

Supporting Figure S7: *patS* upstream sequences

Organism ^a	Upstream sequence ^b	<i>patS</i> →	DIF1 score ^c
Nos21	AATCTTTAAACCTTCTCAACTAGTCTTTTGTGTTCAACTATCCGGATTTTCCGC-TAAAAAACAAGAGAAATTAATAGGG ... (562) ... CCATTGATTTTGAATAATGAAGGCAGCAATGTTTA	5.2	
Nos7107	AATCTTTAAACCTTCTCAACTAGTCTTTTGTGTTAACTATCCGGATTTTTCAGC-TAAAAAATCAAGAGAAATTAATAGGG ... (567) ... TCATTGATTTTGAATAATGAAGGCAGTAATGTTTA	5.4	
Nos7524	CATGATTTAACTTGCTCGTLAGCTTTTGGGATTAACATACTCCGGATTTTGTAGC-AAATAACTGTAGAGAAC-AAATAGGT ... (570) ... CATTGATTTTGGGATTAAGCAGATAACTATGTTTA	5.5	
Nos3756	AAGCCTTTTACTTACCCGATGAGCTTTTGCATATAACTATCCGGATTTTGTAGC-ACHTAACTGTAGAGAAA-AGTAGGG ... (580) ... TCACCGATTTTACGATTATGAAGGCAATTATGTTTA	5.8	
Ana29413	AAGCCTTTTACTTCTCCCATAGCTTTTGCATTTAACTATCCGGATTTTGTAGC-AAATAACTGTAGAGAAA-AAATAGGG ... (559) ... TCACCGATTTTAAAGATTATGAAGGCAATTATGTTTA	5.8	
Ana7120	AAGCCTTTTACTTCTCCCGTLAGCTTTTGCATTTAACTATCCGGATTTTGTAGC-AAATAACTGTAGAGAAA-AGTAGGG ... (559) ... TCACCGATTTTAAAGATTATGAAGGCAATTATGTTTA	5.8	
CylS7417	ATATTGAAGTACCAATTAAGGAAATTTTCCTCTGTTAGTTTCCTGTTTATCCCTCTACTAGCTGTAGATAGTTTTCAGAA (0) ACTTTTAGATACATTGGGGAATATTTTGTGAGC	-	
Nos29133	TGTGTGTTAACTTTTCTACTAGCTTTTGTACTGCTAAATACTCCGGATTTTGTCTCTTAATGAGAGTAGAGAACCAATCAGG ... (133) ... TTTTTCACTTATTTGGGTTGAGATAACAATGTTTA	3.1 (8.5)	
Nod9414	AAGATCTAACTTTTAACTACAATTTTGTGTTTAACTATCCGGATTTTCTCTAAATAACTGTAGAGAAATTAGTAGGG ... (595) ... TCATTGATTTTGGATAATGAAGACAACAATGTTTA	6.3	
Tol7601	... ATTTAGCAACAATAAAGGAAGGCTAGAATGTTTA	2.5 (<2)	
Cal7507	... TCACCTATTCTGAATAATTAAGATAACAATGTTTA	2.3 (3.1)	
Mch7126	ATACCTTTAAAGATTTCCGACCCCTTAGGGSTATCCGCTAATCCGCAATTTTATTGTCACAAATTAAGAAAATTCGGCAAA ... (581) ... TGATTAATTTTGAATAAAAAAGATAACAATGTTTA	2.9 (4.1)	
Has12170	... AAAATATTTAGTAAAAGCTGGTAACATAAATGACA	2.5 (<2)	
Tol9009	... AAAATGGCTAATAAAAACTGGTAACATAAATGACA	2.1 (<2)	
Cal336-3	TGTGCAGAGTATATTTCTGGCAAAATTTGCTGTTTAACTATCCGGATTCACACATTAACAATCGTAGAGAAAGTAAATACAG ... (243) ... CACAGTATTTTATTGAGTTGAGATAGCCAAATGTTT	4.2	
Cal6303	TGCTGTACTCGTATATTTACATACTCAAATAAACTATATTTCCGGATGATTACTGATAATCACAGTTTATTTAGTATGTAAC ... (717) ... AACATAGTTTATCTGTAGAGAGGAGTAAAATGATT	4.9	
Cal7103	CTACCAGTTCGGAGAATTTCTTAAAAACAGTCAAAAAAAATCCCTGATGTCACAAGAGCCAGGGTGTATTGGACTTGATGT ... (86) ... AACTATTAATAAATCGAGGAAAGAAAAATGATG	5.4	
Chl6912	TTGCTATTTAAATGTATTCATAAAATTTCASTGTTTAAAGAAATCCGGATTTATTGCC-CAAACTTTGTAGAGAAATGAATACGG ... (236) ... GCAACAATATTTGGGTAAGAGGAATAACATGAGAA	4.8	
Chl9212	TTGCTATTTAAATGTATTCATAAAATTTCASTGTTTAAAGAAATCCGGATTTATTGCC-CAAACTTTGTAGAGAAATGAATACGG ... (234) ... GCAACAATATTTGGGTAAGAGGAATAACATGAGAA	4.8	
Fis3754	AGTGGTTTAACTTTTATAGCTAAATTTCASTGTTTCACTATCCGGATTTAATGTC-TGAGTTTGTAGAGAAATGAATACGT ... (444) ... GTCAATAAATTTAGATAAGAGAGTTAACAATGCTTA	5.1	
Fis7521	AGTGGTTTAACTTTTATAGCTAAATTTCASTGTTTCACTATCCGGATTTAATGTC-AATTTATTACATTTTCTTTGGCAACA ... (77) ... GTCAATAAATTTAGACAAGAGAGTTAACAATGCTTA	3.1 (3.4)	
Fis9339	ATGGTTTAACTTTTATAGCTAAATTTAAAGTATTTCAAAAATCCGGATTTTCATTGA-GAAATTTGTAGAGAAATGAATACGG ... (443) ... ATCACTAAGTTTAGATAAGATAAATAACAATGTTTA	4.6	
Hap220	ATGGTTTAACTTTTATAGCTAAATTTAGGATTTCAAAAATCCGGATTTGTTTAA-TAAATTTGTAGAGAAATGAGTAAAT ... (437) ... ATCACTAGTTTAGATAAGAGAAAATAAATAATGCTTA	4.6	
Scy7110	... AATATTTGTAACAATACTTGGTAAAAAATGACA	3.3 (<2)	
Tol521301	... ATATTTGTAACAATACTTGGTAAAAAATGACA	2.2 (<2)	
Scy61278	... TAATTTTTCGAAGTCAATGGGTGTAATCAAGATGAAA	2.0 (<2)	
Mas10914	AGAGGGTTGCGGTGATGTTCTTCGTTTATTATAAGTATTAACTCCGGACATAATATA-AGCCTTTAAACGTTATTGAGTAAACA ... (126) ... AAAAATGTAAGTGAGGTTTGGTAATTAAGATGACT	3.1 (<2)	
Tol511288	... AGACAATCAGCCATTCATCATATATCAATATGTTTA	3.2 (<2)	
RicHH01	... TAGTATAAAATAATTCCTTGAATTCATATGTTCT	(<2) (<2)	
Riv7116	... TAAATTTAGATTGGGACATACATAATCATGAGA	3.2 (<2)	

^a Organism nicknames are associated with the full names in Fig. 3A.

^b Highlighting conventions are the same as in Supporting Fig. S5. In addition, possible translational start codons are rendered in red font. Sequences that are rendered in gray font are of interest but did not pass the DIF1 score threshold. The nucleotide frequencies are shown graphically in Fig. 7A.

^c The primary DIF1 score is calculated from a PSSM using the 54 DIF1 regions shown in Fig. 4 as the training set, on a log base 10 scale. The secondary DIF1 score (in parentheses) is calculated from a PSSM using the DIF1 regions upstream from *patS* genes shown above with primary DIF1 scores better than 4. A hyphen indicates the analysis wasn't done for that sequence.