

Supporting Figure S8A: *patX* upstream sequences in heterocyst-forming cyanobacteria

| Organism ^a | Upstream sequence ^b | <i>patX</i> → | DIF1 |
|-----------------------|--|---------------|------|
| Nos21 | AATTTCTTCTGTCAAACAGAAATACATTTTGC--CAAACCTCCGGATACACATTCAGAAAGCTCAGTAGCAGTGAATATTAGAGTTC-ACAAGAANTATCTGCC-ATGCGT | | 5.0 |
| Nos7107 | AATTTCTTCTGTA AAAACAGAAATACATTTTGC--CAAACCTCCGGATACACATTCAGAAAGCTCAGTAGCAGTGGATATGAGAGTTC-ACAATAAATATCTGCT-ATGCGT | | 5.5 |
| Nos7524 | ATTTTTTTTTCGTA AAAACACAACTACATTTTTC--ATAATCTCCGGACTCAGCCCATAGAAATCAGTAGCAATAGATAAGAGAGTTCACAAAAGAGATTCCTATG-ATGCGT | | 5.4 |
| Nos3756 | TTAATTTTTCGTA AAAAGCACACTACATTTTTC-AGAATCTCCGGACTCGTCACATAAAAACCTCAGTAGCAATAAATAAGAGA-TTACAAAAGAGCTTCTAAA-ATGCGT | | 5.8 |
| Ana29413 | ATAGGTTTTTCGTA AAAAGCACACTACATTTTTC-AGATTTCCGGACTCGTTATATAAAAACCTCAGTAGCAATAGATAAGAGAAATTTACAAAAGAGACTTTTATG-ATGCGT | | 5.0 |
| Ana7120 | ATGGATTTTTCGTA AAAAGCACACTACATTTTTC-AGATTTCCGGACTCGTTATATAAAAACCTCAGTAGCAATAGATAAGAGAAATTTACAAAAGAGACTTTTATG-ATGCGT | | 5.0 |
| Ana310F | ACTTTTATTCAGTA AAAATACAAATACATTTTTC--GCAAAATTCCCGGATAGACATATAAAATCAAGGATTAATGAAGTAGAGAGTTCATAAATAGACATCTACT-ATGCGT | | 6.0 |
| Ana102 | ACATTTATTCAGTA AAAATACAAATACATTTTTC--GCAAAATTCCCGGATAGACATATAAAGGTTAAGTATTAGTGAAGTAGAGAGTTCATAAATAGACATCTACT-ATGCGT | | 6.0 |
| Ana90 | ACATTTATTCAGTA AACTACAAATACATTTTTC--CCGAATTCCCGGACAGACATATAAAAAGTTAAGTATTAGTGAAGTAGAGAGTTCATAAATAGACATCTACT-ATGCGT | | 5.6 |
| Ana7108 | ATTTTTTTTTCAGTCAAAATACAAATACATTTTTC--ACAAATTCCCGGATGATCATACAAAAAGTTAAGTATCAGTGAATTAAGAGAGTTCATAAATAGACATCTACT-ATGCGT | | 6.0 |
| Ana7122 | ATTTTTTTTTCAGTCAAAATACAAATACATTTTTC--ATAAATTCCCGGATGATCATACAAAAAGTTAAGTATCAGTGAAGTAGAGAGTTCATAAATAGACATCTACT-ATGCGT | | 4.5 |
| AnaAzo | ATTTTTTTTTCAGTCAAAATACAAATACATTTTTC--ATAAATTCCCGGATCATTATCTAAAAGCTAAGTATCAGTGAAGTAGAGAGTTCATAAATAGACATCTACT-ATGCGT | | 5.2 |
| CylR505 | TGGTTC AATTAGATTAAGTCAATCGATTTCTTCAAATTCCCGGATTCATGGGTGATAGGTAAGTATAAGTTAACAGAGAGTTCATAAATAGACATCTACT-ATGCGT | | 5.5 |
| RapD9 | CAGTTAAATAAAGATTAAGTCAATAGATTTCTTCAAGATTCATCCGGATTCATAGGTGATAGACAAGTAAAAGCTAATCAGATAGTTCTATAAATAGACATCTACT-ATGCGT | | 5.8 |
| CylS7417 | TTTTTTTTCAGTGCAGCACAAATACATTTTTC--ATAAATTCCCGGATGCCCATACAAAACTAAGTATCAGTGAATTTGAGAGTTCATAAATAGACATCTACT-ATGCGT | | 6.1 |
| Nos29133 | GTTTATTTTTCAGTAAAACAGAAATACATTTTTC--ATTAATTCCCGGATATTAATCAAAAAGCTAAGTATAAGTGAAGTGGAGAGTTCATAAATAGACATCTACT-ATGCGT | | 5.3 |
| Nos29133-b | CAAAAAGAAAATATTTTTGCCATTATATTTAG--ATATGTTCACGGATAATTTAAAFATACTGTGTATTAGTTAATA---GAGATAAAATAGAGCTATTAAC-ATGCCT | | - |
| Nod9414 | ATTTTTTTTTCAGTA AAAACACAACTACATTTTTC--CTCAATATTCGGACAACTCCCAAAAACCAGTAGCAGTAGACAAGAGAAATTTTAAAATAAACATCTGTT-ATGCGA | | 5.2 |
| Tol7601 | ATTTTTTATTCAGTA AAAAAAACACACTGTTTTTA--GGGATTTCCGGATGCAGTAATAAAAAGCTAAGTAA--GGTGGAAAAGAGAGTTCATAAATAAACATCTGCT-ATGCGT | | 5.3 |
| Cal7507 | CTTCATAATAAGTA AAAAAAACAGTACATTTTTC--AGAAAAATTCGGATGCACTAATAAAAAGCTCAGTAGTGTGAATTTGAGAGTTCATAAATAAACATCTGCT-ATGCGT | | 6.5 |
| Mch7126 | TTTGATATTTTGTAAAACAAATACATTTTTC--AGAAAAATTCGGATGCACTAATAAAAAGCTCAGTAGTGTGAATTTGAGAGTTCATAAATAAACATCTGCT-ATGCGT | | 5.8 |
| Has12170 | CATTTTTTTTGTCAAAAACAAATGCTTTTTGT---CAAATTCGGACGCAGGAAAAAAAAGCTAAGTATAAGTGAAGTGGAGAGTTCATAAATAAACATCTACT-ATGCGT | | 4.9 |
| Tol9009 | CATTTTTTTCAGTAAAACAGTACATTTTGT---CAAATTCGGACGCAGTAAAAAAAAGCTAAGTATCAGTGAAGTGGAGAGTTCATAAATAAACATCTACT-ATGCGT | | 4.1 |
| Cal336-3 | AGAAAATTTAAGTAAAACAAATACATTTTTC--CAAATTCGGATGCACTAATAAAAAGCTCAGTATCAGTGAATGAAGAGTTCATAAATAAACATCTAGAT-ATGCGT | | 5.0 |
| Cal6303 | AATTTCCGTGTA AAAAACAAATACATCAAAA--ATAAACCATCCGGATACACAAACTAAATCTAAGTATCAGTGAATTTGAGAGTTCAGAAAATAAACATTTGCT-ATGCGT | | 3.6 |
| Cal7103 | TTTTTGTGCAAGTA AAAAACAAATACATTTTTC--GAATCATCCGGATGCACTAATAAAAAGCTAAGTACTAGTGGATGAGAGTTCATAAATAAACATCTGCT-ATGCGT | | 4.2 |
| Fis3754 | ATATTTTTTTCGTAAAACAAATACATTTTTC--TTACTAAATTCGGAGAGCAGTAAAAGAAAGCTCAGTATCAGTGAATGAAGAGTTCATAAATAAACATCTGCT-ATGCGT | | 4.5 |
| Fis7521 | ATATTTTTTTCGTAAAACAAATACATTTTTC--TTACTAAATTCGGAGAGCAGTAAAAGAAAGCTCAGTATCAGTGAATGAAGAGTTCATAAATAAACATCTGCT-ATGCGT | | 4.6 |
| Fis9339 | AATTTTTTTTCGTAAAACAAATACATTTTTC--TGACGAATTCGGAGAGCAGTAAAACAAAGCTCAGTATCAGTGAATGAAGAGTTCATAAATAAACATCTGCT-ATGCGT | | 4.7 |
| Hap220 | TATTTTTTTTCGTAAAACAAATACATTTTTC--GACGAATTCGGAGAGCAGTAAAACAAAGCTCAGTATCAGTGAATGAAGAGTTCATAAATAAACATCTGCT-ATGCGT | | 4.4 |
| Scy7110 | GTTTTTTTTCGTAAAACAAATACATTTTTC--ATAAATTCGGATGCACAAAAAGAAAAGCTCAGTATCAGTGAATGAATAGTTCATAAATAAACATCTCATT-ATGCGT | | 5.2 |
| Tol521301 | ATTTTTTTTTCGTAAAACAAATACATTTTTC--ATAAATTCGGATGCACAAAAAGAAAAGCTCAGTATCAGTGAATGAATAGTTCATAAATAAACATCTCATT-ATGCGT | | 5.4 |
| Scy61278 | AAAATTTAACGTAAAATAAATACATTTTTC--AGGATTTCCGGATGCAGTAAATAAAAAGCTCAGTATTAGTGAATGAAGAGTTCATAAATAAACATCTTCTT-ATGCGT | | 5.9 |
| Mas10914 | AAATCTGTGCAAGTA AAAAAAACAAATACATTTTTC--ATAAATTCGGATGCACTAATAAAAAGCTCAGTATCAGTGAATGGAGAGTTCATAAATAAACATCTACT-ATGCGT | | 5.1 |
| Tol511288 | AATTTTGTGCAAGTA AAAAAAACAAATACATTTTTC--CTAAATTCGGATGCAGTAAAAGAAAAGCTCAGTATCAGTGAATGGAGAGTTCATAAATAAACATCTGCT-ATGCGT | | 5.4 |
| RicHH01 | TAGTTTTTAATCTAGCAGAAAACCTTTGG--GTAATTCGGATGCAGTGGAAAAGCTCAGTATCAGTGAATGGAGAGTTCATAAATAAACATCTGCT-ATGCAT | | 5.1 |
| Riv7116 | TTTAATTTAGAGTA AAAAAAACAAATACATTTTTC--TAAATTCGGATGCAGTAAATAAAAAGCTCAGTATCAGTGAATGGATAGTTCATAAATAAACATCTACT-ATGCGT | | 4.1 |
| Mas008 | AAAACAGAAAATTTGAAAACAAATACATTTTTC--ACAATAATCCGGATGCACTAATAAAAAGCTCAGTATCAGTGAATTTGAGAGTTCATAAATAAACATCTTATT-ATGCGT | | 5.0 |

Supporting Figure S8B: *patX* upstream sequences in non-heterocyst-forming cyanobacteria

| Organism ^a | Upstream sequence ^b | <i>patX</i> → | DIF1 |
|-----------------------|---|---------------|------|
| Lep7376 | GCGATCACAAAATTCCTTCAAGAAAAAGATTCTAGAAAATTC CCGG AGATACACCCCTCCTTTTCAGTAAAAAGAACTGTACC-GTTTAAAGTCCATTAGGGCAATCTGCT ATG ACT | | 4.6 |
| Spi9445 | ATTTTTGAGAACTAGAGATCAGGGGATGTGAAAAATTTG CCCTGA TGAATAGCAAGAACTCCGTATCAGGAAATAGAAA-----CTGACACGGATTAATTTGCT ATG CAC | | 5.2 |
| Mic7113 | CAATCCTGTGAATATCCAATAAATTGCGACGAAAGTTTT CCCTGA CGCCCTCATAAGAACTCCGTATCAGGTAATAGAAA-----TCCGATACGGACATCTGCT ATG AGC | | 4.6 |
| Cri9333 | AACTAAATAACTAAAAATAAACTCTGTTTGGAGTTTT CCGG AATCTATTAGCCCTAAATCAGAAACAGGTAATAGCAA-----CCTTAGAGCAGGTATCTGCT ATG CCT | | 5.6 |
| Osc10802 | GGTGAGAGCCGGTTGAGTCTGAGAAAAATTGACAAACTG CCCTGA TGGACTGGTTGTTGATGCGGTATCAGAAAGGAAAC-----TCTGATACGGACATCTC ATG CGG | | 5.2 |
| Osc10802-b | AAAGACCGGCATCCGGACAGTTGACAAAATCGCTCTTTG CCGG AATGGTTGCGCCTCTGTGCGGTAAAGAGAAATGGAAAG-----CTCTGACACGGATAGTTCT ATG CGC | | - |
| Osc10802-c | GTGGACACATCGAAAACCTGGCTGTGTGCCAGTTTTTTA CCAGA TGGCGCTGAGGTTTCTGCGAATCAGCGGATAGAAA-----GTTGATTCAAGGAGATTTCT ATG CGC | | - |
| Osc6304 | ACCCCTACAAAAACAATTTATTTTTCTATCTTTCTTG CCAGA TAGAGCCTAAGGAATGCGTATTAGGGAATATCAG-----TTATCCTAAGCAATTCCTTT ATG CGC | | - |
| Osc6304-b | ACCCCTACAAAAACAATTTATTTTTCTATCTTTCTTG CCAGA TAGAGCCTAAGGAATGCGTATTAGGGAATATCAG-----TTATCCTAAGCAATTCCTTT ATG CGC | | - |
| Osc6304-c | GACCCGAGAGGGTTTACCCCAAATAATTTAAAAAATAG CCGG AATGTTTTGTA-GATATTCGTAGAAAGAAATAACAG-----TCAAAAAAGGAAATATCT ATG AGC | | 4.7 |
| Gei7105 | GGGATGTCATAAAAATCTAACCGGTGCTTCAGATCG CCCTGA CGCCATATTCGGAGACCTTAAACAGGGAACAGAAC-----ACCTGAAACAGATTCTCAAAT ATG CGT | | 3.1 |
| Gei7105-b | GTATTCGTTCAAAAAATTAATTTGGTCAAACAGGTTTC CCGG AATCGCAGCGAGATAGAGAGCTAACTGGAAATAGAAA-----CTCAACGAAGGGATTCTATACC ATG CGC | | 4.6 |
| PhoOSCR | AACCCCTCAGGACAATGGTACGACTGAAGGAATTC CCCTGA TCGCCCCATCGGAAATCTGTAAACAGGAATAAGTA-----CCTGAAACAGCTTCTCCGATT ATG CGT | | 5.0 |
| PhoOSCR-b | GTTGAACAATATCCAAGTAGATCACAATAAATTTGAATTT CCCGA TAGATCCCTCAGATCTATGAGATCAGAAATGAGC-----TATCACTCAAGGGGCTACTTT ATG AGC | | 5.6 |
| MicFGP2 | AGGTTTTTTGTAGAGATCCGAAAATTCACAGATATCTA CCCTGA TAGATCAGGCGAAGTCCGTATCAGTAATAGACA-----CTGACACGGATACTACCC ATG CAA | | 4.7 |
| MicFGP2-b | ACTTACTAATTGATGATTTACCCGCAATTTATTTTTTA CCCTGA TAGCCGAGTACAGACTGCGTATCAAGAAAGGGGAAA-----ATACAAACACCAAATTTAT ATG CGC | | - |
| Osc7112 | GGGTTTTTTGACAGATCCGAAAATTCACAGATATCTA CCCTGA TAGACCCGGCGAAGTCCGTATCAGTAATAGACA-----CTGACACGGATACTACCC ATG CAA | | 4.7 |
| Osc7112-b | ACTTACTAATTGATGACTTCGCCACAATTTATTTTTTA CCCTGA TAGCAGAGTACAGACTGCGTATCAAGAAAGGGGAAA-----ATACAAACACCAAATTTAT ATG CGC | | - |
| Tri101 | CTTTCTATATATGAAGTACAAAATTTATTTCTAATA CCCTGA TTTGTCTCTAACAGTCCGTATCAGAAATGAGACA-----CTCTCAGACGGACGTAACCT ATG CAA | | 4.4 |
| Pla126-8 | GGGTTTATAGCCGGATTTAACAAAAACTTTTGACAGTTA CCCTGA TCTAAATATTAACGGTTTGTAAAGGAATGGACA-----GACTTAAAGGGACACTTTAT ATG CAA | | 4.8 |
| Pla126-8-b | CAGGACTCAAAAGCAATCTAGGAAAGTTTGTCTTTTTTA CCGG AATCAATTTCA-TTCAACTTGTATTAGAGAATGGACA-----CTCTAATACGGACAACCTCT ATG CGA | | 4.6 |
| Pla406 | GGGTTTATAGCCGGATTTAACAAAAACTTTTGACAGTTA CCCTGA TCTAAATATTAACGGTTTGTAAAGGAATGGACA-----GACTTAAAGGGACACTTTAT ATG CAA | | 4.8 |
| Pla406-b | CAGGACTCAAAAGCAATCTAGGAAAGTTTGTCTTTTTTA CCGG AATCAATTTCA-TTCAACTTGTATTAGAGAATGGACA-----CTCTAATACGGACAACCTCT ATG CGA | | 4.6 |
| Pla15 | GGGTTTATAGCCGGATTTAACAAAAACTTTTAACAGTTA CCCTGA TCTAAATATTAACGGTTTGTAAAGGAATGGACA-----GACTTAAAGGGACACTTTAT ATG CAA | | 5.0 |
| Pla15-b | CAGGACTCAAAAGCAATCTAGGAAAGTTTGTCTTTTTTA CCGG AATCAATTTCA-TTCAACTTGTATTAGAGAATGGACA-----CTCTAATACGGACAACCTCT ATG CGA | | 4.6 |
| Pla407 | GGGTTTATAGCCGGATTTAACAAAAACTTTTGACAGTTA CCCTGA TCTAAATATTAACGGTTTGTAAAGGAATGGACA-----GACTTAAAGGGACACTTTAT ATG CAA | | 4.8 |
| Pla407-b | CAGGACTCAAAAGCAATCTAGGAAAGTTTGTCTTTTTTA CCGG AATCAATTTCA-TTCAACTTGTATTAGAGAATGGACA-----CTCTAATACGGACAACCTCT ATG CGA | | 4.6 |
| Arth39 | ACCAGTACCGTAGGGGTGAACAAAAAATATCAGATTA CCCTGA TATAGGCCATTAGATTGGTATCAGGAATGGACG-----CTTAAATACGGACATTTCC ATG AAA | | 5.6 |
| Arth39-b | ACTTTACTTGTGCTTAGTTATCCAAATTTGCGATTTTTA CCGG AATCGCCAATA-TCCCGCTGGTATTAGATAGTGGAAA-----CTTTCACACGGACAACCTCCCT ATG CGT | | 4.1 |
| Arth328 | TCAAGTATTGTAGGGCTGAAGGAAAAAATATCAGATTA CCCTGA TGTAGGCCATTGGGTGGTATCAGGAATGGACG-----CTTAAATACGGACATTTCT ATG AAA | | 5.3 |
| Arth328-b | TGATTAATGATTTCCTAGTTATCCCAATTTGCGATTTTT CCGG AATAGCTGATA-TCCCGCTGGTATTAGATAGTGGAAA-----CTTTGATAAGGACATTTCCCT ATG CGT | | 3.9 |
| Arth8005 | TCAAGTATTGTAGGGCTGAAGGAAAAAATATCAGATTA CCCTGA TGTAGGCCATTGGGTGGTATCAGGAATGGACG-----CTTAAATACGGACATTTCT ATG AAA | | 5.4 |
| Arth8005-b | TGATTAATGATTTCCTAGTTATCCCAATTTGCGATTTTT CCGG AATAGCTGATA-TCCCGCTGGTATTAGATAGTGGAAA-----CTTTGATAAGGACATTTCCCT ATG CGT | | 4.2 |
| Lyn8106 | GATTACAGAATAGAATCTCTATCTAAATGTAGAAAAT CCCTGA TGTTTCAGTCAAGTCTCCGTAAAACACTTAAAGA... (97) ...TTATTACGGACATTTCC ATG AAA | | 5.6 |
| Lyn8106 | AACGCGAGATCAATTTTTGTAGAAAAATTTTAAATCTTA CCCTGA TCTCATCAACTTGGGTTCGTATCAAGAAATGGACA-----GGTTTATTACGGACATTTCC ATG AAA | | 3.7 |
| Lyn8106-b | ATCCTAAATAGTAAATGATAAACCACTTGATTTA CCCGA TCTCGAACA-AATTGTCTGTATTAGAGAATGGACA-----TACTTAGACGGACAGACT ATG CGG | | 4.3 |
| LynBLJ | GATTACAGAATAGAATCTCTATCTAAATGTAGAAAAT CCCTGA TGTTTCAGTCAAGTCTCCGTAAAACACTTAAAGA... (96) ...TAATTACGGACATTTCC ATG AAA | | 5.2 |
| LynBLJ | TAACGCGAGAGCAATTTTTGTAGAAAAATTTTCACTTA CCCTGA TCTGATCAACTTAGGTTGGTATCAAGAAATGGACA-----GGTTAATTACGGACATTTCC ATG AAA | | 3.1 |
| LynBLJ-b | AGCATTTAGGATTAATTAATAAACCACTTGATCATT CCCGA TCTCGAACA-AATTGTCTGTATTAGAGAATGGACA-----GACTCATAACGGACAGACT ATG CGT | | 4.4 |
| Gei7407 | TGTCATCGGGCAACCAAGTGTGGGGTTTTGGTTTTG CCCTGA TGACCATCGAGAAAGATCCGTAAACAGGGAATAGAAA-----CTCTGCTACGGACCTTTCC ATG AGC | | 5.0 |
| Osc12 | CTTATTTCCCATTTCCCTATTCCAGCGGTTTCCAAAA CCCTGA TAGGCAATCCAAACTCCGTATCAGGGAATGAGAC-----CTTTGATACAAACAGTT ATG GAC | | 5.2 |
| Lep6306 | AGGTGCGACTACCGAATTTGAGTTTTCAGGCATTTTTG CCCTGA TGTGATGCCAAAATTTGGAAAGGGGATCAAC-----CACTGAGGACGTAATA ATG CGC | | 4.6 |
| Lep2104 | AGGTGCGACTACCGAATTTGAGTTTTGCGCATGTTTTG CCCTGA TGTGATCACTTAGCCCTAGCCGTAAGGGGTATATCAC-----ACTGAGGACGTA ATG CGC | | 4.8 |
| Lep3755 | AGGTATGACTACCGAATTTGACTTAATCCATGATTTCG CCCTGA TCTAACCCCAAAGCGGTAAGGGGTATATCAC-----ACTGAGGACGTA ATG CGC | | 5.1 |
| Lep6406 | CAACGGCCCTAGATTCCAGCCGGACTGCCCTGAAT CCCTGA ACCCCTCCTTCTGTGCTGTAACAGGGAATAGCAA-----CACAGAGAGGGTCTTTTAC ATG TCT | | 3.9 |
| Syn7335 | ATTTACGAATAGACTAGTTCAACTCTTCAATTTAGAAC CCCTGA CTTCGCACA-AATTTGATGATGATAGTGGGTATAGT---CTTGATTAGAATCTCTGAGTAT ATG TCTA | | 3.0 |
| LepHIJ | AAATCTAAGCTCCATCCACATTTTTTTCACACGATTT CCCTGA GTTATTTAATTAATTTCTGATATCAGAAATGAGCA-----ATACTACTCTATTTCTGACCC ATG AGA | | 5.5 |
| ProH9006-a | GCTTCTGGAAATCACCCAGAACCTTTGGGATTTGGGATTT CCCTGA AGTAACTTTGTCAGCCCTATC-CACTCCCTAT-ACTCTTCTTTCTTCGGAGTCA ATG AAC (<2) | | |
| ProH9006-b | CCGCGATTCTCGAAATTTGCTAAGGTTCCATAA CCCTGA TCTTCTATTGATTTTTCTGTATGGGGGAATGAATCCATTAGACAGCTAAAGGACCACAAAC ATG GCC | | 6.5 |

-
- ^a Organism nicknames are associated with the full names in Fig. 3B. Organisms followed by asterisks present pairs of DIF1 regions (both shown).
- ^b Highlighting conventions are the same as in Supporting Fig. S6. In addition, NtcA-binding sites are highlighted in dark green. The pink hyphen in the sequence of Tol7601 highlights what may be a deletion that occurred in a string of eight A residues as the result of a sequencing error. The nucleotide frequencies are shown graphically in Fig. 8B and 8C.
- ^d The primary DIF1 score is calculated from a PSSM using the 54 DIF1 regions shown in Supporting Fig. S6 as the training set, on a log base 10 scale.