environmental microbiology

Environmental Microbiology (2019) **21**(7), 2544–2558 doi:10.1111/1462-2920.14647

Elements of the heterocyst-specific transcriptome unravelled by co-expression analysis in *Nostoc* sp. PCC 7120

Manuel Brenes-Álvarez,[1] **Jan Mitschke**,[2†] **Elvira Olmedo-Verd**,[1] **Jens Georg**,[2] **Wolfgang R. Hess**,[2,3] **Agustín Vioque**[1] and **Alicia M. Muro-Pastor**[1*]

[1]*Instituto de Bioquímica Vegetal y Fotosíntesis, Consejo Superior de Investigaciones Científicas and Universidad de Sevilla, E-41092, Sevilla, Spain.*

[2]*Genetics and Experimental Bioinformatics, Faculty of Biology, University of Freiburg, D-79104, Freiburg, Germany.*

[3]*Freiburg Institute for Advanced Studies, University of Freiburg, D-79104, Freiburg, Germany.*

Summary

Nitrogen is frequently limiting microbial growth in the environment. As a response, many filamentous cyanobacteria differentiate heterocysts, cells devoted to N_2 fixation. Heterocyst differentiation is under the control of the master regulator HetR. Through the characterization of the HetR-dependent transcriptome in *Nostoc* sp. PCC 7120, we identified the new candidate genes likely involved in heterocyst differentiation. According to their maximum induction, we defined E-DIF (early in differentiation) and L-DIF (late in differentiation) genes. Most of the genes known to be involved in the critical aspects of heterocyst differentiation or function were also classified into these groups, showing the validity of the approach. Using fusions to *gfp*, we verified the heterocyst-specific transcription of several of the found genes, antisense transcripts and potentially trans-acting sRNAs. Through comparative sequence analysis of promoter regions, we noticed the prevalence of the previously described DIF1 motif and identified a second motif, called DIF2, in other promoters of the E-DIF cluster. Both motifs are widely conserved in heterocystous cyanobacteria. We assigned *alr2522* as a third member, besides *nifB* and *nifP*, to the CnfR regulon. The elements identified here are of interest for understanding cell differentiation, engineering of biological nitrogen fixation or production of O_2-sensitive molecules in cyanobacteria.

Received 31 January, 2019; revised 6 April, 2019; accepted 1 May, 2019. *For correspondence. E-mail alicia@ibvf.csic.es; Tel. +34-954489521; Fax +34-954460165. †Present address: University Medical Center Freiburg, D-79106 Freiburg, Germany

Introduction

In the absence of combined nitrogen sources, many filamentous cyanobacteria, including *Nostoc* sp. PCC 7120 (a.k.a. *Anabaena* sp. PCC 7120) the model organism used in this work, differentiate a one-dimensional pattern of specialized cells called heterocysts, that are devoted to the fixation of atmospheric nitrogen (N_2) (Muro-Pastor and Hess, 2012). Heterocyst differentiation relies on a transcriptional program that is exclusive of the cells becoming heterocysts (Flaherty *et al*., 2011; Mitschke *et al*., 2011; Flores *et al*., 2018). Major landmarks of the process include the sequential synthesis and deposition of specific heterocyst envelopes as well as the synthesis and maturation of nitrogenase, the enzymatic complex for N_2 fixation. In parallel, precise coordination between the nitrogen-fixing heterocysts and vegetative cells becomes established because these two cell types depend on each other in order to sustain growth of the filament as a whole. Growth at the expense of N_2 involves not only the morphological and metabolic transformation of selected vegetative cells to become heterocysts, but also the precise regulation of the pattern and timing of differentiation (Flores and Herrero, 2010). Thus, a high number of different regulatory factors can be expected to be involved in these processes.

Classical approaches involving transposon or chemical mutagenesis followed by screening for Fox^- mutants, unable to grow using only molecular nitrogen under aerobic conditions, allowed the identification of major players in the process of heterocyst differentiation, such as the master regulator HetR and genes encoding proteins involved in the synthesis and deposition of envelope polysaccharides and glycolipids (*hep* and *hgl* genes, respectively), or in the synthesis and maturation of nitrogenase (*nif* genes) (see e. g., Wolk *et al*., 1988; Buikema and Haselkorn, 1991b). Some elements, such as *hetR*, were repeatedly discovered

indicating the near-saturation coverage achieved in these experiments (Buikema and Haselkorn, 1991a; Black *et al*., 1993). Because the approaches involving identification of Fox^- mutants are based on inability to grow using N_2, they are not suitable to identify elements that participate in more subtle aspects of differentiation, such as patterning or time of commitment. Similarly, it is also difficult to identify protein-coding genes or regulatory transcripts if their functions are redundant or if they cannot be mutated because they are essential for growth even in the presence of combined nitrogen. Therefore, it is likely that many of the involved genes and factors have remained unidentified. Thus, alternative approaches are required to fully uncover the genetic elements that participate in the process of differentiation.

Here, we identified genes and sequence elements potentially involved in heterocyst differentiation on the basis of time-resolved transcriptional profiling of the wild-type and a *hetR* mutant, unable to differentiate heterocysts (Buikema and Haselkorn, 1991a), using high-density microarrays that include coding and non-coding regions of the genome. We found two groups of genes with temporally distinct HetR-dependent transcriptional responses that we have named as E-DIF (peaking early in differentiation) and L-DIF (peaking late in differentiation) according to the respective maximum expression in response to nitrogen deficiency. As proof of concept, these groups include known *hep, hgl* and *nif* genes, but also several new elements that became clustered with them on the basis of their transcriptional regulation. Most importantly, because our approach uses labelled RNA directly in the hybridization of microarrays, it avoids the pitfalls of cDNA synthesis. In addition, because the customized microarrays are not restricted to protein-coding genes, we also identified heterocyst-specific antisense transcripts and potentially trans-acting RNAs that might exert regulatory roles in these differentiated cells.

Searching for sequences conserved in the promoters of either of the two DIF clusters and in their homologues in heterocystous cyanobacteria led to the discovery of a new motif in the E-DIF group (that we called DIF2), together with the more prevalent DIF1 motif that has been previously identified in several HetR-dependent promoters with heterocyst-specific expression (Mitschke *et al*., 2011). Mutational analysis clearly demonstrated the connection between this DIF2 motif and heterocyst-specific expression. Finally, a previously defined motif mediating the CnfR-regulated expression of a promoter upstream of *nifB* (Vernon *et al*., 2017) or *nifP* (Tsujimoto *et al*., 2016) was also identified in a third transcriptional unit in the L-DIF group. Our results show that substantially more genes, elements and factors might be involved in the genetic program that leads to the differentiation of cyanobacterial heterocysts than anticipated so far.

Results

Transcriptomic analysis of the responses to nitrogen deprivation

Using differential RNA-Seq, we previously described the early (8 h) transcriptional responses of *Nostoc* sp. PCC 7120 to nitrogen deprivation (Mitschke *et al*., 2011). Transcriptional start sites (TSS) and their N-responsive expression dynamics in the wild type and a *hetR* mutant strain, which accumulates high levels of a non-functional HetR protein (Risser and Callahan, 2007) and is unable to differentiate heterocysts, were analysed (Mitschke *et al*., 2011). Based on these transcriptomic data, we designed high density microarrays that allowed the direct hybridization of labelled RNA samples. In contrast to previous approaches, these arrays contain probes for all possible transcripts originating at TSS categorized not only as driving the transcription of protein-coding genes (gTSS) but also of antisense transcripts (aTSS) and non-coding transcripts, sRNAs (nTSS) (see Supporting Information for details). To define the HetR-dependent transcriptome as an approach to find new elements potentially involved in heterocyst differentiation and function, microarrays were hybridized to RNA samples from the wild type strain and the *hetR* mutant along a time course of nitrogen deficiency (0, 6, 12 and 24 h after nitrogen step down). In order to construct a co-expression network, we have also analysed RNA samples from wild-type cells growing using N_2 (steady state) and 8 h after the addition of ammonium and another two samples from the wild-type strain corresponding to the conditions of our previous dRNA-Seq analysis (0–8 h after nitrogen step-down).

Expression data were compared between two samples as follows: every sample was compared to the absolute reference (WT-0 from the -N time-course experiment, i.e., ammonium-grown cells) and every sample was compared to its own reference (for instance *hetR*-12 vs. *hetR*-0). In order to reduce noise and false-positives, only transcripts with a difference in their absolute expression values higher than $\log_2 = 1.25$ entered the clustering analysis of transcriptional dynamics (see Methods and Supporting Information Dataset S1). The $\log_2 > 1.25$ threshold secured that only genes that showed a pronounced transcriptional response were included in the clustering analysis. Ten clusters were resolved (Supporting Information Dataset S2 and Fig. S1). Six clusters included elements that were induced (#3, #5, #9) or repressed (#2, #7, #10) both in the wild type and the *hetR* mutant, therefore suggesting these changes corresponded to global responses to nitrogen deficiency not specifically related to heterocyst differentiation. Among these, for instance, clusters #9 and #10 included previously identified responses directly regulated by NtcA (Mitschke *et al*., 2011), such as the inducible gene *glnA* (encoding glutamine synthetase; cluster #9) (Frías

et al., 1994) or the repressed gene *gifA* (encoding the glutamine synthetase inactivating factor IF7; cluster #10) (Galmozzi *et al*., 2010). Two clusters contained elements that, regardless the time point, showed either a negative (#1) or positive (#8) change in the *hetR* mutant with respect to the wild type, and therefore were not nitrogen regulated.

In order to identify HetR-dependent responses, we selected for further analysis the two clusters that contained elements whose expression was induced in response to nitrogen deficiency in the wild-type strain but not in the *hetR* mutant. HetR-dependent inducible transcripts appeared in two different clusters, that we have named as E-DIF (early in differentiation, #6) and L-DIF (late in differentiation, #4) according to the temporal profile of their induction in the wild-type strain (Fig. 1 and Supporting Information Tables S1, S2). Information on the heterocyst-specific expression or implication in heterocyst differentiation or function existed for 27 of the 72 elements assigned to the E-DIF cluster and for 45 of the 86 elements of the L-DIF cluster. Interestingly, 14 of the E-DIF members and 5 of the L-DIF members were classified as putative antisense or trans-acting sRNAs (Fig. 1 and Supporting Information Tables S1, S2). The enrichment of previously characterized heterocyst-related genes in these two clusters supports the idea that the previously unknown elements could be relevant. The finding of several non-coding RNAs suggests a frequent role of these versatile regulatory molecules in heterocyst differentiation.

Transcripts in the E-DIF cluster were induced at an earlier time after nitrogen deprivation (6 h) than transcripts in the L-DIF cluster, which appeared induced in the 12- or even the 24-h samples (Fig. 1A and B). The transcriptional profiles of *hepA*, *hglE*, *hetN* and *nifP*, four previously characterized genes, illustrate this different timing of maximum expression (Fig. 1C). Consistent with their known sequential roles in the differentiation process (reviewed in Flores *et al*., 2018), the *hep* (heterocyst polysaccharide) genomic islands appeared in the E-DIF cluster, while *hgl* (heterocyst glycolipids) and *nif* genes appeared in the L-DIF cluster (Fig. 1B). Expression data for the two previously characterized *hep* and *hgl* genomic islands are visualized as genome plots in Supporting Information Fig. S2, together with the positions of several transposon insertions that yielded a Fox^{-} phenotype (Fan *et al*., 2005; Huang *et al*., 2005). The complete results of the microarray analyses are presented as whole genome plots in Supporting Information data files 1–4.

The E-DIF cluster of HetR-dependent early responses to nitrogen deprivation

Elements that appeared in the E-DIF cluster are shown in Fig. 1B (left panel). In addition to *hep* genes (*hep* island *alr2822* to *alr2841* and genes *alr3698* to *alr3700*) and *devH*, encoding a transcriptional regulator involved in synthesis of the *hgl* layer (Ramírez *et al*., 2005) the genes encoding NtcA (global nitrogen regulator) and HetR (specific differentiation regulator), both previously described as induced in heterocysts (Black *et al*., 1993; Olmedo-Verd *et al*., 2006), also appeared in the E-DIF cluster, indicating that they both were similarly induced in a HetR-dependent way in early stages of differentiation. In fact, when integrated in plasmid alpha in single copy, a *gfp* fusion to the promoter driving transcription from position −271 with respect to the *hetR* open reading frame, the major inducible TSS of *hetR* corresponding to genomic position 2821366f (Rajagopalan and Callahan, 2010; Mitschke *et al*., 2011), was transcribed in prospective heterocysts before any morphological sign of differentiation was observed (Supporting Information Fig. S3). Quantification of the green fluorescence indicated that the expression of GFP was already strong in cells that were only initiating their transformation into heterocysts, as indicated by the presence of chlorophyll-derived red autofluorescence, characteristic of vegetative cells and immature prospective heterocysts (Supporting Information Fig. S3B). Such early expression is consistent with previous observations for other elements, such as NsiR1 (Muro-Pastor, 2014) or *sigC* (Muro-Pastor *et al*., 2017) that also appeared in the E-DIF cluster.

Initial stages of heterocyst differentiation involve the establishment of a precise pattern of one heterocyst every 10 to 15 vegetative cells in the filaments. One gene involved in such patterning is *patS*, encoding a small peptide (that includes the sequence RGSGR) produced in heterocysts that represses differentiation of neighbouring vegetative cells (Yoon and Golden, 1998). Consistent with its role in early inhibition of differentiation, *patS* appeared in the E-DIF cluster. Recently, another gene encoding a RGTGR-containing peptide that would act as a second, PatS-like inhibitor of heterocyst differentiation, has been described and named as *patX* (Elhai and Khudyakov, 2018). According to our data, *asl2332*, the gene encoding PatX, appears also within the E-DIF cluster, suggesting that the, perhaps redundant, inhibitory effects of both PatS and PatX would take place in early stages of differentiation. *patC*, described as participating in patterning of heterocysts (Corrales-Guerrero *et al*., 2014), also appears in the E-DIF cluster.

Transcription of *alr1174* and *alr3808*, encoding rubrerythrin A and a Dps protein, respectively, has been previously described to be induced in heterocysts (Zhao *et al*., 2007; Li *et al*., 2015). According to our data, these two genes, likely involved in protection against reactive oxygen species, appear in the E-DIF cluster, and therefore are transcribed during the initial stages of heterocyst differentiation.

The E-DIF cluster (as well as the L-DIF cluster, see below) contains several nitrogen-starvation inducible

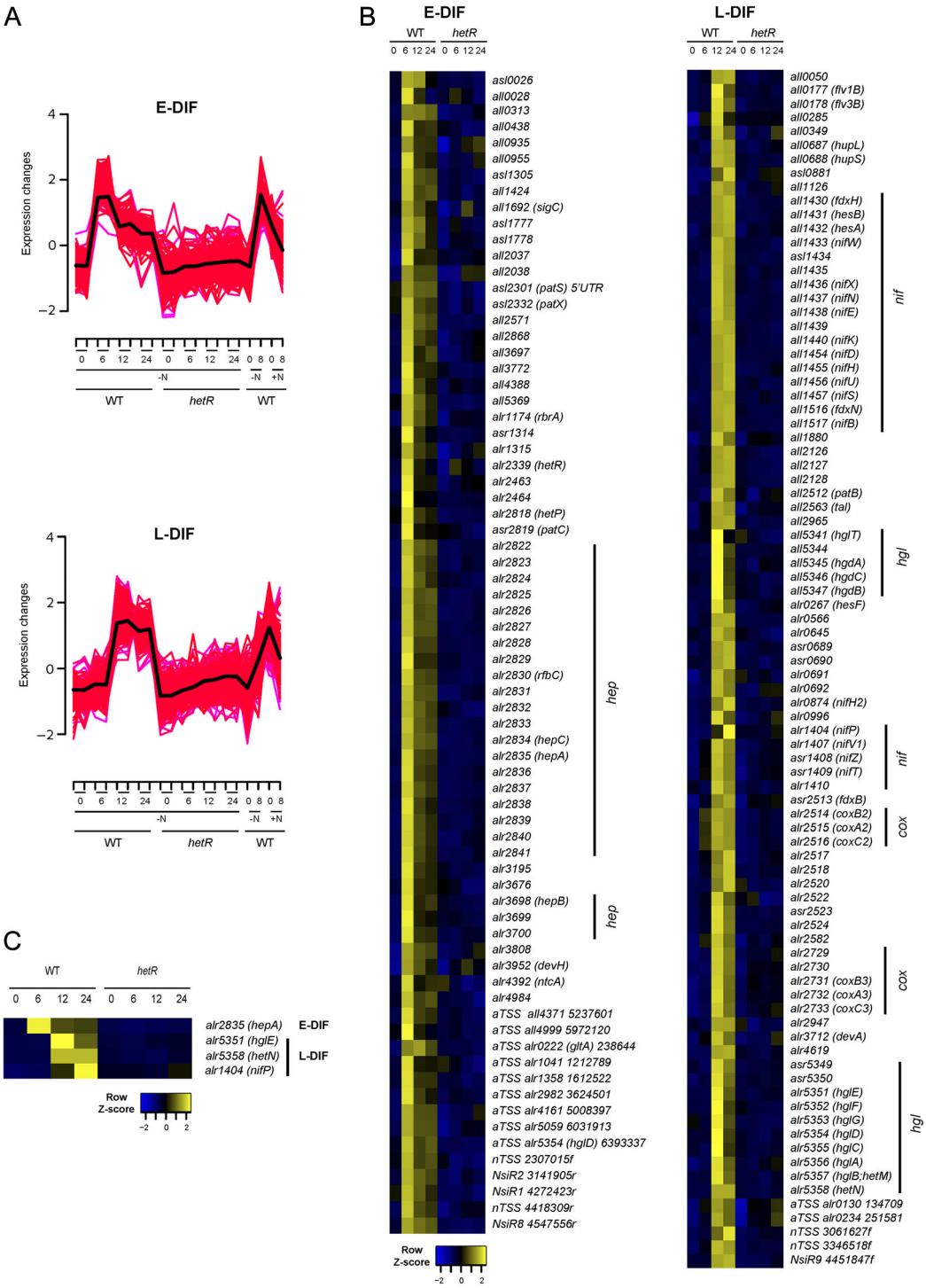

Fig. 1. The E-DIF and L-DIF clusters of HetR-dependent inducible transcripts.
A. Expression changes of transcripts classified as early in differentiation (E-DIF; upper panel) and late in differentiation (L-DIF; lower panel) along the 20 samples analysed by microarray hybridization. Samples corresponded, from left to right, to a time course after nitrogen depletion (-N; 0, 6, 12 and 24 h after nitrogen removal; two biological replicates for each sample) of the wild type and the *hetR* mutant, as well as one nitrogen step down experiment (-N; 0 and 8 h after nitrogen removal) and a nitrogen addition experiment (+N; 0 and 8 h after addition of ammonium to N_2-growing cells) of the wild-type strain. Expression changes are indicated as Z-scores.
B. Expression of elements included in both clusters (E-DIF, left panel; L-DIF, right panel) along the time course (indicated in hours) of nitrogen depletion in the wild type and the *hetR* mutant. Several previously described transcriptional units and clusters of related genes are indicated by vertical black bars.
C. Expression of four selected previously known transcripts exhibiting different temporal profiles of induction after nitrogen depletion. In B-C, colour key represents row Z-scores for expression of each listed element, and therefore the values are not absolute but relative.

non-coding RNAs, including NsiR1, an early marker of heterocyst differentiation (Ionescu *et al*., 2010; Muro-Pastor, 2014), NsiR2 (Mitschke *et al*., 2011), whose heterocyst-specific expression is shown here by *gfp* fusion to its promoter (Supporting Information Fig. S4) or NsiR8 (Brenes-Álvarez *et al*., 2016), as well as several transcripts produced in antisense orientation with respect to annotated genes. As an example of an N-regulated, HetR-dependent, antisense transcript, we have further analysed transcription from the aTSS at position 6031913r, antisense to *alr5059* encoding a putative peptidyl prolyl cis-trans isomerase (Fig. 2). Northern blot analysis demonstrated that this HetR-dependent antisense transcript accumulated already 6 h after nitrogen removal (Fig. 2B). Its heterocyst-specific transcription from position 6031913r was verified by construction of a fusion to *gfp.* Quantification of GFP signal and red autofluorescence indicated that the strongest GFP expression took place in cells that, according to the presence of red autofluorescence, were still at an early stage of differentiation (Fig. 2C-E), consistent with the appearance of this antisense transcript in the E-DIF cluster.

The L-DIF cluster of HetR-dependent late responses to nitrogen deprivation

Elements that appeared in the L-DIF cluster of transcriptional responses are shown in Fig. 1B (right panel). Whereas the dynamics of expression of genes in the E-DIF cluster was very homogeneous, with induction already in the 6-h sample (Fig. 1A), the expression of transcripts in the L-DIF cluster appeared in several, slightly different, temporal profiles. While in all instances the expression was low in the 6-h sample, highest expression was observed in the 12-h sample, the 24-h sample, or both as illustrated by the examples in Fig. 1C.

Similar to the E-DIF cluster, previously characterized genes appeared in the L-DIF cluster. These included *hgl* and *nif* genes, as well as *hetN*, encoding a putative ketoacyl reductase involved in the maintenance of the heterocyst pattern along the filaments (Callahan and Buikema, 2001). In addition, genes *flv1B* and *flv3B* encoding two heterocyst-specific flavodiiron proteins (Ermakova *et al*., 2013), and the *cox2* and *cox3* operons encoding cytochrome oxidases (Valladares *et al*., 2007), all associated to oxygen protection in heterocysts, were assigned to the L-DIF cluster.

Concerning the heterocyst-specific uptake hydrogenase, in addition to genes *hupS-hupL*, genes *asr0689* to *alr0692*, located between *all0688* (*hupS*) and the *hyp* operon (*hypF* to *hypB*) encoding maturation proteins, all appeared in the L-DIF cluster. This suggests that the products of these genes might be functionally related and is further consistent with the definition of an 'extended' *hyp* operon including genes *asr0689* to *alr0693* (Agervald *et al*., 2008). The predicted products of *asr0689*, *asr0690* and *alr0691* are annotated as unknown or hypothetical proteins. However, *alr0692* would encode a protein with homology to NfuA involved in the maturation of Fe-S proteins in *E. coli* under oxidative stress and iron-starvation conditions (Angelini *et al*., 2008).

As an example of a previously undescribed gene whose expression with an L-DIF profile suggests a role in heterocyst differentiation, we have validated HetR dependent, heterocyst-specific expression of *alr0645* (Fig. 3). Alr0645 possesses four predicted transmembrane domains within a phage holin superfamily IV domain involved in transport of proteins across the cytoplasmic membrane. Northern blot analysis identified transcripts of about 1 kb with increased abundance in the 12-h sample and strongest accumulation 24 h after combined nitrogen removal (Fig. 3B). We verified the HetR-regulated expression from position 748886f by primer extension analysis (Fig. 3C), which is consistent with a previously annotated gTSS (Mitschke *et al*., 2011). Heterocyst-specific transcription of *alr0645* was analysed in a promoter fusion to *gfp.* Quantification of GFP signals and red autofluorescence indicated that the strongest expression of GFP took place in cells that, according to the absence of red autofluorescence, were already at an advanced stage of differentiation (Fig. 3D-F), consistent with assigning *alr0645* to the L-DIF cluster.

Similar to the E-DIF cluster, the L-DIF cluster contains several non-coding transcripts, including NsiR9 (Brenes-Álvarez *et al*., 2016). Consistent with its position in the L-DIF cluster, heterocyst-specific transcription from the promoter of NsiR9, analysed in a fusion to *gfp*, took place in heterocysts at an advanced stage of differentiation (Supporting Information Fig. S4).

Conserved sequences associated to heterocyst-specific expression in the E-DIF and L-DIF clusters

The promoter regions (750 bp) upstream from all genes included in each of the two DIF clusters and their homologues in genomes of heterocystous cyanobacteria (Shih *et al*., 2013) were searched for conserved motifs using DREME (see Methods for details). The position-specific occurrences of sequences matching the conserved motifs identified by DREME were analysed by MAST in the promoters of each cluster (from −300 to +50 with respect to the corresponding TSS). Only the promoter for the first gene in each transcriptional unit was addressed (see Methods for details on selection of TSS to be included in the analysis, and Supporting Information Tables S1 and S2).

For the E-DIF group, sequences corresponding to the DIF1 motif previously defined as TCCGGA (Mitschke

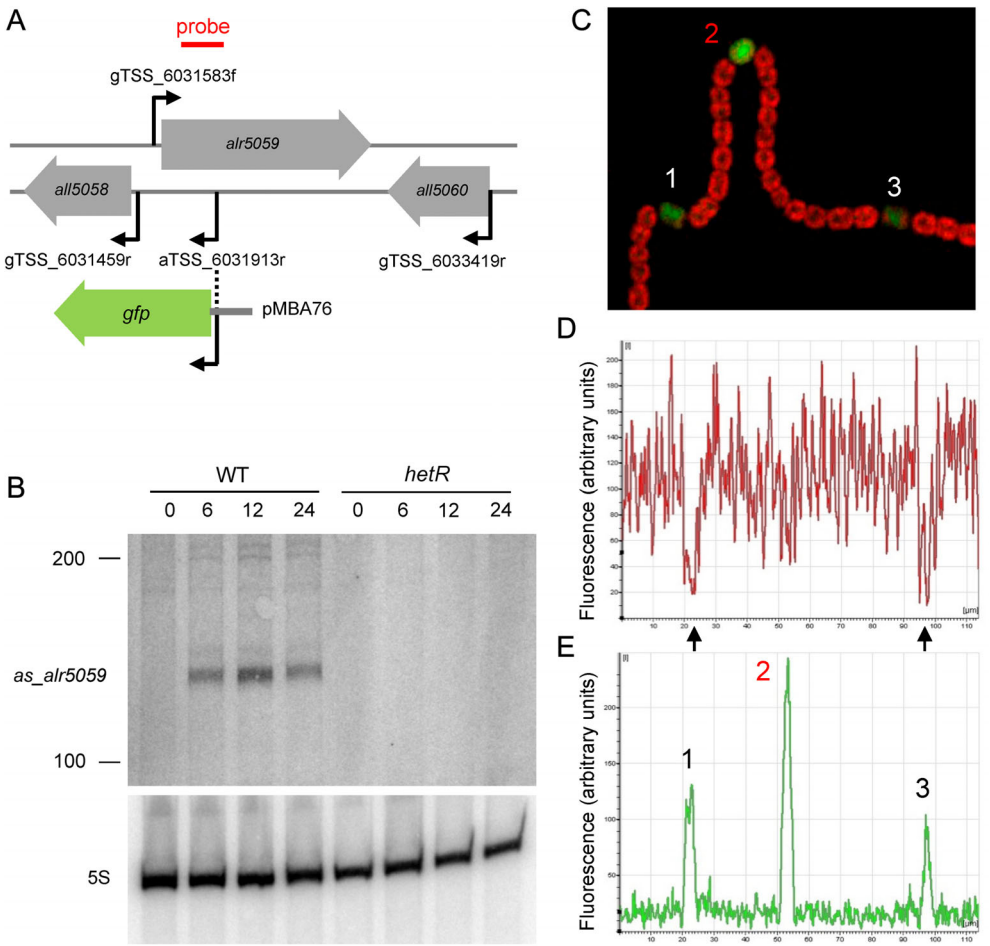

Fig. 2. *as_alr5059* is an N-regulated, HetR-dependent antisense transcript in the E-DIF cluster.
A. Scheme of the chromosomal region covering *all5058* to *all5060*. gTSS annotated for the three genes according to (Mitschke et al., 2011) are indicated by bent arrows, as well as the aTSS inside *alr5059*. A scheme of the fragment fused to *gfp* in pMBA76 is shown, as well as the fragment used as RNA probe in Northern blot (red bar).
B. Northern blot analysis of *as_alr5059* in *Nostoc* sp. PCC 7120 and mutant strain 216 (*hetR*) (upper panel) at different time points (indicated in hours) after nitrogen removal. Hybridization to a probe for 5S rRNA was used as loading control (bottom panel). Sizes are indicated in nt.
C. Expression of P*as_alr5059-gfp* along nitrogen-fixing filaments of *Nostoc* sp. PCC 7120 bearing pMBA76. Confocal fluorescence image (red and green channels merged) of a filament growing on top of nitrogen-free medium is shown together with quantification of the signals for the red (autofluorescence, D) and green (GFP, E) channels. Mature heterocysts (#1 and #3) are indicated in white (C) or black (E). Immature heterocyst (#2) is indicated in red. Black arrows in (D) indicate positions of gaps in red autofluorescence, indicative of mature heterocysts.

et al., 2011) were found around position −35 with respect to the TSS in 30 out of 45 promoters (two mismatches allowed) (Supporting Information Fig. S5), including perfect matches in the promoters of *patS* (−580), *hetR* (−271, Supporting Information Fig. S3), or the promoter for the antisense transcript to *alr5059* described above (Fig. 2). The phylogenetic conservation of DIF1 motifs across genomes from heterocystous cyanobacteria is illustrated by the occurrence of the sequence TCCGGA around position −35 in the promoters for one sRNA (NsiR1) and one previously unknown inducible gene (*alr3676*) in the E-DIF group (Supporting Information Fig. S6). Consistent with recent observations (Elhai and Khudyakov, 2018; Wegelius *et al*., 2018), a conserved G (likely associated to the −10 box of these promoters) is located 16–18 nt downstream of the DIF1 motif (Supporting Information Fig. S5 and S6).

In addition to the DIF1 motif that is prevalent in the promoters of the E-DIF cluster, a second motif appeared slightly upstream from position −35. This motif was identified twice by MAST in the promoters of two sRNAs, NsiR2 (Mitschke *et al*., 2011) and NsiR8 (Brenes-Álvarez *et al*., 2016), as well as in the promoters corresponding to open reading frames *alr2463*, *asr1314* and to *as_all4999* (Fig. 4A). The motif, which we have named as DIF2, is a hexameric direct repeat separated by two nucleotides, with the consensus sequence TTGGCAWWTTGGCA. The DIF2 motif is phylogenetically conserved in the promoters of all homologues of NsiR2, NsiR8 and *alr2463* chosen as examples (Supporting Information Fig. S7).

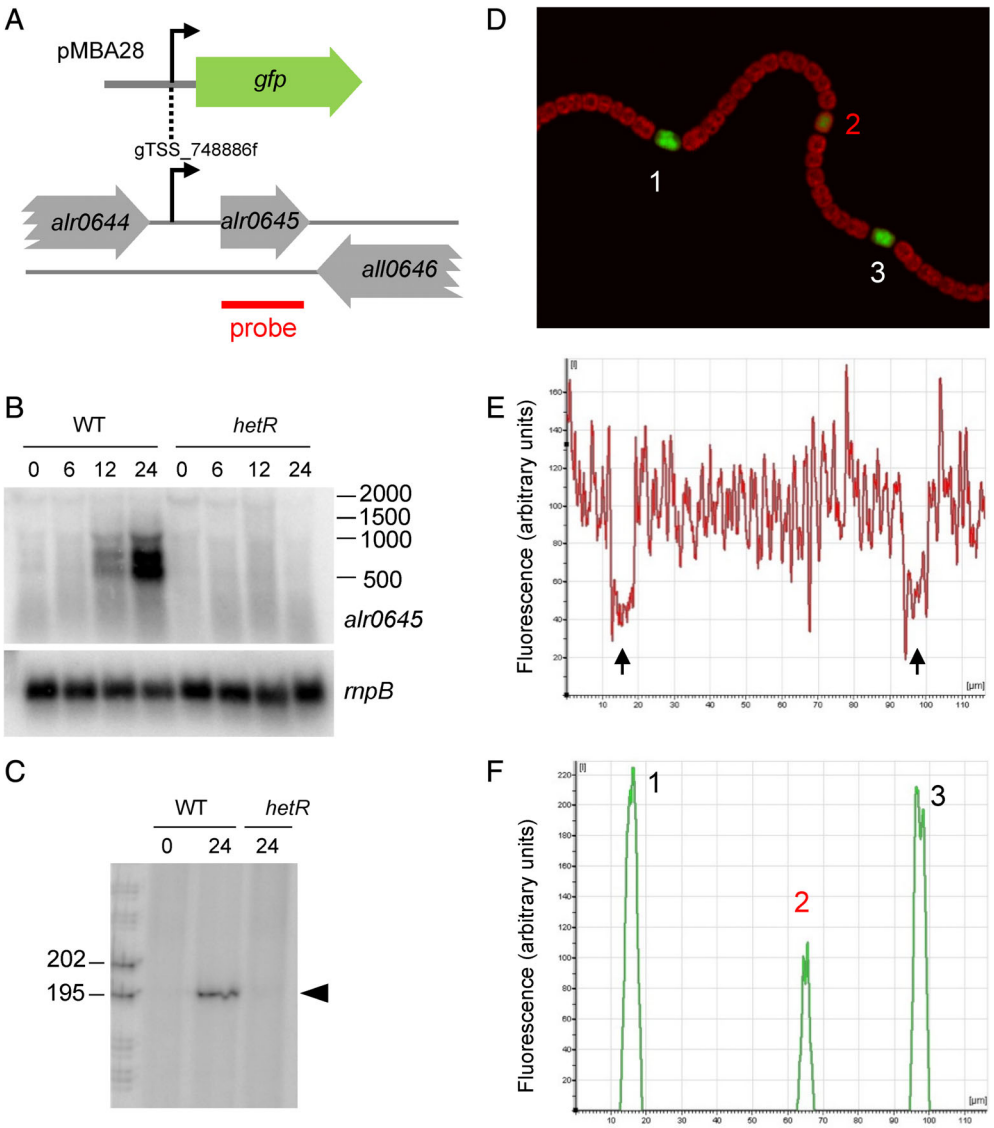

Fig. 3. *alr0645* is an N-regulated, HetR-dependent gene in the L-DIF cluster.
A. Scheme of the chromosomal region covering *alr0644* to *all0646*. gTSS for *alr0645* according to (Mitschke *et al.*, 2011) is indicated by a bent arrow. The fragment fused to *gfp* in pMBA28 is shown, as well as the fragment used as DNA probe in Northern blot (red bar).
B. Northern blot analysis of *alr0645* in *Nostoc* sp. PCC 7120 and mutant strain 216 (*hetR*) at different times after combined nitrogen removal (upper panel). Hybridization to a probe for *rnpB* was used as a loading control (lower panel).
C. Primer extension analysis (oligonucleotide alr0645_3, Supporting Information Table S3) of transcripts starting at position 748886f in RNA isolated from the wild type strain (WT) or the *hetR* mutant strain after the cells were transferred to nitrogen-free medium for 24 h.
D. Expression of $P_{alr0645}$-*gfp* along nitrogen-fixing filaments of *Nostoc* sp. strain PCC 7120 bearing pMBA28. Confocal fluorescence image (red and green channels merged) of a filament growing on top of nitrogen-free medium is shown. Quantification of the signals along the filament is shown separately for the red (E) and green (F) channels. Mature heterocysts (#1 and #3) are indicated in white (D) or black (F). Immature heterocyst (#2) is indicated in red. Black arrows in (E) indicate the positions of gaps in red autofluorescence, indicative of mature heterocysts.
Sizes are indicated in nt.

To test the functional relevance of the DIF2 motif, two mutations were introduced that altered the first repeat (pSAM321) or increased the spacing between repeats (pSAM322) in the promoter of NsiR2 that was fused to *gfp* in pSAM320, bearing a promoter fragment from position −117 to +19 with respect to the TSS (Supporting Information Table S4 and Fig. S8). Each variant was introduced into wild type *Nostoc* sp. PCC 7120 and integrated into plasmid alpha by homologous recombination. Hence, the resulting strains bear two copies of the NsiR2 promoter, the native one in the chromosome, driving transcription of NsiR2, and one of the variants fused to *gfp* integrated in plasmid alpha. Transcription from both NsiR2 promoters (that starts at two closely spaced positions) was analysed simultaneously by primer extension using two different primers, one specific for the

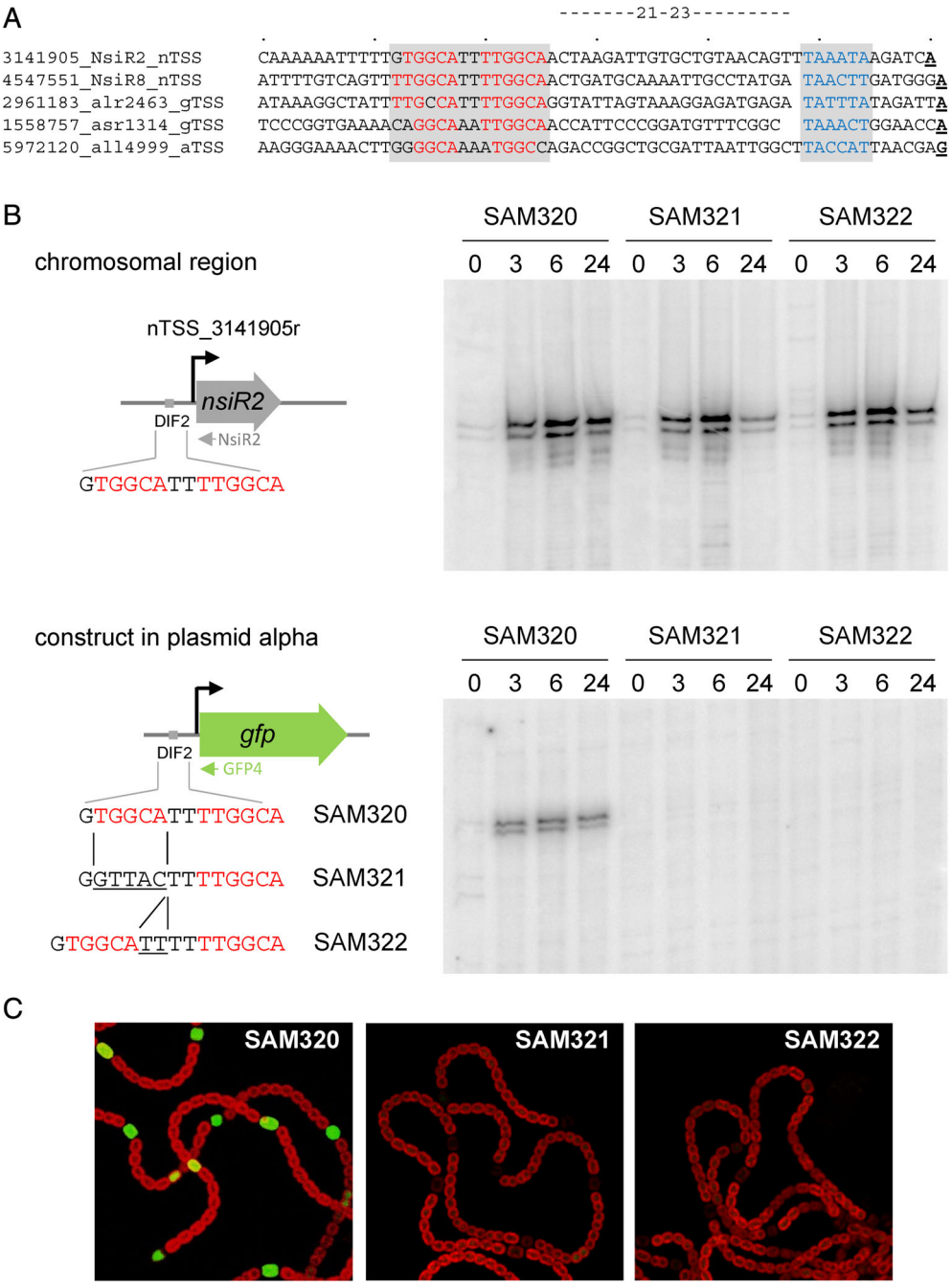

Fig. 4. DIF2 motif associated to heterocyst-specific expression of promoters of the E-DIF group.
A. Scheme of HetR-dependent promoter sequences in the E-DIF group that contain a putative DIF2 motif. Sequences matching the TTGGCA repeats are indicated in red, and putative −10 sequences are indicated in blue.
B. Primer extension analysis of the TSS of NsiR2 in strains-bearing plasmids pSAM320, pSAM321 or pSAM322 integrated in plasmid alpha. Schemes of the wild-type chromosomal *nsiR2* region and the different mutant variants of the additional *nsiR2* promoter fused to *gfp* (integrated in plasmid alpha) are shown on the left. The positions of the oligonucleotides used as primers are indicated as arrows in the scheme of the chromosomal region (in grey for NsiR2) or in the scheme of the construct integrated in plasmid alpha (in green for GFP4). Primer extension analysis with primer NsiR2 (upper panel) or GFP4 (lower panel) was carried out using RNA isolated from strains SAM320, SAM321 and SAM322 after the cells were transferred to nitrogen-free medium for the number of hours indicated.
C. Expression pattern of GFP from the wild-type promoter of NsiR2 (strain SAM320), or from the two mutated variants in strains SAM321 or SAM322. Images correspond to merged red (autofluorescence) and green (GFP) fluorescence.

chromosomal copy, and another specific for the *gfp* fusion integrated in plasmid alpha (Supporting Information Fig. S8). The results of the primer extension (Fig. 4B) and confocal fluorescence microscopy (Fig. 4C) showed that the 135-nt DNA fragment fused to *gfp* in pSAM320 was sufficient to promote early heterocyst-specific expression, and that either the mutation of the upstream repeat or the increase in the spacing between the two repeats led to lack of GFP expression.

In contrast to these findings for the E-DIF cluster, the comparison of promoters in the L-DIF cluster only led to the identification of conserved motifs that corresponded to the sequences associated to the CnfR (a.k.a. PatB, *alr2512*, also in the L-DIF cluster)-regulated expression of the *nifB* promoter of *Anabaena variabilis* ATCC 29413 (Vernon *et al*., 2017) and the *nifB* and *nifP* promoters of the non-heterocystous cyanobacterium *Leptolyngbya boryana* (Tsujimoto *et al*., 2016) (Supporting Information Fig. S9A). In addition to *nifB* and *nifP*, we identified the CnfR-associated sequence motif also in the promoter driving transcription of a tricistronic operon (*alr2522* to *alr2524*) with unclear function. *all2522* has been assigned to the core genome of heterocyst-forming cyanobacteria (Simm *et al*., 2015) and is up-regulated in a mutant over-expressing the sigma factor SigJ that confers resistance to desiccation by activating genes involved in polysaccharide synthesis (Yoshimura *et al*., 2007). *asr2523* and *alr2524* appeared among genes down-regulated in mutants of sigma factors SigC or SigE (Ehira and Miyazaki, 2015). Phylogenetic conservation of CnfR-associated sequence motifs in the promoter regions of *nifB*, *nifP* and *alr2522* homologues across genomes of heterocystous cyanobacteria is shown in Supporting Information Fig. S9B together with the TSS determined for *nifB* in *Nostoc* sp. PCC 7120 (Mulligan and Haselkorn, 1989; Mitschke *et al*., 2011) and in *Nodularia spumigena* CCY9414 (Voβ *et al*., 2013) as well as the TSS determined for the *nifP* homologue in *Nodularia spumigena* CCY9414 (Voβ *et al*., 2013).

Discussion

Heterocyst differentiation in filamentous cyanobacteria involves substantial transcriptome remodelling. Most of these changes take place specifically in the cells that start differentiating as heterocysts. In contrast to general responses to nitrogen deficiency, mainly aimed at nitrogen scavenging and under control of NtcA, transcriptional responses specifically involved in heterocyst differentiation additionally depend on the master regulator HetR. In this study, we have identified two groups (early and late) of HetR-dependent (DIF) genetic elements (genes, sRNAs, antisense transcripts) involved in heterocyst differentiation and/or function. The approach we have followed is based on the clustering of similar transcriptional profiles with changes above a certain threshold. This approach allows the identification of essential elements, which are not possible to mutate and of functionally redundant or accessory elements that would not yield a distinct phenotype when mutated. Nevertheless, such elements can have a role in heterocyst differentiation or function.

Using GFP as a reporter, we demonstrate that transcripts included in the DIF clusters defined here are expressed specifically in heterocysts. This is consistent with their dependence on HetR and corroborates the different temporal profiles that led to their inclusion in the E-DIF or L-DIF groups. Both groups include previously known genes but also genes of unknown function. A schematic representation of all elements identified in both DIF groups defined here is shown in Fig. 5. Genes previously related to heterocyst differentiation and function are shown in the left side of the scheme, whereas elements for which previously no function has been defined in the context of heterocyst differentiation are shown to the right.

NtcA and HetR are mutually dependent in the context of differentiation (Muro-Pastor *et al*., 2002). Both genes are transcribed from complex promoter regions that include multiple, constitutive and inducible TSS (Olmedo-Verd *et al*., 2006; Rajagopalan and Callahan, 2010; Mitschke *et al*., 2011). The observation that their transcripts accumulate with E-DIF dynamics suggests their heterocyst-specific promoters make a major contribution to their transcriptional response to the absence of combined nitrogen. In both cases, the regulated TSS (position −271 for *hetR* and − 180 for *ntcA*) is controlled by a promoter containing a DIF1 motif (Supporting Information Fig. S5). The observation that the expression of *ntcA* itself correlates with the expression of heterocyst-specific genes is consistent with the higher amount of NtcA protein observed in heterocysts versus vegetative cells (Sandh *et al*., 2014). Increased amounts of NtcA in developing heterocysts might contribute to the differential regulation of NtcA-dependent genes in heterocysts versus vegetative cells. In this context, increased transcription of NtcA-induced genes, such as *glnA* or *nblA*, in developing heterocysts, could lead to higher levels of certain proteins that are otherwise not heterocyst-specific, resulting, for instance, in higher glutamine synthetase activity or phycobilisome degradation in those specialized cells. Similarly, regulatory loops involving sRNAs regulated by NtcA, such as NsrR1 (Álvarez-Escribano *et al*., 2018), could operate more strongly in heterocysts versus vegetative cells.

The *patS* and *hetN* genes appeared in the E-DIF and L-DIF clusters, respectively. This is consistent with recent observations concerning the relative timing of their expression (Di Patti *et al*., 2018), and the roles in early establishment (PatS) or maintenance (HetN) of the

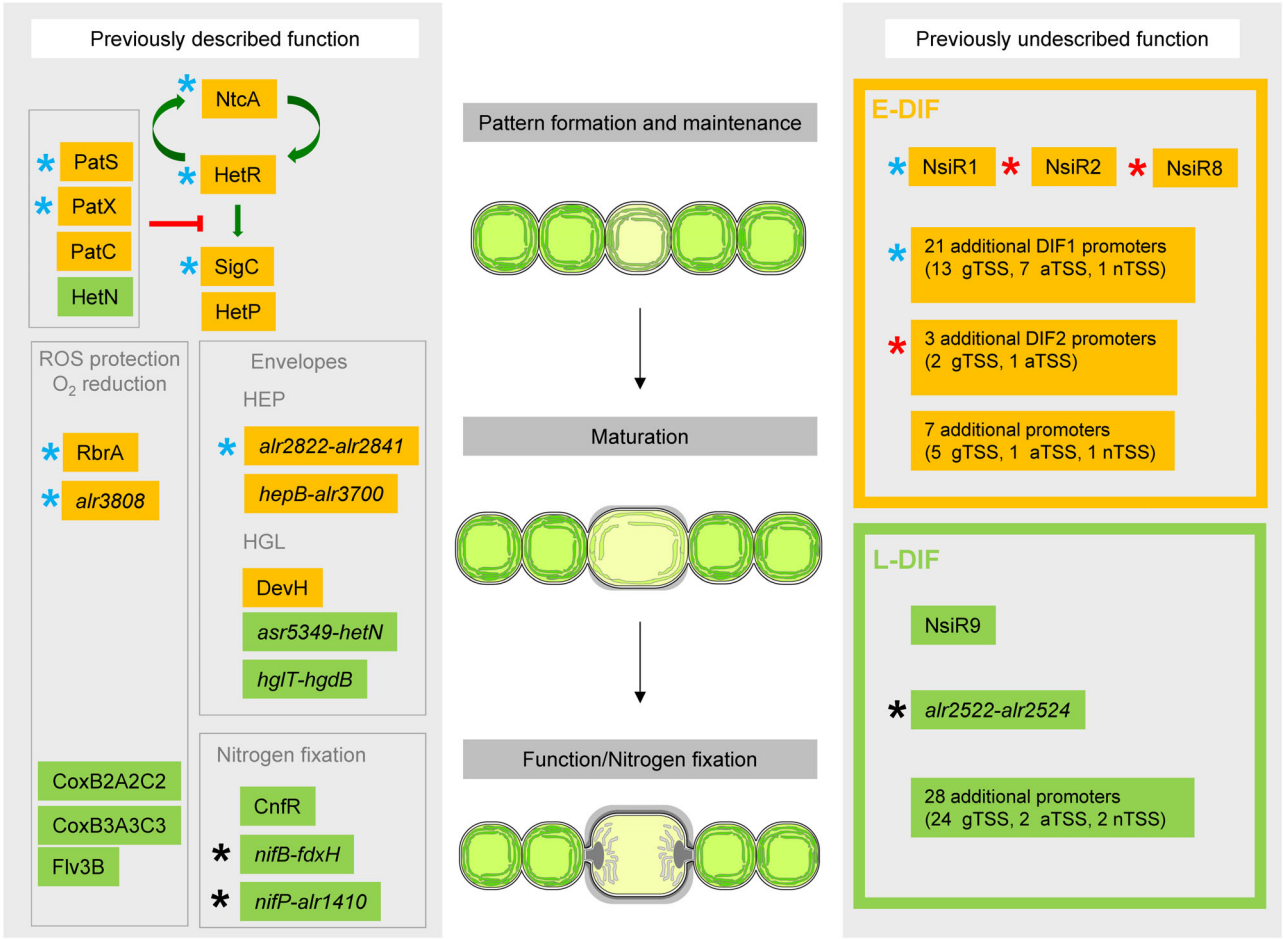

Fig. 5. The HetR-dependent transcriptome. Previously characterized genes and genetic elements defined here are summarized according to their assignment to the E-DIF (orange) and L-DIF (green) groups defined in this work. Genes appearing in the two groups are shown along a temporal sequence of differentiation and according to their previously known (left side) or unknown (right side) implication in the differentiation of heterocysts. Elements-bearing promoters with a recognizable conserved sequence motif are indicated with blue (DIF1), red (DIF2) or black (CnfR) asterisks. Note that this scheme is restricted to genes appearing in the two DIF clusters analysed in this study, and therefore refers to genes with transcriptional $\log_2$FC > 1.25. Also, although heterocyst-specific expression has been demonstrated for sRNAs NsiR1, NsiR2, NsiR8 and NsiR9, a possible role in heterocyst differentiation remains to be demonstrated, therefore their position in the right side of the scheme. gTSS, aTSS, nTSS: TSS corresponding to genes, antisense RNAs or non-coding RNAs, respectively.

pattern (Yoon and Golden, 1998; Callahan and Buikema, 2001). Recently, it has been demonstrated that *patS* and *hetN* are not functionally equivalent when their promoter regions are switched, therefore suggesting their respective timing of expression is critical for proper function (Rivers *et al*., 2018). Two other genes described as participating in patterning of heterocysts, *patC* (Corrales-Guerrero *et al*., 2014) and *patX* (Elhai and Khudyakov, 2018), also appeared in the E-DIF group. Therefore, according to the transcriptional profiles described here, up to three elements would act as negative regulators during initial steps of differentiation in *Nostoc* sp. PCC 7120. This observation illustrates possible functional redundancy among critical players governing cell differentiation processes.

Timing of expression of genes encoding proteins related to oxygen protection suggests an early role during development for elements like RbrA, a cyanobacterial rubrerythrin (Zhao *et al*., 2007) or the Dps protein encoded by *alr3808* (Li *et al*., 2015), both in the E-DIF group (Fig. 1B). In contrast, the heterocyst-specific flavodiiron proteins encoded by *flv1B* and *flv3B* (Ermakova *et al*., 2013; Ermakova *et al*., 2014) or the cytochrome oxydases encoded by *coxBAC2* and *coxBAC3* (Valladares *et al*., 2007), all in the L-DIF cluster, would likely be involved in functional aspects of mature heterocysts.

The relatively homogeneous temporal dynamics observed for genes in the E-DIF cluster points to a limited number of participating regulatory factors. Accordingly, the search for regulatory sequences that appear statistically enriched in the E-DIF cluster led to the identification of the DIF1-motif (TCCGGA) (Mitschke *et al*., 2011) previously associated to early heterocyst-specific expression (Muro-Pastor, 2014; Muro-Pastor *et al*., 2017). A new

motif (TTGGCAWWTTGGCA), which we have called DIF2, is present in the promoters of several transcripts (including sRNAs NsiR2 and NsiR8) that also belong to the early heterocyst-specific transcriptome. The slightly different temporal profiles observed for induction of L-DIF elements, likely due to the participation of several different regulators, is consistent with a limited identification of conserved promoter elements in this group. Nevertheless, the array of conserved sequences defined as CnfR-regulated promoters for *nifB* (Vernon *et al*., 2017) and *nifP* (Tsujimoto *et al*., 2016) could also be identified in the promoter driving transcription of another gene cluster with late DIF profile (*alr2522-alr2524*), suggesting these genes might also be under the control of CnfR in *Nostoc* sp. PCC 7120. The presence of DIF1, DIF2 or CnfR motifs in the corresponding promoters is indicated with asterisks in the elements shown in Fig. 5.

Phylogenetic conservation of the occurrence of DIF1, DIF2 or CnfR motifs in certain genes suggests the conservation of the transcriptional machinery involved in regulated transcription of the corresponding promoters across heterocystous cyanobacteria. Whether specialized sigma factors, repressors or activators are involved in the regulation of heterocyst-specific promoters described here is currently unknown. Genes encoding sigma factors SigC, SigE and SigG are transcribed heterocyst-specifically (Aldea *et al*., 2007), although the enhanced transcription of *sigG* in heterocysts might be a consequence of cell envelope stress during differentiation (Bell *et al*., 2017). Furthermore, *sigC* is transcribed in very early stages of differentiation (Muro-Pastor *et al*., 2017). Whereas *sigC* or *sigE* mutants show a slight delay in differentiation (Khudyakov and Golden, 2001; Mella-Herrera *et al*., 2011) and both SigC and SigE are required for normal expression of heterocyst-specific genes (Mella-Herrera *et al*., 2011; Ehira and Miyazaki, 2015), no single sigma factor is exclusively associated to heterocyst development, likely due to functional redundancy. Identification of the proteins involved in heterocyst-specific regulated transcription is a challenge for future research.

The N-responsive transcriptome of *Nostoc* sp. PCC 7120 includes abundant non-coding transcripts (Flaherty *et al*., 2011; Mitschke *et al*., 2011). Both DIF clusters defined here include, in addition to protein-coding genes, several antisense and putative sRNAs, suggesting that cell-specific post-transcriptional regulation is relevant to properly adjust expression levels of certain genes specifically in developing heterocysts. Both trans-acting sRNAs and antisense RNAs have been described as regulators of cyanobacterial metabolism, including acclimation to changing iron availability (Georg *et al*., 2017), carbon supply (Eisenhut *et al*., 2012) or light intensity (Georg *et al*., 2014). Nitrogen-regulated sRNAs are involved in feed-forward regulatory loops providing additional levels of response to nitrogen deficiency (Klähn *et al*., 2015; Álvarez-Escribano *et al*., 2018). Similarly, heterocyst-specific sRNAs (NsiR1, NsiR2, NsiR8 or NsiR9) and antisense transcripts identified in the E-DIF and L-DIF groups of transcriptional responses could modulate the stability and/or translation of their corresponding target transcripts specifically in developing heterocysts, thus contributing to the metabolic rearrangements that take place in these specialized cells (Olmedo-Verd *et al*., 2019).

Classical approaches involving transposon mutagenesis had identified major players of the process of heterocyst differentiation. Some other genes, such as nitrogenase genes, can be identified on the basis of their similarity to genes previously described in other nitrogen-fixing organisms. However, genes that remain annotated as unknown have not been identified in the context of mutagenesis experiments or by homology to genes with known function. And yet, the transcriptional profiles of the new elements identified in this work strongly suggest they participate in some, perhaps subtle, aspects of heterocyst differentiation or function of the nitrogen-fixing machinery. Knowledge of these additional factors might be crucial to optimize experimental approaches to engineer biological nitrogen fixation using nitrogenases of cyanobacterial origin (Liu *et al*., 2018) or to use cyanobacteria as hosts for the production of O_2-sensitive molecules (Avilan *et al*., 2018).

Experimental procedures

Strains and growth conditions

Cultures of *Nostoc* sp. PCC 7120 WT and the *hetR* mutant 216 bearing point mutation S179N (Buikema and Haselkorn, 1991a) were bubbled with an air/CO_2 mixture (1% v/v) and grown photoautotrophically at 30°C in BG11C medium (Rippka *et al*., 1979) lacking $NaNO_3$ but containing 6 mM NH_4Cl, 10 mM $NaHCO_3$ and 12 mM N-[tris(hydroxymethyl)methyl]-2-aminoethanesulfonic acid-NaOH buffer (pH 7.5). Additionally, cells were grown in the same medium but in the absence of any combined nitrogen source ($BG11_0C$). To induce nitrogen deficiency, cells were filtered, washed with and resuspended in nitrogen-free medium ($BG11_0C$). RNA samples were isolated from cells taken at T = 0 h (WT-0 and *hetR*-0) and T = 6, 8, 12, 24 h after removing combined nitrogen from the media. It should be noted that precise timing of heterocyst differentiation depends on laboratory conditions and the previous growth state of cells. All nitrogen step-down experiments shown here were performed with exponentially growing cells. Additionally, for the experiment involving addition of ammonium to N_2-fixing cells, 10 mM NH_4Cl, and 20 mM N-[tris(hydroxymethyl)methyl]-

2-aminoethanesulfonic acid-NaOH buffer (pH 7.5) were added to cells previously grown in the absence of combined nitrogen. Construction of strains bearing plasmids is described in Supporting Information methods. Plasmids and oligonucleotides used are described in Supporting Information Tables S3, S4 and S5.

RNA extraction, microarrays, Northern blots and primer extension assays

Total RNA was isolated using hot phenol as described (Mohamed and Jansson, 1989) with modifications (see Supporting Information). Probes in the microarrays were designed in antisense orientation allowing the direct hybridization to labelled RNA, without conversion to cDNA. The microarray contains multiple probes for every protein-coding region, and for every previously detected transcript with more than 50 reads in the data set of Mitschke *et al*., (2011). In addition, all intergenic regions without known transcription were covered by probes on both strands. Details of the labelling and hybridization protocol can be found in Voβ and Hess (2014) and in Supporting Information. Microarrays were hybridized with samples from two experiments. The first one contains two biological replicas of a time course (0, 6, 12 and 24 h) after the removal of combined nitrogen using ammonium-grown cells of the wild type and the *hetR* mutant. The second experiment is a single-time point experiment after the removal of combined nitrogen from cells growing in the presence of ammonium (0, 8 h) or after the addition of ammonium to cells growing at the expense of atmospheric nitrogen, N_2 (0, 8 h). The microarray data can be accessed in the GEO database under the accession GSE120377. Northern blot hybridization and primer extension analysis of 5′ ends were performed as previously described (Muro-Pastor *et al*., 1999; Steglich *et al*., 2008).

Computational methods

Raw data were processed with the *limma* (Ritchie *et al*., 2015) R package to extract the differentially expressed probes. We performed the following comparisons of expression between two samples: every condition versus the absolute reference (WT 0 h from the -N experiment = ammonium-grown cells) and every condition versus its own reference (for instance, *hetR*-12 vs. *hetR*-0). A total of 8919 genetic elements (genes, putative sRNAs, putative asRNAs, etc.) that exhibited a $\log_2$FC (fold change) >1.25 in at least one of the comparisons were selected for further analysis (Supporting Information Dataset S1). A co-expression network was constructed based on the microarray data. We used a cutoff of 0.92 of the absolute value of Pearson correlation between the expression profile of one feature and the expression profiles of any other feature in the microarray. If two features have a correlation >0.92 (absolute value), they are connected by an edge in the network. Using these parameters, the network contains 7134 features. Finally, we performed a clustering analysis of the 7134 features in the co-expression network using the *mfuzz* R package (Kumar and Futschik, 2007) and obtained 10 clusters using a *'fuzzifier'* (m parameter) of m = 1.25. We selected a strict confidence threshold of 0.85 to consider a gene placed within the core of any given cluster. Cytoscape was used for the visualization and edition of the network (Supporting Information data file 5). We ran DREME (Bailey, 2011) with default parameters, except e-value (e = 10) and maximum number of motifs to find (m = 12) to search for conserved sequence motifs in each of the two clusters of HetR-dependent responses (#4 and #6), while the other two clusters bearing HetR-independent responses (#9 and #10) served as a negative control data set. In each prediction, we used the pool of sequences corresponding to one cluster as input, while sequences of the remaining 3 clusters served as a negative control data set. The resulting motifs were then searched in a window from −300 to +50 of the respective TSS in the E-DIF (#6) or L-DIF (#4) clusters using MAST (Bailey and Gribskov, 1998) with the parameter settings ev = 1000, –remcorr and –mt 2e-3. To map the identified motifs relative to the set of TSS, TSS positions were taken from reference (Mitschke *et al*., 2011) (E-DIF cluster), reference (Flaherty *et al*., 2011) (L-DIF cluster), as well as from the expression of probes in the microarrays. In addition, some TSS was experimentally determined by primer extension. Only the TSS of the first gene was considered for transcriptional units involving several genes. Additional details on computational methods can be found in Supporting Information methods.

Fluorescence microscopy

Images of filaments subjected to nitrogen deficiency on plates were taken four days after plating in media without combined nitrogen. The accumulation of GFP was analysed and quantified using a Leica TCS SP2 confocal laser scanning microscope as described (Muro-Pastor, 2014). GFP was excited at 488 nm by an argon ion laser, and the fluorescent emission was monitored by collection across windows of 500–538 nm (GFP imaging) and 630–700 nm (cyanobacterial autofluorescence). Images were treated with ImageJ 1.45 s software (Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, MD, http://imagej.nih.gov/ij/, 1997–2018).

Acknowledgements

We thank James Golden for making raw RNA-Seq data available to us prior to its publication. This work was

supported by grants BFU2013-48282-C2-1 from Ministerio de Economía y Competitividad, and BFU2016-74943-C2-1-P from Agencia Estatal de Investigación (AEI), Ministerio de Economía, Industria y Competitividad, both cofinanced by the European Regional Development Fund, to AMMP and by the German Federal Ministry for Education and Research (BMBF) program de.NBI-Partner (Grant 031 L0106B) to WRH. MBA is the recipient of a predoctoral contract from Ministerio de Educación, Cultura y Deporte, Spain (FPU014/05123 and EST16-00088).

Conflict of Interest

The authors declare no conflict of interest.

References

Agervald, A., Stensjö, K., Holmqvist, M., and Lindblad, P. (2008) Transcription of the extended hyp-operon in *Nostoc* sp. strain PCC 7120. *BMC Microbiol* **8**: 69.

Aldea, M.R., Mella-Herrera, R.A., and Golden, J.W. (2007) Sigma factor genes *sigC*, *sigE*, and *sigG* are upregulated in heterocysts of the cyanobacterium *Anabaena* sp. strain PCC 7120. *J Bacteriol* **189**: 8392–8396.

Álvarez-Escribano, I., Vioque, A., and Muro-Pastor, A.M. (2018) NsrR1, a nitrogen stress-repressed sRNA, contributes to the regulation of *nblA* in *Nostoc* sp. PCC 7120. *Front Microbiol* **9**: 2267.

Angelini, S., Gerez, C., Ollagnier-de Choudens, S., Sanakis, Y., Fontecave, M., Barras, F., and Py, B. (2008) NfuA, a new factor required for maturing Fe/S proteins in *Escherichia coli* under oxidative stress and iron starvation conditions. *J Biol Chem* **283**: 14084–14091.

Avilan, L., Roumezi, B., Risoul, V., Bernard, C.S., Kpebe, A., Belhadjhassine, M., *et al*. (2018) Phototrophic hydrogen production from a clostridial [FeFe] hydrogenase expressed in the heterocysts of the cyanobacterium *Nostoc* PCC 7120. *Appl Microbiol Biotechnol* **102**: 5775–5783.

Bailey, T.L. (2011) DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics* **27**: 1653–1659.

Bailey, T.L., and Gribskov, M. (1998) Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* **14**: 48–54.

Bell, N., Lee, J.J., and Summers, M.L. (2017) Characterization and in vivo regulon determination of an ECF sigma factor and its cognate anti-sigma factor in *Nostoc punctiforme*. *Mol Microbiol* **104**: 179–194.

Black, T.A., Cai, Y., and Wolk, C.P. (1993) Spatial expression and autoregulation of *hetR*, a gene involved in the control of heterocyst development in *Anabaena*. *Mol Microbiol* **9**: 77–84.

Brenes-Álvarez, M., Olmedo-Verd, E., Vioque, A., and Muro-Pastor, A.M. (2016) Identification of conserved and potentially regulatory small RNAs in heterocystous cyanobacteria. *Front Microbiol* **7**: 48.

Buikema, W.J., and Haselkorn, R. (1991a) Characterization of a gene controlling heterocyst differentiation in the cyanobacterium *Anabaena* 7120. *Genes Dev* **5**: 321–330.

Buikema, W.J., and Haselkorn, R. (1991b) Isolation and complementation of nitrogen fixation mutants of the cyanobacterium *Anabaena* sp. strain PCC 7120. *J Bacteriol* **173**: 1879–1885.

Callahan, S.M., and Buikema, W.J. (2001) The role of HetN in maintenance of the heterocyst pattern in *Anabaena* sp. PCC 7120. *Mol Microbiol* **40**: 941–950.

Corrales-Guerrero, L., Flores, E., and Herrero, A. (2014) Relationships between the ABC-exporter HetC and peptides that regulate the spatiotemporal pattern of heterocyst distribution in *Anabaena*. *PLoS One* **9**: e104571.

Di Patti, F., Lavacchi, L., Arbel-Goren, R., Schein-Lubomirsky, L., Fanelli, D., and Stavans, J. (2018) Robust stochastic Turing patterns in the development of a one-dimensional cyanobacterial organism. *PLoS Biol* **16**: e2004877.

Ehira, S., and Miyazaki, S. (2015) Regulation of genes involved in heterocyst differentiation in the cyanobacterium *Anabaena* sp. strain PCC 7120 by a group 2 sigma factor SigC. *Life (Basel)* **5**: 587–603.

Eisenhut, M., Georg, J., Klähn, S., Sakurai, I., Mustila, H., Zhang, P., *et al*. (2012) The antisense RNA As1_flv4 in the cyanobacterium *Synechocystis* sp. PCC 6803 prevents premature expression of the *flv4-2* operon upon shift in inorganic carbon supply. *J Biol Chem* **287**: 33153–33162.

Elhai, J., and Khudyakov, I. (2018) Ancient association of cyanobacterial multicellularity with the regulator HetR and an RGSGR pentapeptide-containing protein (PatX). *Mol Microbiol* **110**: 931–954.

Ermakova, M., Battchikova, N., Allahverdiyeva, Y., and Aro, E.M. (2013) Novel heterocyst-specific flavodiiron proteins in *Anabaena* sp. PCC 7120. *FEBS Lett* **587**: 82–87.

Ermakova, M., Battchikova, N., Richaud, P., Leino, H., Kosourov, S., Isojarvi, J., *et al*. (2014) Heterocyst-specific flavodiiron protein Flv3B enables oxic diazotrophic growth of the filamentous cyanobacterium *Anabaena* sp. PCC 7120. *Proc Natl Acad Sci U S A* **111**: 11205–11210.

Fan, Q., Huang, G., Lechno-Yossef, S., Wolk, C.P., Kaneko, T., and Tabata, S. (2005) Clustered genes required for synthesis and deposition of envelope glycolipids in *Anabaena* sp. strain PCC 7120. *Mol Microbiol* **58**: 227–243.

Flaherty, B.L., Van Nieuwerburgh, F., Head, S.R., and Golden, J.W. (2011) Directional RNA deep sequencing sheds new light on the transcriptional response of *Anabaena* sp. strain PCC 7120 to combined-nitrogen deprivation. *BMC Genomics* **12**: 332.

Flores, E., and Herrero, A. (2010) Compartmentalized function through cell differentiation in filamentous cyanobacteria. *Nat Rev Microbiol* **8**: 39–50.

Flores, E., Picossi, S., Valladares, A., and Herrero, A. (2018) Transcriptional regulation of development in heterocyst-forming cyanobacteria. *Biochim Biophys Acta Gene Regul Mech* (in press). DOI: 10.1016/j.bbagrm.2018.1004.1006.

Frías, J.E., Flores, E., and Herrero, A. (1994) Requirement of the regulatory protein NtcA for the expression of nitrogen assimilation and heterocyst development genes in the cyanobacterium *Anabaena* sp. PCC 7120. *Mol Microbiol* **14**: 823–832.

Galmozzi, C.V., Saelices, L., Florencio, F.J., and Muro-Pastor, M.I. (2010) Posttranscriptional regulation of glutamine synthetase in the filamentous cyanobacterium *Anabaena* sp. PCC 7120: differential expression between vegetative cells and heterocysts. *J Bacteriol* **192**: 4701–4711.

Georg, J., Dienst, D., Schürgers, N., Wallner, T., Kopp, D., Stazic, D., *et al*. (2014) The small regulatory RNA SyR1/PsrR1 controls photosynthetic functions in cyanobacteria. *Plant Cell* **26**: 3661–3679.

Georg, J., Kostova, G., Vuorijoki, L., Schon, V., Kadowaki, T., Huokko, T., *et al*. (2017) Acclimation of oxygenic photosynthesis to iron starvation is controlled by the sRNA IsaR1. *Curr Biol* **27**: 1425–1436 e7.

Huang, G., Fan, Q., Lechno-Yossef, S., Wojciuch, E., Wolk, C.P., Kaneko, T., and Tabata, S. (2005) Clustered genes required for the synthesis of heterocyst envelope polysaccharide in *Anabaena* sp. strain PCC 7120. *J Bacteriol* **187**: 1114–1123.

Ionescu, D., Voβ, B., Oren, A., Hess, W.R., and Muro-Pastor, A.M. (2010) Heterocyst-specific transcription of NsiR1, a non-coding RNA encoded in a tandem array of direct repeats in cyanobacteria. *J Mol Biol* **398**: 177–188.

Khudyakov, I.Y., and Golden, J.W. (2001) Identification and inactivation of three group 2 sigma factor genes in *Anabaena* sp. strain PCC 7120. *J Bacteriol* **183**: 6667–6675.

Klähn, S., Schaal, C., Georg, J., Baumgartner, D., Knippen, G., Hagemann, M., *et al*. (2015) The sRNA NsiR4 is involved in nitrogen assimilation control in cyanobacteria by targeting glutamine synthetase inactivating factor IF7. *Proc Natl Acad Sci U S A* **112**: E6243–E6252.

Kumar, L., and Futschik, M. (2007) Mfuzz: a software package for soft clustering of microarray data. *Bioinformation* **2**: 5–7.

Li, X., Sandh, G., Nenninger, A., Muro-Pastor, A.M., and Stensjö, K. (2015) Differential transcriptional regulation of orthologous *dps* genes from two closely related heterocyst-forming cyanobacteria. *FEMS Microbiol Lett* **362**: fnv017.

Liu, D., Liberton, M., Yu, J., Pakrasi, H.B., and Bhattacharyya-Pakrasi, M. (2018) Engineering nitrogen fixation activity in an oxygenic phototroph. *MBio* **9**: e01029–18.

Mella-Herrera, R.A., Neunuebel, M.R., Kumar, K., Saha, S. K., and Golden, J.W. (2011) The *sigE* gene is required for normal expression of heterocyst-specific genes in *Anabaena* sp. strain PCC 7120. *J Bacteriol* **193**: 1823–1832.

Mitschke, J., Vioque, A., Haas, F., Hess, W.R., and Muro-Pastor, A.M. (2011) Dynamics of transcriptional start site selection during nitrogen stress-induced cell differentiation in *Anabaena* sp. PCC7120. *Proc Natl Acad Sci U S A* **108**: 20130–20135.

Mohamed, A., and Jansson, C. (1989) Influence of light on accumulation of photosynthesis-specific transcripts in the cyanobacterium *Synechocystis* 6803. *Plant Mol Biol* **13**: 693–700.

Mulligan, M.E., and Haselkorn, R. (1989) Nitrogen fixation (*nif*) genes of the cyanobacterium *Anabaena* species strain PCC 7120. The *nifB-fdxN-nifS-nifU* operon. *J Biol Chem* **264**: 19200–19207.

Muro-Pastor, A.M. (2014) The heterocyst-specific NsiR1 small RNA is an early marker of cell differentiation in cyanobacterial filaments. *MBio* **5**: e01079–14.

Muro-Pastor, A.M., and Hess, W.R. (2012) Heterocyst differentiation: from single mutants to global approaches. *Trends Microbiol* **20**: 548–557.

Muro-Pastor, A.M., Brenes-Álvarez, M., and Vioque, A. (2017) A combinatorial strategy of alternative promoter use during differentiation of a heterocystous cyanobacterium. *Environ Microbiol Rep* **9**: 449–458.

Muro-Pastor, A.M., Valladares, A., Flores, E., and Herrero, A. (1999) The *hetC* gene is a direct target of the NtcA transcriptional regulator in cyanobacterial heterocyst development. *J Bacteriol* **181**: 6664–6669.

Muro-Pastor, A.M., Valladares, A., Flores, E., and Herrero, A. (2002) Mutual dependence of the expression of the cell differentiation regulatory protein HetR and the global nitrogen regulator NtcA during heterocyst development. *Mol Microbiol* **44**: 1377–1385.

Olmedo-Verd, E., Brenes-Álvarez, M., Vioque, A., and Muro-Pastor, A.M. (2019) A heterocyst-specific antisense RNA contributes to metabolic reprogramming in *Nostoc* sp. PCC 7120. *Plant Cell Physiol* (in press). DOI: 10.1093/pcp/pcz087

Olmedo-Verd, E., Muro-Pastor, A.M., Flores, E., and Herrero, A. (2006) Localized induction of the *ntcA* regulatory gene in developing heterocysts of *Anabaena* sp. strain PCC 7120. *J Bacteriol* **188**: 6694–6699.

Rajagopalan, R., and Callahan, S.M. (2010) Temporal and spatial regulation of the four transcription start sites of *hetR* from *Anabaena* sp. strain PCC 7120. *J Bacteriol* **192**: 1088–1096.

Ramírez, M.E., Hebbar, P.B., Zhou, R., Wolk, C.P., and Curtis, S.E. (2005) *Anabaena* sp. strain PCC 7120 gene *devH* is required for synthesis of the heterocyst glycolipid layer. *J Bacteriol* **187**: 2326–2331.

Rippka, R., Deruelles, J., Waterbury, J.B., Herdman, M., and Stanier, R.Y. (1979) Generic assignments, strain stories and properties of pure cultures of cyanobacteria. *J Gen Microbiol* **111**: 1–61.

Risser, D.D., and Callahan, S.M. (2007) Mutagenesis of *hetR* reveals amino acids necessary for HetR function in the heterocystous cyanobacterium *Anabaena* sp. strain PCC 7120. *J Bacteriol* **189**: 2460–2467.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015) *limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**: e47.

Rivers, O.S., Beurmann, S., Dow, A., Cozy, L.M., and Videau, P. (2018) Phenotypic assessment suggests multiple start codons for HetN, an inhibitor of heterocyst differentiation, in *Anabaena* sp. strain PCC 7120. *J Bacteriol* **200**: e00220–18.

Sandh, G., Ramstrom, M., and Stensjö, K. (2014) Analysis of the early heterocyst Cys-proteome in the multicellular cyanobacterium *Nostoc punctiforme* reveals novel insights into the division of labor within diazotrophic filaments. *BMC Genomics* **15**: 1064.

Shih, P.M., Wu, D., Latifi, A., Axen, S.D., Fewer, D.P., Talla, E., *et al*. (2013) Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc Natl Acad Sci U S A* **110**: 1053–1058.

Simm, S., Keller, M., Selymesi, M., and Schleiff, E. (2015) The composition of the global and feature specific cyanobacterial core-genomes. *Front Microbiol* **6**: 219.

Steglich, C., Futschik, M.E., Lindell, D., Voβ, B., Chisholm, S.W., and Hess, W.R. (2008) The challenge of regulation in a minimal photoautotroph: non-coding RNAs in *Prochlorococcus*. *PLoS Genet* **4**: e1000173.

Tsujimoto, R., Kamiya, N., and Fujita, Y. (2016) Identification of a cis-acting element in nitrogen fixation genes recognized by CnfR in the nonheterocystous nitrogen-fixing cyanobacterium *Leptolyngbya boryana*. *Mol Microbiol* **101**: 411–424.

Valladares, A., Maldener, I., Muro-Pastor, A.M., Flores, E., and Herrero, A. (2007) Heterocyst development and diazotrophic metabolism in terminal respiratory oxidase mutants of the cyanobacterium *Anabaena* sp. strain PCC 7120. *J Bacteriol* **189**: 4425–4430.

Vernon, S.A., Pratte, B.S., and Thiel, T. (2017) Role of the *nifB1* and *nifB2* promoters in cell-type-specific expression of two Mo nitrogenases in the cyanobacterium *Anabaena variabilis* ATCC 29413. *J Bacteriol* **199**: e00674–16.

Voβ, B., and Hess, W.R. (2014) The identification of bacterial non-coding RNAs through complementary approaches. In *Handbook of RNA Biochemistry*, Hartmann, R.K., Bindereif, A., Schön, A., and Westhof, E. (eds). Weinheim, Germany: WILEY-VCH, pp. 787–800.

Voβ, B., Bolhuis, H., Fewer, D.P., Kopf, M., Möke, F., Haas, F., *et al*. (2013) Insights into the physiology and ecology of the brackish-water-adapted cyanobacterium *Nodularia spumigena* CCY9414 based on a genome-transcriptome analysis. *PLoS One* **8**: e60224.

Wegelius, A., Li, X., Turco, F., and Stensjö, K. (2018) Design and characterization of a synthetic minimal promoter for heterocyst-specific expression in filamentous cyanobacteria. *PLoS One* **13**: e0203898.

Wolk, C.P., Cai, Y., Cardemil, L., Flores, E., Hohn, B., Murry, M., *et al*. (1988) Isolation and complementation of mutants of *Anabaena* sp. strain PCC 7120 unable to grow aerobically on dinitrogen. *J Bacteriol* **170**: 1239–1244.

Yoon, H.-S., and Golden, J.W. (1998) Heterocyst pattern formation controlled by a diffusible peptide. *Science* **282**: 935–938.

Yoshimura, H., Okamoto, S., Tsumuraya, Y., and Ohmori, M. (2007) Group 3 sigma factor gene, *sigJ*, a key regulator of desiccation tolerance, regulates the synthesis of extracellular polysaccharide in cyanobacterium *Anabaena* sp. strain PCC 7120. *DNA Res* **14**: 13–24.

Zhao, W., Ye, Z., and Zhao, J. (2007) RbrA, a cyanobacterial rubrerythrin, functions as a FNR-dependent peroxidase in heterocysts in protection of nitrogenase from damage by hydrogen peroxide in *Anabaena* sp. PCC 7120. *Mol Microbiol* **66**: 1219–1230.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Appendix S1: Supporting pdf file, including supplementary methods, supplementary Figs S1 to S9, supplementary Tables S1 to S5 and description of supplementary data files 1 to 5 (to be downloaded from http://hdl.handle.net/10261/179430; DOI:10.20350/digitalCSIC/8631).

Dataset S1: 8919 elements with a $\log_2$FC larger than 1.25 (in any of the comparisons made).

Dataset S2: 2463 elements that enter the 10 clusters (confidence >0.85) obtained after *mfuzz* analysis.