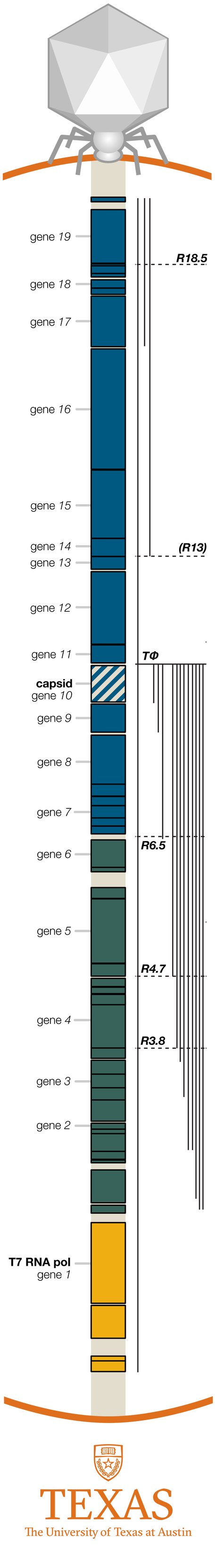


Disrupted proteome in a virus attenuated by codon deoptimization

Benjamin R. Jack, Daniel R. Boutz, Matthew L. Paff, Bartram L. Smith, James J. Bull, Claus O. Wilke

Department of Integrative Biology, University of Texas at Austin, Austin, TX

How does codon deoptimization affect the viral life cycle?

A general means of viral attenuation involves the extensive replacement of synonymous codons in the viral genome. The mechanistic underpinnings of this approach remain unclear, however. Using quantitative proteomics and RNA sequencing, we explore the molecular basis of attenuation in a strain of bacteriophage T7 whose major capsid gene was engineered to carry 182 suboptimal codons. We do not detect transcriptional effects from recoding. Proteomic observations reveal that translation is halved for the recoded major capsid gene, and a more modest reduction applies to several co-expressed downstream genes. Viral burst size, like capsid protein abundance, is also decreased by half. **Together, these observations suggest that codon deoptimization reduced translation of an essential polycistronic transcript and diminished virion assembly, leading to a decline in burst size and viral fitness.**

Strains of bacteriophage T7 used in this study

recoded

fitness: 35.7 doublings per hour

Approximately half of the codons in the major capsid gene (*10*) have been replaced with sub-optimal codons. This recoding resulted in a ~180 fold decline in fitness [1].

evolved

fitness: 38.7 doublings per hour

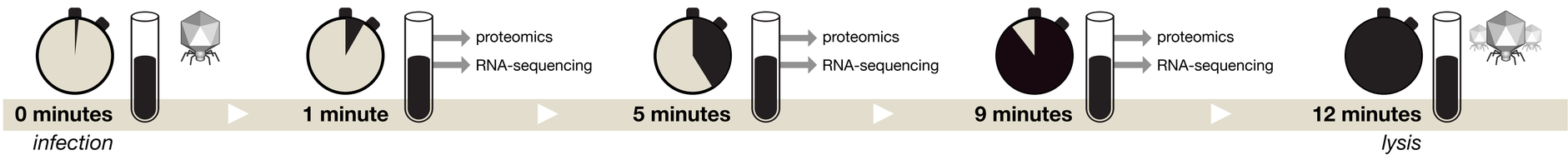
The recoded strain (above) after it has evolved for 1000 generations. The evolved strain recovers some fitness but still remains attenuated relative to the wildtype [1].

wildtype

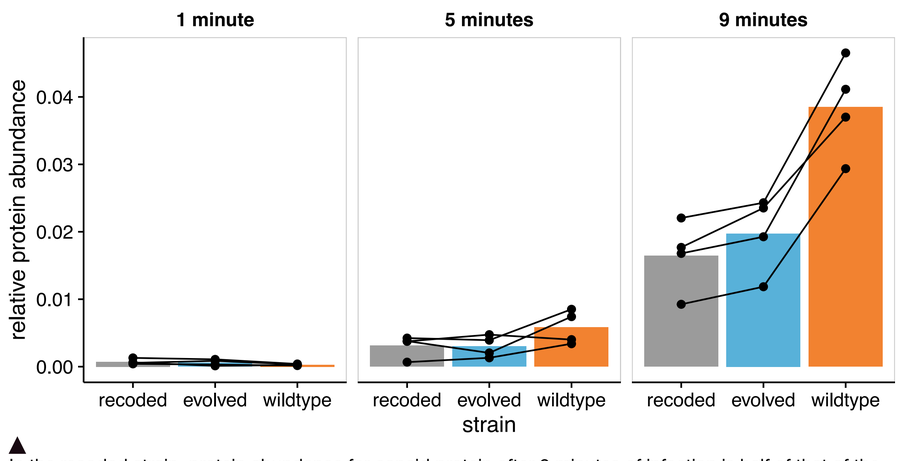
fitness: 43.2 doublings per hour

Strain of T7 adapted to laboratory conditions. This strain was used as the basis for the recoded strain [1].

Proteomics and transcript data collected at 1, 5, and 9 minutes after infection



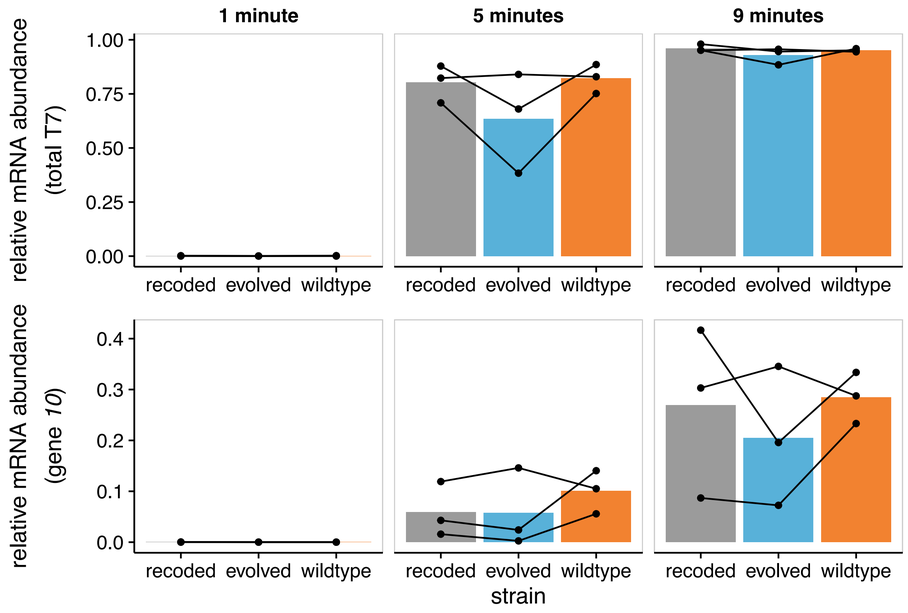
Codon deoptimization reduces capsid protein abundance



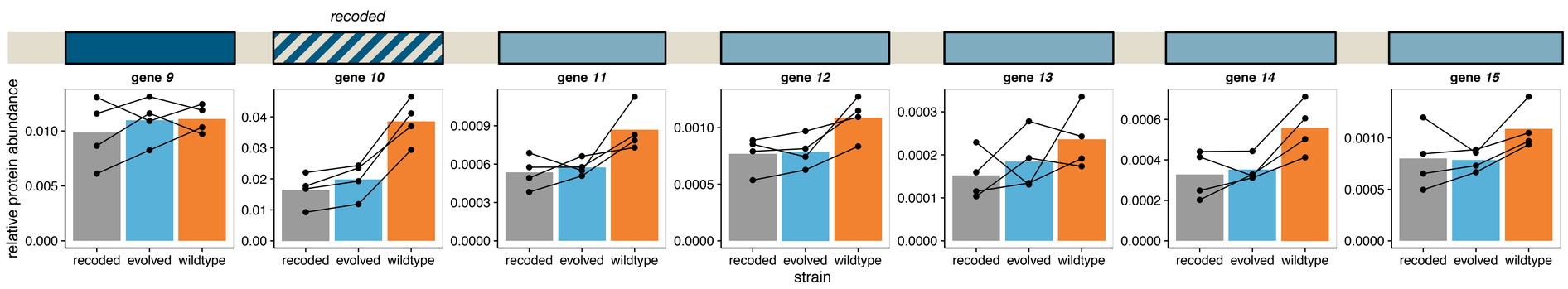
▲ In the recoded strain, protein abundance for capsid protein after 9 minutes of infection is half of that of the wild type ($p < 0.05$, paired *t* test). The evolved strain also has significantly lower levels of capsid protein after 9 minutes. Each point represents a single measurement, and lines connect biological replicates.

T7 transcripts increase rapidly during infection, constituting ~95% of mRNA by 9 minutes after infection **(A)**.▶ Gene *10* transcripts increase over time, consistent with previously reported expression patterns **(B)**. In both **(A)** and **(B)**, there are no detectable strain-specific differences in transcript abundances.

Recoding has no detectable effect on transcript abundances



Codon deoptimization reduces protein abundances of downstream, but not upstream genes



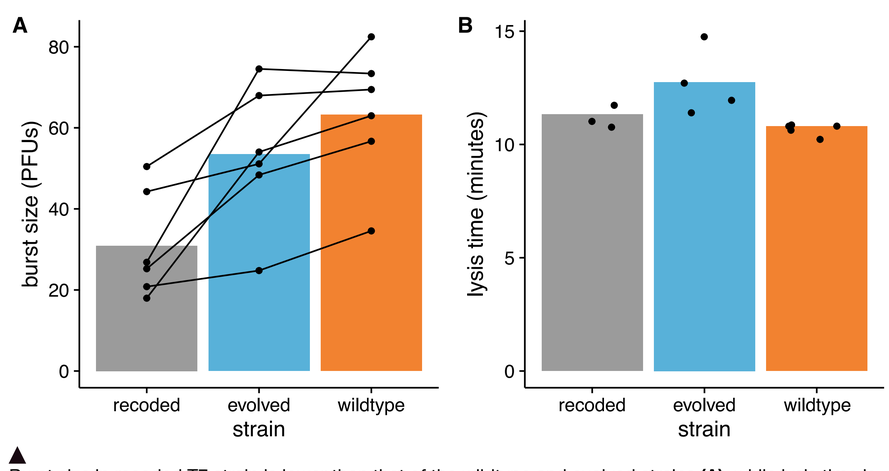
▲ The scaffold protein (9) shows no change in protein abundance from recoding of gene *10*. However, five proteins encoded downstream of *10* exhibit reduced abundances. These proteins are tail tubular proteins (11 and 12), probable virion-associated protein (13), and two internal virion proteins (14 and 15).

The T7 genome produces many polycistronic transcripts

class 1 (early) class 2 (DNA replication) class 3 (virion assembly)

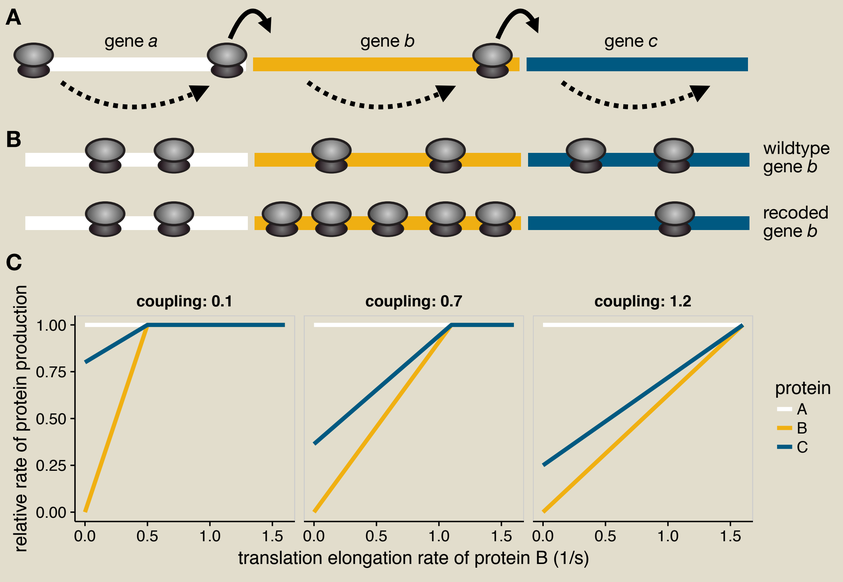
◀ Colored vertical blocks represent genes in the T7 genome, and each class is shown in a different color. Vertical lines represent possible transcripts. Dashed horizontal lines represent RNAse cleavage sites, where *R3.8*, *R4.7*, *R6.5*, and *R18.5* are strong cleavage sites. *(R13)* is a weak RNAse cleavage site. The solid horizontal line represents the terminator *TΦ*. Genes *11* and *12* are only ever expressed as a product of read-through of *TΦ*, indicated by the read-through transcript that extends the length of the genome. Only transcripts containing class III genes are shown. Not all read-through products are shown. Data are from Dunn and Studier [2]. Gene *10* was the gene recoded in our experiments.

Recoding reduces fitness via decreased burst size



▲ Burst size in recoded T7 strain is lower than that of the wildtype and evolved strains **(A)**, while lysis time is indistinguishable for all three strains **(B)**.

Protein abundances support model of translational coupling



When genes are translationally coupled, the translation rate of upstream genes can affect translation rates of downstream genes. We assume three genes (*a*, *b*, and *c*) are expressed in a polycistronic transcript and are translationally coupled, such that the translation initiation rates of one gene depend partially on the rate of ribosomes reaching the stop codon of the previous gene **(A)**. Upon recoding gene *b*, we hypothesize that ribosomes will accumulate on the gene *b* transcript and downstream ribosomal densities will decrease **(B)**. As translation elongation rates increase in *b*, the rate of translation of downstream genes also increases. From left to right, as coupling increases, production of protein C becomes more sensitive to translation elongation rates of gene *b* **(C)**.

Acknowledgements

We thank I. J. Molineux for helpful discussion and comments. This work was supported in part by National Institutes of Health Grants R01 GM088344, National Science Foundation Cooperative agreement no. DBI-0939454 (BEACON Center), and Army Research Office Grant W911NF-12-1-0390. The Texas Advanced Computing Center provided high-performance computing resources. Test tube icon designed by Nimal Raj.

References

1 Bull JJ, Molineux IJ, Wilke CO (2012) Slow fitness recovery in a codon-modified viral genome. Mol Biol Evol 29 : 2997–3004

2 Dunn JJ, Studier FW (1983) Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. J Mol Biol 166 : 477–535

TEXAS
The University of Texas at Austin