FIGURE LEGENDS

Fig. 1. Model and proposed realization of a reaction-diffusion system. (**A**) Generic Turing/Meinhardt model. A molecule **R** regulates some action of interest. It also exerts positive feedback on its own synthesis or activity and increases the synthesis or activity of a second molecule, **S**, a suppressor of **R**. **S** is able to diffuse, while **R** remains at the site of synthesis. (**B**) Model of regulation of heterocyst differentiation in *Anabaena* PCC 7120. HetR plays the role of the R morphogen, here activating heterocyst differentiation. PatS and HetN collectively play the role of the S, suppressing HetR activity at different times during development. NtcA monitors nitrogen status. (**C**) Outcome of the proposed model -- filaments of *Anabaena* PCC 7120 grown in the absence of fixed nitrogen. Green cells are vegetative cells, capable of photosynthesis. Cells stained with Alcian Blue are heterocysts, incapable of photosynthesis but sites of nitrogen fixation. Photo courtesy of AV Matveyev.

Fig. 2. Overview of phylogenetic tree, rooted by *Gloeobacter violaceus* PCC 7421. See *Experimental procedures* section for details on construction and Fig. 3 for an expanded version of the tree. Strains highlighted in red are unicellular, blue heterocyst-forming, green other filamentous strains, and pink picocyanobacteria. The tree can be interpreted in multiple ways with regards to switches between unicellularity and filamentarity. What is shown is not the interpretation with the fewest switches but one that proceeds from the hypothesis that filamentarity arose only once. Group numbers correspond to those used by Howard-Azzeh *et al.* (2014). Arrows indicate proposed acquisition of proteins, and stars indicate proposed loss of proteins.

Fig. 3. Phylogenetic tree of cyanobacterial genomes and presence of key regulatory proteins. See Fig. 2 for phylogenetic context and the *Experimental procedures* section for a description of how the tree came about. Colored circles on the right indicate the presence in the genome of A=NtcA, R=HetR, N=HetN, S=PatS, X=PatX. Small circles indicate protein assignments for which there is doubt owing to synteny concerns, and triangles indicate proteins with differences with respect to RGSGR of HetN or PatX or two of the conserved residues of HetR (see text). Asterisks indicate genomes that have a surprising presence of HetR or a surprising absence of PatX. Yellow circles and orange circles in the tree indicates bootstrap support of at least 90% and 67%, respectively. Group numbers correspond to those used by Howard-Azzeh *et al.* (2014). (**A**) Heterocyst-forming cyanobacteria. One strain (*Raphiodopsis brookii* D9) does not make functional heterocysts, but shows evidence by electron microscopy of differentiation of terminal cells (Stucken *et al.*, 2010). (**B**) Non-heterocyst-forming filamentous cyanobacteria and unicellular cyanobacteria.

Fig. 4. Alignment of HetR proteins. 91 HetR protein sequences were aligned, including 40 filamentous, heterocyst-forming strains (highlighted blue), 43 filamentous, non-heterocyst-forming strains (highlighted green, except for 9 secondary copies as defined in the text, which are highlighted gray), and 7 unicellular strains (highlighted red). The order of the organisms is the same as in the phylogenetic tree of HetR (Supporting Fig. S2), and their full names are given in Fig. 3. Columns in which no more than three mutational events are evident from the primary filamentous HetR sequences are colored in green when the amino acid agrees with the consensus, cyan when it is a conservative substitution as defined by a BLOSUM80 (Henikoff

and Henikoff, 1992; Chao and Zhang, 2009) score of 2 or greater, gray when the substitution is not conservative but nonetheless represents a substitution of one of the six most hydrophobic residues (Monera *et al.*, **1995**) with another, or otherwise pink. The three lines at the top of the alignment (and repeated at the bottom) indicate residues for which there is evidence concerning functional importance. The top line (H) indicates whether at least one mutant residue affects heterocyst differentiation (red if the frequency of heterocysts markedly decreases, blue if it markedly increases, green if it does not change). The second line (A) indicates whether at least one mutant residue affects an in vitro assay for DNA binding (red) or PatS binding (blue). The residue is green if the assay of the mutant HetR gives a similar result as wild-type HetR. The bottom line (S) indicates whether an analysis of the structure of a crystalized HetR protein indicates binding of the residue to DNA (red), to PatS (blue), or to another HetR subunit (gray) The full alignment is given in Supporting Fig. S3, along with evidence for the functional assertions.

Fig. 5. Amino acid sequences of candidate PatS proteins**.** Amino acid sequences suspected of encoding a functional PatS protein were identified as described in the text, using sequence and contextual cues. The RGSGR motif is highlighted in dark green and a conserved preceding glutamate residue in light green. Gray and blue highlighted letters indicate hydrophobic and very hydrophobic amino acid side chains at pH7, respectively (Monera *et al.*, **1995**). Red letters indicate alternative start sites, with possible N-terminal extensions indicated in gray font. Despite the seemingly straightforward experiment of Corrales-Guerrero et al (2013), there is considerable doubt regarding the most active start codon for *patS* of *Anabaena* PCC 7120 (Ana7120), never mind the other instances of *patS*. See Supporting Table S7, which provides evidence that translation starts primarily at the valine codon (producing an 11-amino acid PatS) and to a lesser extent at the second methionine codon (producing a 13-amino acid PatS). The lower case italicized sequence of *Fischerella thermalis* PCC 7521 (Fis7521) is the virtual translation that removes a one-nucleotide insertion relative to the sequences of other *Fischerella*. Arrows indicate the position and direction of flanking genes (not to scale): dihydroorotase (blue), patatin (green), and HetY (red). Sequences lacking contextual support are marked with asterisks. *Mastigocoleus testarum* BC008 (Mas008) has two identical sequences. Organismal abbreviations are explained in Fig. 3.

Fig. 6. Amino acid sequences of candidate PatX proteins. Amino acid sequences suspected of encoding a functional PatX protein were identified as described in the text, using sequence and contextual cues. Amino acids are colored as described in Fig. 5. In addition, proline residues are highlighted in yellow. Arrows indicate the position and direction of flanking genes (not to scale): *hetR* (blue), *sepJ* (cyan), FAD-dependent oxidoreductase (DH; green), conserved hypothetical (Hyp; light pink), methyltransferase (MTase; dark pink), and *glnA* (red). Sequences lacking contextual support are marked with asterisks. Nicknames of the organisms are followed by a symbol indicating the presence of a N-terminal signal peptide sequence as predicted by SignalP (see *Experimental procedures*): # (present, strict conditions), + (present, but only if possible transmembrane regions are ignored), o (absent), and ~ (within 10% of threshold). Organismal abbreviations are explained in Fig. 3. (**A**) Heterocyst-forming cyanobacteria. (**B**) Non-heterocyst-forming filamentous cyanobacteria and unicellular cyanobacteria.

Fig. 7. Conserved amino acid residues in PatS and PatX. Sequence logos representing the amount of information (associated with the degree of conservation) are shown for sequences of (**A**) PatS, (**B**) PatX (heterocyst-forming cyanobacteria), and (**C**) PatX (other cyanobacteria). Residues are colored blue (positively charged), red (negatively charged, green (polar), and black (non-polar). Variable spacing between clusters of aligned amino acids are shown. See Figs. 6 and 7 for amino acid sequences of each protein.

Fig. 8. Conserved nucleotides upstream from PatS, PatX, and DIF1-motif-containing regions. Sequence logos representing the amount of information (associated with the degree of conservation) are shown for sequences upstream from (**A**) *patS*, (**B**) *patX* (heterocyst-forming cyanobacteria), (**C**) *patX* (other cyanobacteria), and (**D**) HetR- and N-regulated genes of *Anabaena* PCC 7120. Variable spacing between clusters of aligned nucleotides are shown as are DIF1 and NtcA-binding motifs. The approximate position of transcriptional initiation for the appropriate gene from *Anabaena* PCC 7120 (Mitschke *et al.*, 2011) is shown by an arrow. See Supporting Figs. S5 – S7 for nucleotide sequences of each upstream sequence.

Fig. 9. Kinetics of P_{patX}-*gfp* and P_{patS}-*gfp* reporters expression in *Anabaena* PCC 7120.
Wild-type *Anabaena* PCC 7120 carrying pRIAM970 with the reporter fusion P_{patX}-*gfp* or pAM830 (Yoon and Golden, 1998) with the reporter fusion P_{patS}-*gfp* was grown on neomycin-containing BG-11 plate and then transferred to combined nitrogen-free liquid BG-11_0 medium. Micrographs were taken 6 h and 16 h after nitrogen step down. Images correspond to phycobilisome-induced red autofluorescence (left panel) and GFP fluorescence (right panel). Note the reduced autofluorescence in developing heterocysts at 16 h. Arrowheads point to cells with high GFP fluorescence, presumably (at 6 h) potential (but not committed) proheterocysts and (at 16 h) developing proheterocysts (note the reduced autofluorescence in some of the indicated cells).

Fig. 10. Ectopic overexpression of *patX* on diazotrophic growth and heterocyst differentiation in *Anabaena* PCC 7120.
Wild-type *Anabaena* PCC 7120 carrying a control plasmid pAM1956 (A, C) or P_{petE}-*patX*-containing pRIAM810 (B, D) was grown on nitrate-containing BG-11 plate with 25 µg/ml of neomycin and then transferred in combined nitrogen-free liquid BG-11_0 (A, B) or on solid BG-11_0 (C, D) medium. Micrographs were taken 3 days (A, B) and 4 days (C, D) after nitrogen step down. Arrowheads indicate heterocysts.