

Supplementary Information for

Unraveling RubisCO Form I and Form II Regulation in an Uncultured Organism from a Deep-Sea Hydrothermal Vent via Metagenomic and Mutagenesis studies

Stefanie Böhnke and Mirjam Perner*

*To whom correspondence should be addressed: mirjam.perner@uni-hamburg.de

This file contains:

Supplementary Results and Discussion

Supplementary Figures 1 - 3

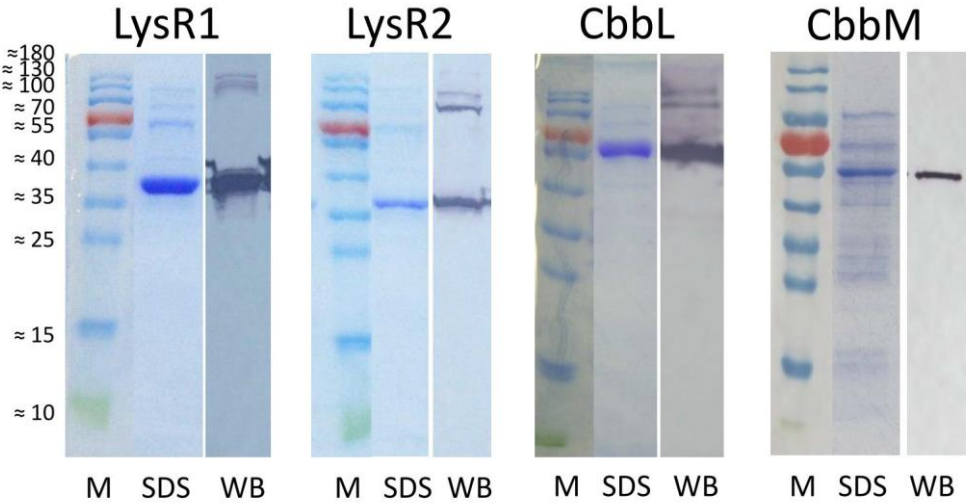
Supplementary Table 1

Supplementary Results and Discussion

Polar effects caused by transposon insertions. Polar effects of the *cbbM* and *cbbL* gene deletions were tested for the downstream genes *cbbO-m/cbbQ-m* and *cbbO-1/cbbQ-1*, respectively (Supplementary Figure 3). In the *cbbM* mutant *cbbQ-m* and *cbbO-m* transcription was disturbed. In $\Delta cbbL$ only *cbbQ-1* transcripts were significantly downregulated. However, we consider the actual polar effects on downstream genes not of significant importance for understanding the regulatory machinery behind RubisCO expression on our metagenomic fragment for the following reasons: If indeed *cbbQ-m/cbbO-m* downregulation was detrimental to RubisCO activity then one would expect (i) a loss of activity in the *cbbM* mutant relative to the undeleted fragment, and (ii) no or little change in RubisCO activity when these genes were deleted. However, neither is the case. RubisCO activity rises 5-fold when *cbbM* is knocked out and deletion of *cbbQ-m* and *cbbO-m* results in a considerable RubisCO activity loss (Figure 2A). Since deletions in the *cbbQ-m/cbbO-m* analogues *cbbQ-1* and *cbbO-1* do not result in any change of activity, CbbQ-m/CbbO-m appear to be sufficiently present to activate the RubisCO enzyme. The same principle applies for the *cbbQ-1* gene, downregulated likely as a consequence of the insertion in the *cbbL* gene.

Supplementary Figures

Supplementary Figure 1



Over expression of RubisCO and RubisCO associated genes. The purified eluates of the over expressed RubisCO transcriptional regulators LysR1 and LysR2 and the RubisCO form I (CbbL) and form II (CbbM) are visualized on SDS-polyacrylamide gels (SDS) and with Western Blot (WB). The prestained protein ladder, 10-180 kDa (Thermo Fisher Scientific, Waltham, MA, USA) was used as a marker (M).

Supplementary Figure 2

						IR1		
1	ATGTCACTTG	AGTTGACGAG	GTTGGCGATG	TTTCGGTGAA	CAAAAATAGT	TCCTGGATCC	ATCTTAACCA	GCTCGTTTGC
	-----	-----	----	<i>carbonic anhydrase</i>	----	-----	-----	-----
81	AGGAACTCGG	CTATCGGAGC	AACCTATCCA	GAGATATTCG	GGTTTTTGT	GGTGAGATAG	GGTTTCAAAA	AAGTCGGGAC
	-----	-----	----	<i>carbonic anhydrase</i>	----	-----	-----	-----
161	GAATGGCGTT	GATTTTCATCC	ACCCAATCAC	GGTTGTTTTG	TAACAGCTGC	TCGATTGATG	TATCGCAGTC	ACATTGATGG
	-----	-----	----	<i>carbonic anhydrase</i>	----	-----	-----	-----
						IR2		
241	CACATTAAGA	ATCCTTTTGG	AGAAGTTGTT	TTGGGGATAC	AAAAATGCC	TTAAATTGTA	AGGGCATTTA	CTTTAAAGCT
	-----	-----	-----	-----	-----	-----	-----	-----
321	TCTTCAAAAT	GGGAATTATA	CACGTTTTTA	TTGTGTGATT	TTCATAAAGA	CTTGAGGTAA	TTTTTCTGGC	AACTTGTTGA
	-----	-----	---	-----	-----	<i>cbbO-m</i>	-	-----
401	CATGGTCAAT	GACGGTATAG	TGGTTGTCAA	AAATGGTCTC	AACGTATTCA	TCGGCATTTCG	GATCTAAGGT	AATGCAGTAA
	-----	-----	-----	-----	<i>cbbO-m</i>	-----	-----	-----
481	GAATAGATCC	CTTTGCTTTT	AAGTTCTTCG	ACCGCTTTAT	GGGTGTCTTG	AATTAAAACC	TGTGGGTCTT	TTGTGTCAAT
	-----	-----	-----	-----	<i>cbbO-m</i>	-----	-----	-----
561	ATCCGCCGGC	TCACCATCGG	TTAAAATCAG	CATCAGTTTT	TTCTCAGCTT	GCTGTGCTTC	CAGATAGTGT	GCGGCATGAC
	-----	-----	-----	-----	<i>cbbO-m</i>	-----	-----	-----
641	GCATGGCGGC	TCCCATTCTGA	GTCGAATAAG	AGGCTTCCAT	CGCGGCAATG	CGAGACTTGA	CCTCATCGCC	ATAATGCTCT
	-----	-----	-----	-----	<i>cbbO-m</i>	-----	-----	-----
721	GAATAGCCTT	TGATATGCTG	ATAGCGCACT	TCATGACGCG	TGTCCGAGCA	AAAGCCCGCA	ATGGCGAACT	TATCCCCTAG
	-----	-----	-----	-----	<i>cbbO-m</i>	-----	-----	-----
801	CTGTTCAATG	GTCCACGCAG	TGATGGCAAG	GGCTTCTTCA	CTTAATTCCA	GTAAAGTTTG	GCCCGTTTCC	TGATTGCGTT
	-----	-----	-----	-----	<i>cbbO-m</i>	-----	-----	-----

```

881  CATTCAACGA TTGTGAGGTG TCGACCAGTA ACATGACGGC AATGTTTCGG CTGTCGGTAG TATGGCTGTA GTTGATTCGA
-----
          cbb0-m
-----

961  GGTCCGGTG CTTGCCCGCT TTTAAAGTCA ATGACGGAAC GAATCGCGAC GTCCAAATCC AACTCTTCCC CTTCTTCCTG
-----
          cbb0-m
-----

1041 AAACCGAATG CGTTTTTTGT TTTGAGGTTT GAGGGCTTCA ATCATTTTTT TCAAGCGTTT GGCCAAACCG TCATGTTTTT
-----
          cbb0-m
-----

1121 CCATCAGGCG GTCAATTTTG TTTGAATCTC CAGAAGGGTG TAAGCGTTCA TAAACCGTCG CCCAGTCTGG ACGATAGCTT
-----
          cbb0-m
-----

1201 TCGGAGATAT AATCCCATTC GTCATAGTGG CGGGGAGGTA AGCTGTTTGC GTCGTCGACC ACTTTTTTCAT CGGAGTGATA
-----
          cbb0-m
-----

1281 TTCGTTGTTA TGGAAATCTT CCGCTTCGTC ATTTTCTTCG TGATGGAACC AGATAAAGCG ATTGTCATCT CGATAGCTGA
-----
          cbb0-m
-----

1361 TTTCAGTATC GTCAAAGAAA ACGTTTGGCA AGCTGTCGGT TTTTTTGCGC GATTTGACAT AGAAGTTCGT GCCCAGTTTG
-----
          cbb0-m
-----

1441 GCCATGTCTT GTGTGGTGGG GCGTCGCCT TTTTCAGCCA TAATGGCATT GAACTGGGTG CGGAAGTCTT CGATCAATGG
-----
          cbb0-m
-----

1521 GTTTTCTGGG TCAAATGTTT CATCCATTAA GCGCGAGAC AGACGGGTGG ATCGGTATCG TAAACAGGCC TGTGGGTTCG
-----
          cbb0-m
-----
          IR3
-----

1601 GATCGCACGC GCCTTTTTCC GGCAGTGGGT GGAGTGCAAG AAACAAAGGT TTTAATCCGG GGAAGCGTTG AATGGCCAAT
-----
          cbb0-m
-----

1681 CTTTCCACTC GGCAGTCTTC AAAGGTGGAA ATAAAGAGCT GCATGTGAGG GGCGAAATTA TCGGCCATCA ACTTGGTTGA
-----
          cbb0-m
-----

```

1761	CCATTCTTTG	TGCGCCATCA	TATGCGCCAG	CATGGCGCGA	TAGCGATCCA	AGCCGCTAAC	ACCATTTTCA	TCTTCATACA
	-----	-----	-----	-----	<i>cbb0-m</i>	-----	-----	-----
1841	CATCTGGCAC	GGCAATGAGT	TCTTCATCCA	AATACGGAAC	GGGTTTACGA	AGTTGATCGA	ATGCGGTGGA	GAAAACCGAA
	-----	-----	-----	-----	<i>cbb0-m</i>	-----	-----	-----
1921	AAAGGGCGGT	CACAATGCCA	CATGCAGTCT	TTTAACATAT	CCAAATGACG	TTCCACGTCT	TTGAACGTTG	TGCCTTGACG
	-----	-----	-----	-----	<i>cbb0-m</i>	-----	-----	-----
				IR4				
2001	TTCACGTTGA	ATGATGCTTT	TGGCATCGTG	CGATTCTAAG	CTGAAATAGG	CAATTTGCTG	ATCGGGCGCA	TCATTATAGT
	-----	-----	-----	-----	<i>cbb0-m</i>	-----	-----	-----
2081	TACGCGCGCC	ATAATCAATA	AAATTACGGA	TTCCGGCCAG	GCTTAATTTG	GAAATCAGTT	GGGGCATGGA	TTCCAGTAAA
	-----	-----	-----	-----	<i>cbb0-m</i>	-----	-----	-----
2161	GGAATCATGC	CCGGACTTTC	ATACAAAGAG	TGATGACCAT	GAATGACGGT	TTGCGTTTTG	TCGACGTATT	CATCAATGAT
	-----	-----	-----	-----	<i>cbb0-m</i>	-----	-----	-----
								IR5
2241	GTCTAGATAG	TGCTGAAGGT	CTTCCAGCGT	ATCGATGCGT	CGGCAGACAC	TGCTGAGTGA	GGCCAAAAAA	GGAATTAAGG
	-----	-----	-----	-----	<i>cbb0-m</i>	-----	-----	-----
2321	CTTTTTTGTG	AGGGCTGCGC	GCTAACTTAT	AGCCGTAGTC	GGCCACCATT	TTCATGGTGC	CTTTTCCAAT	ATGAGAGGCG
	-----	-----	-----	-----	<i>cbb0-m</i>	-----	-----	-----
2401	ATTTCCGGCA	TGATCTCAAG	GTAGACCAGA	ACCGGCTCTA	CGCCCATGCC	TATCTTACAA	AGGAAGTTGG	CGCCATCAAT
	-----	-----	-----	-----	<i>cbb0-m</i>	-----	-----	-----
2481	ATAGGCAGTC	ACCCCTTCGT	CTGAAAGTTG	CTGGGCAGCC	TCTTCGATAC	ACGCTGGAAA	AACTTCCAAC	GCTTTGGGAA
	-----	-----	-----	-----	<i>cbb0-m</i>	-----	-----	-----
2561	AATTACAGGT	GAATTTGTCT	TGATACTCAG	AAACGACTTC	TGTTATGACG	TCTGCATTCA	TAATCTAACT	CATTTGATTG
	-----	-----	-----	-----	<i>cbb0-m</i>	-----	-	-

Unraveling RubisCO Regulation

2641 AATGATAAAT CGTCACCGCT CATGTTTGGG CGGTGGCTTA GCAAAAATAA CCAGGCCTGG CTGTTTTTTAG CCAAAAATCA
 --- *cbbQ-m* --

2721 TTTCAATGGC GTTATCAAGG GTCTGACGAA TATCGGCATC GTCGGTAATG GGACGAACCA AAGCCATTTT ACAGGCTTCA
 ----- *cbbQ-m* -----

2801 ATCGGCGCAA TACCTTGGTT AATCAGGGTG GCGGCATACA CCATTAGACG CGTTGAAATC CCTTCATCCA GCCCGTGGCC
 ----- *cbbQ-m* -----

2881 TTTCAGGTTA CGAGCGGTTT CACCGATCTT AACCAGTTTT TGGGCCGTAT CAGAGTCGAC ATTCCCTTCC TTTTGTAAGA
 ----- *cbbQ-m* -----

2961 TGTGCGCTTC GACTTCTGGT AGAGCATAAT CAAAGTCCAA TGCACAAAAA CGTTGTTTGG TTGATTGCTT TAAATCTTTC
 ----- *cbbQ-m* -----

3041 ATCAGTGATT GATAACCCGG GTTGTATGAA ATCACTAACT GAAAATCAGG GTGGGCTTTA ATCAACTCGC CTTTTTTATC
 ----- *cbbQ-m* -----

3121 TAAAGACAAT TCACGACGAT GGTCCGTTAA CGCATGAATG ACAACCATGG TGTCTTGACG TGCTTCCACA ATTTTCATCTA
 ----- *cbbQ-m* -----

3201 AATAACAGAT GGCACCATAA CGTGCGGCCA GCGTTAAAGG CCCATCCACC CAGCGAGTGC CGTTGGCATC CAGCAAATAA
 ----- *cbbQ-m* -----

3281 CGCCCAACCA GGTCTGATGC AGTGATGTCT TCATTACACG AAACGGTAAT AATCGGTTTG CCTAATTTCC AGGCCATGTG
 ----- *cbbQ-m* -----

3361 TTCGACAAAG CGGGATTTAC CACAGCCCGT CGGACCTTTC ACCATGACCG GTAAGCGAGC GGCATAAGCC GCTTCATAAA
 ----- *cbbQ-m* -----

3441 GTTCAACCTC ATTGGATTGA GCATCATAAA ATGGTTCGTT TTCAATCTTG TATTGTGTAA TGTCCATATC AGTTCCCCGT
 ----- *cbbQ-m* -----

-10 box *cbbQ-m* -35 box *cbbQ-m* Unraveling RubisCO Regulation

```

                                     IR6
                                     LysR bs2cbbQ-m
3521 AACGTTTATG GTAGGTATTT TCAAACAGAG TGTGAAAATA CCGATGAATA AATTAAAGCG ATATTAAAGT CATTTAGGTT
                                     -----
                                     TnA - n7 - TnA

3601 TCTGGTTTAC CGTAAGACAG GTAAAAGAGA AACTAAATA ACCTTGTGTT TGGCTAGGTT TTTGGTTTGG GTGTAAAAA

                                     LysR bs1cbbQ-m
3681 AACCCCTGCG CGAACGCAGA GTTTCTTGG GTGTTTAAAT AAACATTCGT TAAAGAATGG CTTATTTGTG TACGCCAAC
                                     - - - - -
                                     T nA -n7 - T nA
                                     ----- cbbM -----

3761 TTCTCTCTCC AACCTGGGTA GATTGTGTCT GCATCTGCAG GGAAAGATTC GAATGCGCGA GCGAACTCTT TGTGATCTTT
-----
                                     cbbM

3841 TGCGAATTCG ATTGGATCTG CACCTGATTT CCAGCACTCG TAAGATTGAC GAAGTGACGT TGCACCTGCC GCTGGAGAAT
-----
                                     cbbM

3921 CGATGTGACC GTAAGAACCA CCACCAGAAG TGTTGATAAC GTTCCCGTGA CCTAGGTTTT CGAAGAAACC AGGTAGACGT
-----
                                     cbbM

4001 AGTGCGTTCA TACCACCAGA GATGATAGGA GTCGTTGGCT TCATACCATT CCATTTCTGG TAGAAAGCTG GCCCTTGACA
-----
                                     cbbM

4081 CTCATCACGC TCGATCATGT ATGCGATGTT CTTATCAGAA GCATCACCTT CCATTTTACC GTAACCCATT GTGCCACGT
-----
                                     cbbM

4161 GGATAACCAGA CGCACCTTGT AGACGAGAGA TTTTAGCAAG AACGAATGCC GTGTAACCAC GTTTCGCTGA AGGCGACGTG
-----
                                     cbbM

4241 ATCGCACCAT GACCTGCACG GTGATAATGT AAGTACTGTG AAGGATAGTT ACGACGTGCT GTCGTAATCA TACCAGGACC
-----
                                     cbbM
IR7
4321 ACCAACATAA CCATCAACTA GGAAAGCAAC CTGAGGCGCG TCAGCACCGA AAGTTTCTAG GATGTAATCT GCACGGTAGC
-----
                                     cbbM

```


	IR7							
4401	ACATTTTCATG	GTGGTCATCT	GCAGTGATGT	TTGCTGAGAA	TAAC TTGGTT	TCACCCGTTT	CATCTTGAGC	ACGCTTCATT
	-----	-----	-----	-----	<i>cbmM</i>	-----	-----	-----
4481	GCGTCAGCAA	CAAGCGGAAT	GACTTTTTTTC	ATTGGGCAGA	ATGTTTGGTT	ACCTTGTGGC	TCATCGTTTT	TGATGAAATC
	-----	-----	-----	-----	<i>cbmM</i>	-----	-----	-----
4561	ACCACCTAGC	CAGAACTGGT	ATGCTGCGTC	AGCAAATGGC	TCAGGACGTA	GACCTAGCTT	AGGCTTGATG	ATTGTACCTG
	-----	-----	-----	-----	<i>cbmM</i>	-----	-----	-----
4641	CGATGTAACC	ACCGTTGTCT	TTATCACGAC	CAAGCATTTT	CCACATGTTT	GTGATATCAA	CAGCAGGACC	GTCAAATAGG
	-----	-----	-----	-----	<i>cbmM</i>	-----	-----	-----
4721	TGAAGTTTAG	TTTCTGGTAC	CCAGAAATCA	AACATTTGAA	GGTTTTTAAC	GTCACCCATG	CCTTGGTTGT	TACCAATGGC
	-----	-----	-----	-----	<i>cbmM</i>	-----	-----	-----
4801	AAGAGTCAAG	AAAGAAACGA	TCATTGTTTT	ACCGTCAAGC	CCGTTACGAT	CGAAAAGATC	GAAAGGGTAG	GCAACCTTCA
	-----	-----	-----	-----	<i>cbmM</i>	-----	-----	-----
4881	TGATTCCGTT	TGCTTCGTCG	ATGTCATAAA	CAATCGCGTC	AACGCCTTTC	GTGAAGTCAT	CTGTTGTACA	AACTTCAACG
	-----	-----	-----	-----	<i>cbmM</i>	-----	-----	-----
4961	TTTGTTCCAG	TTGAAGATTC	CGCTGCGATG	TGTGCAGCAA	CTTCTAAGTA	ACCGTAACCA	GCAGCAGGTT	CCATTGTGTA
	-----	-----	-----	-----	<i>cbmM</i>	-----	-----	-----
5041	AGCAACAAGA	ATGTGATTCT	GACCAGCAAT	CAGATCTTCT	TCTTTTAGGC	TTAAATCAGC	ATAACGATTC	GA CTGATCCA
	-----	-----	-----	-----	<i>cbmM</i>	-----	-----	-----
		IR8					LysR bs3 _{cbmM}	LysR bs2 _{cbmM}
5121	TGTTTTATAC	CTCTATA	TAT TGGGACAACA	GCTCAGGAAA	GTGTGTCCAT	AATTTTTGTT	AAGTCATTGT	AAAAACA ACT
-			-10 box <i>cbmM</i>		-35 box <i>cbmM</i>		Tn A - n7 - T nA	T nA - n8 -
			-----		-----		LysR bs1 _{cbmM}	
5201	TAACCCCGAA	ATTTGAATGG	TCATTATAGA	GGTTGTTT CAG	ATAACTTCAA	ATCAATATTT	TTGATTAAAT	TGATAGACTA
	TnA		IR8				TnA - n7 - TnA	

5281	TTATCTATTC	TAGTGAGGTG	CGATTTAAAT	GATGCGTAAT	AATCTGCAAA	GTTGGGTTAG	CGAAATTGGT	TAAACGCCTG
						----	--	<i>lysR2</i>
5361	TTCGCTGAGT	TTTTCGCCTT	CGATTTTCAG	TAACTCAATG	AGTTTTTCTG	CCGCAATAGA	CAGTACTTTG	CCTTTTGGGT
	-----	-----	-----	-----	<i>lysR2</i>	-----	-----	-----
5441	AAGCCAGGTA	CCAATGACGA	TTGATTGGAA	ATCCTTTGAC	ATTGAGGATG	CTGATTTTTC	CTTGATTAAT	TTCTTCCATT
	-----	-----	-----	-----	<i>lysR2</i>	-----	-----	-----
5521	AAAGTGGGAA	CCGATGCAAC	CGTGATTCCA	AGGTTTTGTA	GTAAGCCGAG	TCGAATAGGT	TCGTTGCCAC	CCAGCACCAT
	-----	-----	-----	-----	<i>lysR2</i>	-----	-----	-----
5601	TTTAATATTG	GGTTGAAATT	CAAATTCTCC	AAAGACCTTT	TCAATTTGAG	CCCGAATACC	TGAACCCGTT	TCGCGCATTG
	-----	-----	-----	-----	<i>lysR2</i>	-----	-----	-----
5681	AAAAAGGTTT	ATTTACCAGG	TCTTGAATTG	AGAGTTGCTT	TCCAACGAGA	GGGTGATCAC	TGTTGGCCAA	AAAAGCCAAT
	-----	-----	-----	-----	<i>lysR2</i>	-----	-----	-----
5761	GGGTTGACGG	AGAGCTGAGA	AGATTCCACA	TTTAAGTCTT	CTGGCGGTTG	TCCGAGTAAA	TAAAAGTCGT	CCTTGTTATT
	-----	-----	-----	-----	<i>lysR2</i>	-----	-----	-----
5841	ATTAATACGT	TCGATCAAAC	TTTCTTTGTT	ACCAACCCGC	ATGATGATGG	TGACGTCTGG	GTAGGATTGC	GTGAAATTTT
	-----	-----	-----	-----	<i>lysR2</i>	-----	-----	-----
5921	GAATGACTTT	AGGCACAAAA	TATTGCGCCG	TTGAGATTAC	CGTAAATTTG	ACGCTGCCCC	CAGAAAACCC	CTTGAGGTGA
	-----	-----	-----	-----	<i>lysR2</i>	-----	-----	-----
	<i>LysR bs2_{lysR2}</i>							
6001	TTGATTTTCT	GTTTCAGCAAT	CGACAGTTGC	TGCAAGATAT	TATTGGCGGA	ATCGTAGAGT	GCTTTACCGG	CTTCTGTTAA
	-----	-----	-----	-----	<i>lysR2</i>	-----	-----	-----
	TnA	- n8	- TnA					
6081	ATGGGCTTTA	CGACCGATTT	GTTCAAATAA	AGGACTTTCC	AAAATTTTCGG	CCAGTTTCTT	AACCTGCATG	GAAACCGTTG
	-----	-----	-----	-----	<i>lysR2</i>	-----	-----	-----

```

6161 GTTGAGTCAG GTGAAGCTCT TCAGCTGCTT TGGAGAAGCT GTTATGTCGG GCAATGCTTT CAAAGACTTG AAGCTGTCTT
-----
                lysR2
6241 AATGTGGCGT TTCGTGCCAG GATGGAAATT TTTTCAGGCA TAATGATTCC TTAGCGCGTG TCCGACATGG CCAGCAATTT
-----
                lysR2
                TnA - n7 - TnA
6321 GTTAATGTCG ACCGATTTCA GAAATTCCAA AAAGGCGACC GCTGGTGGCA TGAGTACCTT ATCTTGCATT TTGGCAATGT
-----
                lysR1
6401 ACCAATCGCG ACTCAATGGA AACCCTTGAA CGTCAAGAAC GGTTAAGTGA TCGTATTTTG CTTCGATGCG CACCGCGTGT
-----
                lysR1
6481 CGGGCAATGA CAGAAATACC CAGGCCTGCC ATAACCCCTT GCTTGATTGC TTCGGTACTG CCCAGTTCCA TGTAGGGCTG
-----
                lysR1
6561 GATAGAATAC CCTTCTTCGG CAAAGCGTTC TTCTGTGGCA ATACGGATGC CGGACCCGGC TTCTCGCATT AAAAAAGTTTT
-----
                lysR1
6641 CATTTACAAG GTCGCTCAAT GGGATTTGTT TATGTTTTGC AAGTGGGTGA TTGGGCGGAG CGGCCACAAT CAATTCATTC
-----
                lysR1
6721 TTGAAGAATG GGAAAGCTTC CATCTTTAAT TCCGGTGGTA CTCTTCCCAT AATAGAGAGG TGATATTGGT TTTGTTTAAG
-----
                lysR1
6801 CTCATCTATG ATGCGTCTTC TATTAACAAC TGTTATGGAA GGAATGACAT CTGGGTATTT ATTTAAAAAT GCTTTCAAAA
-----
                lysR1
6881 CATAAGGAAT AAAAAATTTG GCTGGGGTTA CAACCGCTAA TTTCAAGTTCC CCTTCAATTT TATTTTGATC TGTCTTAATG
-----
                LysR bs1cbbL
                TnA - n7 - TnA
                lysR1
                TnA -
6961 TTGGTGTAA GGGTCTTGAG CTCTTCTAGA ATGGTGGTGC AGGTCTTAAA CATGTGTTCG CCTGCGGATG TTAAATAGAG
-----
                LysR bs2cbbL
                N8 - TnA

```

7041 TTTTTTCCG ATCACTTCGA TCAATTTAAT TTCGTTATTT TCTTCCAGAC GCTTAACTTG AACGGATACG GCAGGTTGAG

 ----- lysR1 -----

7121 AAAGGTTGAG TTCTTCGGCT GCTTTTGTGT AGCTTAAGTG CCTTGAACA GCCTCGAAAA TCCTTATCTG CTGGGCGGTT

 ----- lysR1 -----
 ----- -35 box cbbL -----
 ----- LysR bs3_{cbbL} -----
 ----- -10 box cbbL -----

7201 ATATGTAAGT TTTGCATTTG GTTTTCCTAA TTGGTAAACT CATAAATGGT **TGATA**TATAA **AGTAAATGAG** **AGGAATCATA**

 ----- lysR1 -----
 ----- -35 box lysR1 -----
 ----- T nA - n8 - TnA IR9 -----
 ----- -10 box lysR1 -----
 ----- LysR bs2_{lysR1} -----

7281 **TACATTTGTT** **TATGAATTGT** **ATAACAAATA** **CTCA**ATATAT CACCAGATTA TTTTAGATAA GATACGAATC ACTGAAAGCG

 ----- IR9 -----
 ----- TnA - n7 - TnA -----
 ----- LysR bs1_{lysR1} -----

7361 GAAGTGTTTT TAAAACGCAA TTTAAGAATC GCTTTCTCTA ACTCTTTTCA TGATTAAACT AAAACGAAGA GTAAACCAT

 ----- TnA - n8 - TnA -----

7441 GGCTAAGACT TATAACGCCG GTGTAAAAGA ATACCGCGAA ACGTATTGGA TGCCAGAATA TGAGCCTAAA GACTCAGATT

 ----- cbbL -----

7521 TTCTAGCATG TTTTAAAGTA ATCCCACAGG ATGGTGTTCC ACGTGAAGAA ATCGCAGCAG CTGTTGCGGC TGAATCTTCA

 ----- cbbL -----

7601 ACGGGTACAT GGACAACGTG TTGGACTGAC TTGTTAACAG ATCTTGACTA TTATAAAGGT CGCGCATACA AAATTGAAGA

 ----- cbbL -----

7681 CGTTCCTGGT GACGATGCAG CATTTTATGC TTTCATCGCA TACCCAATCG ATCTATTCGA AGAAGGTCTT GTCGTTTCTG

 ----- cbbL -----

7761 TAATGACATC TTTAGTTGGT AACGTATTCG GATTTAAAGC ACTACGTGCT TGTCGTCTAG AAGATATCCG TTTCCCTCTA

 ----- cbbL -----

7841 GCGTATGTGA TGACGTGTGG TGGACCACCA CACGGTATCC AGGTAGAGCG TGACAAAATG GATAAGTATG GTCGTCCAAT

 ----- cbbL -----

7921	GCTAGGTTGT	ACTATCAAGC	CTAAGCTAGG	TCTATCGGCT	AAAAACTACG	GTCGTGCTGT	ATATGAGTGT	CTACGTGGTG
	-----	-----	-----	-----	<i>cbbL</i>	-----	-----	-----
8001	GTCTTGACTT	CACGAAAGAT	GATGAAAACG	TTACTTCTCA	GCCGTTTCATG	CGTTGGAGAG	ATCGTTTCCT	ATTCTGTCAA
	-----	-----	-----	-----	<i>cbbL</i>	-----	-----	-----
8081	GATGCAATCG	AAAAATCACA	AGCAGAAAACG	GGTGAGCGTA	AAGGTCACTA	CCTAAACTGT	ACAGCTGGTA	CGCCAGAAGA
	-----	-----	-----	-----	<i>cbbL</i>	-----	-----	-----
8161	AATGTACGAG	CGTGCTGAGT	TCGCTAAAGA	AATCGGTACT	CCGATCATCA	TGCACGATTA	CCTAACAGGT	GGTTTCACTG
	-----	-----	-----	-----	<i>cbbL</i>	-----	-----	-----
8241	CTAACACAGG	TCTTGCGAAC	TACTGTCGTA	AGAATGGTCT	GTTGTTACAC	ATTCACCGCG	CAATGCACGG	TGTAATTGAC
	-----	-----	-----	-----	<i>cbbL</i>	-----	-----	-----
8321	CGTAACCCAC	ATCACGGTAT	TCACTTCCGT	GTTTTAACGA	AAGCACTACG	TCTATCGGGT	GGTGATCACT	TACTACTCAGG
	-----	-----	-----	-----	<i>cbbL</i>	-----	-----	-----
8401	TACTGTTGTT	GGTAAGCTTG	AAGGTGACCG	TGAAGCAACT	CTAGGTTGGA	TCGATCTTAT	GCGTGATTCA	TTCATTCTTG
	-----	-----	-----	-----	<i>cbbL</i>	-----	-----	-----
8481	AAGATCGTTC	ACGTGGAATC	ATGTTGACCG	AAGACTTCGG	TGCGATGCCT	GGTGTTATGC	CAGTTGCTTC	TGGTGGTATC
	-----	-----	-----	-----	<i>cbbL</i>	-----	-----	-----
8561	CACGTATGGC	ATATGCCAGC	ACTTGTTTCT	ATCTTCGGTG	ATGACTCTGT	TCTTCAGTTC	GGTGGTGGTA	CGCTAGGTCA
	-----	-----	-----	-----	<i>cbbL</i>	-----	-----	-----
8641	CCCATGGGGT	AACGCTGCGG	GTGCGGCTGC	GAACCGTGTT	GCTTTGGAAG	CGTGTGTACA	AGCACGTAAC	GAAGGTAAAG
	-----	-----	-----	-----	<i>cbbL</i>	-----	-----	-----
8721	AAGTCGAGAA	AGAAGGTAAA	GAAATCCTTA	CTAACGCTGC	TAAACATAGC	CCAGAACTTA	AGATCGCAAT	GGAAACTTGG
	-----	-----	-----	-----	<i>cbbL</i>	-----	-----	-----

Unraveling RubisCO Regulation

8801 AAAGAAATCA AGTTCGAATT CGACACAGTT GATAAGCTAG ATGTTAAGCA TAAGTAATAG ATAATTTAGG AGAATATAAT

 ----- *cbbL* -----
 TnA - n8 - TnA

8881 TATGAGTATT CAAGATTACC CATCTCGTCT TTCAGATCCA CAATCTCGTA AGGCTGAGAC GTTCTCTTAC CTTCCAAAAA

 ----- *cbbS* -----

8961 TGACGGCTGA GCAAATCAAA GCACAAGTTC AGTACATCAT TGACAAAGGT TGGAACCCTG CAATCGAACA TTCAGAGCCA

 ----- *cbbS* -----

9041 GAGAATGCGT TCTCTTACTA CTGGTATATG TGGAAGCTTC CTATGTTTCGG TGAAACTGAC GCTGATGCCG TATTGGCAGA

 ----- *cbbS* -----

LysR bs1_{*cbbQ-1*}
 9121 AGTTGATGCA TGTATCAAAG CTAACCCAAA CAACCACGTT CGTTTGGATTG GTTATGATAA CTATGCACAG TCTCAAGGTG

 TnA - n8 - TnA ----- *cbbS* -----
 -35 box *cbbQ-1*

9201 CGAACATGCT TATTAAGCGT GGTGACATGT AATTTTAATT AAAATTACAT GACCTTACCC CTCAGGC^{TT} GCCTGAGGG^{GG}

 ----- *cbbS* -----
 -10 box *cbbQ-1* IR10

9281 **CCAAATT**TTG AGGGGAGATT AAACATCGGG TTTAATGCTT CTCTCGCCCT GGTTTACTTA GGCAGCTTAT ACCTCCTGCT

IR11
 9361 TAAGTAACGG CCAGCTAGCC **CCGAA**ACTAG CTGGTTTTCA ATCCATTTAG ATTT**ACTTAT** **TTTAATCAAG** **TAATG**AAATC

 ----- LysR bs2_{*cbbQ-1*} ----- IR12
 TnA - n7 - TnA

9441 TAAATATTTT GAATGCTACT TGCCAATGGA GAAAAAACAT GTCGGAAATT AACGCATTAG ATCCTAGTCA ATACCTTATT

 ----- *cbbQ-1* -----

9521 AAAGACGAGC CATTTTATCG TCCAGTTGCT GATGAAGTAG AGCTGTTTGA ATCGGGTTAT GCTTCAAGAA TGCCTATCAT

 ----- *cbbQ-1* -----

IR13
 9601 GTTAAAAGGG CCAACAGGGT GTGGTAAATC TCGTTTTGTT GAGTATATGG CACATAAACT CGGTAAGCCA TTAAT**CACTG**

 ----- *cbbQ-1* -----

IR13

```

9681 TGGCGTGTA TGAAGATATG ACCGCTTCAG ATTTGGTAGG TCGCTTCCTA ATCGATAATTA ATGGTACGCG CTGGCAGGAT
-----
           cbbQ-1

9761 GGTCCTTTAA CCGTCGCGGC ACGTATCGGT GCCATCTGTT ATTTGGATGA GGTGGTTGAA GCCCGTCAAG ACACCACGGT
-----
           cbbQ-1

9841 TGTTATTAC CATTAAACGG ATCACCGTCG TGTTTTACCT CTTGATAAAA AAGGTGAGTT GGTTAGAAGCG CACCCTGATT
-----
           cbbQ-1

9921 TCCAGTTGGT TATTTCTTAT AACCCGGGTT ATCAGTCTAT GATGAAAGAC CTTAAGCAAT CGACGAAGCA GCGTTTCGGT
-----
           cbbQ-1

10001 GGGTTTATAG TTGACTACCC TGAAGAATCT GTTGAAGCCG ATATCGTTGC GAAAGAAGCG GGCATTGATA AAGAAACGGC
-----
           cbbQ-1

10081 TGAAAAGCTG GTTCAAATTG CACAGCGTTC ACGTAACCTA AAAGGGCAGC GTCTGGATGA AGGGATTTCA ACACGTCTAT
-----
           cbbQ-1

10161 TGGTGTATGC GGCACAGCTA ATCAATAAAG GTGTGGATCC TAAAAAAGCG TGTACTATGG CCTTGATTAC GCCGTTGACA
-----
           cbbQ-1

10241 GATGATGACG ATATTCTAGA TACTTTGTTA ACCACTGTTG ATACTTTCTT CGAGTAATAA AAATCAGCTC AGTCAACCGG
-----
           cbbQ-1

10321 TTAAATTTA GGTCGCTTTT GCGGCCTAAT TTTTCACGGT GACTTGTAGA TATTATTGAT GATTAAATA AGAGAAAGAG
-----
           -35 box cbb0-1           -10 box cbb0-1

10401 CTATGAGTTT TGATATTGAA GAAGTCAAAG CATCCTTGAT TGAAAGTGTG CCACAACCTGG AAGAAATGCT TGATTCATTG
-----
           cbb0-1

10481 ATCAAGAAG CTTCGCATTA TATGAATGAA GCCTCTCGTG AAACATGGTT GCAAAACGCA CAAGGAATTG CTTACCTAGG
-----
           cbb0-1

```

```

10561 AAAAGGTCAA CAGGTTGTCG TTAGCTATCT TGAGGCAGTG CCGCAAGTTA TTGCTAGGAT TGATGATGAG ATTCTGGATG
-----
IR14
10641 ATATCCTTGA AACCGTCATG AAATTGTCTT CGGTGACCTC TGGAGAAGTG GTGTCTCTTG TATTGGATTC TTTGCCTGTT
-----
IR14
10721 GTGTCTGAAA GAACAGGAGA TATCGACCTT TTAAGACAAT ATTTAGCGCT CGTTTACCAG ATTGGTTCCA AAACGCCACG
-----
10801 CGGAATGCGT CCGATGTTGA GCAATATTGA CGAGTTGATG TCAAAACTGA CTGTTTCAGG CTTAAGACGC TGGGCACAGT
-----
10881 GGGGTGCGCA AGCGCATGCT CGAAACTTCC AAGCTCAGAT TGATTACTTT GGACTGGCAT CGGAAGATTC CAAAGCGGTT
-----
10961 TTCCAGCAAC AGCGTAAAGG CTCGTTGTTC ATCGATTATC ATCGACCAAT CAATTTTTTAT TTGAGAGCTT TCTGGGCACG
-----
11041 TGATTTCTTT ATTCGTCCTG CGGCGGCAGA TTACGATGAT TTTAAGCCTT ATTTTGAAAA TATGGCCATG CATCTACCTG
-----
11121 ACGCATTGAA TGATTTAGGC GAGATTAAAG GGGGCGAGTT GTATCGCGCA ATGGCCGCCC ATATGGCATC GCATTTAGCG
-----
11201 TATACCAAAG AAGCCATTTT GATGGAGCAG TTGAACCCCC AGCAAATGTT CTTTATTGAA TTGATTGAAG ATGCGCGTGT
-----
11281 GGAATATAAC GCCATTAAAA ATTTCCCAGG CCTGAAGGGG CTATGGAAAA AAGTCATTAA AGCCAGTATG GAAGCCTCCG
-----
11361 AGCTGCCTGA AAAGTCGACC GCTTATCGCT TAGAACAGCT TGCAGTAAAA TTGATGGATG TTAAGCATGA TTTGCAAGAT
-----

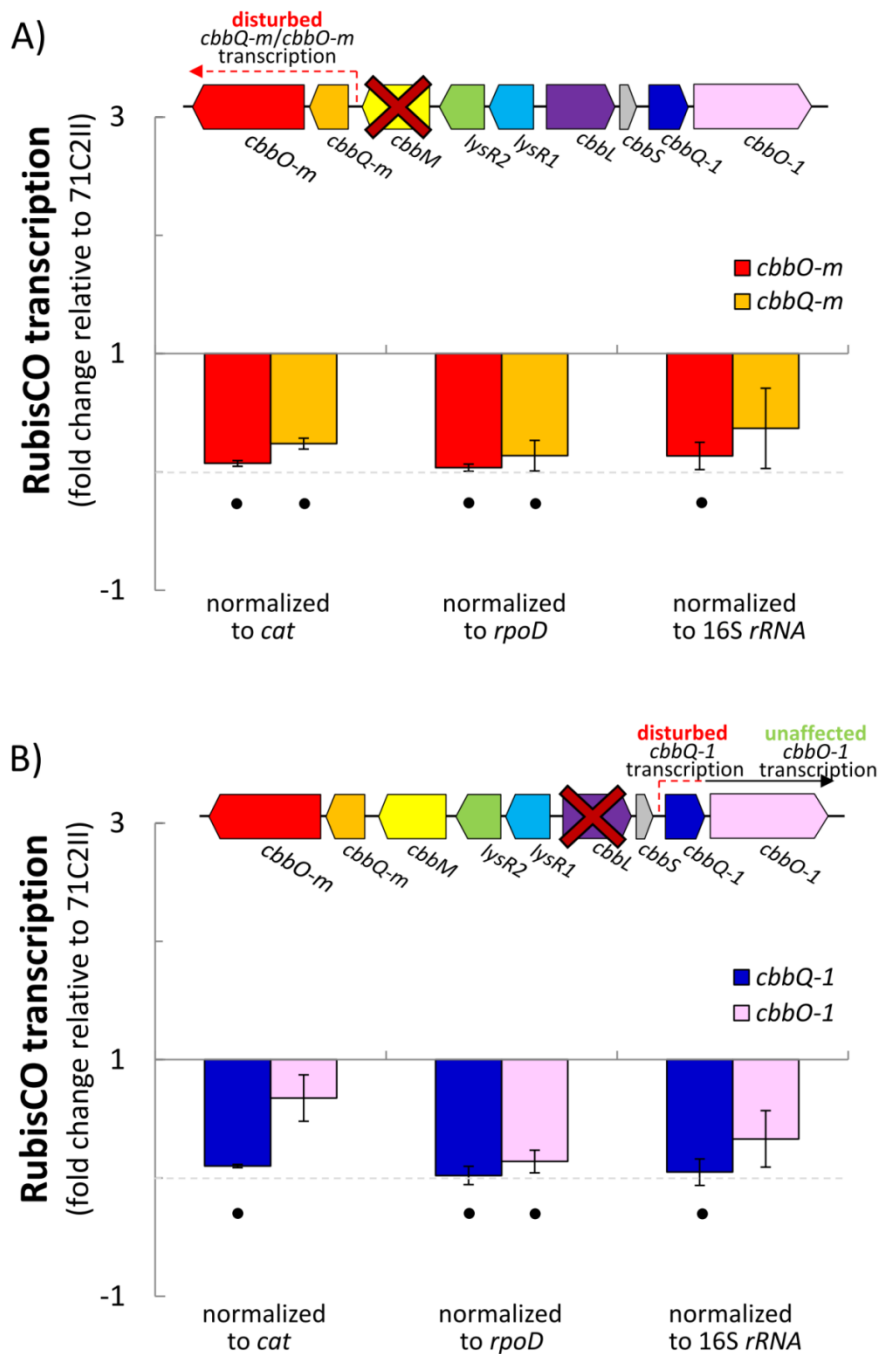
```


11441	GAGCAAATGA	TGGTGGTGGC	GGAACGTTTC	CATAATGAAA	TTGAAGAAAA	TCTGGACAAC	GAGAAATGGT	CTTGGGATTT
	-----	-----	-----	-----	<i>cbb0-1</i>	-----	-----	-----
11521	GGGGATTCTG	CTTTACAATG	TCTTGAATAA	AGCGACTTCT	AAGTGGGAAT	CTCTTACGGA	GATCAGTCAA	CAACGTTTTG
	-----	-----	-----	-----	<i>cbb0-1</i>	-----	-----	-----
11601	GCTATCGTGA	TGACAACCGT	TTGGTTTGGG	CATCCGACGA	ATGGGCCGAA	ATGGAAGGGG	GCGGTGCACC	ACATCAAGAA
	-----	-----	-----	-----	<i>cbb0-1</i>	-----	-----	-----
			IR15					
11681	ACCGTTCGTA	AAAATGTTTC	ATTGATGGAA	ATGATCAATG	AAATTGATTG	TGAACTGGTG	GATGTGGATC	ATGAAGAAGT
	-----	-----	-----	-----	<i>cbb0-1</i>	-----	-----	-----
							IR16	
11761	TTGGGTACTA	GGCTCGGAGC	TTTATCCTTA	CGAGGATAAT	GGACTGTCTT	ATAATGAGAT	GGAAGGCATT	GAGCCCGTTT
	-----	-----	-----	-----	<i>cbb0-1</i>	-----	-----	-----
11841	CAGATCCTTT	CCATTATCAT	GAGTGGGATT	ACCGCGTTCA	GTTGAATCGT	CCAAACTGGG	TCACGCTGTA	TGAGCACCGT
	-----	-----	-----	-----	<i>cbb0-1</i>	-----	-----	-----
11921	GCGAAAAAAG	GCGACCCTCA	GTTATATAAC	CGAATTCTAG	ATCAGAATAA	AGGAATTGCG	CATCGTATTA	AGCAAATCGT
	-----	-----	-----	-----	<i>cbb0-1</i>	-----	-----	-----
							IR17	
12001	TGATAAGTTG	CAAGCGGTTG	GTTTGCAGCG	TATTCGCCGA	ATTGAAGATG	GAGACGAACT	GGATTTAAAT	GCCTGCGTTG
	-----	-----	-----	-----	<i>cbb0-1</i>	-----	-----	-----
12081	AAGCAATTAC	GTCAATTCGT	ATGGGGCATG	AACCTGATCC	TCGCATTACG	ATGAAAAATG	TGATTTCGAG	CCGTGAAGTA
	-----	-----	-----	-----	<i>cbb0-1</i>	-----	-----	-----
12161	TCGGTGGTCG	TGTTGCTTGA	TTTATCAGAA	TCGACTAATG	AAATGGTTGA	TGGTGGCGAT	AAAACGTGCC	TAGAAGTAAC
	-----	-----	-----	-----	<i>cbb0-1</i>	-----	-----	-----
12241	GCAGGAGGCG	GCAATCTTGG	TGTCGCATGC	CATCAATGGT	ATCGGCGATA	AGTTTGCTGT	TCATGGTTTC	TCATCAGATG
	-----	-----	-----	-----	<i>cbb0-1</i>	-----	-----	-----

12321	GGCGTCATGA	CTTGCAATAC	ACTCGATTCA	AACAGTTTGA	CGAACCTTTT	GATCAAGATG	TTCATTACAG	TCTTGCGGGA
	-----	-----	-----	-----	<i>cbb0-1</i>	-----	-----	-----
12401	ATGAAAGGTG	GCCTCTCAAC	TCGTATGGGC	GGCGCCATGC	GCCATGCGGG	TAGTTACCTT	GAGAAACAAT	CCAGTAAGCA
	-----	-----	-----	-----	<i>cbb0-1</i>	-----	-----	-----
12481	GAAGCTGTTG	TTGGTCATTA	CCGATGGTGA	ACCAGCGGAT	ATTGAC	IR18 GAAA AAGACGGTCA	ATATTTAAAA	CAAGATGCCA
	-----	-----	-----	-----	<i>cbb0-1</i>	-----	-----	-----
12561	AGAAAGCCGT	TGAAGAACTT	CAGGCGAAAG	GGGTTTATTC	GTATTGCTTA	ACGATCGATC	AGTATGCCGA	TAAGTATGTG
	-----	-----	-----	-----	<i>cbb0-1</i>	-----	-----	-----
12641	CATAATATTT	TTGGTCAAAA	CCGTTATGCG	ATTGTGGATA	ATGTCTTGAA	GCTGCCTGAA	AAATTACCGC	AATTATTTGC
	-----	-----	-----	-----	<i>cbb0-1</i>	-----	-----	-----
12721	AAACTTAACC	ACCTAACTTG	GAACGAAAAA	TAAGATGAAA	GCACAACCGT	TTATTGAAGC	CGTCAAGCAG	TTAACTGATG
	-- <i>cbb0-1</i>	----		-----	-----	hypothetical protein	-----	-----
12801	ATGATTTTCA	GCTTATTCTT	GAAGGAAGTG	CCATTATCAT	TGAAAATGAT	GTGGCATTAA	CAACAGGAAG	GGCCGACAGT
	-----	-----	-----	-----	hypothetical protein	-----	-----	-----
12881	GCTTATGTTA	TTTATGAACT	GGGAGAAGAC	CCTTTTACAT	CTTCTGACGA	GATTAAAGCG	TTTTTAATAC	AGAATGCCGA
	-----	-----	-----	-----	hypothetical protein	-----	-----	-----
12961	AGCGCTTTTA	AAAGAGTATT	ACCAGTTTAA	TCCGGTCAGC	CGTCAATATT	TTGATCGCAG	TTT	
	-----	-----	-----	-----	hypothetical protein	-----	---	

Structural features predicted for the DNA of the metagenome derived RubisCO gene cluster. Exact positions of predicted promoter boxes are indicated by black dashes over each associated base and a heading with -10 box or -35 box. Putative LysR binding sites (bs) are highlighted in green. Inverted repeats (IR) are numbered consecutively from IR1 to IR18. Each IR is highlighted in grey and red, whereby the grey part shows the repeat and its reverse complement, putatively forming the stem, and the bases in between, in red, represents the putative loop forming area. Only IR with at least 8 bp in the potential stem region and loops with a maximum of 100 nt were considered.

Supplementary Figure 3



Fold changes of *cbbO-m*, *cbbQ-m*, *cbbQ-1* and *cbbO-1* transcription. Fold changes ($2^{-\Delta\Delta C_t}$) of A) *cbbO-m* (red) and *cbbQ-m* (orange) transcripts expressed from the transposon clone $\Delta cbbM$ (22II) and of B) *cbbQ-1* (blue) and *cbbO-1* (pink) transcripts expressed from the transposon clone $\Delta cbbL$ (24II). The data shown is normalized to three different reference genes, the *cat* gene, which is encoded on the fosmid vector and reflects its copy number, the 16S rRNA and *rpoD* housekeeping genes. RubisCO transcription is relative to the clone 71C2II, containing the intact RubisCO gene cluster (*cbbO-mQ-mM lysR2lysR1 cbbLSQ-1O-1*). Bars and error bars denote mean values and +/- standard errors. Black dots denote that values are significantly different (p-value ≤ 0.05).

Supplementary Tables

Supplementary Table 1

Supplementary Table 1. Insertion positions of tested double transposon clones with 13 kb inserts.

inserted genes	clone designation	insertion position [aa]	total orf length [aa]	specific RubisCO activity [nmol 3-PGA*min ⁻¹ *mg ⁻¹]
/	71C2II	/	/	49 ± 6
<i>ΔcbbM</i>	22II	167	459	265 ± 48
<i>ΔcbbM ΔcbbO-m</i>	22II1A3	208	757	130 ± 22
	22II1G1	34		178 ± 24
<i>ΔcbbM ΔcbbQ-m</i>	22II2H4	135	266	81 ± 7
<i>ΔcbbM ΔlysR2</i>	22II3A3	28	314	126 ± 39
<i>ΔcbbM ΔlysR1</i>	22II2B2	79	308	163 ± 29
<i>ΔcbbM ΔcbbL</i>	22II1C5	15	472	0.5 ± 7
	22II2B4	377		-1.5 ± 11
<i>ΔcbbM ΔcbbQ-1</i>	22II2C9	159	272	205 ± 48
<i>ΔcbbM ΔcbbO-1</i>	22II2A10	75	777	235 ± 32
	22II4H6	503		234 ± 10
<i>ΔcbbL</i>	24II	38	472	18 ± 7
<i>ΔcbbL ΔcbbO-m</i>	24II2H3	482	757	17 ± 5
	24II2G7	212		13 ± 5
	24II1H11	82		14 ± 5
<i>ΔcbbL ΔcbbQ-m</i>	24II1F9	147	266	17 ± 5
<i>ΔcbbL ΔcbbM</i>	24II2F3	301	459	-4 ± 5
<i>ΔcbbL + ΔncrcbbM-lysR2</i>	24II1G2	103 ¹	214 ²	8 ± 3
<i>ΔcbbL ΔlysR2</i>	24II1H7	46	314	48 ± 8
<i>ΔcbbL ΔlysR1</i>	24II2F11	263	308	16 ± 7
	24II1H1	234		17 ± 7
<i>ΔcbbL + ΔncrcbbS-cbbQ-1</i>	24II2G1	222 ¹	245 ²	9 ± 2
<i>ΔcbbL ΔcbbQ-1</i>	24II6H6	10	272	2 ± 2
	24II5G11	43		6 ± 1
	24II6G12	367		17 ± 3
<i>ΔcbbL ΔcbbO-1</i>	24II5G9	416	777	19 ± 4
	24II5D3	618		13 ± 1

¹ insertion positions in non-coding regions given in nucleotides² total length of non-coding region given in nucleotides