



**HAL**  
open science

## Interaction network among factors involved in heterocyst-patterning in cyanobacteria

Xiaomei Xu, Raphaël Rachedi, Maryline Foglino, Emmanuel Talla, Amel Latifi

► **To cite this version:**

Xiaomei Xu, Raphaël Rachedi, Maryline Foglino, Emmanuel Talla, Amel Latifi. Interaction network among factors involved in heterocyst-patterning in cyanobacteria. *Molecular Genetics and Genomics*, 2022, 10.1007/s00438-022-01902-5 . hal-03670191

**HAL Id: hal-03670191**

**<https://amu.hal.science/hal-03670191>**

Submitted on 17 May 2022

**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.

# Molecular Genetics and Genomics

## Interaction network among factors involved in heterocyst-patterning in cyanobacteria --Manuscript Draft--

<b>Manuscript Number:</b>	MGAG-D-21-00741R2	
<b>Full Title:</b>	Interaction network among factors involved in heterocyst-patterning in cyanobacteria	
<b>Article Type:</b>	Original Article	
<b>Corresponding Author:</b>	Amel Latifi, Ph.D Aix-Marseille Universite Marseille, FRANCE	
<b>Corresponding Author Secondary Information:</b>		
<b>Corresponding Author's Institution:</b>	Aix-Marseille Universite	
<b>Corresponding Author's Secondary Institution:</b>		
<b>First Author:</b>	Xiaomei Xu	
<b>First Author Secondary Information:</b>		
<b>Order of Authors:</b>	Xiaomei Xu	
	Raphaël Rachedi	
	Maryline Foglino	
	Emmanuel Talla	
	Amel Latifi, Ph.D	
<b>Order of Authors Secondary Information:</b>		
<b>Funding Information:</b>	Agence Nationale de la Recherche (ANR-21-CE20-0025-01)	Pr. Amel Latifi
<b>Abstract:</b>	<p>The genetically regulated pattern of heterocyst formation in multicellular cyanobacteria represents the simplest model to address how patterns emerge and are established, the signals that control them, and the regulatory pathways that act downstream. Although numerous factors involved in this process have been identified, the mechanisms of action of many of them remain largely unknown. The aim of this study was to identify specific relationships between 14 factors required for cell differentiation and pattern formation by exploring their putative physical interactions in the cyanobacterium model <i>Nostoc</i> sp. PCC 7120 and by probing their evolutionary conservation and distribution across the cyanobacterial phylum. A bacterial two-hybrid assay indicated that 10 of the 14 factors studied here are engaged in more than one protein-protein interaction. The transcriptional regulator PatB was central in this network as it showed the highest number of binary interactions. A phylum-wide genomic survey of the distribution of these factors in cyanobacteria showed that they are all highly conserved in the genomes of heterocyst-forming strains, with the PatN protein being almost restricted to this clade. Interestingly, eight of the factors that were shown to be capable of protein interactions were identified as key elements in the evolutionary genomics analysis. These data suggest that a network of 12 proteins may play a crucial role in heterocyst development and patterning. Unravelling the physical and functional interactions between these factors during heterocyst development will certainly shed light on the mechanisms underlying pattern establishment in cyanobacteria.</p>	
<b>Response to Reviewers:</b>	<p>Dear editor, in this revised version of our manuscript we introduced all the corrections asked by the editor and Reviewer 1. The modifications in the text are marked in yellow, We thank the reviewers and the communicating editor for their valuable reading of our publication. We hope that this revised manuscript will be suited for publication.</p>	

[Click here to view linked References](#)

# Interaction network among factors involved in heterocyst-patterning in cyanobacteria

Xiaomei Xu<sup>1\*</sup>, Raphaël Rachedi<sup>1\*</sup>, Maryline Foglino<sup>1</sup>, Emmanuel Talla<sup>1\*\*</sup>, Amel Latifi<sup>1\*\*</sup>

<sup>1</sup>: Aix Marseille Univ, CNRS, IMM, LCB, Laboratoire de Chimie Bactérienne, Marseille, France

\*: these authors contributed equally to the work

\*\* : correspondence:

Pr Amel Latifi

[latifi@imm.cnrs.fr](mailto:latifi@imm.cnrs.fr)

Dr Emmanuel Talla

[talla@imm.cnrs.fr](mailto:talla@imm.cnrs.fr)

**Key words:** Cell differentiation, Cyanobacteria, Genomics, Pattern establishment, Protein-Protein interaction.

**Running title:** Protein-protein interactions and heterocyst pattern

30 **Abstract**

1  
2  
3 31 The genetically regulated pattern of heterocyst formation in multicellular cyanobacteria represents the  
4  
5 32 simplest model to address how patterns emerge and are established, the signals that control them, and  
6  
7 33 the regulatory pathways that act downstream. Although numerous factors involved in this process have  
8  
9  
10 34 been identified, the mechanisms of action of many of them remain largely unknown. The aim of this  
11  
12 35 study was to identify specific relationships between 14 factors required for cell differentiation and  
13  
14 36 pattern formation by exploring their putative physical interactions in the cyanobacterium model *Nostoc*  
15  
16 37 *sp.* PCC 7120 and by probing their evolutionary conservation and distribution across the cyanobacterial  
17  
18 38 phylum. A bacterial two-hybrid assay indicated that 10 of the 14 factors studied here are engaged in  
19  
20 39 more than one protein-protein interaction. The transcriptional regulator PatB was central in this network  
21  
22 40 as it showed the highest number of binary interactions. A phylum-wide genomic survey of the  
23  
24 41 distribution of these factors in cyanobacteria showed that they are all highly conserved in the genomes  
25  
26 42 of heterocyst-forming strains, with the PatN protein being almost restricted to this clade. Interestingly,  
27  
28 43 eight of the factors that were shown to be capable of protein interactions were identified as key elements  
29  
30 44 in the evolutionary genomics analysis. These data suggest that a network of 12 proteins may play a  
31  
32 45 crucial role in heterocyst development and patterning. Unravelling the physical and functional  
33  
34 46 interactions between these factors during heterocyst development will certainly shed light on the  
35  
36 47 mechanisms underlying pattern establishment in cyanobacteria.  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

## 54 Introduction

55 Spatial patterns of cellular differentiation occur in many developmental organisms across the tree of life,  
56 and in most cases, key aspects of the molecular interactions that are involved in patterning are poorly  
57 understood. In this work, we questioned the possible involvement of protein-protein interactions in  
58 patterning. Our study model is the development of heterocysts in multicellular and diazotrophic  
59 cyanobacteria. Indeed, in prokaryotes, heterocyst formation in multicellular cyanobacteria represents  
60 the simplest model to address one-dimensional patterns based on repeated peaks of diffusible  
61 morphogens. These peaks are generated by the activity of a transcriptional regulator that activates the  
62 expression of the gene encoding its own inhibitor (the diffusible morphogen); thus, this loop has been  
63 called the reaction-diffusion model (Fig. 1a) (Turing 1952).

64 Heterocysts are nearly anoxic cells that form semi-regularly along the filaments and function to host the  
65 oxygen-sensitive nitrogenase. Therefore, the developmental program in these heterocyst-forming  
66 bacteria allows the occurrence of two antagonistic metabolic activities (i.e., oxygenic photosynthesis  
67 and nitrogen fixation) in the same organism (Meeks and Elhai 2002; Kumar et al. 2010; Herrero et al.  
68 2016). Much of the knowledge about cell differentiation in cyanobacteria stems from studies of the  
69 model bacterium *Anabaena/Nostoc* sp. PCC 7120 (termed *Nostoc* hereafter). Combined nitrogen  
70 starvation triggers the activation of the expression of the *hetR* gene, which encodes the master and  
71 specific transcriptional activator of cell differentiation (Buikema and Haselkorn 1991). HetR is essential  
72 for cell differentiation, as its absence abolishes differentiation and its overexpression induces the  
73 differentiation of multiple contiguous heterocysts (MCH) under combined nitrogen starvation and  
74 allows differentiation even under non-permissive conditions (Buikema and Haselkorn 1991). The  
75 regulation of *hetR* transcription is complex and involves the global regulator NtcA and a positive  
76 autoregulation loop mediated by HetR itself (Black et al. 1993; Frias et al. 1994; Rajagopalan and  
77 Callahan 2010). The inhibition of HetR activity in the neighboring cells of the heterocysts initiates and  
78 maintains the pattern of differentiation throughout the growth of the filaments. Three genes (*patS*, *patX*,  
79 and *hetN*), the expression of which is induced in developing cells, negatively control this process (Black  
80 and Wolk 1994; Yoon and Golden 1998; Yoon and Golden 2001; Higa et al. 2012; Corrales-Guerrero

81 et al. 2014b; Elhai and Khudyakov 2018; Khudyakov et al. 2020). Deletion of the *patS* gene also leads  
1 to  
2 82 to the formation of multiple contiguous heterocysts (Fig. 1b). A common feature of the PatS, PatX, and  
3  
4 83 HetN proteins is that they contain a conserved hexapeptide (ERGSGR for PatS and HetN, and HRGTGR  
5  
6 84 for PatX). The RGSGR peptide inactivates the DNA-binding activity of HetR (Huang et al. 2004;  
7  
8 85 Feldmann et al. 2011). These peptides are considered as diffusible morphogens that ensure patterning  
9  
10 86 by lateral inhibition (Corrales-Guerrero et al. 2013). Recently, HetR was shown to interact with a  
11  
12 87 pentapeptide-repeat-containing protein called HetL at the same interface used by the ERGSGR peptide  
13  
14 88 (Xu et al. 2020). Moreover, the transcription of *hetL*, which is a heterocyst-specific gene, is activated  
15  
16 89 by HetR (Xu et al. 2020). *hetL* overexpression suppresses the inhibitory effect of PatS (Liu and Golden  
17  
18 90 2002) and the interaction of HetL with HetR provides protection against PatS (Xu et al. 2020), thus  
19  
20 91 explaining how the developing cell becomes immune to PatS/HetN/PatX inhibition in the heterocyst  
21  
22 92 (Fig. 1a).

23  
24  
25  
26  
27 93 Correct patterning requires the function of several genes that have been termed *pat* genes (Fig. 1b). The  
28  
29 94 phenotype of mutants of these genes provides an important indication of their role and establishes a  
30  
31 95 starting point for the investigation of how the pattern is established. These genes include *patA* and *patL*,  
32  
33 96 the mutation of which results in the formation of rare heterocysts only at the ends of the filaments (Liang  
34  
35 97 et al. 1992; Liu and Wolk 2011). PatA is a two-domain protein with a C-terminal region similar to the  
36  
37 98 CheY-receiver domain and a conserved N-terminal domain called PATAN (Makarova et al. 2006). PatL  
38  
39 99 is a pentapeptide-repeat-containing protein with two transmembrane domains (Liu and Wolk 2011). In  
40  
41 100 addition to sharing similar mutant phenotypes, the PatA and PatL proteins interact, as shown by yeast-  
42  
43 101 two-hybrid assays (Liu and Wolk 2011). A cluster of three genes (*hetZ*, *patU3*, and *patU5*) that are  
44  
45 102 expressed in the heterocysts and encode proteins of unknown function modulates heterocyst formation  
46  
47 103 and pattern establishment (Zhang et al. 2007). Inactivation of *patU3* leads to an MCH phenotype and  
48  
49 104 restores heterocyst formation in a *patA* mutant background (Zhang et al. 2007). A *hetZ* mutant of *Nostoc*  
50  
51 105 is unable to differentiate heterocysts (Zhang et al. 2007; Videau et al. 2018). HetZ action has been  
52  
53 106 proposed to be involved in its interaction with PatU3 and, to a lesser extent, with HetR (Videau et al.  
54  
55 107 2018; Zhang et al. 2018; Du et al. 2020). In *N. punctiforme*, the deletion of the *patN* gene, which encodes  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

108 a putative membrane protein, increases heterocyst frequency and decreases the number of vegetative  
109 cells between two heterocysts, thus leading to a multiple singular heterocyst (MSH) phenotype (Risser  
110 et al. 2012) (Fig. 1b). A similar phenotype has been observed in a *patN* mutant of *Nostoc* (Abed et al.  
111 2009). PatN was proposed to modulate the competency of a vegetative cell to differentiate into a  
112 heterocyst (Risser et al. 2012); therefore, it also affects patterning. Interestingly, the MSH phenotype  
113 has been observed in a mutant of the *patB* (*cnfR*) gene, the product of which is predicted to be a  
114 transcriptional regulator with a DNA-binding helix-turn-helix at its C-terminal extremity and two  
115 putative ferredoxin domains at its N terminus (Jones et al. 2003). In *Anabaena variabilis*, the homolog  
116 of PatB (CnfR1) activates the transcription of the heterocyst specific nitrogenase-encoding genes (Pratte  
117 and Thiel 2016). Furthermore, in the non-heterocyst-forming cyanobacterium *Leptolyngbya boryana*,  
118 CnfR is essential for the transcription of the *nif* genes and for the diazotrophic growth under micro-oxic  
119 conditions (Tsujiimoto et al. 2014). In *Nostoc*, the transcription of the *nif* genes has been shown to be  
120 abolished in a *patB* (*cnfR*) mutant (Kurio et al. 2020) and the DNA-binding sites of this regulator has  
121 been predicted in the promoter of the *nif* operon (Brenes-Alvarez et al. 2019). The function of CnfR/PatB  
122 as activator of the *nif* genes seems therefore conserved in cyanobacteria, but the manner in which PatB  
123 exerts this negative effect on heterocyst spacing in *Nostoc* is unknown. Two other *pat* genes, *patC* and  
124 *patD*, both encoding proteins of unknown functions, seem to negatively regulate patterning, as their  
125 deletion increases heterocyst frequency (Corrales-Guerrero et al. 2014a; Wang et al. 2019).

126 In addition to the Pat proteins, other factors involved in the early steps of the differentiation process can  
127 affect pattern formation. The *hetC* gene, which encodes a putative bacterial ABC protein transporter, is  
128 involved in the regulation of the early steps of heterocyst differentiation (Khudyakov and Wolk 1997;  
129 Corrales-Guerrero et al. 2014a; Videau et al. 2015). *hetC* inactivation mutants exhibit a pattern of spaced  
130 cells that are still able to divide (Khudyakov and Wolk 1997; Xu and Wolk 2001). These cells are likely  
131 to represent an early stage of heterocyst differentiation; thus, HetC has been suggested to be required  
132 for the transition of heterocysts to their terminal non-dividing state (Xu and Wolk 2001). *hetF*, which  
133 encodes a putative protease, is another gene that is essential for heterocyst formation, as a *hetF* mutant

134 does not form heterocysts and has altered morphology, whereas *hetF* overexpression results in an MCH  
135 phenotype (Wong and Meeks 2001; Risser and Callahan 2008).

136 Shortly after the initiation step and pattern establishment, the decision to differentiate is irreversibly  
137 fixed, which is known as commitment. The HetP protein (encoded by the *alr2818* gene) and its  
138 homologs (encoded by the *asl1930*, *alr3234*, and *alr2902* genes) have been proposed as important  
139 players in the regulation of this step (Videau et al. 2016). The inactivation of the *hetP* gene affects the  
140 differentiation process downstream of initiation and patterning, as the mutant develops heterocysts,  
141 albeit at a lower frequency and at a later time than the wild type strain (Fernandez-Pinas et al. 1994;  
142 Higa and Callahan 2010). Unlike *hetP*, inactivation of the genes encoding the 3 HetP homologs does  
143 not impair cell differentiation, indicating that they are not required for heterocyst formation or patterning  
144 (40). Overexpression of *hetP* bypasses the absence of HetZ and vice versa, which indicates a functional  
145 overlap between these factors (Zhang et al. 2018). This observation highlights the connection between  
146 patterning and commitment and places HetZ at the intersection between the two steps.

147 In summary, pattern establishment has clearly been shown to be genetically controlled, and the  
148 involvement of several factors in this process has been demonstrated (Fig. 1b). However, their function  
149 and their potential molecular interactions remain poorly known. Because protein–protein interaction  
150 approaches have been shown to be useful for identifying protein functions and understanding several  
151 cellular pathways, we used a bacterial two-hybrid assay (BACTH) to address the possible binary  
152 interactions among the 14 factors involved in the early, commitment, and patterning stages in *Nostoc*.  
153 In addition, we performed a phylum-wide analysis of the conservation of these factors in 212  
154 cyanobacterial genomes. A comparative genomics analysis of the genes encoding these factors across  
155 cyanobacteria supported the contention that they likely were present in the common ancestor of  
156 cyanobacteria and that they evolved through gene loss in clades of cyanobacteria that do not form  
157 heterocysts. Taken together, our data suggests that a set of 12 proteins (DevH, HetC, HetF', HetL, HetP,  
158 HetR, HetZ, PatA, PatB, PatL, PatN, and PatU3), which encoding genes are conserved in heterocyst-  
159 forming cyanobacterial genomes, are associated through multiple cross-talk interactions. Our findings  
160 provide a framework for a better understanding of heterocyst pattern establishment.

## 161 **Materials and methods**

### 162 **Bacterial strains and growth media**

163 The *E. coli* K-12 strain DH5 alpha (Invitrogen) was used in all the cloning steps. BACTH  
164 complementation assays were carried out with the *E. coli cya* strain BTH101 (F<sup>-</sup>, *cya-99*, *araD139*,  
165 *galE15*, *galK16*, *rpsL1* (Str<sup>r</sup>), *hsdR2*, *mcrA1*, *mcrB1*) (Karimova et al. 1998). Bacteria were grown at  
166 30°C in Luria-Bertani (LB) broth supplemented with ampicillin at 100 µg/ml and kanamycin at 50 µg/ml  
167 when needed. Screening for the ability to complement the *cya* mutation was performed on LB agar plates  
168 supplemented with 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-Gal; 40 µg/ml), isopropyl-β-D-  
169 galactopyranoside (IPTG; 0.5 mM), and the appropriate antibiotics. The *E. coli* K-12 Lemo21(DE3)  
170 strain (Biolabs) was used for protein production and purification. Spectinomycin (50 µg/ml) was used  
171 for the selection of pCDF Duet plasmid (Novagen) and derivatives. Kanamycin (100 µg/ml) for the  
172 selection of pLIC07 plasmid (Michel-Souzy et al. 2018) and derivatives.

### 173 **Plasmid constructions**

174 For BACTH analysis: the genes involved in heterocyst differentiation and patterning were amplified  
175 from genomic DNA of *Nostoc* using the primers listed in Table S1. The recombinant plasmids were  
176 constructed by inserting gene sequences of interest in pKT25 (T25 at N-terminus), pKNT25 (T25 at the  
177 C-terminus), pUT18 (T18 at the C-terminus) and PUT18C (T18 at the N-terminus) (Karimova et al.  
178 1998). All the open reading frames (ORFs) of the analyzed genes, except *patN*, were inserted between  
179 PstI and EcoRI restriction sites of the BACTH plasmids. *patN* ORF was inserted between PstI and KpnI  
180 restriction sites.

181 For pull-down assays: the *hetR*, *hetL* or *hetF*' sequences fused at their 3' extremities to the sequence  
182 coding the 89-106 amino acids of the human influenza hemagglutinin protein (HA-tag) were synthesized  
183 by Genewiz (<https://www.genewiz.com/en-GB/>). The obtained genes were inserted between NdeI and  
184 XhoI restriction sites of the pCDF Duet plasmid. The *patB* coding sequence was fused to a 6xHis tag at  
185 its 5' extremity. For this, the *patB* ORF was amplified by PCR using pLIC07-*patB*-Fwd and pLIC07-

186 patB-Rev primers (Table S1) and inserted between the BsaI restriction sites of the pLIC07 plasmid using  
1 the SLIC method (Li and Elledge 2012).  
2  
3

4  
5 188 All the constructs were analyzed by sequencing. The plasmids used in this study are listed in Table S2.  
6

### 7 8 189 **SDS-PAGE and immunodetection** 9

10 190 Proteins were separated on 4-20% polyacrylamide gradient gels (NuSep) and transferred onto  
11  
12 191 nitrocellulose membranes using a semidry blotting apparatus. Membranes were blocked with 3% bovine  
13  
14 192 serum albumin (BSA) in TBST (10 mM Tris, 150 mM NaCl, 0.05% Tween 20). Membranes were  
15  
16 193 incubated with monoclonal anti-HA (Enzo) or anti-His (Qiagen) antibodies, diluted at 1:5000 and  
17  
18 194 1:2000, respectively in TBST, 3% BSA. After three-step washes in TBST, the membranes were  
19  
20 195 incubated with peroxidase-coupled anti-mouse antibodies (1:5000, Promega). The detection was  
21  
22 196 performed using the enhanced chemiluminescence reagents (Cytiva). The membranes were scanned  
23  
24 197 using ImageQuant TL analysis software (GE Healthcare).  
25  
26  
27  
28

### 29 198 **Coproduction of proteins in *E. coli* and Nickel-affinity chromatography** 30 31

32 199 Lemo21 (DE3) competent cells were co-transformed with pCDF Duet, pLIC07, and their derivatives.  
33  
34 200 The recombinant clones were grown until reaching an optical density at 600 nm of 0.5 at 37 °C in LB  
35  
36 201 medium with appropriate antibiotics (50 µg/ml spectinomycin, 100 µg/ml kanamycin). Induction was  
37  
38 202 performed with 0.5 mM IPTG for 3 hours at 30 °C. Bacteria were collected by centrifugation and broken  
39  
40 203 using a French press (10,000 p.s.i.) in cold buffer (100 mM Tris-HCl, pH 7.3, 400 mM NaCl, 0.5 mg/ml  
41  
42 204 lysozyme, 20 µg/ml DNase, 1X cOmplete™ EDTA-free protease inhibitor cocktail (Roche). The lysate  
43  
44 205 was cleared by ultracentrifugation. The cleared lysate was loaded onto a 1-ml HiFliQ Ni-NTA FPLC  
45  
46 206 column (Generon) using an ÄKTA Prime apparatus. After a washing step (10 ml of buffer: 100 mM  
47  
48 207 Tris-HCl, pH 7.3, 400 mM NaCl), the immobilized proteins were eluted in buffer: 100 mM Tris-HCl,  
49  
50 208 pH 7.3, 400 mM NaCl, 500 mM imidazol. The loaded flow-through, and elution fractions were analyzed  
51  
52 209 by SDS-PAGE and immunodetection.  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

## 211 **Bacterial two hybrid assays**

1  
2  
3 212 Bacterial two-hybrid assays were performed following the procedure described by Karimova et al  
4  
5 213 (1998) (Karimova et al. 1998). Briefly, after co-transforming the BTH101 strain with the two plasmids  
6  
7 214 expressing the T18- and T25- fusions, LB plates containing ampicillin and kanamycin were incubated  
8  
9 215 at 30° C for 2 days. For each assay, 10 independent colonies were inoculated in 3 ml of LB medium  
10  
11 216 supplemented with ampicillin, kanamycin, and 0.5 mM IPTG and incubated at 30 °C overnight.  
12  
13  
14 217  $\beta$ -galactosidase activity was then determined as previously described (Zubay et al. 1972).  
15  
16

## 17 218 **Bioinformatics analysis**

### 19 219 **Datasets**

20  
21  
22 220 The genome data of 212 cyanobacterial strains available in March 2019 were downloaded from NCBI  
23  
24 221 ftp site (<ftp://ftp.ncbi.nih.gov/genomes/>) and constituted the primary data source (Table S3, Sheet 1).  
25  
26 222 This data source includes all complete genomes in the highest levels of assembly, their refseq category  
27  
28 223 (reference or representative genome, or none) as well as their annotation features and taxonomy lineages.  
29  
30  
31 224 The ability of each strain to form heterocysts was provided by the literature (Rippka R 1979; Henson  
32  
33 225 B.J 2004; Shih et al. 2013; Komárek J 2014; Hammerschmidt et al. 2021). Hidden Markov Models  
34  
35 226 (HMMs) of protein family profiles (Pfam (version 33.0, February 2020) and Pgap (version 4, November  
36  
37 227 2020) were downloaded from <http://ftp.ebi.ac.uk/pub/databases/Pfam/releases/> and  
38  
39 228 <ftp://ftp.ncbi.nlm.nih.gov/hmm/4.0/> ftp sites, respectively. Proteins involved in cell differentiation in  
40  
41  
42 229 *Nostoc* used as reference seed proteins (retrieved from Uniprot database, [www.uniprot.org](http://www.uniprot.org)), are  
43  
44 230 indicated in Table S4.  
45

### 46 231 **Identification of homologs**

47  
48 232 HMMER package (Mistry et al. 2013) and HMM domain profiles were first used to establish the list of  
49  
50  
51 233 seed functional domains associated to reference seed proteins (Table S4). Alignments with a score  
52  
53 234 higher than the Pfam trusted thresholds were considered as significant seed domains. Then, HMMER3  
54  
55 235 (Mistry et al. 2013) and self-written Perl scripts were used to search for protein homologs (with reference  
56  
57 236 seed domains) in the complete genomes. For each reference protein, the presence of one or more seed  
58  
59  
60 237 domains was a requisite to identify putative homologs (Table S4). As described above, alignments with  
61  
62  
63  
64  
65

238 a score higher than the Pfam trusted thresholds were also considered significant. Note that HetP protein  
1  
2 239 and its 3 homologs in *Nostoc* were considered as a unique seed protein HetP since they share an identical  
3  
4 240 functional domain (Table S4). In contrary, the two seed proteins HetF and HetF' were analyzed  
5  
6 241 independently since their domain organizations are different (Table S4). For each reference protein,  
7  
8 242 putative homologs were subsequently analyzed with the same software in order to determine the  
9  
10 243 potential presence of additional functional domains. In-house Perl scripts were used to define the domain  
11  
12 244 organization of each homolog. Putative homologs were then selected for the strictly presence of seed  
13  
14 245 domains (i.e., in the right order and without additional domain), and therefore led to the set of Hmm-  
15  
16 246 homologs. To select specific homologs, additional criteria were added for the selection of HetC, HetL  
17  
18 247 and PatL. For this, we considered that to be a HetC homolog, the gene must be located in the same  
19  
20 248 genomic context as *hetP* (based on the co-localization of these two genes in the genome of *Nostoc*, see  
21  
22 249 Table S4) with a maximal distance of 5 genes upstream or downstream from *hetP*. HetL and PatL  
23  
24 250 homologs were limited to protein homologs sharing similar features with their seed sequences [a protein  
25  
26 251 length between 0.8 and 1.25 times when compared to the reference seed proteins (Table S4) and a  
27  
28 252 number of repeats fixed to 3-5, and 4-6, respectively]. Since PatN and PatU3 do not exhibit a known  
29  
30 253 functional domain, a Blast (Altschul et al. 1990) analysis search was performed using all reference seed  
31  
32 254 proteins as queries and an *E*-value of  $10^{-5}$ . Blast alignments with a MinLrap [Alignment size / Minimum  
33  
34 255 size (Query, Subject)]  $\geq 0.8$  and MaxLrap [Alignment size / Minimum size (Query, Subject)]  $\geq 0.8$   
35  
36 256 were considered as significant and constituted the set of Blast-homologs. For PatU3 and PatN, for which  
37  
38 257 no seed functional domain was found, MinLrap and MaxLrap were set to 0.7. Hmm-homolog and Blast-  
39  
40 258 homolog sets were finally combined in order to yield the final set of homologs for each reference protein.  
41  
42 259 The location of transmembrane domains in the proteins was performed using the TMHMM software  
43  
44 260 (Krogh et al. 2001).

#### 261 **Statistical analysis.**

262  $\beta$ -galactosidase activities: the values presented are means of 3 independent assays containing each 10  
263 independent colonies. The results are expressed as the means  $\pm$  standard variation.

264 Genomics analysis: one-sample *t*-test as well as boxplot, heatmap and distribution graphs were done  
1  
2 265 using the *R* (R Core Team, 2018). For boxplot analysis, taxonomic orders with low number of genomes  
3  
4 266 (i.e., Chroococciopsidales, Gloeobacterales, Gloeoemargaritales, Pleurocapsales, Spirulinales, and  
5  
6 267 unclassified Cyanobacteria) were fused into a single group (named Others) before computation (Table  
7  
8 268 S3, Sheet 4).

## 270 **Results**

271 To characterize the physical interactions between proteins that are involved in heterocyst differentiation  
18  
19 272 and pattern establishment in *Nostoc*, 13 full-length proteins (i.e., DevH, HetC, HetF, HetL, HetP, HetR,  
20  
21 273 HetZ, PatA, PatB, PatD, PatL, PatN, and PatU3) were tested systematically for pairwise interactions  
22  
23 274 using BACTH. We also included the All1730 ORF, which potentially encodes a HetF homolog (named  
24  
25 275 HetF') and the expression of which (i.e., all1730) has been shown to be induced in the same manner as  
26  
27 276 HetF after combined nitrogen starvation, as demonstrated by RNA-deep-sequencing experiments  
28  
29  
30 277 (Flaherty et al. 2011).

31  
32  
33 278 As for the BACTH assay the T18 and T25 domains must be in the cytosol, we first determined the  
34  
35 279 functional and structural domain composition of the studied proteins, as described in the Materials and  
36  
37 280 Methods. This allowed the definition of a subset of membrane proteins (HetC, HetF, PatL, and PatN)  
38  
39 281 and cytosolic factors (DevH, HetF', HetL, HetP, HetR, HetZ, PatA PatB, PatD, and PatU3) (Fig. 2,  
40  
41  
42 282 Table S4). Each protein was fused to a part of the catalytic domain of the chimeric adenylate cyclase  
43  
44 283 (T25 or T18) of *Bordetella pertussis*, at the C and N termini, when possible. According to the predicted  
45  
46 284 topology of the membrane proteins, insertions were made at one terminus to have the T18 and T25  
47  
48 285 domains in the cytosol. Therefore, membrane proteins (HetC, HetF, PatL, and PatN) were fused to T18  
49  
50  
51 286 and T25 at the extremity of their domain, which was predicted to be in the cytoplasm (Fig. 2). Direct  
52  
53 287 interactions were screened by transforming all possible pairwise combinations into the *E. coli* BTH101  
54  
55 288 strain (500 in total). Protein interactions were probed as explained in the Materials and Methods section.  
56  
57 289 In addition to the negative T25–T18 negative-control combination, all constructs were tested using a  
58  
59  
60 290 relevant control vector (as an example: T25-HetR+ NT18 and NT25 + T18-HetR served as negative

291 controls for HetR) and only results above the control values were considered as being positive in the  
1 remainder of the study (Fig. 3 and 4). This control is important for the elimination of self-activation  
2 292 values. In addition to the BACTH T18-Zip and T25-Zip positive control, the well-established  
3  
4 293 interactions between HetR subunits and HetL subunits, and between HetR and HetL, served as positive  
5  
6 294 internal controls. The pairwise combinations that yielded positive values are displayed in histograms  
7  
8 295 (Fig. 3 and 4), whereas negative results are grouped in Table 1.  
9  
10  
11  
12  
13

### 14 297 **Network analysis of early factors**

15  
16  
17 298 The major regulator of heterocyst differentiation, HetR, displays a specific fold that has been proposed  
18  
19 299 to allow its interaction with other factors (Kim et al. 2011; Hu et al. 2015). In fact, HetR has been shown  
20  
21 300 to interact with homologs of HetP (Videau et al. 2016), with the serine/threonine kinase Pkn22 (Roumezi  
22  
23 301 et al. 2019), and with HetL (Xu et al. 2020). The potential interaction between HetR and other factors  
24  
25 302 acting early after the initiation step and during pattern establishment was analyzed. Among the factors  
26  
27 303 that are induced early, HetR, HetL, and HetF' were found to interact with the PatB transcription factor,  
28  
29 304 as the  $\beta$ -galactosidase activities obtained were 10- to 40-fold higher than those of the negative controls  
30  
31 305 (Fig. 3a, b). In addition, interactions between HetL (Fig. 3b) and HetF' and multimerization of HetF'  
32  
33 306 (Fig. 3a) were observed. A weaker interaction (80  $\beta$ -galactosidase units) between HetF' and DevH was  
34  
35 307 also identified (Fig. 3a). Because HetF' has not been studied to date, these results suggest the possible  
36  
37 308 implication of this protein in the differentiation process, which is worth exploring. Among this group of  
38  
39 309 factors, the assays involving HetC and HetF all yielded negative results (Table 1)  
40  
41  
42  
43  
44

### 45 310 **Network analysis of commitment and patterning factors**

46  
47  
48 311 Among the three factors (HetP, HetZ, and PatU3) that formed the “commitment group”, a moderate but  
49  
50 312 significant interaction was observed between HetP and PatB (Fig. 4a). The highest  $\beta$ -galactosidase  
51  
52 313 values (400 units) were obtained for the HetP-T25/PatB-T18 combination. In addition, BACTH data  
53  
54 314 confirmed the previously published interaction between HetZ and PatU3 (Du et al. 2020), as two  
55  
56 315 combinations (T25-HetZ/PatU3-T18 and PatU3-T25/T18-HetZ) showed average  $\beta$ -galactosidase  
57  
58 316 activities of 300 units (Fig. 4a).  
59  
60  
61  
62  
63  
64  
65

317 The patterning factors consist of PatA, PatD, PatB, PatL, and PatN. Our analysis also included DevH in  
1  
2 318 this group, because it is required for *patB* transcriptional expression (Kurio et al. 2020). In addition to  
3  
4 319 the interactions with HetR, HetL, HetF', and HetP mentioned above, PatB formed multimers and  
5  
6 320 interacted with PatL in the PatL-T25/PatB-T18 combination (Fig. 4b). Moreover, PatA interacted with  
7  
8 321 PatL, which is in agreement with published data obtained using the yeast-two-hybrid approach (Liu and  
9  
10  
11 322 Wolk 2011). PatD and PatN did not show any interaction in our assays (**Table 1**).

### 14 323 **The patterning protein network of *Nostoc* as defined by BACTH**

16  
17 324 Among the 14 factors analyzed here, 10 were engaged in protein-protein interactions. In total, 16  
18  
19 325 interactions were detected, including self-associations (i.e. multimerization) (Fig. 5). Five proteins  
20  
21 326 (HetR, HetL, HetF', PatB, and PatL) were able to associate with multiple partners (at least two). The  
22  
23 327 interactions identified form a network that connects the elements involved in the initiation stage of  
24  
25 328 differentiation, commitment, and patterning. As expected, early and commitment factors mostly interact  
26  
27 329 among themselves, whereas the patterning factors, in particular PatB (with five interactions), appear to  
28  
29 330 be the central network and, therefore, serve as the master link between the three developmental stages  
30  
31 331 (Fig. 5). In addition, HetL and HetF' displayed three interactions each, suggesting that these proteins  
32  
33 332 are also important key players in the interaction network.  
34  
35  
36  
37

### 38 333 **Validation of the interaction between PatB and the early factors**

39  
40  
41 334 To validate our BACTH procedure, we sought to confirm some of the protein-protein interactions  
42  
43 335 described above by another approach. Since PatB appears to be central in the network, we decided to  
44  
45 336 analyze its ability to interact with the early factors by pull-down assays. PatB was produced as a His<sub>6</sub>-  
46  
47 337 tagged fusion and its capacity to bind nickel column was verified by affinity chromatography  
48  
49 338 purification (Fig. S1a). HetR, HetL and HetF' fused to the HA-tag were produced as shown by the  
50  
51 339 immunoblots of Fig. S1b. To proceed with the pull-down assay, His<sub>6</sub>-PatB and each of the HA-tagged  
52  
53 340 proteins were coproduced in *E. coli* and protein purifications were performed using Nickel-affinity  
54  
55 341 chromatography. The analysis of the elution fractions by SDS-PAGE followed by immunoblotting with  
56  
57 342 anti-His<sub>6</sub> and anti HA antibodies showed a specific co-elution of HetR, HetL and HetF' with PatB (Fig.  
58  
59  
60  
61  
62  
63  
64  
65

343 6, left panels). To control for non-specific binding of the HA-tagged proteins, HetR-HA, HetL-HA and  
1  
2 344 HetF'-HA produced in the absence of His<sub>6</sub>-PatB were purified following the same procedure. The data  
3  
4 345 presented in Fig. 6, right panels show that none of them was detected in the elution fractions, indicating  
5  
6 346 that they were unable to bind to the column. Altogether these results reveal direct interactions between  
7  
8 347 PatB and the three early factors and thus confirm the data obtained by BACTH.  
9

## 10 11 348 **Evolutionary conservation of genes underlying heterocyst formation and patterning**

12  
13  
14 349 To address the general relevance in cyanobacteria of the interaction network described above, a large-  
15  
16 350 scale (212 complete genomes in total) *in silico* identification and analysis of the set of the 14 protein  
17  
18 351 factors studied was performed (see Materials and methods and Table S4). Note that PatS, PatX and PatC  
19  
20 352 proteins were not included in our *in silico* study. The reasons are that these proteins are too short, do not  
21  
22 353 contain any known functional domain and in particular, *patC* was only present in *Nostoc* PCC7120 and  
23  
24 354 *Trichormus variabilis* NIES-23 genomes. These proteins are thus not compatible with significant  
25  
26 355 sequence comparisons based on functional domains search.  
27

28  
29  
30 356 Interestingly, this genome survey detected 63 (for PatN) to 779 (for HetL) homologs within the genomes  
31  
32 357 (Table S3, Sheets 2 and 3). This large variation in the total number of homologs for each reference  
33  
34 358 protein reflects the wide variation in the number of copies of each gene per genome. In fact, the  
35  
36 359 distribution of the number of copies within a genome (Table S3, Sheet 3; Fig. S2) showed that: (i) PatN,  
37  
38 360 HetP, PatB, HetF', HetR, PatU3, HetZ, and HetF did not have genes encoding the corresponding  
39  
40 361 homologs in a high number of genomes ( $\geq 80$ ); (ii) when present, they were on average present in  $\sim 1$   
41  
42 362 copy (for PatN, PatB, HetF', HetR, and PatU3) or  $\sim 2$  copies (for HetZ and HetF); and (iii) HetL group  
43  
44 363 of homologs displayed up to 4 copies per genome on average, with 13 copies encoding orthologs for  
45  
46 364 HetL in *Microcoleus* sp. PCC 7113. Moreover, HetL, which is encoded by a gene that is present in  
47  
48 365 multiple copies within most of the cyanobacterial genomes, harbored a low copy number of homologs  
49  
50 366 in Prochloraceae, with no copies detected in some *Prochlorococcus marinus* strains (e.g., *P. marinus*  
51  
52 367 str. NATL2A, *P. marinus* str. MIT9515, and *P. marinus* str. MIT 9211). Interestingly, genes encoding  
53  
54 368 HetR homologs were present in 2 or 3 copies in some genomes (e.g., *Nodosilinea nodulosa* PCC 7104  
55  
56 369 or *Prochlorothrix hollandica* PCC 9006, respectively). More surprisingly, most of the Nostocaceae  
57  
58  
59  
60  
61  
62  
63  
64  
65

370 and Aphanizomenonaceae strains displayed homolog(s) for each group of reference proteins, except for  
371 few organisms (e.g., *Cylindrospermopsis raciborskii* CS-505 and *Dolichospermum circinale*  
372 AWQC310F) for which no homolog of PatU3, PatB, and HetF was found.

373 Next, we asked whether a relationship exists between the number of homologs in an organism and its  
374 taxonomic lineage and, therefore, its ability to form heterocysts. For this purpose, a heatmap graphical  
375 representation of the number of homologs within organisms, according to the taxonomic orders, was  
376 performed (Fig. 7). Our analysis showed that except for HetC, nearly all the Nostocales contain at least  
377 one gene encoding a homolog of each reference group, with a variable number of homologs per  
378 organism, as described above. Interestingly, PatN seemed to be specifically present in Nostocales  
379 exclusively. To a lesser extent, HetP and PatB also appeared to be poorly represented outside the  
380 Nostocales group (Fig. 7). The distributions of the genes encoding HetC, HetF', HetP, HetZ, PatU3,  
381 PatB and PatL homologs over the cyanobacterial lineage were similar to the one of *hetR*, suggesting that  
382 they might share a similar evolutionary history from the ancestor of cyanobacteria. Considered this  
383 observation and the central and essential role of HetR within the heterocyst differentiation process, it  
384 can be suggested that HetC, HetF'; HetP, HetZ, PatU3; PatB, PatL, and PatN might also act as key  
385 factors for the early, commitment, and patterning stages, respectively, in **Nostocales**. Fig. 7 also shows  
386 that HetL, DevH, PatA, and PatD are conserved over the entire cyanobacterial taxonomic lineages,  
387 which suggests that they may be involved in biological processes that are common to all these  
388 prokaryotes. Unveiling the mechanisms beyond their speciation for heterocyst formation in the  
389 Nostocales will be a significant progress in this field.

390 Another important question that needed to be addressed was whether the distribution of the studied  
391 genes in Nostocales was significantly distinct compared with other taxonomic orders. For this, we  
392 computed for each reference group of proteins the average number of genes encoding homologs per  
393 genome for each cyanobacterial order (Materials and methods). These data are shown in the boxplots of  
394 Fig. 8a (see also Table S3, Sheet 4). With the exception of *patN*, the other analyzed genes exhibited a  
395 normal distribution among cyanobacterial orders, according to the Shapiro–Wilk test. As expected, HetL  
396 and DevH exhibited a high number of homologs on average, regardless of the cyanobacterial taxonomic

397 orders. Interestingly, the presence of HetC, HetP, PatA, PatB, PatL and of PatN, was significantly higher  
398 ( $P$ -value  $< 0.01$ ) in Nostocales compared with other taxonomic orders (Fig. 8a). The high prevalence of  
399 these six studied genes in Nostocales indicates their functional importance in heterocyst-forming  
400 cyanobacteria, and are also considered as key factors genes. In addition, the data presented in this Fig.  
401 highlight again the specific occurrence of PatN in genomes of Nostocales strains.

402 Altogether, when combined the two comparative genomics analyses (Fig. 7 and 8a), the overall key  
403 factors consist of a network encompassing 10 proteins (HetC, HetR, HetF', HetP, HetZ, PatU3, PatA,  
404 PatB, PatL, and PatN) that subsequently refers to the patterning factor network of heterocyst-forming  
405 cyanobacteria from an evolutionary point of view.

406

## 407 Discussion

408 In this study, we combined protein–protein interaction and genomics analyses to determine the physical  
409 and evolutionary relationships that exist between the factors involved in heterocyst development and  
410 patterning. BACTH has proven to be a powerful approach for probing interactions among the proteins  
411 involved in the same process. For instance, it allowed the deciphering of the interactions that occur  
412 between the membrane proteins involved in cell division in *E. coli* (Karimova et al. 2005), in cell wall  
413 synthesis in *Corynebacterium glutamicum* (Jankute et al. 2014), and in spore coat formation in *Bacillus*  
414 *subtilis* (Krajcikova et al. 2017). In eukaryotes, a yeast-two-hybrid system was used to determine the  
415 physical interactions between proteins involved in vulva development in *Caenorhabditis elegans*  
416 (Walhout et al. 2000), and this study was the starting point for the elucidation of the global interactome  
417 of this organism (Li et al. 2004). In a similar way, our study may help reveal the global interactome in  
418 *Nostoc*. The fact that the interactions observed by BACTH for two of the analyzed proteins (HetL and  
419 HetR) were confirmed by a BioLayer interferometry (Xu et al. 2020) and that pull-down assay confirmed  
420 the interaction of PatB with 3 factors (Fig. 6) validates the BACTH approach. One of the interesting  
421 perspectives of this work will be the study of all these possible interactions in *Nostoc* throughout the  
422 differentiation process. The development of techniques such as bimolecular fluorescence

423 complementation (BifC) (Kerppola 2006) and mass spectrometry (Gingras et al. 2007) should help  
1  
2 424 achieve this objective. This study represents an important step-forward future works aiming to reveal  
3  
4 425 the interactions occurring in the heterocyst. We propose to refer to the protein network involved in  
5  
6 426 pattern formation as the "patternome".  
7  
8

9 427 The possibility that proteins with transmembrane segments may not be correctly inserted in the  
10  
11 428 membrane of *E. coli* constitutes a limitation of the BACTH assays. We did not observe any interactions  
12  
13 429 for the HetC, HetF, and PatN proteins, which are predicted to be membrane proteins. In the case of  
14  
15 430 HetC, in addition to this prediction, its localization at the inner membrane of the heterocyst has been  
16  
17 431 proposed, as assessed using fusion to the green fluorescent protein (Corrales-Guerrero et al. 2014a).  
18  
19 432 Similarly, membrane localization has been indicated for PatN in *Nostoc punctiforme* (Risser et al. 2012).  
20  
21 433 The possibility that the lack of interactions of these proteins could be related to an incorrect topology of  
22  
23 434 the transmembrane domains in *E. coli* cannot be ruled out. HetF has recently been shown to be part of  
24  
25 435 the divisome of the vegetative cells under specific light conditions (Xing et al. 2021). The fact that no  
26  
27 436 interaction was found with the proteins involved in heterocyst formation and patterning might therefore  
28  
29 437 reflect that it does actually not belong to the protein network formed by these factors.  
30  
31  
32  
33

34 438 Fourteen of the 16 protein interactions observed in our BACTH analysis were detected for the first time.  
35  
36 439 These new data provide additional clues about the molecular mechanisms underlying patterning, as they  
37  
38 440 show that the three steps of the process are likely to be connected through physical interactions between  
39  
40 441 key factors (Fig. 5). Strikingly, our results indicate that the transcriptional regulator PatB interacts with  
41  
42 442 multiple partners, which renders it a central factor in the network (Fig. 3, 4, and 6). Among the cytosolic  
43  
44 443 proteins, PatD alone was shown to not interact with any other factor (Fig. 5). This result agrees with the  
45  
46 444 hypothesis that the effect of PatD on patterning might occur indirectly through general regulators, such  
47  
48 445 as ppGpp (Wang et al. 2019). The BACTH analysis also revealed that PatA interacts with two proteins  
49  
50 446 that are required for cell division, ZipN and SepF (Valladares et al. 2020). As PatA also interacts with  
51  
52 447 PatB and PatL, taken together, these data show that the physical protein interactions described here may  
53  
54 448 impact the evolution of the heterocyst toward its terminal differentiation state (i.e., the arrest of cell  
55  
56 449 division). Contrary to what has been published (Du et al. 2020), we did not observe any interaction  
57  
58  
59  
60  
61  
62  
63  
64  
65

450 between HetR and HetZ (Table 1). The facts that we systematically tested combinations of the fused  
1  
2 451 constructs with the BACTH control vectors as negative controls (Fig. 4a) and that we considered as  
3  
4 452 being positive only values that were higher than those obtained for these controls may explain the  
5  
6 453 discrepancy between our data and the published ones. The genomics analysis indicated that the 14 genes  
7  
8 454 analyzed here are highly conserved in the genomes of heterocyst-forming cyanobacteria (Fig. 7);  
9  
10  
11 455 therefore, an evolutionary history of these genes can be drawn. In fact, the location of almost all genes  
12  
13 456 over the taxonomic lineages suggests that they were present within the genome of the cyanobacteria  
14  
15 457 common ancestor. Throughout the evolution process, some of these genes were lost and/or functional  
16  
17 458 specialization might have occurred for those present in multiple copies (e.g., HetL). The process of gene  
18  
19 459 loss appears to be faster in some lineages than others, with Chroococcaceae, Gloeomargaritaceae,  
20  
21  
22 460 Gloeobacteraceae, and Synechococcaceae being the clades in which the gene losses are the most  
23  
24 461 pronounced. Thus, gene loss occurs in clades encompassing strains that do not exhibit cell  
25  
26 462 differentiation, whereas Nostococales genomes have conserved this set of genes (Fig. 7). Regarding  
27  
28 463 functional specialization, the most emblematic example is that of HetF and HetF', as the present copy  
29  
30 464 of HetF would probably be closer to the ones of the common ancestor. If future investigation unveils a  
31  
32 465 role for HetF' in the differentiation process, then one might suggest that its functional specialization  
33  
34 466 during evolution would have consisted in a loss of the transmembrane region, followed by the  
35  
36 467 acquisition of the Forkhead-associated domain (FHA). As the FHA domain has been maintained during  
37  
38 468 evolution, its importance in the differentiation process and in the interaction of HetF' with PatB, DevH,  
39  
40 469 and HetL may be questioned.

45 470 The heatmap generated here revealed that various levels can be proposed for the evolutive and functional  
46  
47 471 importance of the analyzed proteins in each step of the differentiation process. Among the early factors,  
48  
49 472 HetF', HetR, and HetC are the major evolutionarily conserved proteins; *hetF'* and *hetR* for their single  
50  
51 473 copy in the genome, and *hetC* for a very significant presence bias in favor of the Nostocales compared  
52  
53 474 with the other orders (Fig. 7). The essential role of HetR in *Nostoc* differentiation is fully understood.  
54  
55 475 As mentioned above, both the protein interaction analysis and the genomics survey suggest that the  
56  
57 476 HetF' protein is an interesting factor that deserves investigation. Genetic analyses revealed the  
58  
59  
60  
61  
62  
63  
64  
65

477 involvement of HetC and its N-terminal domain in heterocyst differentiation (Khudyakov and Wolk  
1 1997; Xu and Wolk 2001; Corrales-Guerrero et al. 2014a). However, whether HetC actually acts as an  
2 1997; Xu and Wolk 2001; Corrales-Guerrero et al. 2014a). However, whether HetC actually acts as an  
3  
4 479 ABC transporter remains to be demonstrated. A comparative and evolutionary analysis of HetP, HetZ,  
5  
6 480 and PatU3 showed that all of them are key players in the commitment step, even if there is a significant  
7  
8 481 presence bias in favor of HetP in the Nostocales. It has been proposed that cell-fate commitment in  
9  
10 482 *Nostoc* is modulated by a hierarchy of four homologous HetP proteins, with HetP promoting cell  
11  
12 483 differentiation, whereas the HetP-like proteins seem to exert an opposite control (Videau et al. 2016).  
13  
14 484 These observations combined with the presence of multiple copies of HetP-encoding genes in  
15  
16 485 Nostocales clearly demonstrate that, during the evolutionary time, some HetP paralogs underwent a  
17  
18 486 neofunctionalization process, leading to new functions. The interactions of HetP with PatB and PatU3  
19  
20 487 with HetZ indicate that the protein complexes formed might explain the role of these proteins in  
21  
22 488 connecting the initiation step with subsequent steps. From an evolutionary point of view, PatA, PatB,  
23  
24 489 PatL and PatN appear as the key players among the patterning factors. The presence of PatN exclusively  
25  
26 490 in Nostocales seems to indicate that its action is important for the formation of heterocysts in all  
27  
28 491 organisms of this clade. Therefore, the elucidation of the molecular mechanism of PatN is crucial for  
29  
30 492 the understanding of this process. The significant enrichment of PatA in the Nostocales also makes it a  
31  
32 493 key factor that deserves further attention. In the protein network, PatB occupies a central position as it  
33  
34 494 interacts with factors belonging to the three defined groups and is mostly present in heterocyst-forming  
35  
36 495 cyanobacteria. Little is known about the mechanism of this regulator and about the genes that are  
37  
38 496 controlled by it in *Nostoc*, although it appears from our study that it deserves additional attention. Among  
39  
40 497 all the key players defined here, 8 out of 10 exhibited interactions with each other and with other  
41  
42 498 partners. This shows that protein–protein interactions may play pivotal roles in heterocyst  
43  
44 499 differentiation. Therefore, unveiling their dynamics through this process is a determining direction for  
45  
46 500 the understanding of development in these prokaryotes.  
47  
48  
49  
50  
51  
52  
53

54 501 Given that three factors in the interaction network are transcriptional regulators (HetR, DevH, and PatB),  
55  
56 502 interactions occurring inside the network likely impact the transcriptional program of the developing  
57  
58 503 cell. To obtain insights into this interconnection, we first referred to the literature to build the  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
504 transcriptional network (Table S5). Indeed, HetR affects the transcription of genes belonging to the three  
505 groups (early events, commitment, and patterning) by acting either directly inside the network (effect  
506 on *hetZ*, *hetP*, *patA*, and likely *devH*) or indirectly through the regulation of the expression of the global  
507 regulator NtcA (effect on *hetC* and *patD*) (see references cited in Table S5). DevH directly controls the  
508 transcription of *patB* (Kurio et al. 2020), which was suggested to act on genes involved in the maturation  
509 step (Brenes-Alvarez et al. 2019). Secondly, we combined the network of both protein-protein  
510 interactions and comparative genomics, with the transcriptional one, leading to an integrative model of  
511 the three distinct approaches on heterocyst differentiation genes (Fig. 8b). From this, several facts can  
512 be pinpointed: (i) the activities of the regulators connect the 14 genes through transcriptional  
513 relationships, thus rendering the protein and transcriptional networks interdependent; (ii) HetR, PatB  
514 and PatA remain the factors that exhibit the highest numbers of biological links; (iii) based on the  
515 number of biological links, HetZ and PatB can be considered as the main actors among the commitment  
516 and patterning factors, respectively; (iv) cross-talks link early, commitment and patterning factors with  
517 3 key players (as defined by *in silico* analysis) in each differentiation stage. It is important to remind  
518 here that in addition to these factors, acting inside the heterocyst, the patterning requires HetN, PatS and  
519 PatX which are supposed to generate the diffusible morphogens essential for lateral inhibition along the  
520 filaments (Fig. 1a).

38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
**In conclusion**, our study questioned the possible physical and evolutionary interactions between 14  
proteins involved in heterocyst formation and patterning in *Nostoc*. BACTH approach was used to  
analyze the protein-protein interactions and pull-down assays were used to confirm three of the observed  
interactions. Combining the data obtained to a genomic survey of these 14 genes in cyanobacterial  
genomes reveals that a core network of 12 proteins (all the analyzed proteins factors except HetF and  
PatD) may play a crucial role in heterocyst development and patterning not only in *Nostoc* but within  
all the clade of heterocyst-forming cyanobacteria (Fig. 8b). The results of this study open up interesting  
perspectives for the elucidation of the mechanism of action of all these proteins and the importance of  
their interactions in establishing the pattern.

531

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

532 **Acknowledgments:**

533 The authors thank D<sup>r</sup> Badreddine Douzi and D<sup>r</sup> Romé Voulhoux for helpful discussions regarding pull-  
534 down assays. They also thank Anais Scholivet for technical assistance.

535 **Funding:**

536 X. Xu was funded by a fellowship from the Chinese government. R. Rachedi has a fellowship from the  
537 Région SUD and CNRS. The project was supported by the “Agence Nationale pour la Recherche  
538 Scientifique” (ANR-21-CE20-0025-01) and by funding from AMU.

539 **Ethics declarations:**

540 **Conflict of interest**

541 The authors declare that they have no conflict of interest.

542 **Ethical approval**

543 This article does not contain any studies with human participants or animals performed by any of the  
544 authors.

545 **Informed consent**

546 All the authors have consent for publication.

547 **REFERENCES**

- 1  
2  
3 548 Abed RM, Dobretsov S, Sudesh K (2009) Applications of cyanobacteria in biotechnology. *J Appl*  
4 549 *Microbiol* 106:1-12  
5 550 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol*  
6 551 *Biol* 215:403-410  
7  
8 552 Black TA, Cai Y, Wolk CP (1993) Spatial expression and autoregulation of hetR, a gene involved in the  
9 553 control of heterocyst development in *Anabaena*. *Mol Microbiol* 9:77-84  
10 554 Black TA, Wolk CP (1994) Analysis of a Het- mutation in *Anabaena* sp. strain PCC 7120 implicates a  
11 555 secondary metabolite in the regulation of heterocyst spacing. *J Bacteriol* 176:2282-2292  
12 556 Brenes-Alvarez M, Mitschke J, Olmedo-Verd E, Georg J, Hess WR, Vioque A, Muro-Pastor AM (2019)  
13 557 Elements of the heterocyst-specific transcriptome unravelled by co-expression analysis in  
14 558 *Nostoc* sp. PCC 7120. *Environ Microbiol* 21:2544-2558  
15  
16 559 Buikema WJ, Haselkorn R (1991) Characterization of a gene controlling heterocyst differentiation in  
17 560 the cyanobacterium *Anabaena* 7120. *Genes Dev* 5:321-330  
18 561 Corrales-Guerrero L, Flores E, Herrero A (2014a) Relationships between the ABC-exporter HetC and  
19 562 peptides that regulate the spatiotemporal pattern of heterocyst distribution in *Anabaena*.  
20 563 *PLoS One* 9:e104571  
21  
22 564 Corrales-Guerrero L, Mariscal V, Flores E, Herrero A (2013) Functional dissection and evidence for  
23 565 intercellular transfer of the heterocyst-differentiation PatS morphogen. *Mol Microbiol*  
24 566 88:1093-1105  
25 567 Corrales-Guerrero L, Mariscal V, Nurnberg DJ, Elhai J, Mullineaux CW, Flores E, Herrero A (2014b)  
26 568 Subcellular localization and clues for the function of the HetN factor influencing heterocyst  
27 569 distribution in *Anabaena* sp. strain PCC 7120. *J Bacteriol* 196:3452-3460  
28  
29 570 Du Y, Zhang H, Wang H, Wang S, Lei Q, Li C, Kong R, Xu X (2020) Expression from DIF1-motif  
30 571 promoters of hetR and patS is dependent on HetZ and modulated by PatU3 during  
31 572 heterocyst differentiation. *PLoS One* 15:e0232383  
32  
33 573 Elhai J, Khudyakov I (2018) Ancient association of cyanobacterial multicellularity with the regulator  
34 574 HetR and an RGSGR pentapeptide-containing protein (PatX). *Mol Microbiol* 110:931-954  
35 575 Feldmann EA, Ni S, Sahu ID, Mishler CH, Risser DD, Murakami JL, Tom SK, McCarrick RM, Lorigan GA,  
36 576 Tolbert BS, Callahan SM, Kennedy MA (2011) Evidence for direct binding between HetR from  
37 577 *Anabaena* sp. PCC 7120 and PatS-5. *Biochemistry* 50:9212-9224  
38 578 Fernandez-Pinas F, Leganes F, Wolk CP (1994) A third genetic locus required for the formation of  
39 579 heterocysts in *Anabaena* sp. strain PCC 7120. *J Bacteriol* 176:5277-5283  
40  
41 580 Flaherty BL, Van Nieuwerburgh F, Head SR, Golden JW (2011) Directional RNA deep sequencing sheds  
42 581 new light on the transcriptional response of *Anabaena* sp. strain PCC 7120 to combined-  
43 582 nitrogen deprivation. *BMC Genomics* 12:332  
44 583 Frias JE, Flores E, Herrero A (1994) Requirement of the regulatory protein NtcA for the expression of  
45 584 nitrogen assimilation and heterocyst development genes in the cyanobacterium *Anabaena*  
46 585 sp. PCC 7120. *Mol Microbiol* 14:823-832  
47  
48 586 Gingras AC, Gstaiger M, Raught B, Aebersold R (2007) Analysis of protein complexes using mass  
49 587 spectrometry. *Nat Rev Mol Cell Biol* 8:645-654  
50 588 Hammerschmidt K, Landan G, Domingues Kummel Tria F, Alcorta J, Dagan T (2021) The Order of Trait  
51 589 Emergence in the Evolution of Cyanobacterial Multicellularity. *Genome Biol Evol* 13  
52 590 Henson B.J HSM, Watson L.E, Barnum S.R (2004) Molecular phylogeny of the heterocystous  
53 591 cyanobacteria (subsections IV and V) based on nifD. *Journal of Systematic and Evolutionary*  
54 592 *Microbiology* 54:493-497  
55  
56 593 Herrero A, Stavans J, Flores E (2016) The multicellular nature of filamentous heterocyst-forming  
57 594 cyanobacteria. *FEMS Microbiol Rev* 40:831-854  
58  
59 595 Higa KC, Callahan SM (2010) Ectopic expression of hetP can partially bypass the need for hetR in  
60 596 heterocyst differentiation by *Anabaena* sp. strain PCC 7120. *Mol Microbiol* 77:562-574  
61  
62  
63  
64  
65

597 Higa KC, Rajagopalan R, Risser DD, Rivers OS, Tom SK, Videau P, Callahan SM (2012) The RGSGR  
1 598 amino acid motif of the intercellular signalling protein, HetN, is required for patterning of  
2 599 heterocysts in *Anabaena* sp. strain PCC 7120. *Mol Microbiol* 83:682-693  
3 600 Hu HX, Jiang YL, Zhao MX, Cai K, Liu S, Wen B, Lv P, Zhang Y, Peng J, Zhong H, Yu HM, Ren YM, Zhang  
4 601 Z, Tian C, Wu Q, Oliveberg M, Zhang CC, Chen Y, Zhou CZ (2015) Structural insights into HetR-  
5 602 PatS interaction involved in cyanobacterial pattern formation. *Sci Rep* 5:16470  
6 603 Huang X, Dong Y, Zhao J (2004) HetR homodimer is a DNA-binding protein required for heterocyst  
7 604 differentiation, and the DNA-binding activity is inhibited by PatS. *Proc Natl Acad Sci U S A*  
8 605 101:4848-4853  
9 606 Jankute M, Byng CV, Alderwick LJ, Besra GS (2014) Elucidation of a protein-protein interaction  
10 607 network involved in *Corynebacterium glutamicum* cell wall biosynthesis as determined by  
11 608 bacterial two-hybrid analysis. *Glycoconj J* 31:475-483  
12 609 Jones KM, Buikema WJ, Haselkorn R (2003) Heterocyst-specific expression of *patB*, a gene required  
13 610 for nitrogen fixation in *Anabaena* sp. strain PCC 7120. *J Bacteriol* 185:2306-2314  
14 611 Karimova G, Dautin N, Ladant D (2005) Interaction network among *Escherichia coli* membrane  
15 612 proteins involved in cell division as revealed by bacterial two-hybrid analysis. *J Bacteriol*  
16 613 187:2233-2243  
17 614 Karimova G, Pidoux J, Ullmann A, Ladant D (1998) A bacterial two-hybrid system based on a  
18 615 reconstituted signal transduction pathway. *Proc Natl Acad Sci U S A* 95:5752-5756  
19 616 Kerppola TK (2006) Design and implementation of bimolecular fluorescence complementation (BiFC)  
20 617 assays for the visualization of protein interactions in living cells. *Nat Protoc* 1:1278-1286  
21 618 Khudyakov I, Gladkov G, Elhai J (2020) Inactivation of Three RG(S/T)GR Pentapeptide-Containing  
22 619 Negative Regulators of HetR Results in Lethal Differentiation of *Anabaena* PCC 7120. *Life*  
23 620 (Basel) 10  
24 621 Khudyakov I, Wolk CP (1997) *hetC*, a gene coding for a protein similar to bacterial ABC protein  
25 622 exporters, is involved in early regulation of heterocyst differentiation in *Anabaena* sp. strain  
26 623 PCC 7120. *J Bacteriol* 179:6971-6978  
27 624 Kim Y, Joachimiak G, Ye Z, Binkowski TA, Zhang R, Gornicki P, Callahan SM, Hess WR, Haselkorn R,  
28 625 Joachimiak A (2011) Structure of transcription factor HetR required for heterocyst  
29 626 differentiation in cyanobacteria. *Proc Natl Acad Sci U S A* 108:10109-10114  
30 627 Komárek J KJ, Mareš J, Johansen J.R (2014) Taxonomic classification of cyanoprokaryotes  
31 628 (cyanobacterial genera) 2014, using a polyphasic approach. *Preslia* 86  
32 629 Krajcikova D, Forgac V, Szabo A, Barak I (2017) Exploring the interaction network of the *Bacillus*  
33 630 *subtilis* outer coat and crust proteins. *Microbiol Res* 204:72-80  
34 631 Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein  
35 632 topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305:567-  
36 633 580  
37 634 Kumar K, Mella-Herrera RA, Golden JW (2010) Cyanobacterial heterocysts. *Cold Spring Harb Perspect*  
38 635 *Biol* 2:a000315  
39 636 Kurio Y, Koike Y, Kanesaki Y, Watanabe S, Ehira S (2020) The CRP-family transcriptional regulator  
40 637 DevH regulates expression of heterocyst-specific genes at the later stage of differentiation in  
41 638 the cyanobacterium *Anabaena* sp. strain PCC 7120. *Mol Microbiol* 114:553-562  
42 639 Li MZ, Elledge SJ (2012) SLIC: a method for sequence- and ligation-independent cloning. *Methods*  
43 640 *Mol Biol* 852:51-59  
44 641 Li S, Armstrong CM, Bertin N, Ge H, Milstein S, Boxem M, Vidalain PO, Han JD, Chesneau A, Hao T,  
45 642 Goldberg DS, Li N, Martinez M, Rual JF, Lamesch P, Xu L, Tewari M, Wong SL, Zhang LV, Berriz  
46 643 GF, Jacotot L, Vaglio P, Reboul J, Hirozane-Kishikawa T, Li Q, Gabel HW, Elewa A,  
47 644 Baumgartner B, Rose DJ, Yu H, Bosak S, Sequerra R, Fraser A, Mango SE, Saxton WM, Strome  
48 645 S, Van Den Heuvel S, Piano F, Vandenhaute J, Sardet C, Gerstein M, Doucette-Stamm L,  
49 646 Gunsalus KC, Harper JW, Cusick ME, Roth FP, Hill DE, Vidal M (2004) A map of the  
50 647 interactome network of the metazoan *C. elegans*. *Science* 303:540-543

648 Liang J, Scappino L, Haselkorn R (1992) The *patA* gene product, which contains a region similar to  
1 649 CheY of *Escherichia coli*, controls heterocyst pattern formation in the cyanobacterium  
2 650 *Anabaena* 7120. *Proc Natl Acad Sci U S A* 89:5655-5659  
3 651 Liu D, Golden JW (2002) *hetL* overexpression stimulates heterocyst formation in *Anabaena* sp. strain  
4 652 PCC 7120. *J Bacteriol* 184:6873-6881  
5 653 Liu J, Wolk CP (2011) Mutations in genes *patA* and *patL* of *Anabaena* sp. strain PCC 7120 result in  
6 654 similar phenotypes, and the proteins encoded by those genes may interact. *J Bacteriol*  
7 655 193:6070-6074  
8 656 Makarova KS, Koonin EV, Haselkorn R, Galperin MY (2006) Cyanobacterial response regulator *PatA*  
9 657 contains a conserved N-terminal domain (PATAN) with an alpha-helical insertion.  
10 658 *Bioinformatics* 22:1297-1301  
11 659 Meeks JC, Elhai J (2002) Regulation of cellular differentiation in filamentous cyanobacteria in free-  
12 660 living and plant-associated symbiotic growth states. *Microbiol Mol Biol Rev* 66:94-121; table  
13 661 of contents  
14 662 Michel-Souzy S, Douzi B, Cadoret F, Raynaud C, Quinton L, Ball G, Voulhoux R (2018) Direct  
15 663 interactions between the secreted effector and the T2SS components *GspL* and *GspM* reveal  
16 664 a new effector-sensing step during type 2 secretion. *J Biol Chem* 293:19441-19450  
17 665 Mistry J, Finn RD, Eddy SR, Bateman A, Punta M (2013) Challenges in homology search: HMMER3 and  
18 666 convergent evolution of coiled-coil regions. *Nucleic Acids Res* 41:e121  
19 667 Pratte BS, Thiel T (2016) Homologous regulators, *CnfR1* and *CnfR2*, activate expression of two distinct  
20 668 nitrogenase gene clusters in the filamentous cyanobacterium *Anabaena variabilis* ATCC  
21 669 29413. *Mol Microbiol* 100:1096-1109  
22 670 Rajagopalan R, Callahan SM (2010) Temporal and spatial regulation of the four transcription start  
23 671 sites of *hetR* from *Anabaena* sp. strain PCC 7120. *J Bacteriol* 192:1088-1096  
24 672 Rippka R D, J, Wterbury JB, Herdman M, Stanier RY (1979) Generic assignments, strain stories and  
25 673 properties of pure cultures of cyanobacteria. *J. Gent. Microbiol.* 111:1-61  
26 674 Risser DD, Callahan SM (2008) *HetF* and *PatA* control levels of *HetR* in *Anabaena* sp. strain PCC 7120.  
27 675 *J Bacteriol* 190:7645-7654  
28 676 Risser DD, Wong FC, Meeks JC (2012) Biased inheritance of the protein *PatN* frees vegetative cells to  
29 677 initiate patterned heterocyst differentiation. *Proc Natl Acad Sci U S A* 109:15342-15347  
30 678 Roumezi B, Xu X, Risoul V, Fan Y, Lebrun R, Latifi A (2019) The *Pkn22* Kinase of *Nostoc* PCC 7120 Is  
31 679 Required for Cell Differentiation via the Phosphorylation of *HetR* on a Residue Highly  
32 680 Conserved in Genomes of Heterocyst-Forming Cyanobacteria. *Front Microbiol* 10:3140  
33 681 Shih PM, Wu D, Latifi A, Axen SD, Fewer DP, Talla E, Calteau A, Cai F, Tandeau de Marsac N, Rippka R,  
34 682 Herdman M, Sivonen K, Coursin T, Laurent T, Goodwin L, Nolan M, Davenport KW, Han CS,  
35 683 Rubin EM, Eisen JA, Woyke T, Gugger M, Kerfeld CA (2013) Improving the coverage of the  
36 684 cyanobacterial phylum using diversity-driven genome sequencing. *Proc Natl Acad Sci U S A*  
37 685 110:1053-1058  
38 686 Tsujimoto R, Kamiya N, Fujita Y (2014) Transcriptional regulators *ChlR* and *CnfR* are essential for  
39 687 diazotrophic growth in nonheterocystous cyanobacteria. *Proc Natl Acad Sci U S A* 111:6762-  
40 688 6767  
41 689 Turing A (1952) The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society*  
42 690 B 237:37-72  
43 691 Valladares A, Velazquez-Suarez C, Herrero A (2020) Interactions of *PatA* with the Divisome during  
44 692 Heterocyst Differentiation in *Anabaena*. *mSphere* 5  
45 693 Videau P, Rivers OS, Higa KC, Callahan SM (2015) ABC Transporter Required for Intercellular Transfer  
46 694 of Developmental Signals in a Heterocystous Cyanobacterium. *J Bacteriol* 197:2685-2693  
47 695 Videau P, Rivers OS, Hurd K, Ushijima B, Oshiro RT, Ende RJ, O'Hanlon SM, Cozy LM (2016) The  
48 696 heterocyst regulatory protein *HetP* and its homologs modulate heterocyst commitment in  
49 697 *Anabaena* sp. strain PCC 7120. *Proc Natl Acad Sci U S A* 113:E6984-E6992  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

698 Videau P, Rivers OS, Tom SK, Oshiro RT, Ushijima B, Swenson VA, Philmus B, Gaylor MO, Cozy LM  
1 699 (2018) The hetZ gene indirectly regulates heterocyst development at the level of pattern  
2 700 formation in *Anabaena* sp. strain PCC 7120. *Mol Microbiol*  
3 701 Walhout AJ, Sordella R, Lu X, Hartley JL, Temple GF, Brasch MA, Thierry-Mieg N, Vidal M (2000)  
4 702 Protein interaction mapping in *C. elegans* using proteins involved in vulval development.  
5 703 *Science* 287:116-122  
6 704 Wang L, Lin GM, Niu TC, Zhang SR, Zhang JY, Tang GF, Chen W, Zhang CC (2019) patD, a Gene  
7 705 Regulated by NtcA, Is Involved in the Optimization of Heterocyst Frequency in the  
8 706 Cyanobacterium *Anabaena* sp. Strain PCC 7120. *J Bacteriol* 201  
9 707 Wong FC, Meeks JC (2001) The hetF gene product is essential to heterocyst differentiation and  
10 708 affects HetR function in the cyanobacterium *Nostoc punctiforme*. *J Bacteriol* 183:2654-2661  
11 709 Xing WY, Liu J, Wang ZQ, Zhang JY, Zeng X, Yang Y, Zhang CC (2021) HetF Protein Is a New Divisome  
12 710 Component in a Filamentous and Developmental Cyanobacterium. *mBio*:e0138221  
13 711 Xu X, Risoul V, Byrne D, Champ S, Douzi B, Latifi A (2020) HetL, HetR and PatS form a reaction-  
14 712 diffusion system to control pattern formation in the cyanobacterium *Nostoc* PCC 7120. *Elife* 9  
15 713 Xu X, Wolk CP (2001) Role for hetC in the transition to a nondividing state during heterocyst  
16 714 differentiation in *Anabaena* sp. *J Bacteriol* 183:393-396  
17 715 Yoon HS, Golden JW (1998) Heterocyst pattern formation controlled by a diffusible peptide. *Science*  
18 716 282:935-938  
19 717 Yoon HS, Golden JW (2001) PatS and products of nitrogen fixation control heterocyst pattern. *J*  
20 718 *Bacteriol* 183:2605-2613  
21 719 Zhang H, Wang S, Wang Y, Xu X (2018) Functional Overlap of hetP and hetZ in Regulation of  
22 720 Heterocyst Differentiation in *Anabaena* sp. Strain PCC 7120. *J Bacteriol* 200  
23 721 Zhang W, Du Y, Khudyakov I, Fan Q, Gao H, Ning D, Wolk CP, Xu X (2007) A gene cluster that regulates  
24 722 both heterocyst differentiation and pattern formation in *Anabaena* sp. strain PCC 7120. *Mol*  
25 723 *Microbiol* 66:1429-1443  
26 724 Zubay G, Morse DE, Schrenk WJ, Miller JH (1972) Detection and isolation of the repressor protein for  
27 725 the tryptophan operon of *Escherichia coli*. *Proc Natl Acad Sci U S A* 69:1100-1103

33 726  
34  
35  
36  
37 727

38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

728 **Table 1: Matrix of BACTH data highlighting negative interactions between proteins**  
 1 729 **involved in heterocyst differentiation within *Nostoc*.**

2  
 3 730

5	HetR	HetL	HetC	HetF	HetF'	HetP	HetZ	PatU3	PatB	PatA	PatL	PatN	PatD	DevH
6	HetR													
7	HetL													
8	HetC													
9	HetF													
10	HetF'													
11	HetP													
12	HetZ													
13	PatU3													
14	PatB													
15	PatA													
16	PatL													
17	PatN													
18	PatD													
19	DevH													

26 731

27

28 732

29 733

30 734

31 735

32 736

33 737

34 738

35 739

36

37 740

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

Each rectangle of the matrix represents a tested two hybrid interactions (from horizontal and vertical axes). The  $\beta$ -galactosidase activities correspond to the functional complementation of the T18/T25 adenylate cyclase domains following interaction of the two hybrid proteins.  $\beta$ -galactosidase activity lower than 20 Miller units, or lower than the self-activation controls (ie: the fused-construct tested with the control vector) was considered as negative. Those whose interactions were detected as negative by all possible combinations are marked as dark grey. Those whose interactions were detected at least by one combination are marked as blank (see Fig. 3 and 4). The repeat copy of each protein pair in this table is marked as light grey.

741 **Legends to figures**

1  
2 742 **Figure 1**

3  
4 743 (a) Local activation/long range inhibition model for heterocyst pattern establishment in *Nostoc*. The red  
5 744 areas indicate the concentration of the inhibitory peptides (derived from PatS, HetN and probably PatX).  
6 745 A: activator (HetR). I: immunity protein (HetL). A/I: complex formed by the interaction between HetR  
7 746 and HetL.

8  
9 747 (b) Phenotype of mutants of genes involved in heterocyst formation and patterning. o.e. stands for  
10 748 overexpression of the corresponding genes. Δ stands for deletion of the corresponding genes.

11  
12 749  
13  
14 750 **Figure 2**

15 751 Structural and functional domains located within the reference proteins analyzed in this study.

16  
17 752 The domains predicted by Pfam are indicated by blue boxes. For HetR, the domains correspond to those  
18 753 identified by the resolution of the structure (Hu et al. 2015). The topology of the membrane protein was  
19 754 predicted by the TMHMM software (Krogh et al. 2001).

20  
21 755 Protein sizes are in parenthesis (aa: amino acids). The frame of each protein is colored corresponding  
22 756 to its functional group in our study (bleu: early factors, yellow: commitment, green: patterning).

23 757 TM, Transmembrane domains. Functional domain descriptions are : Crp, Bacterial regulatory proteins,  
24 758 crp family; HTH\_Crp, Crp-like helix-turn-helix domain; Peptidase\_C39, Peptidase C39 family;  
25 759 ABC\_tran, ABC transporter; CHAT, CHAT domain; FHA, FHA domain; Pentapeptide, Pentapeptide  
26 760 repeats (8 or 9 copies); HetP\_family, HetP family heterocyst commitment protein; Peptidase\_S48,  
27 761 Peptidase family S48; HetR\_C, Heterocyst differentiation regulator C-terminal Hood domain;  
28 762 DUF4388, Domain of unknown function (DUF4388); Response\_reg, Response regulator receiver  
29 763 domain; heterocyst\_HetZ, heterocyst differentiation protein HetZ; HTH\_3, Helix-turn-helix; HTH\_31,  
30 764 Helix-turn-helix domain; het\_cyst\_PatD, heterocyst frequency control protein PatD.

31  
32  
33 765

34  
35 766 **Figure 3**

36 767 BACTH analysis of protein-protein interactions of factors involved in early stage of differentiation. (a)  
37 768 Positive assays of HetR or HetF' interacting with other proteins involved in pattern formation. (b)  
38 769 Positive assays for HetL interacting with other proteins involved in pattern formation.

39 770 For both, BTH101 strains were transformed with the indicated plasmids. β-galactosidase activities were  
40 771 measured as described in section "Materials and Methods" and were expressed in Miller units. Strains  
41 772 producing T25-Zip and T18-Zip served as positive control (PC). Strains producing the T25 with T18  
42 773 served as negative control (NC). Strains producing T25-HetR and T18-HetR served as inner positive  
43 774 control. For each combination, the two corresponding self-activation controls were analyzed and shown  
44 775 in gray columns. Blue columns indicate high BACTH signals (more than 300 Miller units); yellow  
45 776 columns indicate intermediate BACTH signals (100-300 Miller units); green columns indicate weak  
46 777 BACTH signals (70-100 Miller units).

47  
48 778  
49

50 779 **Figure 4**

51 780 BACTH analysis of protein-protein interactions of factors involved in commitment (a) and patterning  
52 781 (b).

53 782 BTH101 strains were transformed with the indicated plasmids. β-galactosidase activities were measured  
54 783 as described in section "Materials and Methods" and were expressed in Miller units. Strains producing  
55 784 T25-Zip and T18-Zip served as positive control (PC). Strains producing the T25 with T18 served as  
56 785 negative control (NC). Strains producing T25-HetR and T18-HetR served as inner positive control. For  
57 786 each combination, the two corresponding self-activations controls were analyzed and shown in gray  
58 787 columns. Blue columns indicate high BACTH signals (more than 300 Miller units); yellow columns  
59 788 indicate intermediate BACTH signals (100-300 Miller units).

60  
61  
62  
63  
64  
65

789

1  
2 **790 Figure 5**

3  
4 791 Protein-protein interaction network of 14 factors involved in heterocyst formation and patterning.

5  
6 792 The thickness of the lines is relative to the level of BACTH signals. ■■■■ : indicate high levels ( $\beta$ -  
7 793 galactosidase values above 300 Miller units),■■■■ : indicate intermediate levels ( $\beta$ -galactosidase values  
8 794 between 100-300 Miller units). ——— : indicates weak levels ( $\beta$ -galactosidase between 70-100 Miller  
9 795 units).

10  
11 796

12  
13 **797 Figure 6**

14  
15 798 Interactions between PatB and the early factors analyzed by pull-down assays.  
16 799 Left panels: Nickel-affinity based copurification and immunoblotting experiments of His-tagged PatB  
17 800 co-produced with HetR-HA (a), HetL-HA (b) or HetF<sup>+</sup>-HA (c) in *E. coli*.  
18 801 Right panels: negative controls corresponding to the purification of each HA-tagged protein on a Nickel  
19 802 column, after heterologous production in *E. coli*.  
20 803 M: molecular size of the protein marker (kDa), L: loading material; FT: flow-through; W: wash fraction;  
21 804 E1–E6: elution fractions.  
22 805 The antibodies used are indicated at the top of each immunoblot.

23 806

24  
25  
26 **807 Figure 7**

27 808 Heatmap graphical representation of the number of homologs within organisms, with their associated  
28 809 taxonomic levels (order and family) over early, commitment and patterning factors. Each line  
29 810 corresponds to one cyanobacterial strain. The color scale (0, 1 or  $\geq 2$ ) represents the number of  
30 811 homologs in each genome. Taxonomic name abbreviations are: c, Chroococciopsidales; p,  
31 812 Pleurocapsales; s, Spirulinales; g, Gloeomargaritales

32  
33  
34 813

35  
36 **814 Figure 8**

37 815 (a) Boxplot showing the average number of homolog genes per genome within the cyanobacterial  
38 816 taxonomic orders. Blue dots are related to *Nostocales* order, with \* or \*\* on the top when the statistical  
39 817 significance (one-sample *t*-test, *p*-value) was below 0.05 or 0.01, respectively.  
40 818 (b) Integrative map of protein-protein interactions, transcriptional regulations and comparative  
41 819 genomics of heterocyst differentiation genes.  
42 820 The size of the gray cycles is relative to the total number of biological links, including protein-protein  
43 821 interactions and transcription regulatory actions. Key player proteins (as defined by *in silico* analysis)  
44 822 are underlined. The transcriptional regulations include direct and indirect actions (see Table S4 for  
45 823 details). As described within the main text, HetN, PatS and PatS were not included in our analyses.

46  
47 824

48

49

50

51

52

53

54

55

56

57

58

59

60

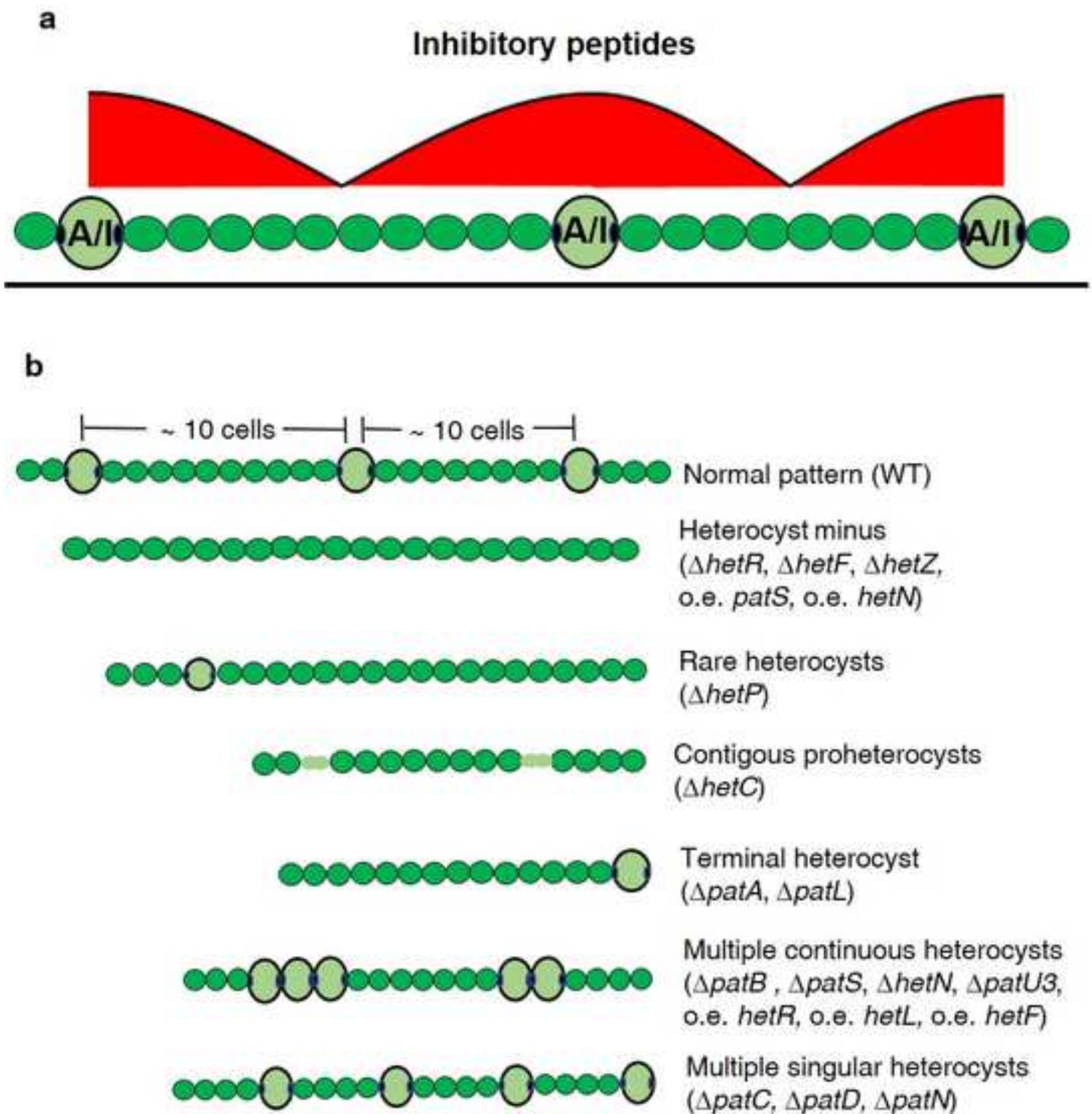
61

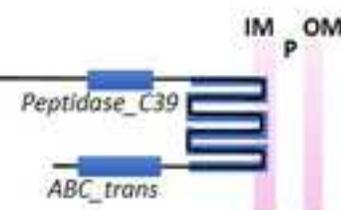
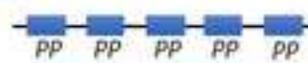
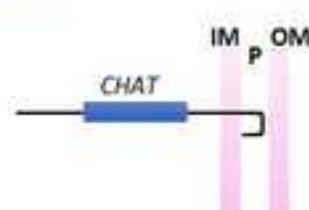
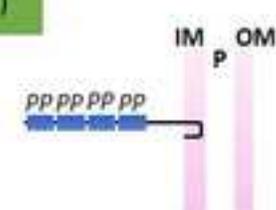
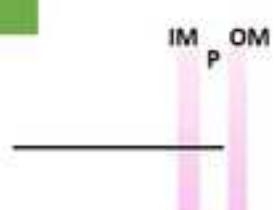
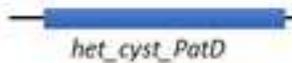
62

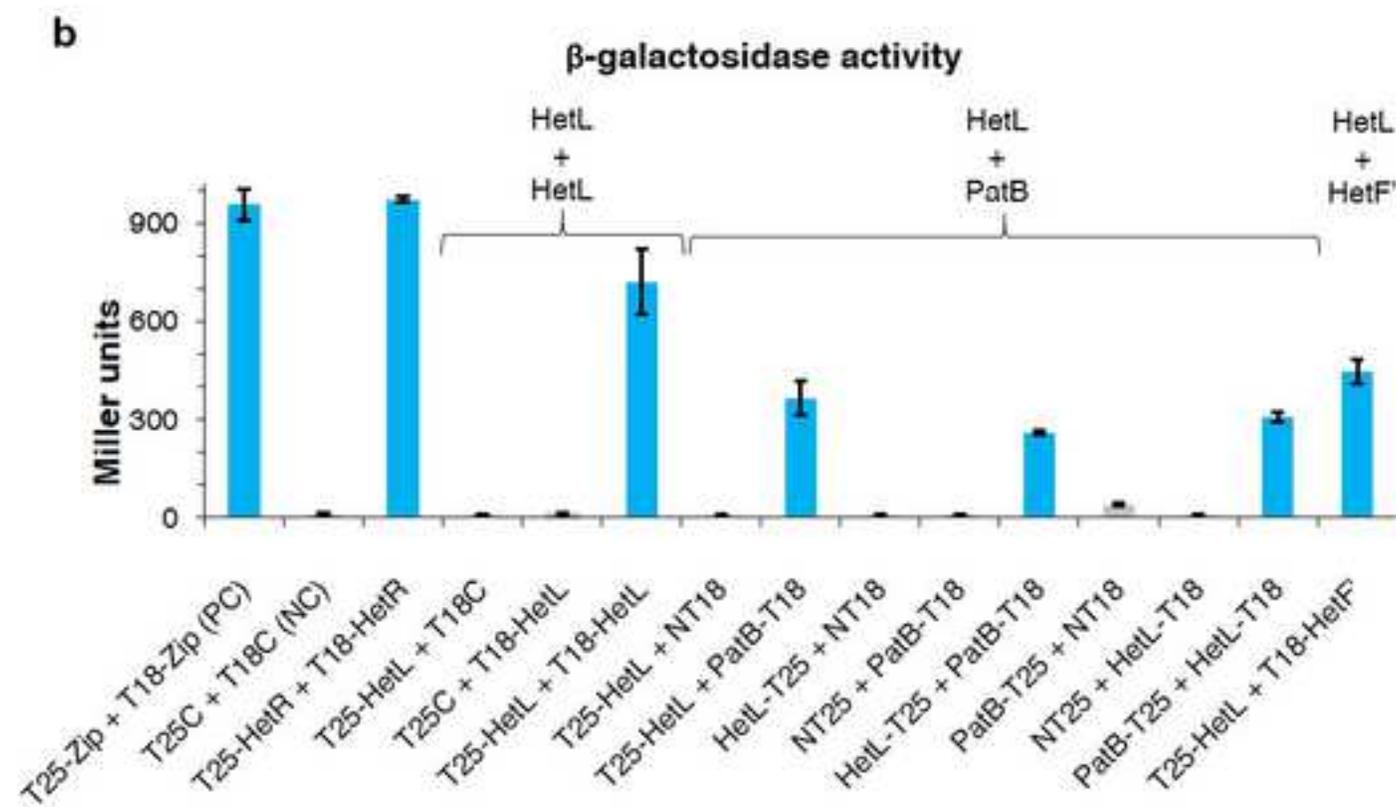
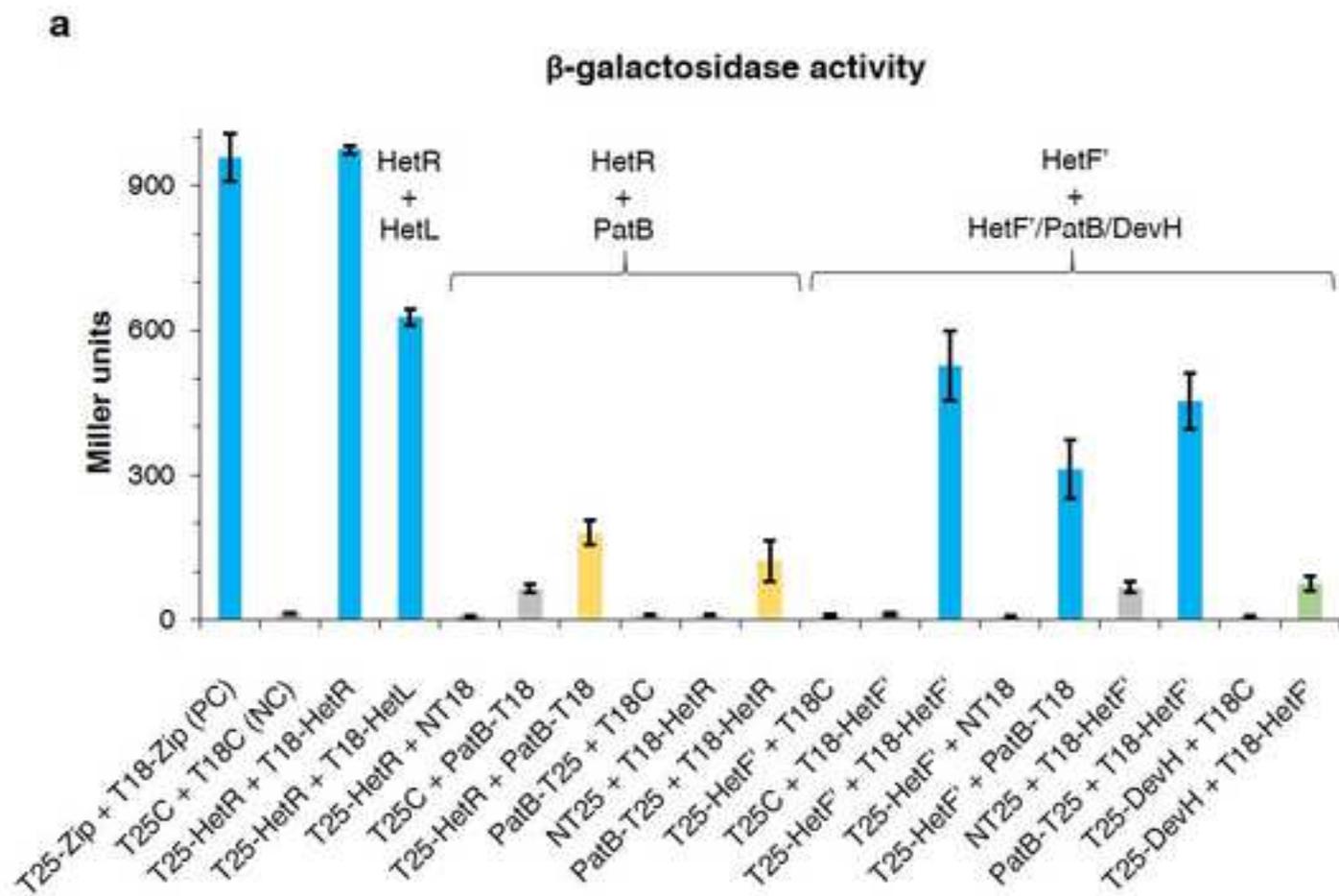
63

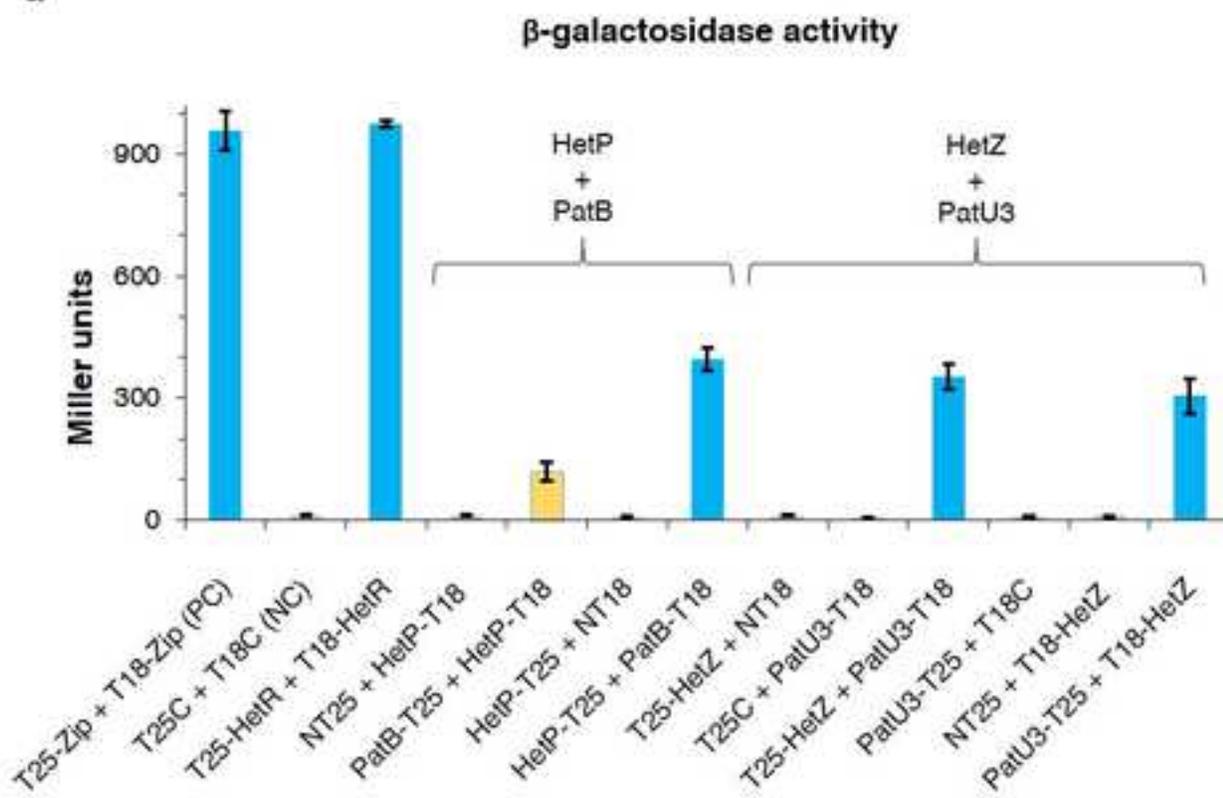
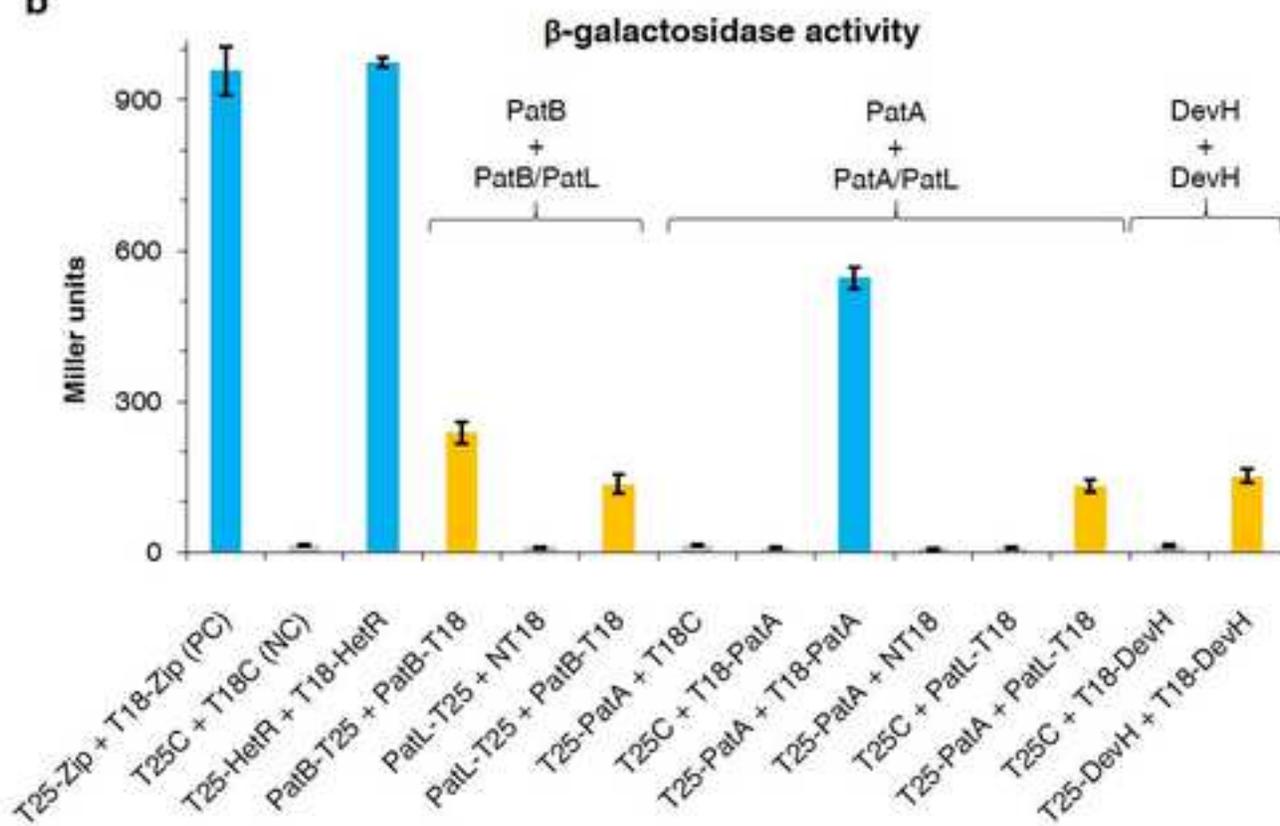
64

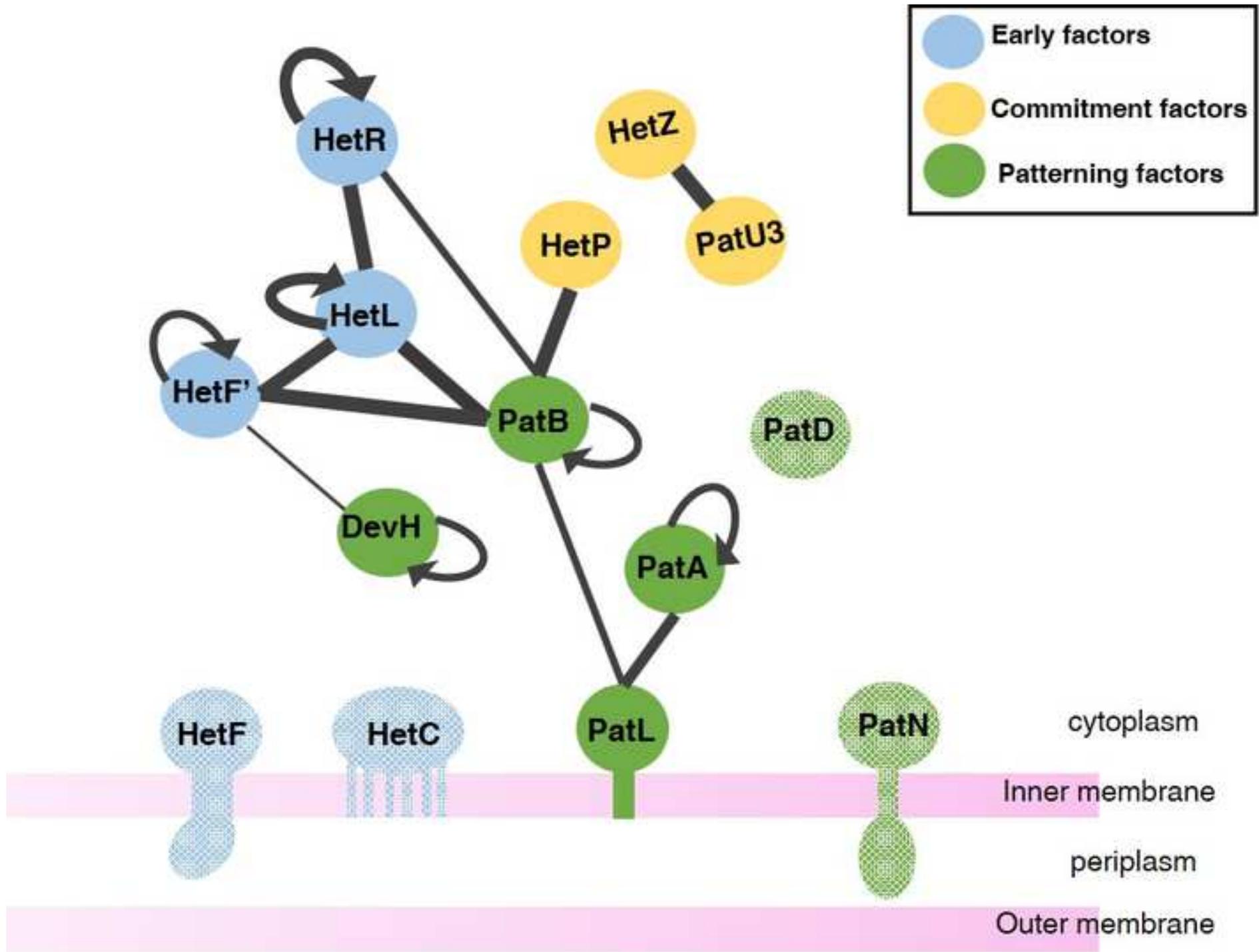
65

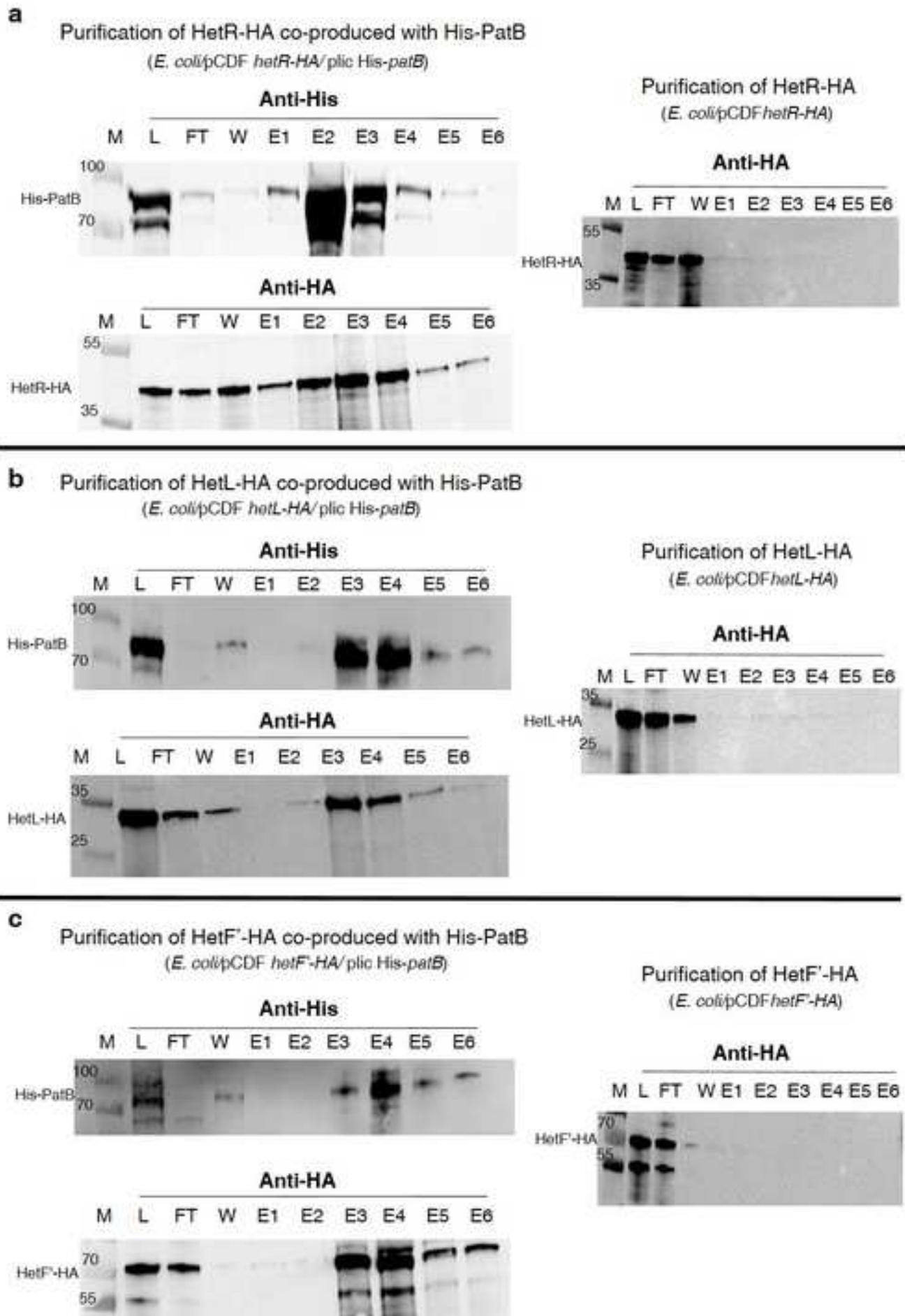


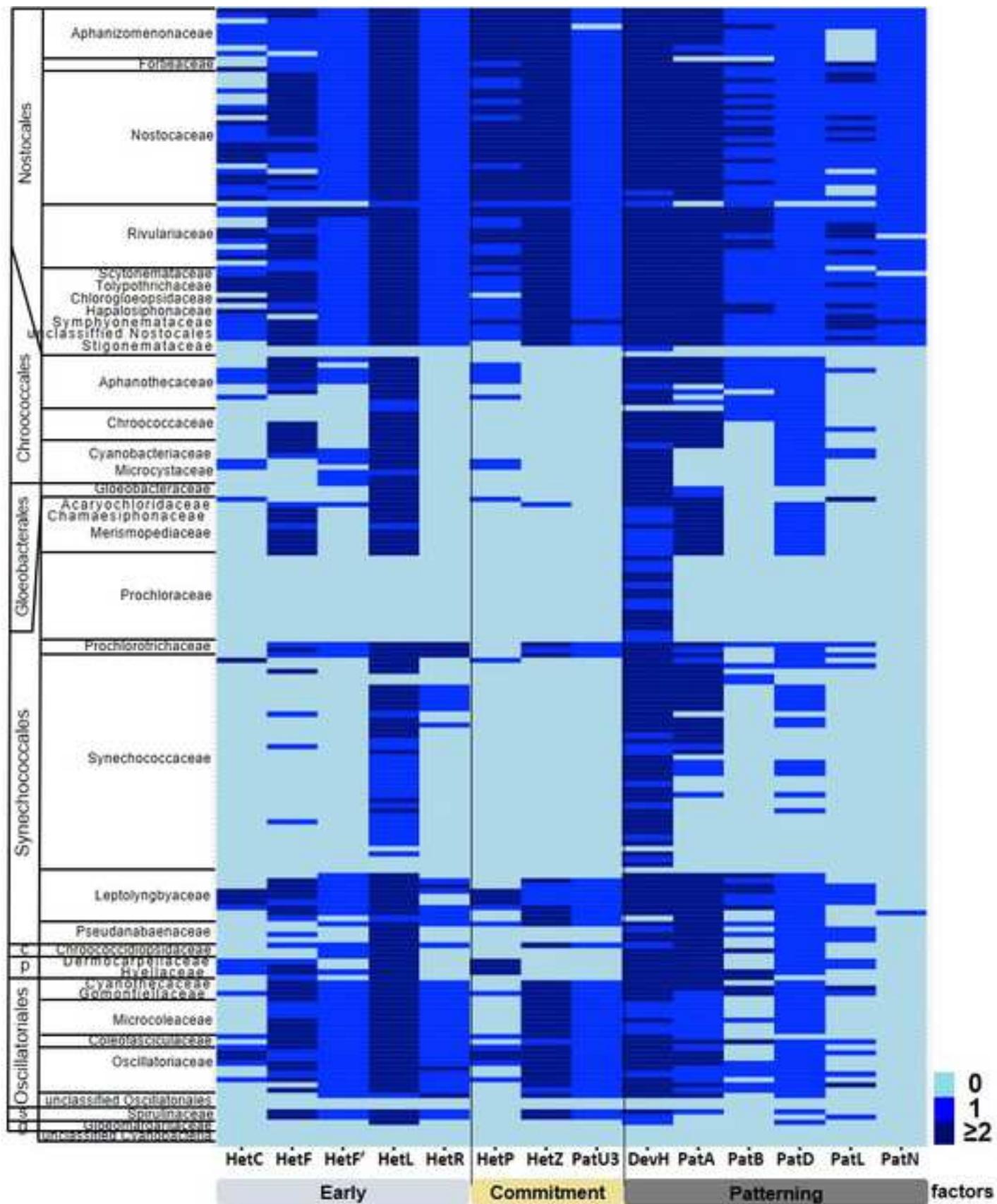
Membrane	Cytosolic	
<p><b>HetC</b> (1044 aa)</p> 	<p><i>Transcriptional regulators</i></p> <p><b>HetR</b> (299 aa)</p> 	<p><i>Others</i></p> <p><b>HetF'</b> (525 aa)</p>  <p><b>HetL</b> (237 aa)</p> 
<p><b>HetF</b> (779 aa)</p> 	<p><b>DevH</b> (239 aa)</p> 	<p><b>HetP</b> (159 aa)</p> 
<p><b>PatL</b> (496 aa)</p> 	<p><b>PatB</b> (529 aa)</p> 	<p><b>HetZ</b> (401 aa)</p>  <p><b>PatA</b> (379 aa)</p> 
<p><b>PatN</b> (215 aa)</p> 		<p><b>PatD</b> (119 aa)</p>  <p><b>PatU3</b> (258 aa)</p> 

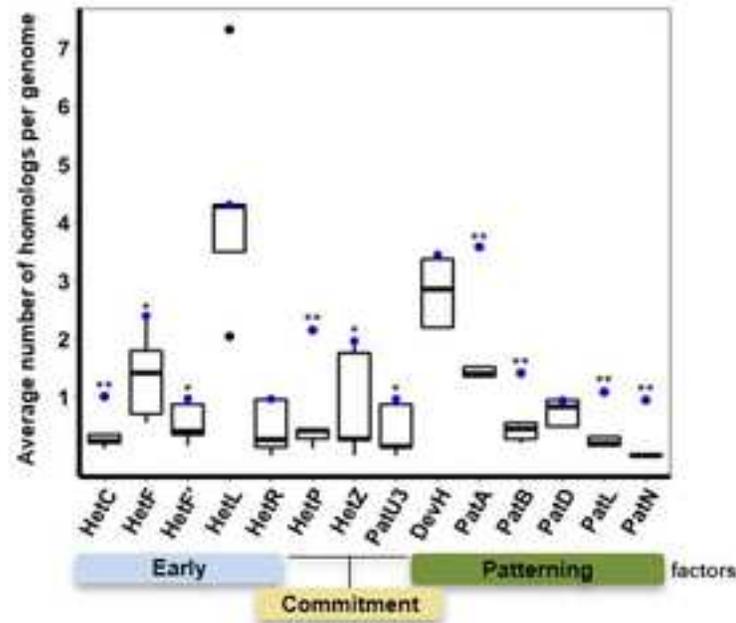
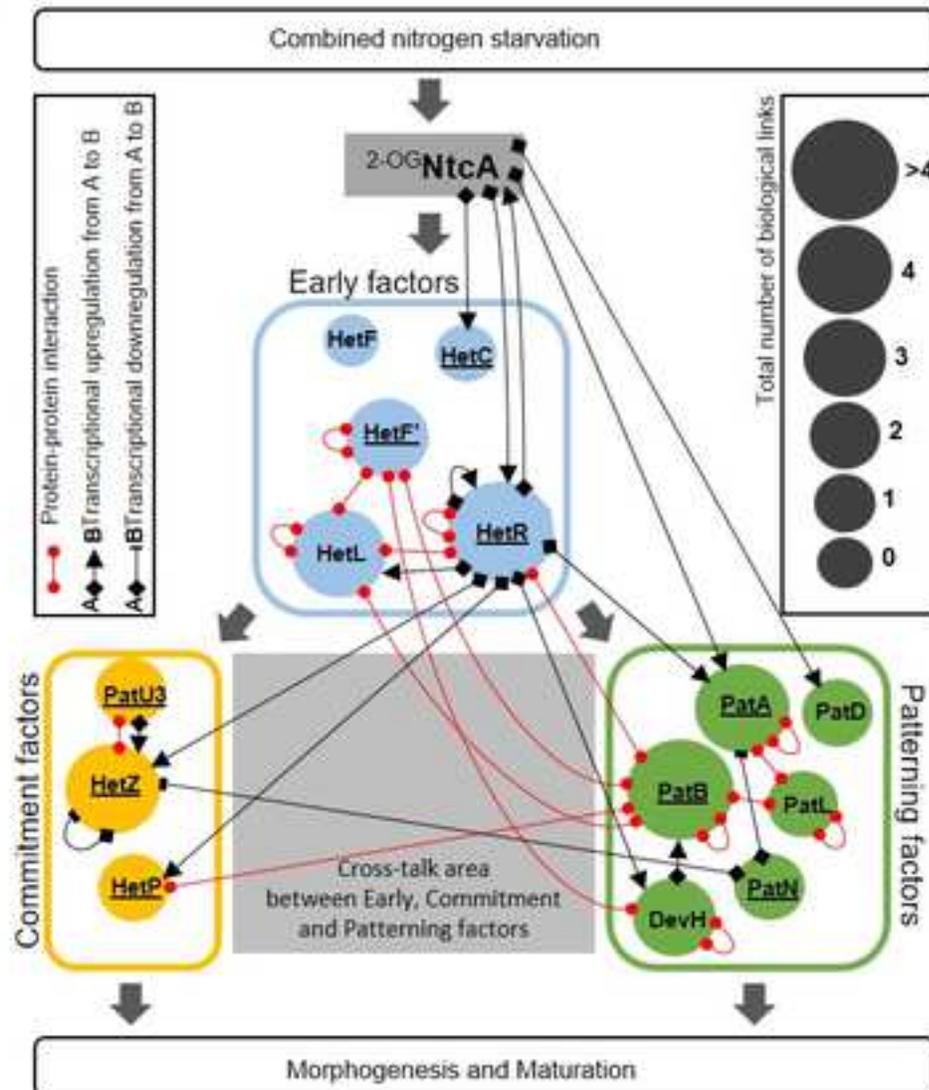


**a****b**







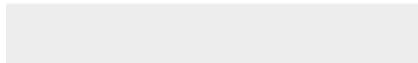
**a****b**



[Click here to access/download](#)

**Supplementary Material**

Supplementary Material-MGG revision.docx





# Change of authorship request form (pre-acceptance)

## Please read the important information on page 4 before you begin

This form should be used by authors to request any change in authorship including changes in corresponding authors. Please fully complete all sections. Use black ink and block capitals and provide each author's full name with the given name first followed by the family name.

Please note: In author collaborations where there is formal agreement for representing the collaboration, it is sufficient for the representative or legal guarantor (usually the corresponding author) to complete and sign the Authorship Change Form on behalf of all authors.

### Section 1: Please provide the current title of manuscript

(For journals: Please provide the manuscript ID, title and/or DOI if available.)

(For books: Please provide the title, ISBN and/or DOI if available.)

<b>Manuscript ID no. in case of unpublished manuscript:</b>
---

<b>DOI in case of published manuscript:</b> MGAG-D-21-00741R1
---

<b>ISBN (for books):</b>
--------------------------

Title: Interaction network among factors involved in heterocyst-patterning in cyanobacteria
---

### Section 2: Please provide the previous authorship, in the order shown on the manuscript before the changes were introduced. Please indicate the corresponding author by adding (CA) behind the name.

	First name(s)	Family name	ORCID or SCOPUS id, if available
1 <sup>st</sup> author	XIAOMEI	Xu	
2 <sup>nd</sup> author	MARYLINE	FOGLINO	
3 <sup>rd</sup> author	EMMANUEL	TALLA	
4 <sup>th</sup> author	AMEL	LATIFI	
5 <sup>th</sup> author			
6 <sup>th</sup> author			
7 <sup>th</sup> author			

Please use an additional sheet if there are more than 7 authors.

## Change of authorship request form (pre-acceptance)

**Section 3: Please provide a justification for change. Please use this section to explain your reasons for changing the authorship of your manuscript, e.g. what necessitated the change in authorship? Please refer to the (journal) policy pages for more information about authorship. Please explain why omitted authors were not originally included and/or why authors were removed on the submitted manuscript.**

The change in the author-list consists in including Raphaël Rachedi as co-first author  
 The reason for this addition is that R. Rachedi performed all the co-expression and copurification and immunoblotting experiments that were performed to answer the reviewers and editor requests (Figure 6 and Figure S1)

**Section 4: Proposed new authorship. Please provide your new authorship list in the order you would like it to appear on the manuscript. Please indicate the corresponding author by adding (CA) behind the name. If the corresponding author has changed, please indicate the reason under section 3.**

	First name(s)	Family name (this name will appear in full on the final publication and will be searchable in various abstract and indexing databases)
1 <sup>st</sup> author	XIAOMEI	XU
2 <sup>nd</sup> author co-first	RAPHAEL	RACHEDI
3 <sup>rd</sup> author	MARYLINE	FOGLINO
4 <sup>th</sup> author	EMMANUEL	TALLA
5 <sup>th</sup> author	AMEL	LATIFI
6 <sup>th</sup> author		
7 <sup>th</sup> author		

Please use an additional sheet if there are more than 7 authors.

## Change of authorship request form (pre-acceptance)

**Section 5: Author contribution, Acknowledgement and Disclosures.** Please use this section to provide a new disclosure statement and, if appropriate, acknowledge any contributors who have been removed as authors and ensure you state what contribution any new authors made (if applicable per the journal or book (series) policy). **Please ensure these are updated in your manuscript - after approval of the change(s) - as our production department will not transfer the information in this form to your manuscript.**

**New acknowledgements:**

No author has been removed from author-list

**New Disclosures (financial and non-financial interests, funding):**

Not applicable

**New Author Contributions statement (if applicable per the journal policy):**

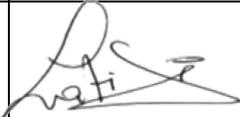
Not applicable

State 'Not applicable' if there are no new authors.

# Change of authorship request form (pre-acceptance)

**Section 6: Declaration of agreement. All authors, unchanged, new and removed *must* sign this declaration.**

**(NB: Please print the form, sign and return a scanned copy. Please note that signatures that have been inserted as an image file are acceptable as long as it is handwritten. Typed names in the signature box are unacceptable.)** \* Please delete as appropriate. Delete all of the bold if you were on the original authorship list and are remaining as an author.

	First name	Family name		Signature	Affiliated institute	Date
1 <sup>st</sup> author	XIAOMEI	XU	I agree to the proposed new authorship shown in section 4 / <del>and the addition/removal*of my name to the authorship list.</del>		Aix-Marseille University CNRS, IMM, Laboratoire de Chimie Bactérienne	03-17th-2022
2 <sup>nd</sup> author	RAPHAEL	RACHEDI	I agree to the proposed new authorship shown in section 4 / <del>and the addition/removal*of my name to the authorship list.</del>		Aix-Marseille University CNRS, IMM, Laboratoire de Chimie Bactérienne	03-17th-2022
3 <sup>rd</sup> author	MARYLINE	FOGLINO	I agree to the proposed new authorship shown in section 4 / <del>and the addition/removal*of my name to the authorship list.</del>		Aix-Marseille University CNRS, IMM, Laboratoire de Chimie Bactérienne	03-17th-2022
4 <sup>th</sup> authors	EMMANUEL	TALLA	I agree to the proposed new authorship shown in section 4 / <del>and the addition/removal*of my name to the authorship list.</del>		Aix-Marseille University CNRS, IMM, Laboratoire de Chimie Bactérienne	03-17th-2022
5 <sup>th</sup> author	AMEL	LATIFI	I agree to the proposed new authorship shown in section 4 / <del>and the addition/removal*of my name to the authorship list.</del>		Aix-Marseille University CNRS, IMM, Laboratoire de Chimie Bactérienne	03-17th-2022
6 <sup>th</sup> author			I agree to the proposed new authorship shown in section 4 / <del>and the addition/removal*of my name to the authorship list.</del>			
7 <sup>th</sup> author			I agree to the proposed new authorship shown in section 4 / <del>and the addition/removal*of my name to the authorship list.</del>			

Please use an additional sheet if there are more than 7 authors.

### Important information. Please read.

- Please return this form, fully completed, to Springer Nature. We will consider the information you have provided to decide whether to approve the proposed change in authorship. We may choose to contact your institution for more information or undertake a further investigation, if appropriate, before making a final decision.
- By signing this declaration, all authors guarantee that the order of the authors are in accordance with their scientific contribution, if applicable as different conventions apply per discipline, and that only authors have been added who made a meaningful contribution to the work.
- Please note, we cannot investigate or mediate any authorship disputes. If you are unable to obtain agreement from all authors (including those who you wish to be removed) you must refer the matter to your institution(s) for investigation. Please inform us if you need to do this.
- If you are not able to return a fully completed form within **30 days** of the date that it was sent to the author requesting the change, we may have to withdraw your manuscript. We cannot publish manuscripts where authorship has not been agreed by all authors (including those who have been removed).
- Incomplete forms will be rejected.