

Genomic context and the collateral effects of gene duplication/amplification during bacterial evolution

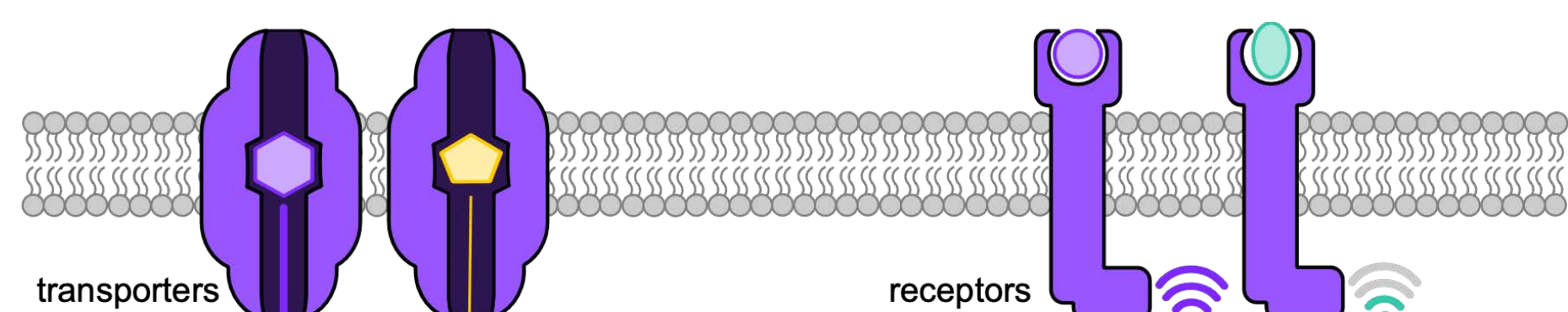


University of Colorado
Boulder

Ryan K. Fritts, Andrew D. Morgenthaler, & Shelley D. Copley

Department of Molecular, Cellular & Developmental Biology, Cooperative Institute for Research in Environmental Sciences,
University of Colorado Boulder, Boulder, CO USA

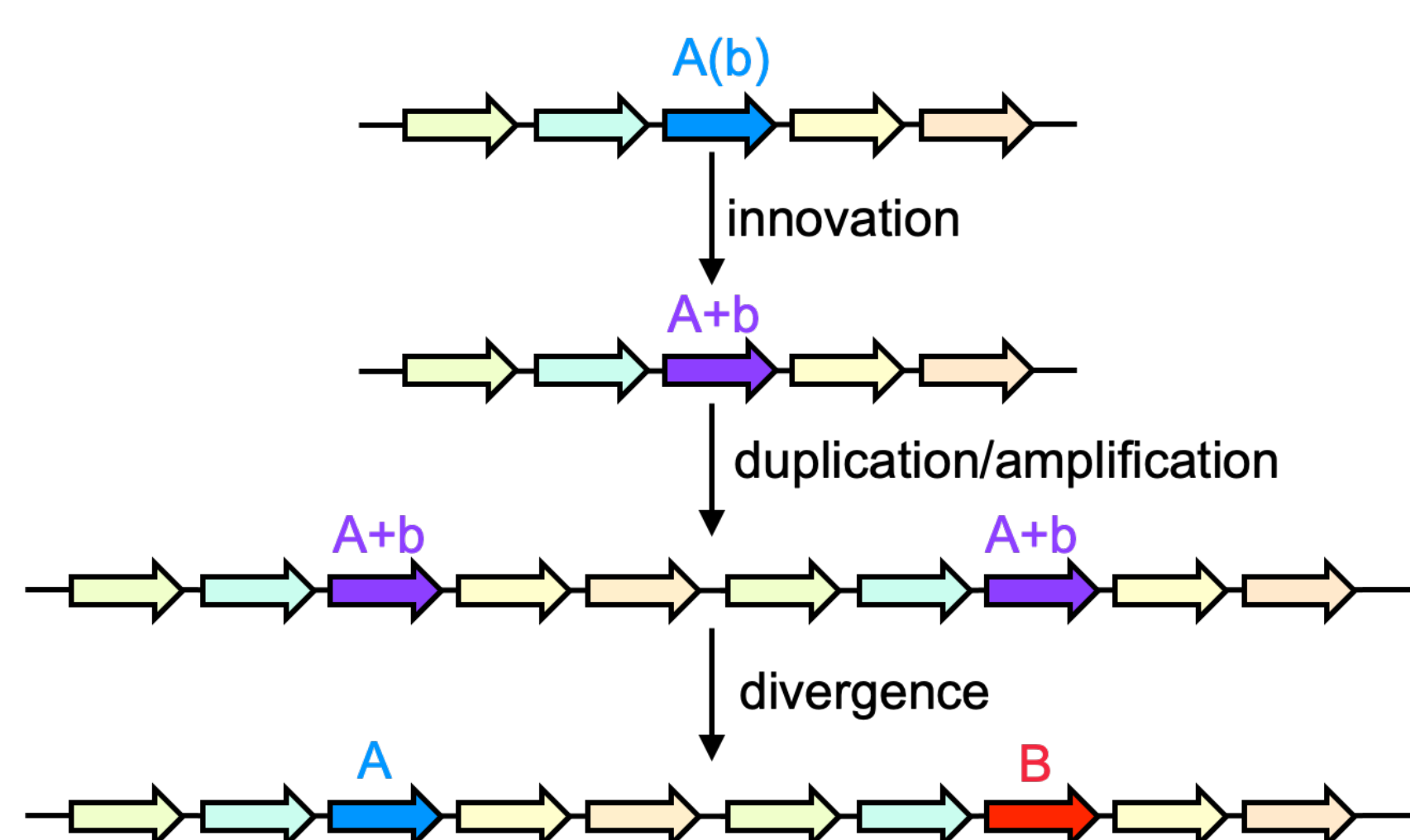
How do new molecular activities evolve?



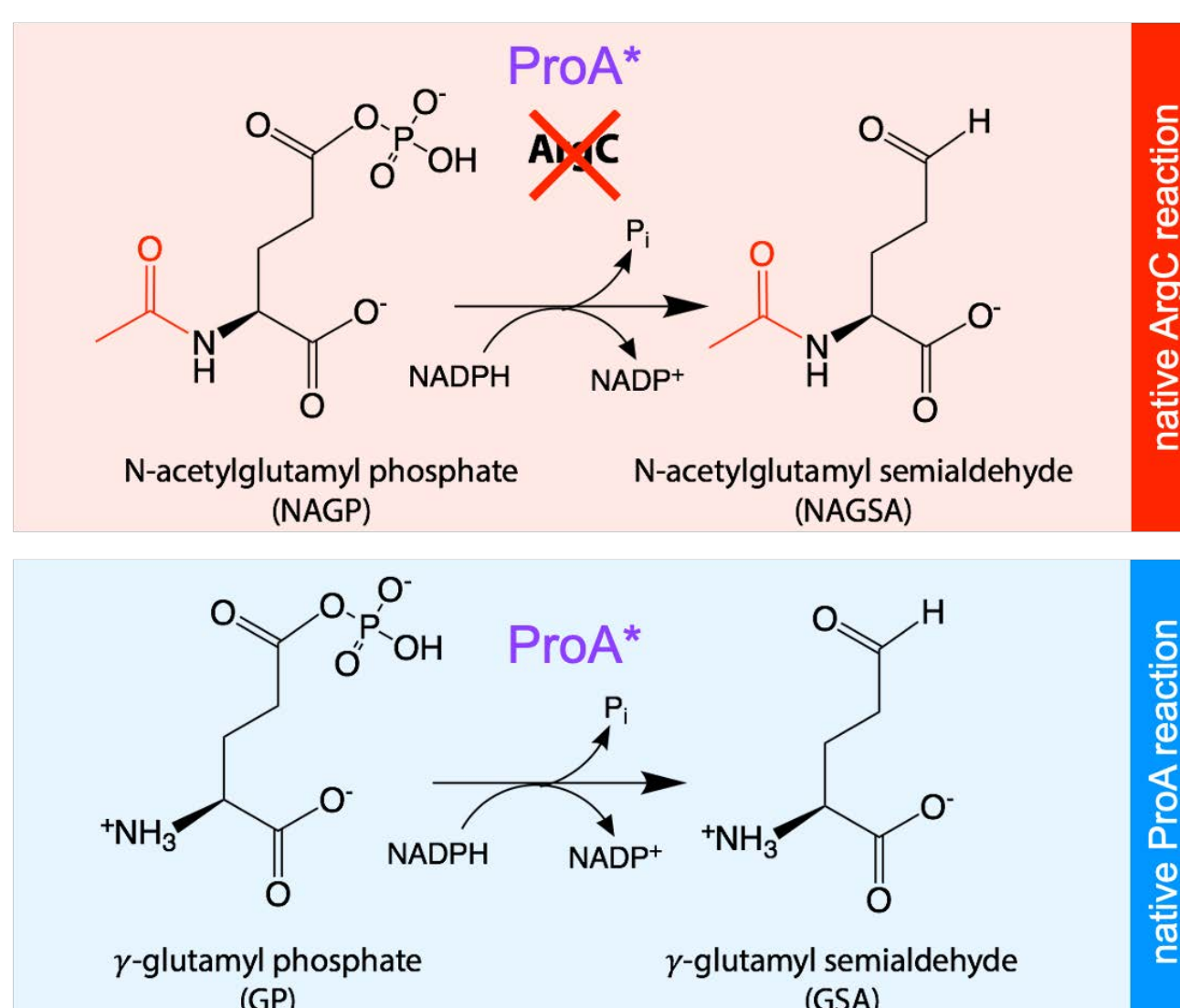
Most proteins have weak side activities

These weak side activities can serve as the basis for evolving a new function

The **Innovation-Amplification-Divergence (IAD)** model of gene evolution
Bergthorsson et al. PNAS. **104**, 17004 (2007); Näsvall et al. Science **338**, 6105 (2012).



A model system to study the dynamics of gene duplication/amplification in *E. coli*



Deletion of *argC* abolishes arginine synthesis

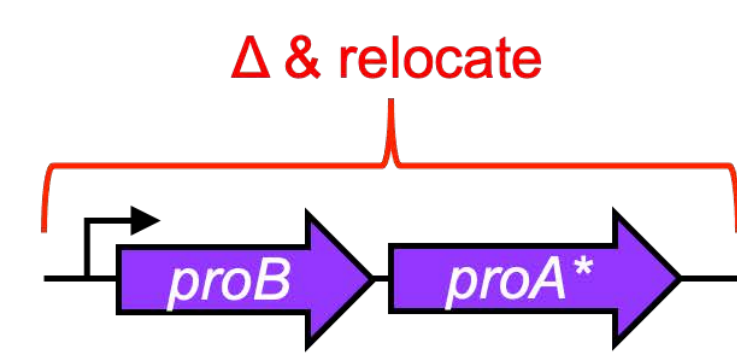
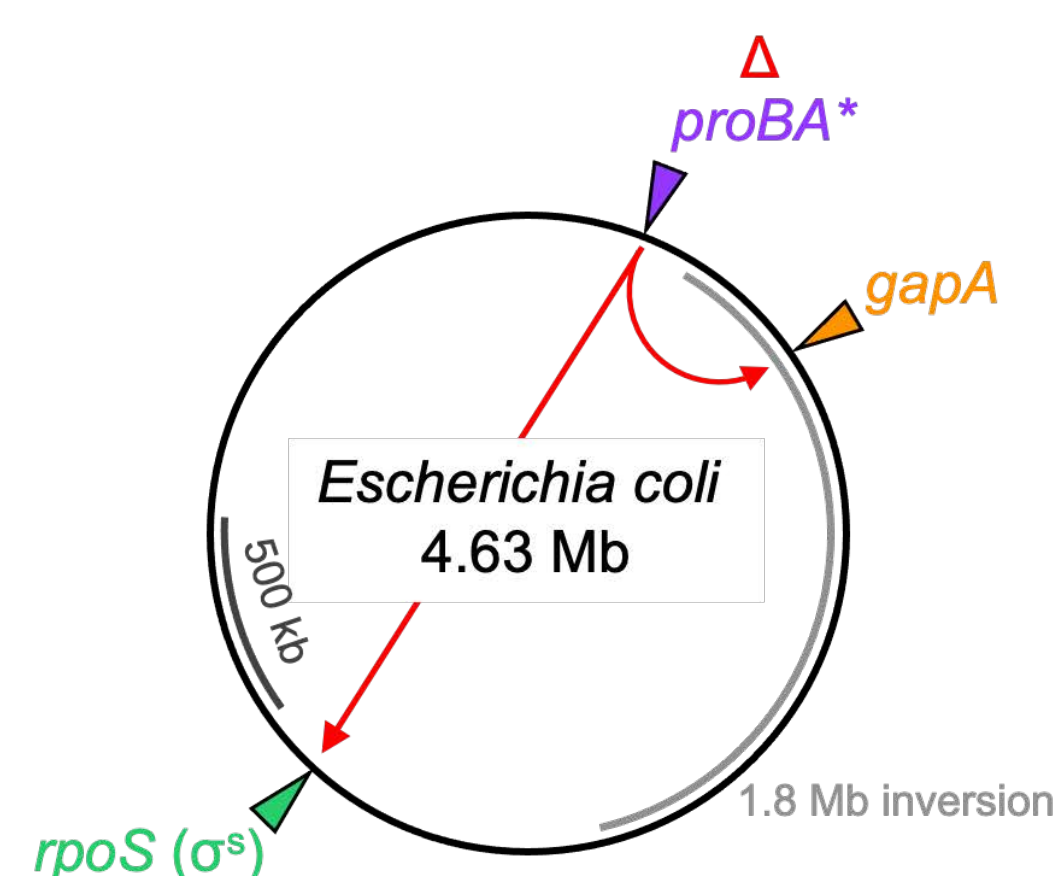
A point mutation in *proA* that causes an E383A substitution (**ProA***) restores arginine synthesis and permits slow growth in minimal medium

The poor catalysis of these two biosynthetic steps by **ProA*** limits growth rate (i.e., fitness)

Relocating the *proBA** operon to new genome sites

Previous work from our lab showed that *proA** typically duplicates/amplifies within 50 generations at its native locus during growth in minimal medium.

Morgenthaler et al. eLife **8**:e53535 (2019); Morgenthaler, Fritts & Copley. Mol. Biol. Evol. **39**, 1 msab289, (2022)

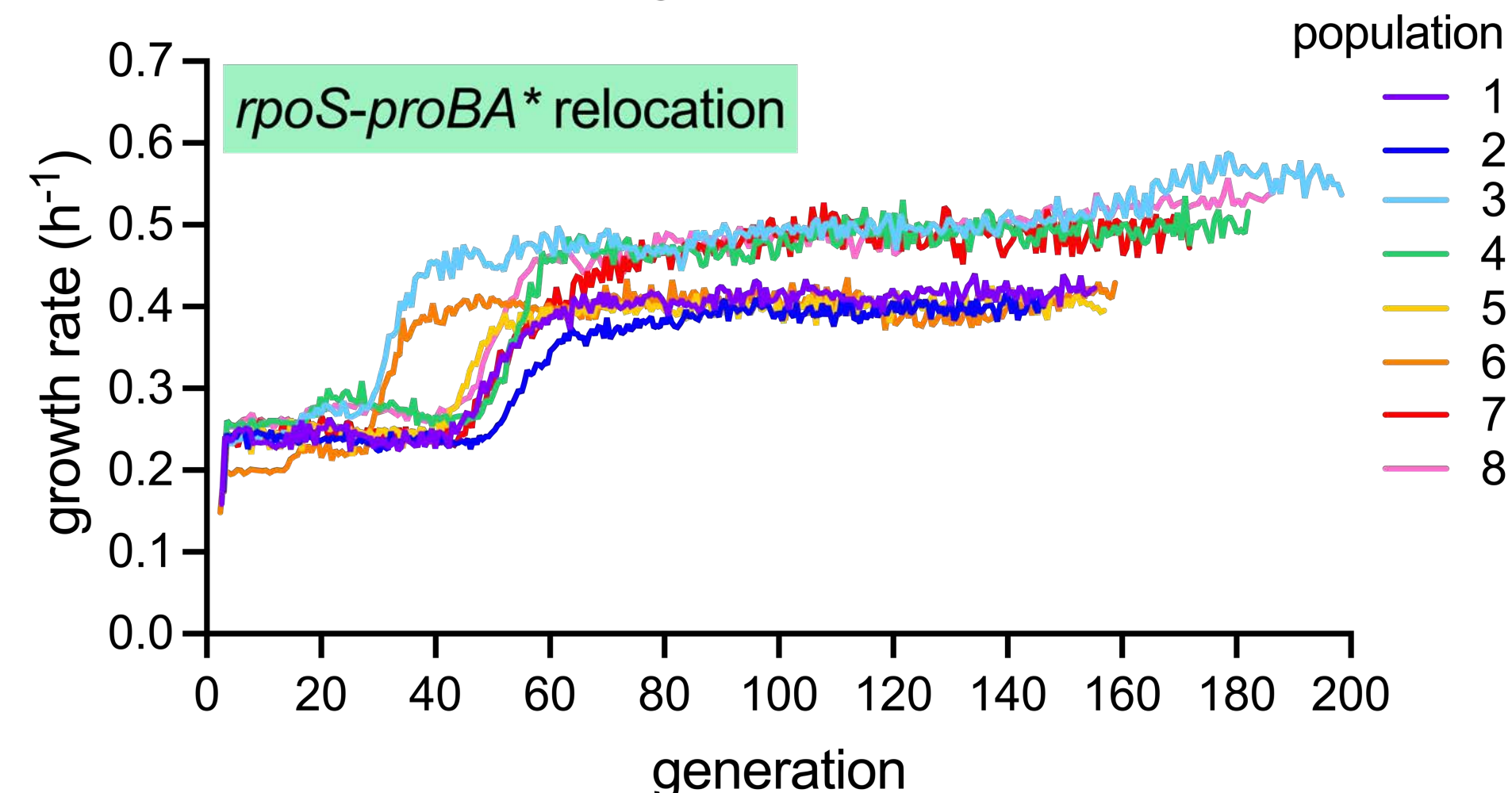
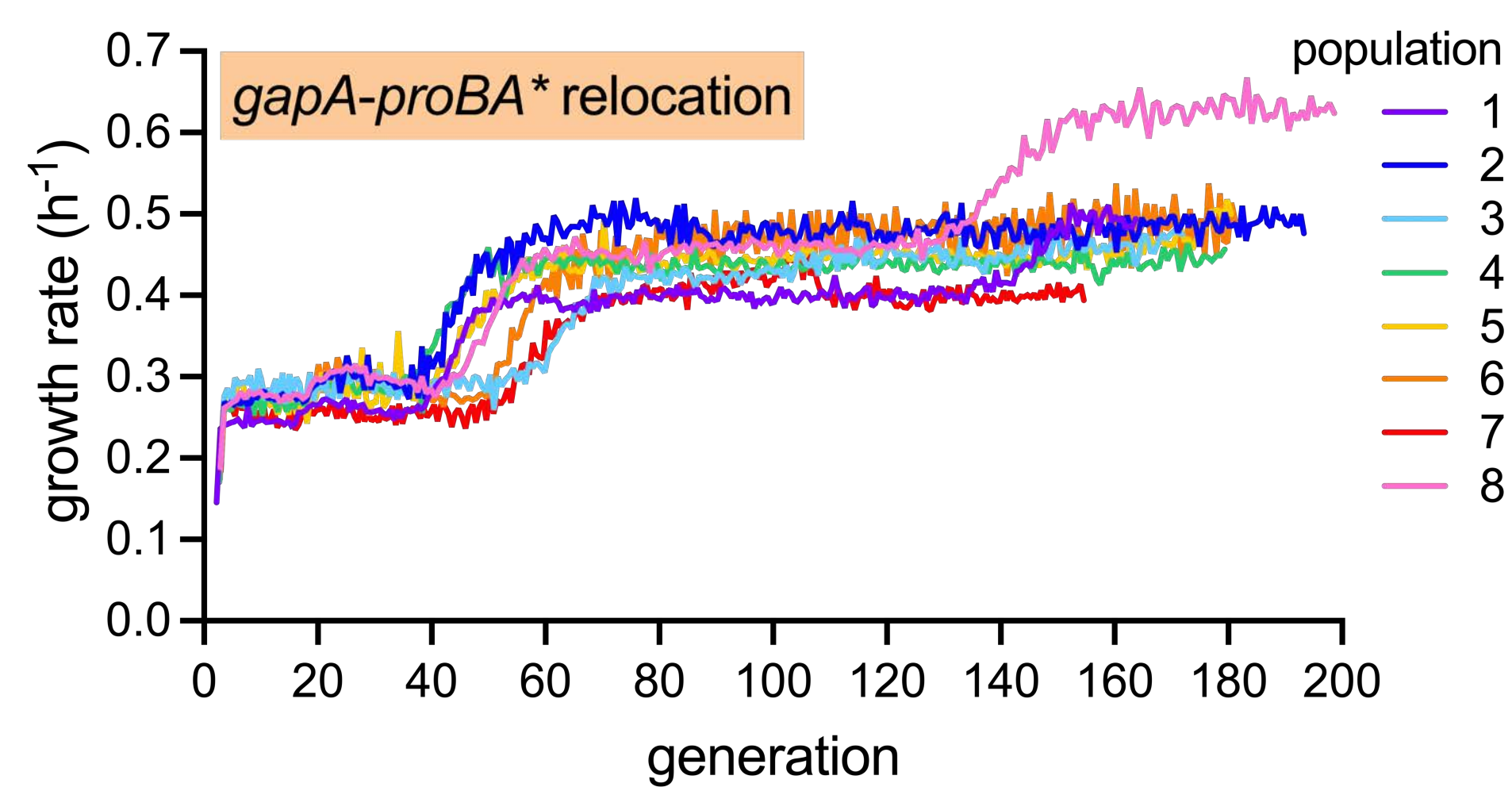
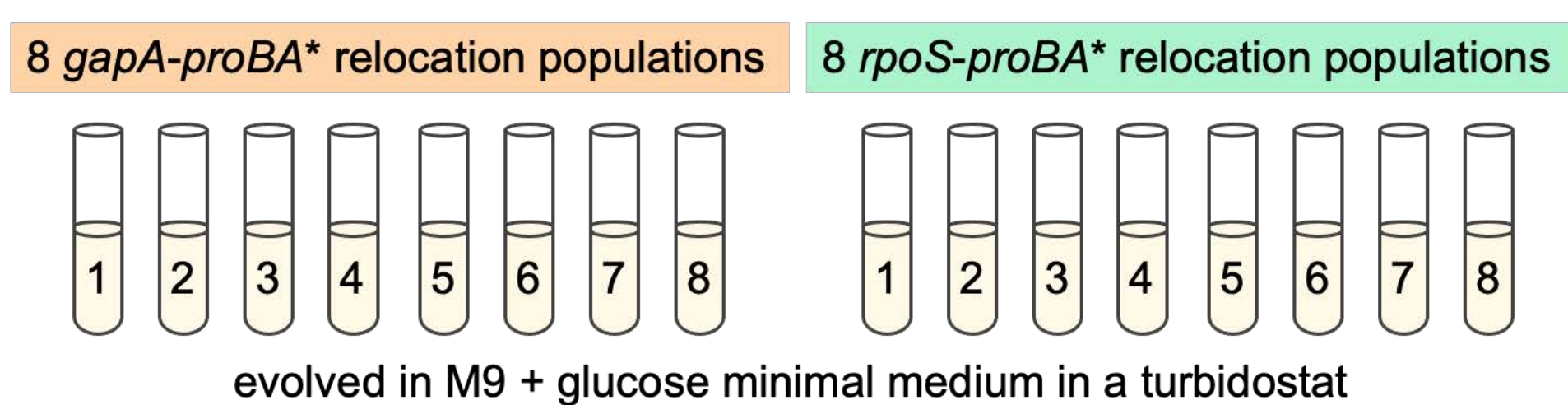


gapA: encodes GAPDH, one of the most highly expressed, essential enzymes

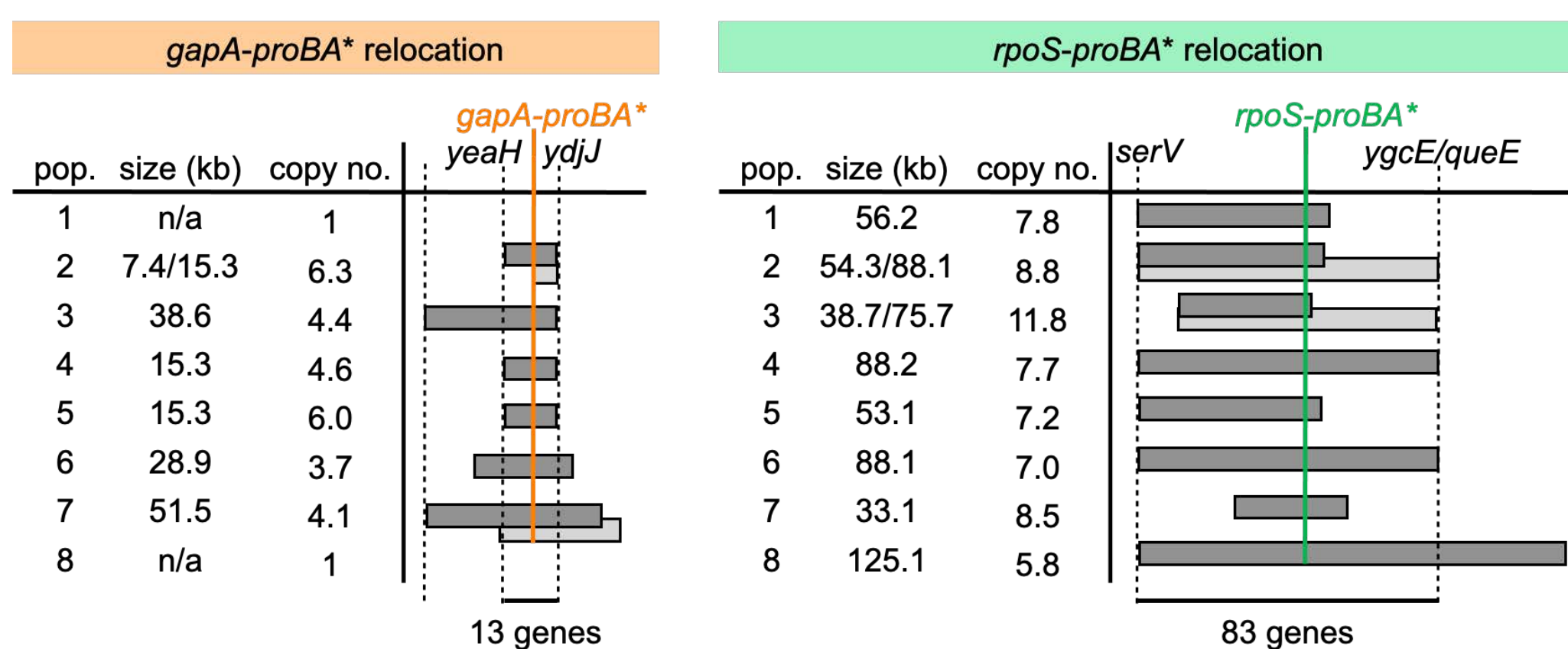
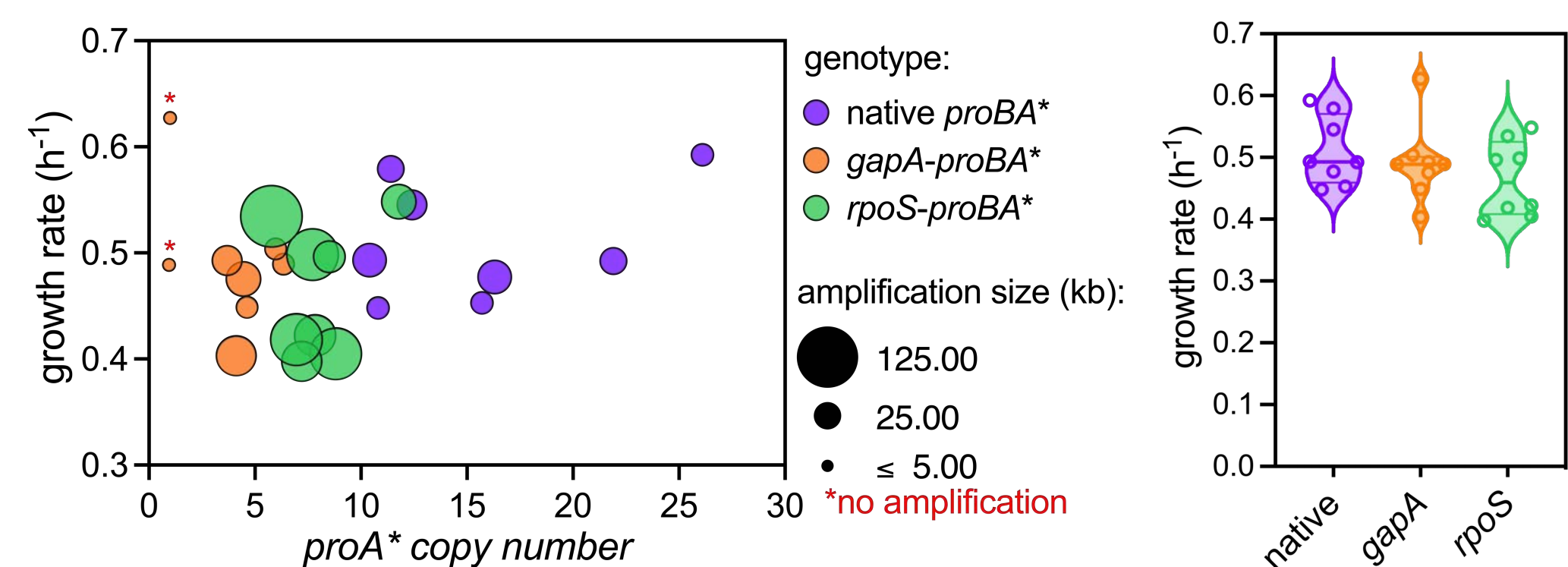
rpoS: encodes RpoS (σ^S), the alternative σ factor for starvation/stationary phase

Relocation of *proBA** next to genes that might perturb physiology and impair fitness if overexpressed following co-amplification.

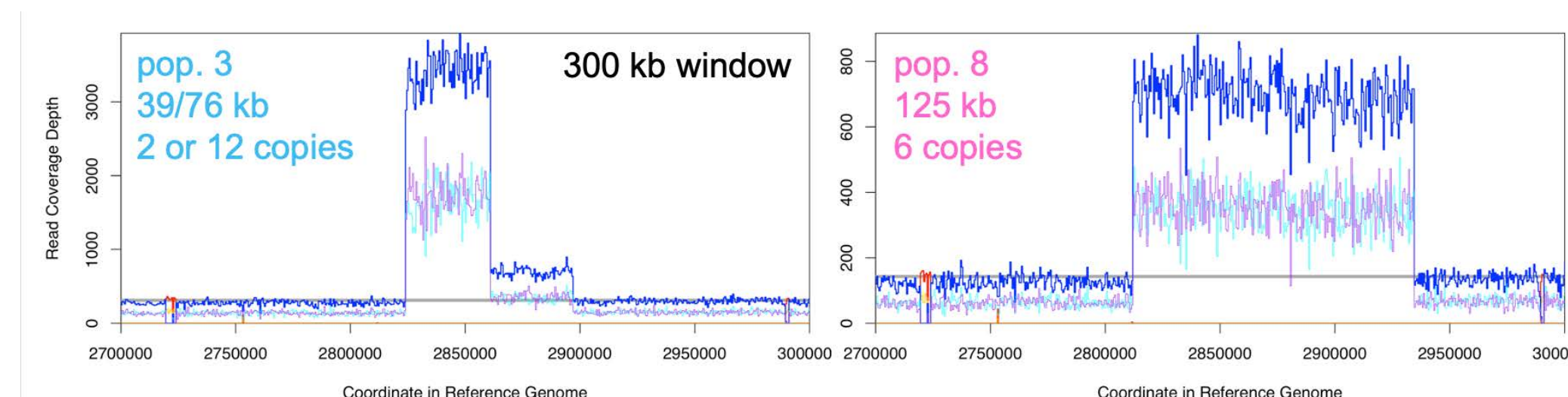
Experimental evolution of *proBA** relocation strains



Growth rate increased within ~60 generations for all relocation populations, in most cases due to *proA** amplification.

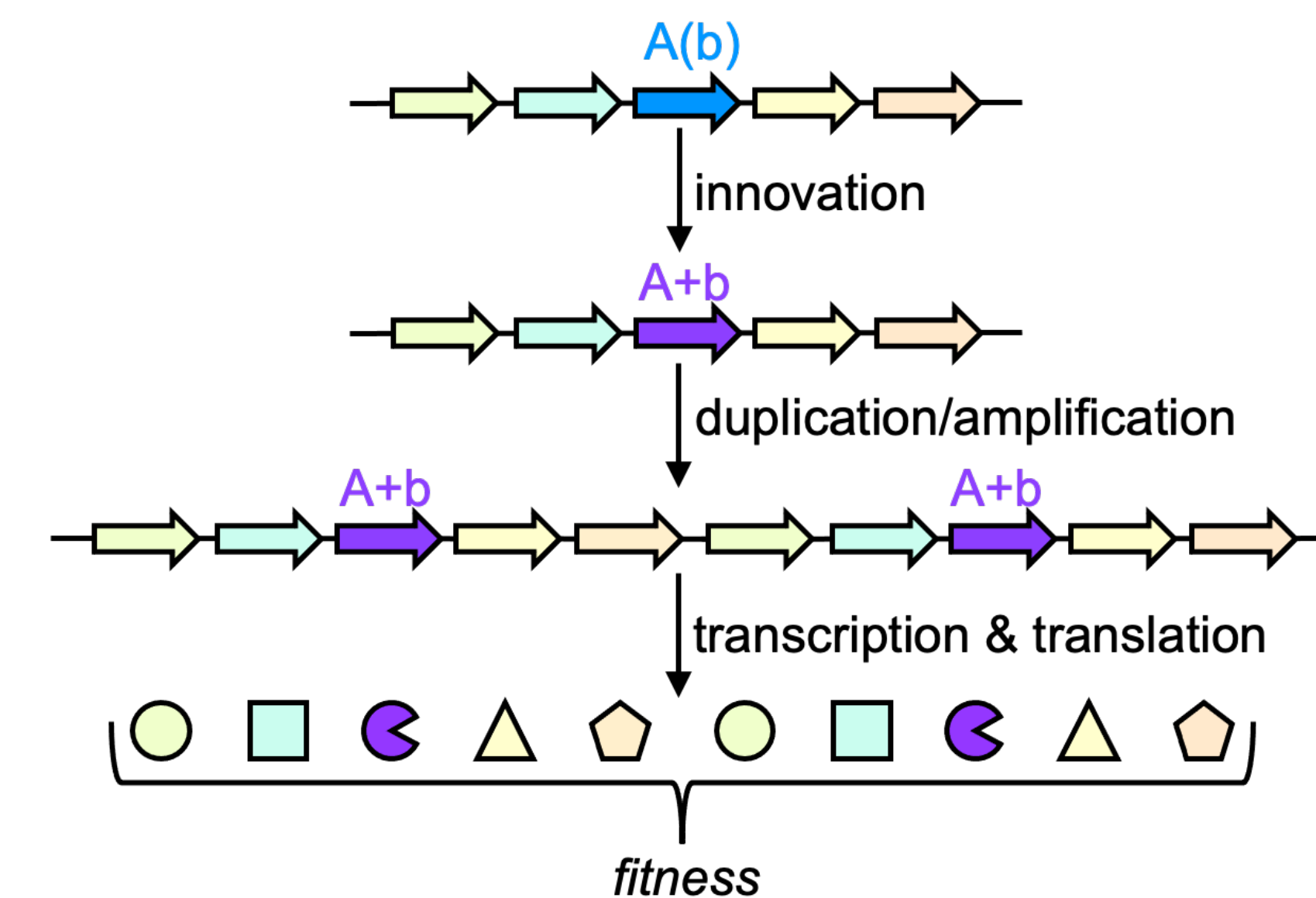


The size and copy number of the amplifications varied across loci, despite each genotype reaching similar mean growth rates. Fitness is not solely determined by *proA** copy number.



RNAseq and label-free proteomics were performed on isolates from *rpoS-proBA** populations 3 and 8, which have co-amplified genes present at 2, 6, or 12 copies. Population 3 has a nested amplification inside a larger duplication.

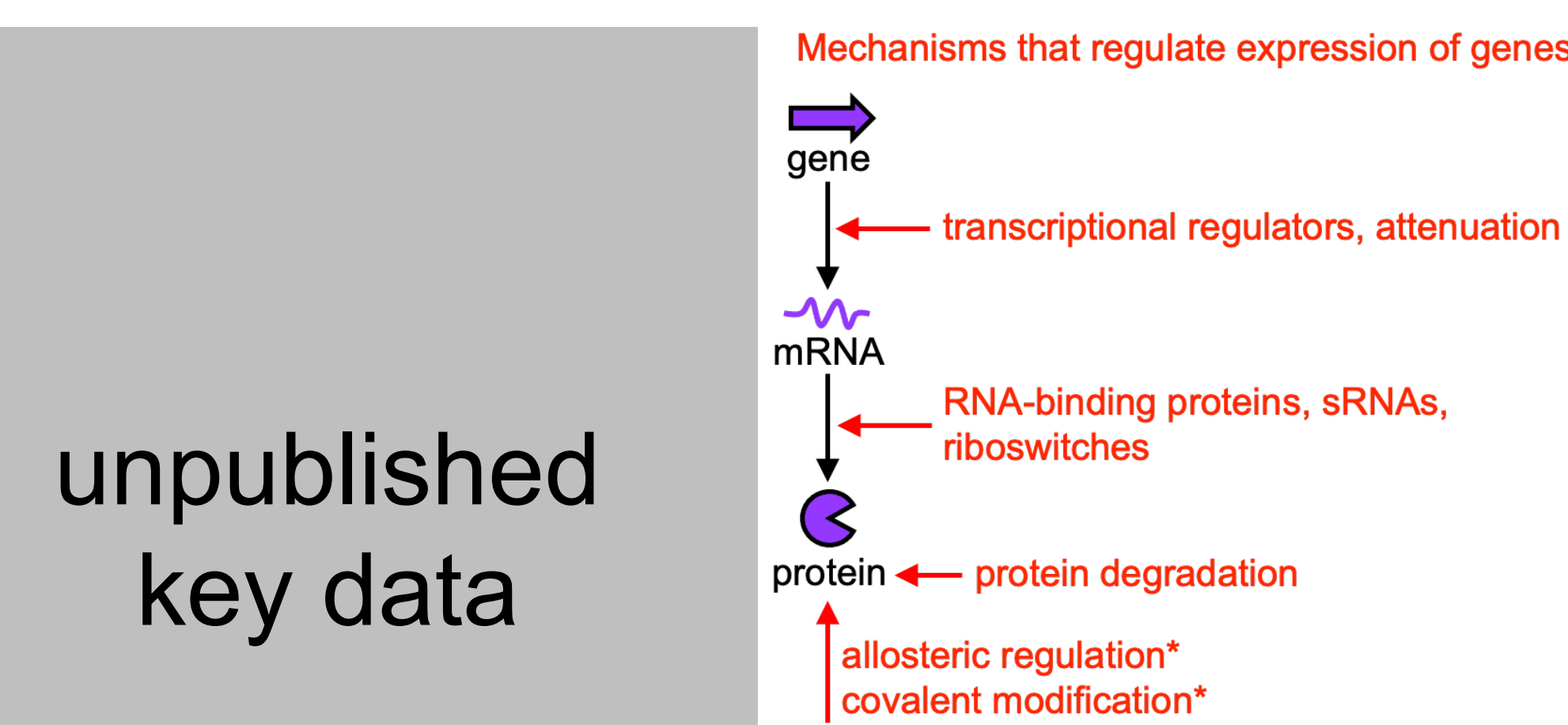
Transcriptomic and proteomic ramifications of gene amplification



The fitness gain derived from an amplification will depend on the benefit of overexpressing the gene under selection and the cost of overexpressing any co-amplified neighboring genes.

unpublished key data

What mechanisms dampen or heighten the expression of co-amplified genes?



unpublished key data

unpublished key data

Acknowledgements & Funding

Thanks to Chris Ebmeier at the Proteomics and Mass Spectrometry Core Facility at CU Boulder for help with proteomics data acquisition and analysis. This work was supported by an NIH NIGMS NRSA F32 fellowship (5F32GM143927) to RKF and an NIH NIGMS R01 grant (5R01GM134044) to SDC

