

Supplemental Material

Supplemental Figures

Supplemental Figure S1.	
Summary of the initial search (CM) for Class I RNA candidates.	2
Supplemental Figure S2.	
Representative RNA-seq coverage plots of Class I RNAs.	3
Supplemental Figure S3.	
Per species ROC evaluation of CM and Classifier thresholds.	4
Supplemental Figure S4.	
Conserved features of Dictyostelia Class I RNAs in representative species.	5
Supplemental Figure S5.	
Class I RNAs are developmentally regulated.	6
Supplemental Figure S6.	
Conserved features identified in representative species are conserved throughout Dictyostelia.	7
Supplemental Figure S7.	
Genomic distribution of Class I RNA loci.	Separate file
Supplemental Figure S8.	
Shared synteny between <i>D. discoideum</i> and <i>D. firmibasis</i> Class I RNA loci.	Separate file
Supplemental Figure S9.	
Conservation of the Class I RNA interacting protein CIBP in Dictyostelia.	8

Supplemental Tables

Supplemental Table S1.	
Genomic resources used in this study.	Separate file
Supplemental Table S2.	
Search summary and expression validation for Class I RNAs from selected dictyostelids.	Separate file
Supplemental Table S3.	
Name, genome location, validation and score of all curated Class I genes for all investigated dictyostelids.	Separate file
Supplemental Table S4.	
Identified Class I RNA candidates in outgroups.	Separate file
Supplemental Table S5.	
Potential CIBP orthologues in outgroup genome sequences.	Separate file
Supplemental Table S6.	
Oligonucleotides used for northern blot analysis.	9

Supplemental Code

Python scripts necessary to reproduce analyses and plots together with example files.	Separate file
---	---------------

Supplemental References

10

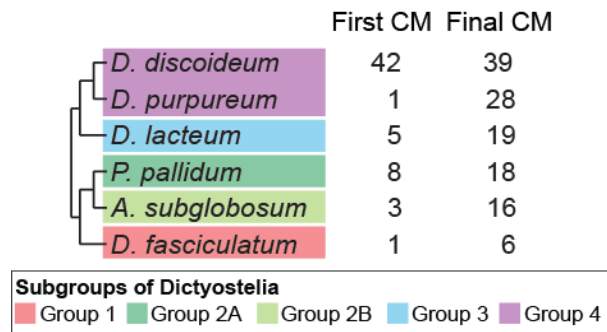


Figure S1. Summary of the initial search (CM) for Class I RNA candidates. Distribution of Class I RNA gene candidates (CM score ≥ 25) identified with the first version of the co-variance model (First CM) and the last version of the CM (Final CM). The first version of the CM was based solely on previously published *D. discoideum* Class I RNAs, while the final version of the CM was based on Class I RNA gene candidates from all six dictyostelids included in the CM build. The final CM search was performed with increased sensitivity (see Methods for details).

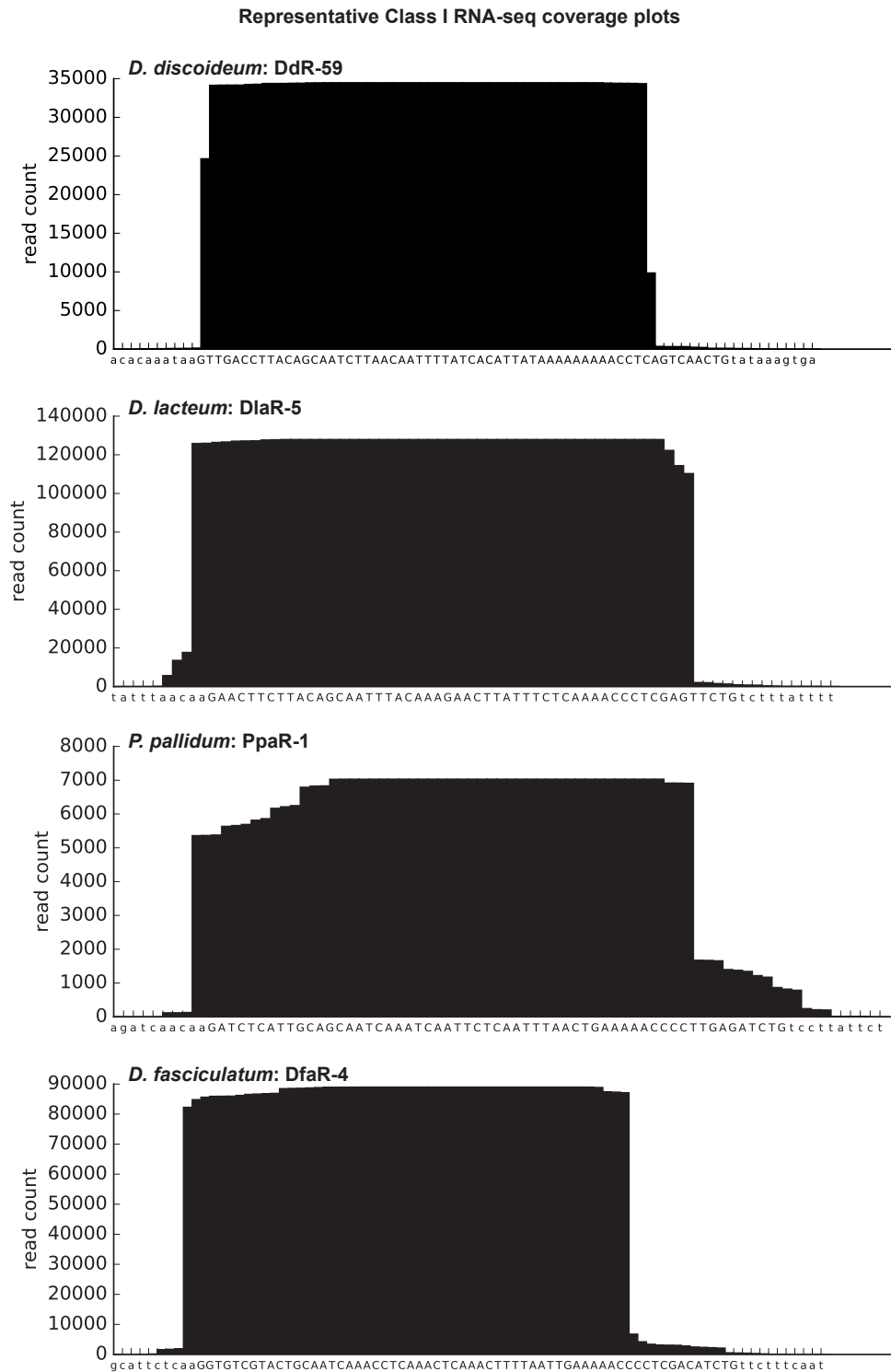


Figure S2. Representative RNA-seq coverage plots of Class I RNAs. Each Class I RNA plot represent the total raw read coverage for all replicates combined. Capital letters (x-axis) denote the nucleotide sequence of the predicted Class I RNA sequence and letters in lower case represent 10 nucleotides upstream and downstream of the 5' and 3' ends. DlaR-5, PpaR-1 and DfaR-4 was also validated by northern blot (see Fig. 2).

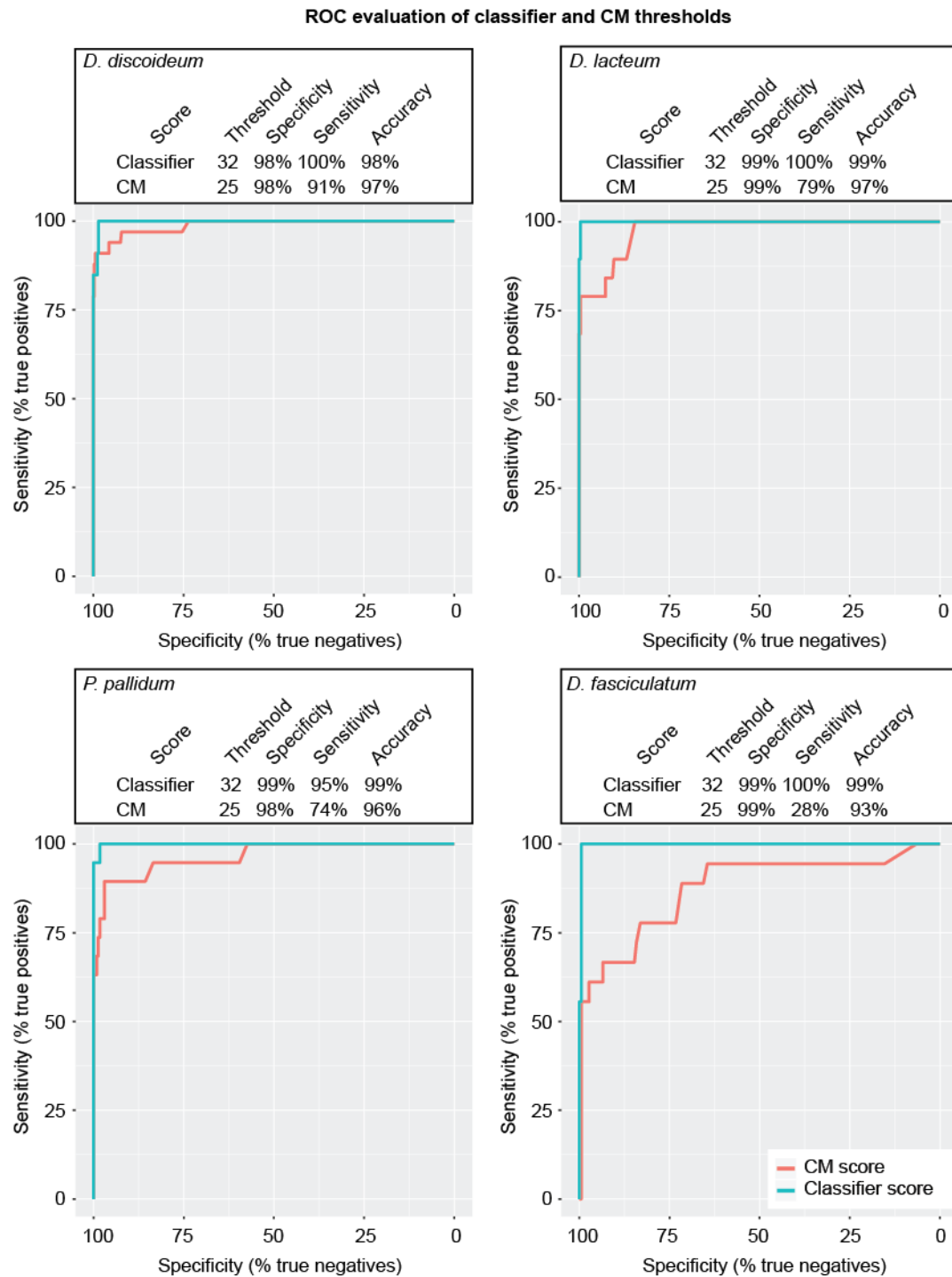


Figure S3. Per species ROC evaluation of CM and Classifier thresholds. Curves are based on the RNA-seq validation of all Class I RNA candidates (CM score ≥ 15) and either classifier score or CM score. The specificity, sensitivity and accuracy at the thresholds used for the classifier (32) and CM (25) is indicated above each plot. The input data used to generate the plots can be found in Supplemental Table S2.

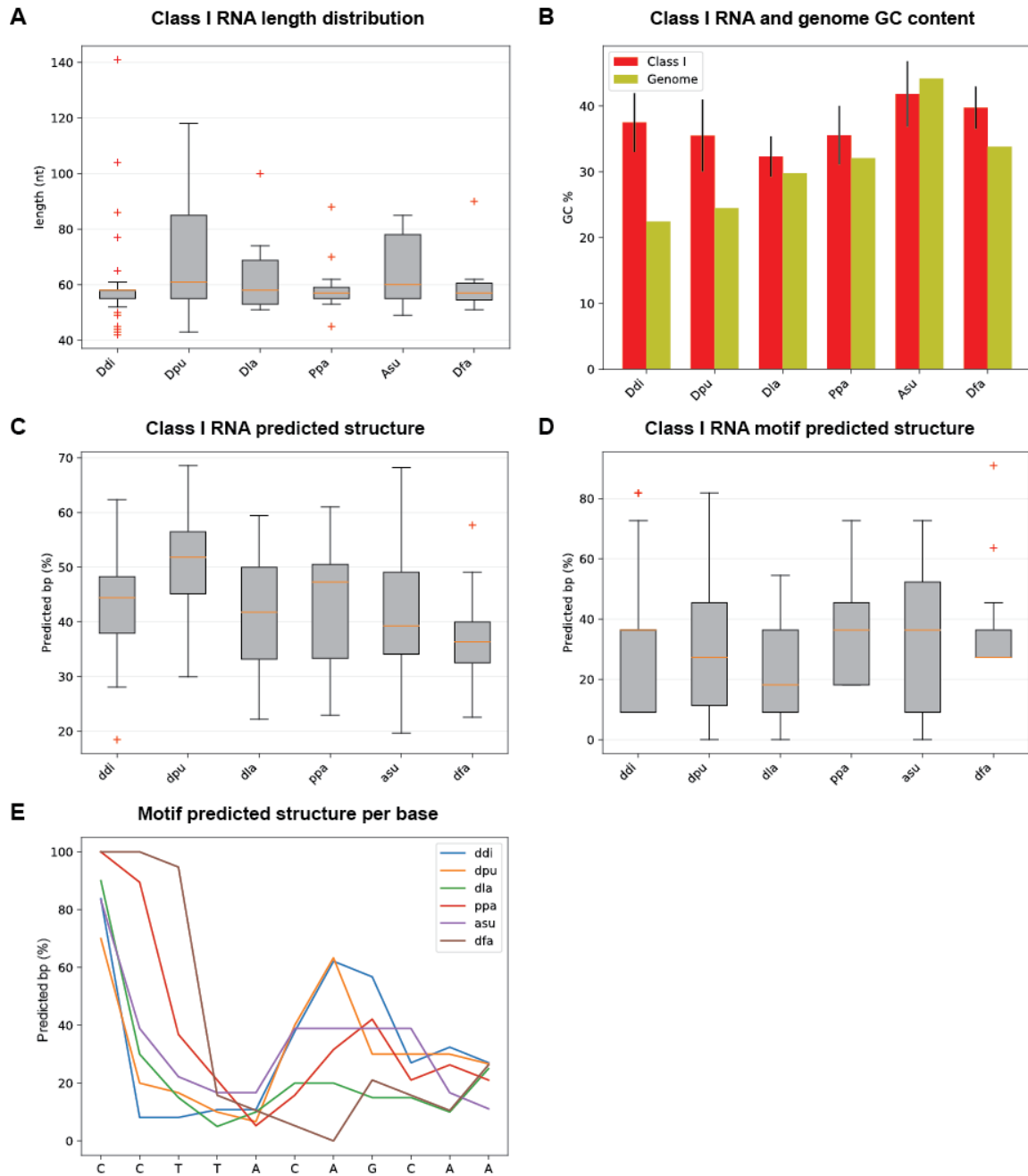


Figure S4. Conserved features of Dictyostelia Class I RNAs in representative species. A,C,D: boxes represent the quartiles of the lengths, whiskers represent the rest of the distribution except for outliers indicated with +. Orange solid line in boxes indicate the median. A) Length distribution of all identified Class I RNAs per species. B) Mean GC content (%) of Class I RNA and respective genome sequence. Error bars indicate standard deviation. C) Percent predicted base pairs of mature Class I RNAs. D) Percent predicted base pairing in the conserved 11 nt motif region. E) Predicted base pairing frequency per nucleotide for the conserved 11 nt motif. *D. discoideum* motif sequence is given on the x-axis and y-axis denotes percent of Class I RNAs that are predicted to be structured at the respective position of the motif. NB that the nucleotides are from the *D. discoideum* motif.

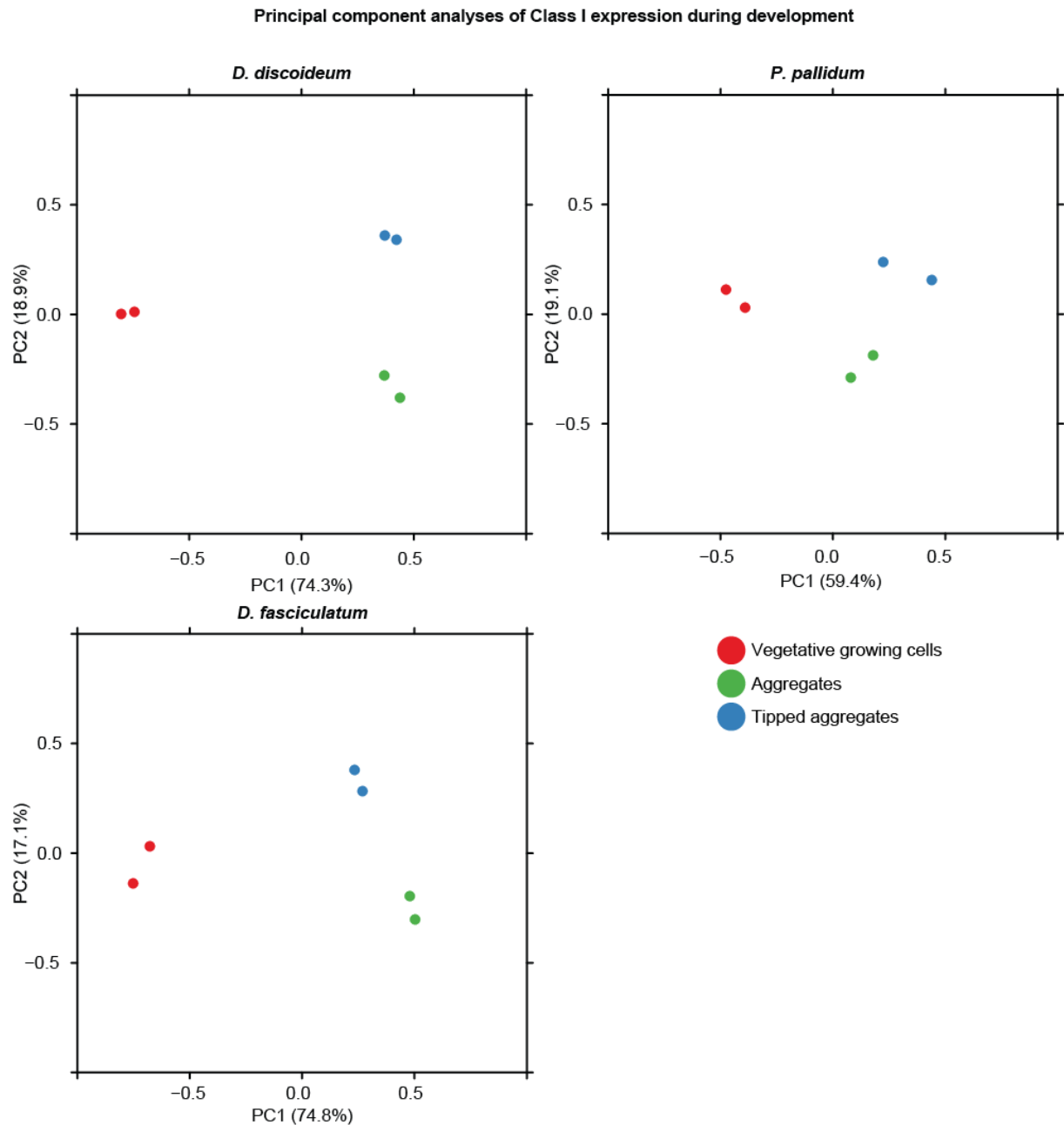


Figure S5. Class I RNAs are developmentally regulated. Principal component analysis based on RNA-seq read counts for Class I RNAs in *D. discoideum*, *P. pallidum* and *D. fasciculatum* (Supplemental Table S2). Developmental time points/structures (biological duplicates) are indicated by different colors.

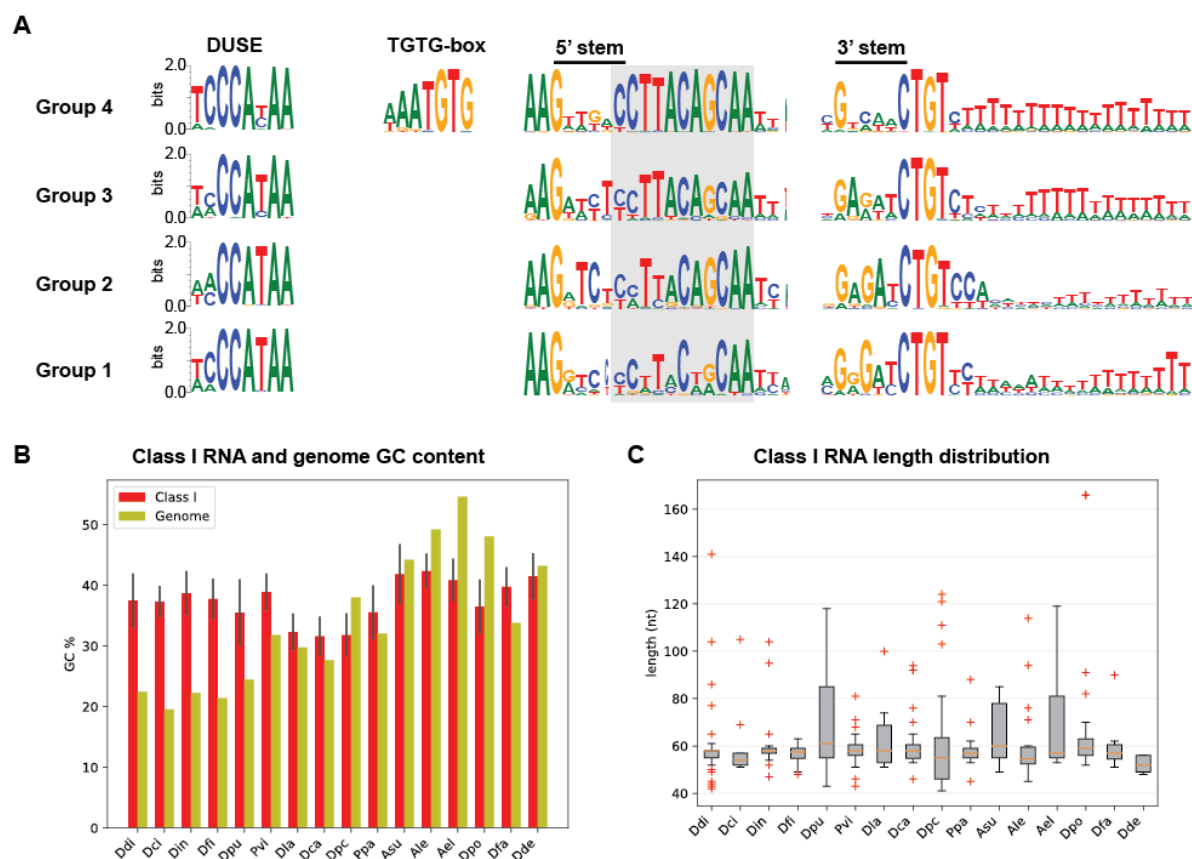
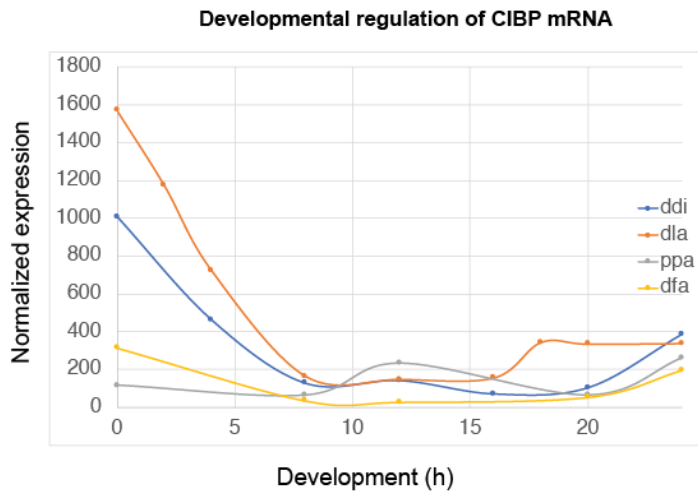


Figure S6. Conserved features identified in representative species are conserved throughout *Dictyostelia*. A) Conservation of DUSE, TGTG-box, 5' and 3' stem, motif sequence (boxed in gray) and downstream region for all Class I RNAs in each major subgroup of *Dictyostelia*. Two *D. citrinum* Class I RNAs were excluded from the alignments due to the lack of flanking genome sequences. B) Mean GC content (%) of Class I RNAs and respective genome sequence. Error bars indicate standard deviation. C) Length distribution of all identified Class I RNAs per species. Boxes represent the quartiles of the lengths, whiskers represent the rest of the distribution except for outliers indicated with +. Orange solid line in boxes indicate the median.

A



B

	CIBP orthologues	RRM's per orthologue	Length
<i>D. citrinum</i>	0	n/a	n/a
<i>D. intermedius</i>	1	2	297 aa
<i>D. firmibasis</i>	1	2	295 aa
<i>D. purpureum</i>	1	1	203 aa
<i>P. violaceum</i>	1	2	273 aa
<i>D. lacteum</i>	1	2	302 aa
<i>D. caveatum</i>	1	2	288 aa
<i>D. polycephalum</i>	1	2	289 aa
<i>P. pallidum</i>	1	2	261 aa
<i>A. subglobosum</i>	2	2	271/275 aa
<i>A. leptosomum</i>	1	2	114 aa
<i>A. elipticum</i>	1	1	61 aa
<i>D. polycarpum</i>	1	2	274 aa
<i>D. fasciculatum</i>	1	2	323 aa
<i>D. diminutivum</i>	2	2	278/282 aa

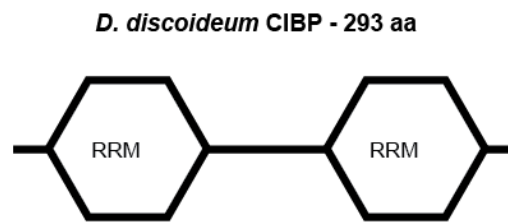


Figure S9. Conservation of the Class I RNA interacting protein CIBP in Dictyostelia. A) Developmental regulation of CIBP mRNA in *D. discoideum* (ddi), *D. lacteum* (dla), *P. pallidum* (ppa), and *D. fasciculatum* (dfa). Developmental time points are normalized to the development of *D. discoideum* (Expression data from (Parikh et al. 2010; Glöckner et al. 2016). B) Number of CIBP orthologues identified in each species, number of RNA recognition motif's (RRM's) per orthologue, and the predicted length of the protein. The domain composition of *D. discoideum* CIBP, which is representative for all CIBP with two RRM's, is schematically depicted to the right.

Table S6. Oligonucleotides used for northern blot analysis. Name, sequence and matching Class I RNA(s) are denoted for each oligonucleotide/probe.

Oligo name	Oligo sequence	Class I RNA
DfaR-4	GTTTGATTGCAGTACGACACC	DfaR-4
AsuR-10	GTAGGATACGTTTGGTTG	AsuR-10
AsuR-13/14	GTTTGAAGTTTGGTAGTGATT	AsuR-13, AsuR-14
PpaR-9	GTTGATTGCTGTAAGGCCACC	PpaR-9
PpaR-1/2	GGGTTTTTCAGTTAAATTGAG	PpaR-1, PpaR-2
DlaR-5	GAGAAATAAGTTCTTTGTAAA	DlaR-5
DlaR-1	GGAGAAAAAAGATCATTTAG	DlaR-1
DpuR-X	CCAATTTTCCCGGAAAGAC	DpuR-5, DpuR-6, DpuR-13, DpuR-16, DpuR-21, DpuR-26
DpuR-7	CTTTGGGAGGGAAAATGTCAA	DpuR-7
DpuR-Y	TTGCTGTAAGGAATTCTTGA	DpuR-2, DpuR-10, DpuR-12, DpuR-6, DpuR-24, DpuR-1, DpuR-26, DpuR-21, DpuR-7, DpuR-8, DpuR-4, DpuR-3, DpuR-17, DpuR-13, DpuR-23, DpuR-18, DpuR-20, DpuR-9, DpuR-16, DpuR-5, DpuR-19, DpuR-14, DpuR-11, DpuR-25

References

- Glöckner G, Lawal HM, Felder M, Singh R, Singer G, Weijer CJ, Schaap P. 2016. The multicellularity genes of dictyostelid social amoebas. *Nature Communications* **7**: 1–11.
- Parikh A, Miranda ER, Katoh-Kurasawa M, Fuller D, Rot G, Zagar L, Curk T, Sucgang R, Chen R, Zupan B, et al. 2010. Conserved developmental transcriptomes in evolutionarily divergent species. *Genome Biology* **11**: R35.