconstruction and validation of the *Corynebacterium glutamicum* metabolic network. *Biotechnol Bioeng* 102(2):583–597
- Koch C, Kroppenstedt RM, Stackebrandt E (1996) Intrageneric relationships of the actinomycete genus *Micromonospora*. *Int J Syst Bacteriol* 46(2):383–387
- Marineo S, Cusimano MG, Limauro D, Cotichio G, Puglia AM (2008) The histidinol phosphate phosphatase involved in histidine biosynthetic pathway is encoded by SCO5208 (*hisN*) in *Streptomyces coelicolor* A3(2). *Curr Microbiol* 56(1):6–13
- Nampoothiri KM, Hoischen C, Bathe B, Mockel B, Pfefferle W, Krumbach K, Sahm H, Eggeling L (2002) Expression of genes of lipid synthesis and altered lipid composition modulates L-glutamate efflux of *Corynebacterium glutamicum*. *Appl Microbiol Biotechnol* 58(1):89–96
- Nett M, Ikeda H, Moore BS (2009) Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat Prod Rep* 26(11):1362–1384
- Osterman A, Overbeek R (2003) Missing genes in metabolic pathways: a comparative genomics approach. *Curr Opin Chem Biol* 7(2):238–251
- Sandoval-Calderon M, Geiger O, Guan Z, Barona-Gomez F, Sohlenkamp C (2009) A eukaryote-like cardiolipin synthase is present in *Streptomyces coelicolor* and in most actinobacteria. *J Biol Chem* 284(26):17383–17390
- Seto H, Jinnai Y, Hiratsuka T, Fukawa M, Furihata K, Itoh N, Dairi T (2008) Studies on a new biosynthetic pathway for menaquinone. *J Am Chem Soc* 130(17):5614–5615
- Seyedsayamdost MR, Traxler MF, Zheng SL, Kolter R, Clardy J (2011) Structure and biosynthesis of amyachelin, an unusual mixed-ligand siderophore from *Amycolatopsis* sp. AA4. *J Am Chem Soc* 133(30):11434–11437
- Staley JT (2010) Comprehending microbial diversity: the fourth goal of microbial taxonomy. *The Bulletin of BISMis* 1(1):1–5
- Tauch A, Schneider J, Szczepanowski R, Tilker A, Viehoyer P, Gartemann KH, Arnold W, Blom J, Brinkrolf K, Brune I, Gotker S, Weisshaar B, Goesmann A, Droge M, Puhler A (2008) Ultrafast pyrosequencing of *Corynebacterium kroppenstedtii* DSM44385 revealed insights into the physiology of a lipophilic corynebacterium that lacks mycolic acids. *J Biotechnol* 136(1–2):22–30
- Tindall BJ, Rossello-Mora R, Busse HJ, Ludwig W, Kampfer P (2010) Notes on the characterization of prokaryote strains for taxonomic purposes. *Int J Syst Evol Microbiol* 60(Pt 1):249–266
- Trujillo ME, Fernández-Molinero C, Velázquez E, Kroppenstedt RM, Schumann P, Mateos PF, Martínez-Molina E (2005) *Micromonospora mirobrigensis* sp. nov. *Int J Syst Evol Microbiol* 55:877–880
- Wagner A (2009) Evolutionary constraints permeate large metabolic networks. *BMC Evol Biol* 9:231
- Yim G, Wang HH, Davies J (2007) Antibiotics as signalling molecules. *Philos Trans R Soc Lond B Biol Sci* 362(1483):1195–1200