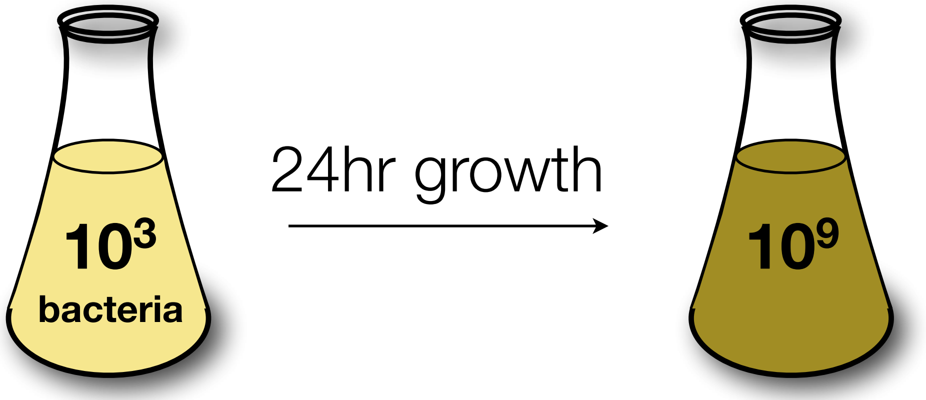
The evolution of bacterial gene regulation

Olin Silander
Institute of Natural and Mathematical Sciences
Massey University

How do novel functions evolve?

Rich media

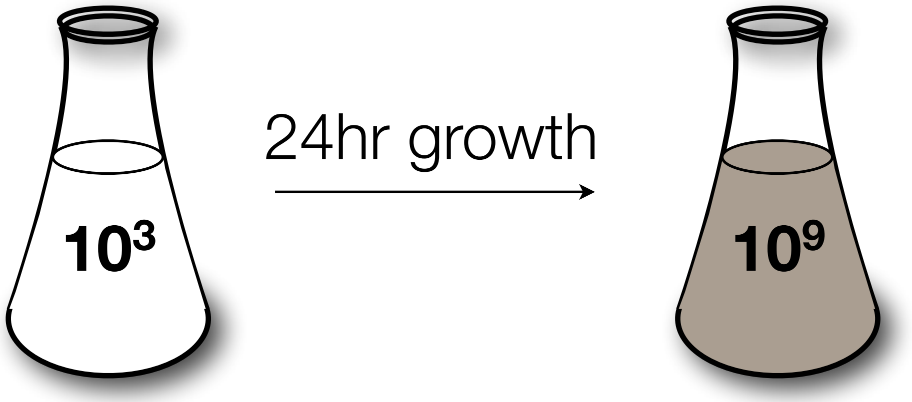
- All amino acids
- Carbohydrates
- All vitamins
- Trace elements



glucose metabolism and synthesis of building blocks **is not** necessary

Minimal media

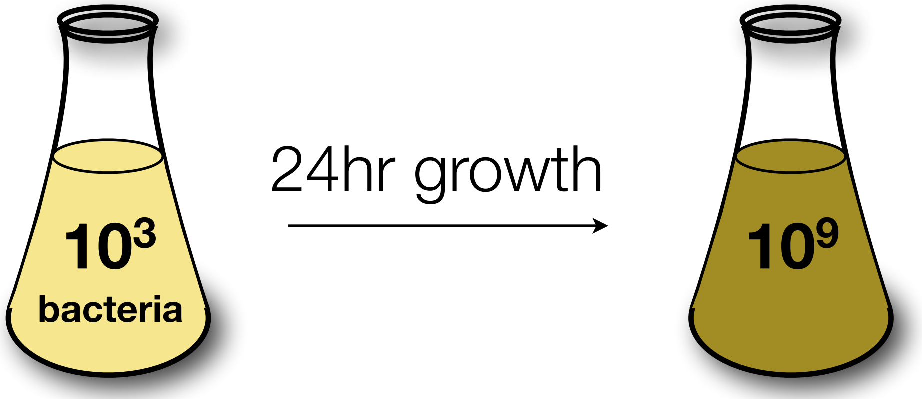
- Carbon (glucose)
- Nitrogen (NH_4Cl)
- Sulfur ($MgSO_4$)
- Phosphorous (Na_2HPO_4, KH_2PO_4)
- Trace elements



glucose metabolism and synthesis of **all** building blocks **is** necessary

Rich media

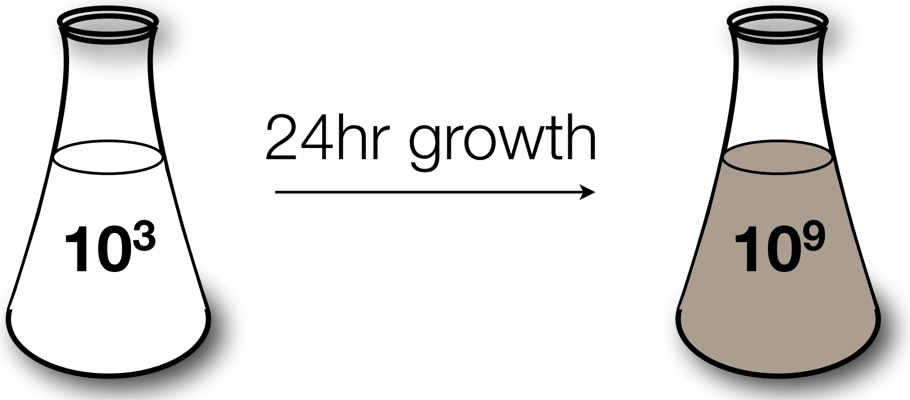
- All amino acids
- Carbohydrates
- All vitamins
- Trace elements



glucose metabolism and synthesis of building blocks **is not** necessary

Minimal media

- Carbon (glucose)
- Nitrogen (NH_4Cl)
- Sulfur ($MgSO_4$)
- Phosphorous (Na_2HPO_4, KH_2PO_4)
- Trace elements

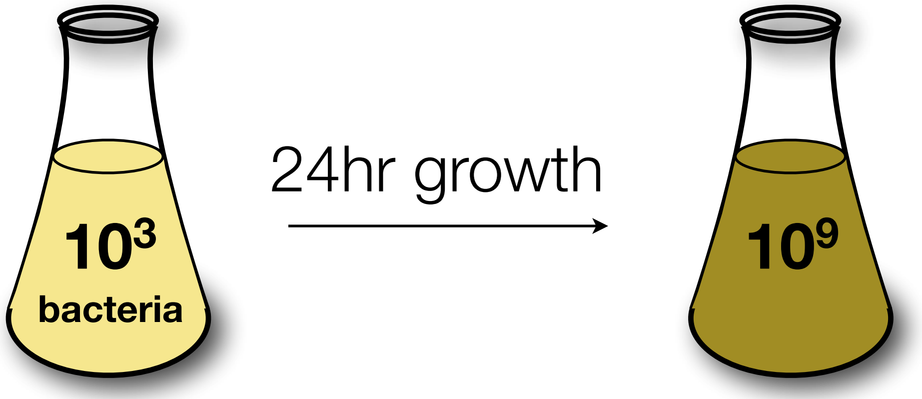


glucose metabolism and synthesis of **all** building blocks **is** necessary

remove a function involved in carbon metabolism or building block biosynthesis

Rich media

- All amino acids
- Carbohydrates
- All vitamins
- Trace elements



10^3 **bacteria**

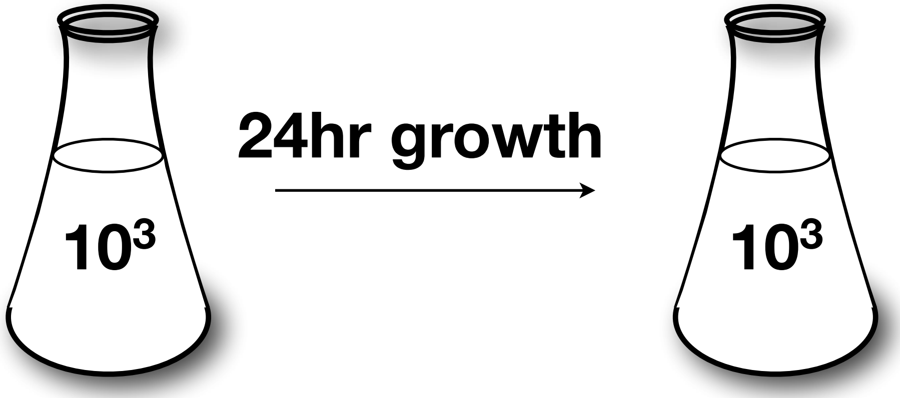
24hr growth

10^9

glucose metabolism and synthesis of building blocks **is not** necessary

Minimal media

- Carbon (glucose)
- Nitrogen (NH_4Cl)
- Sulfur ($MgSO_4$)
- Phosphorous (Na_2HPO_4, KH_2PO_4)
- Trace elements



10^3

24hr growth

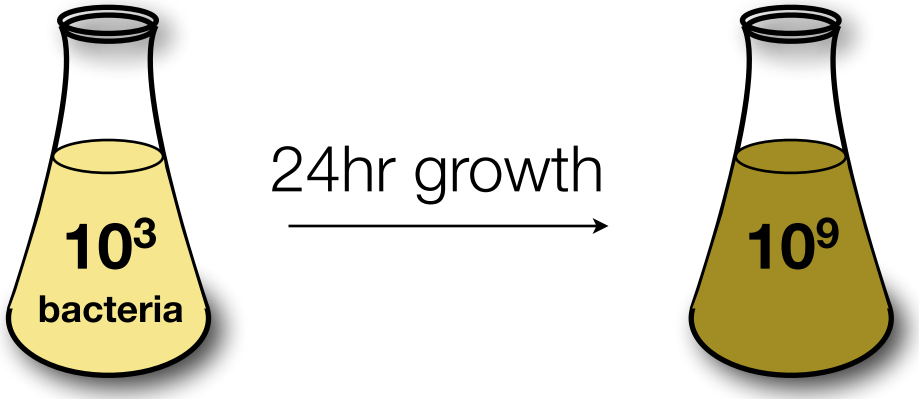
10^3

glucose metabolism and synthesis of **all** building blocks **is** necessary

remove a function involved in carbon metabolism or building block biosynthesis

Rich media

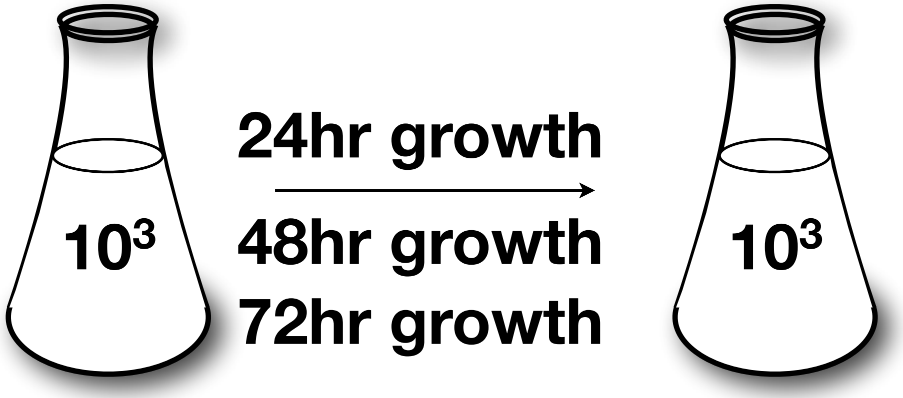
- All amino acids
- Carbohydrates
- All vitamins
- Trace elements



glucose metabolism and synthesis
of building blocks **is not** necessary

Minimal media

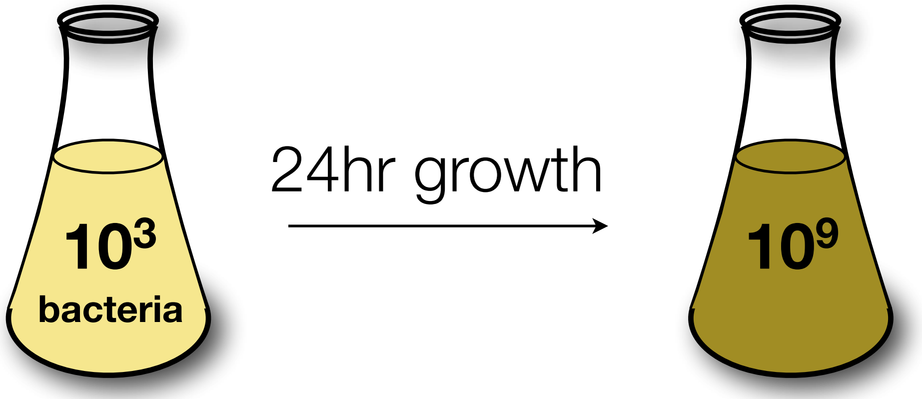
- Carbon (glucose)
- Nitrogen (NH_4Cl)
- Sulfur ($MgSO_4$)
- Phosphorous (Na_2HPO_4, KH_2PO_4)
- Trace elements



glucose metabolism and synthesis
of **all** building blocks **is** necessary

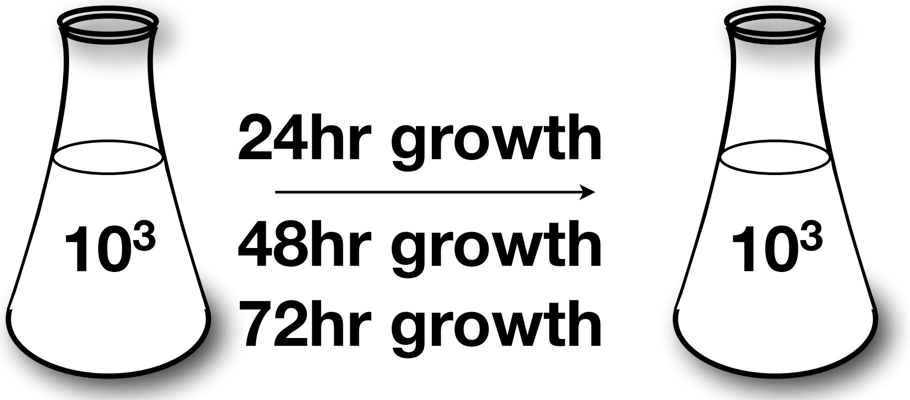
remove a function involved in carbon metabolism or building block biosynthesis

Rich media
- All amino acids
- Carbohydrates
- All vitamins
- Trace elements



glucose metabolism and synthesis
of building blocks **is not** necessary

Minimal media
- Carbon (glucose)
- Nitrogen (NH_4Cl)
- Sulfur ($MgSO_4$)
- Phosphorous (Na_2HPO_4, KH_2PO_4)
- Trace elements



glucose metabolism and synthesis
of **all** building blocks **is** necessary

remove a **gene** that is conditionally
essential in minimal (M9) glucose media

Can this function re-evolve?

How de we know if a gene is conditionally essential in M9 glucose?

How de we know if a gene is
conditionally essential in M9 glucose?

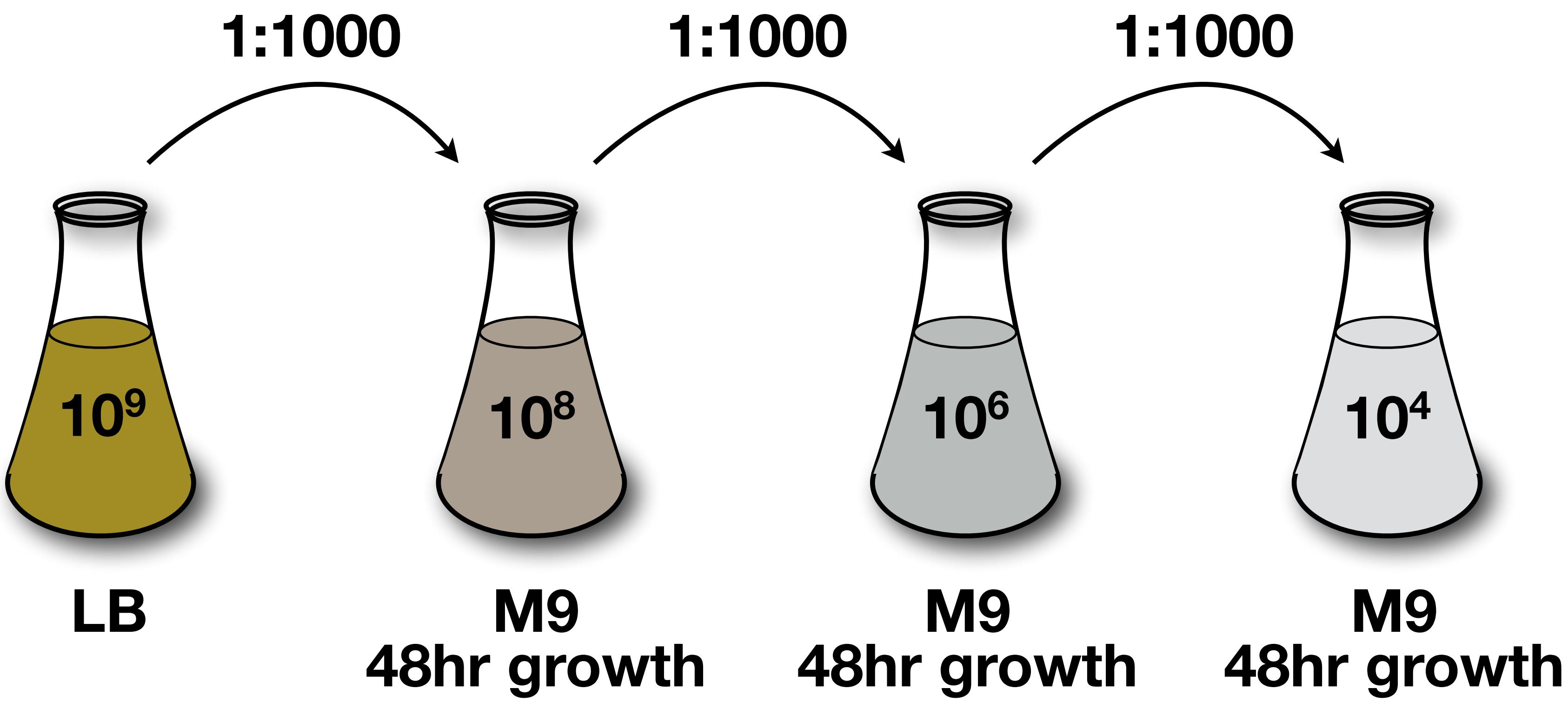
Hypothesis: there is an “underground” metabolism
of enzymes that can perform the functions of
other enzymes, but much less efficiently

Is the function of this enzyme enough to allow growth?

How de we know if a gene is
conditionally essential in M9 glucose?

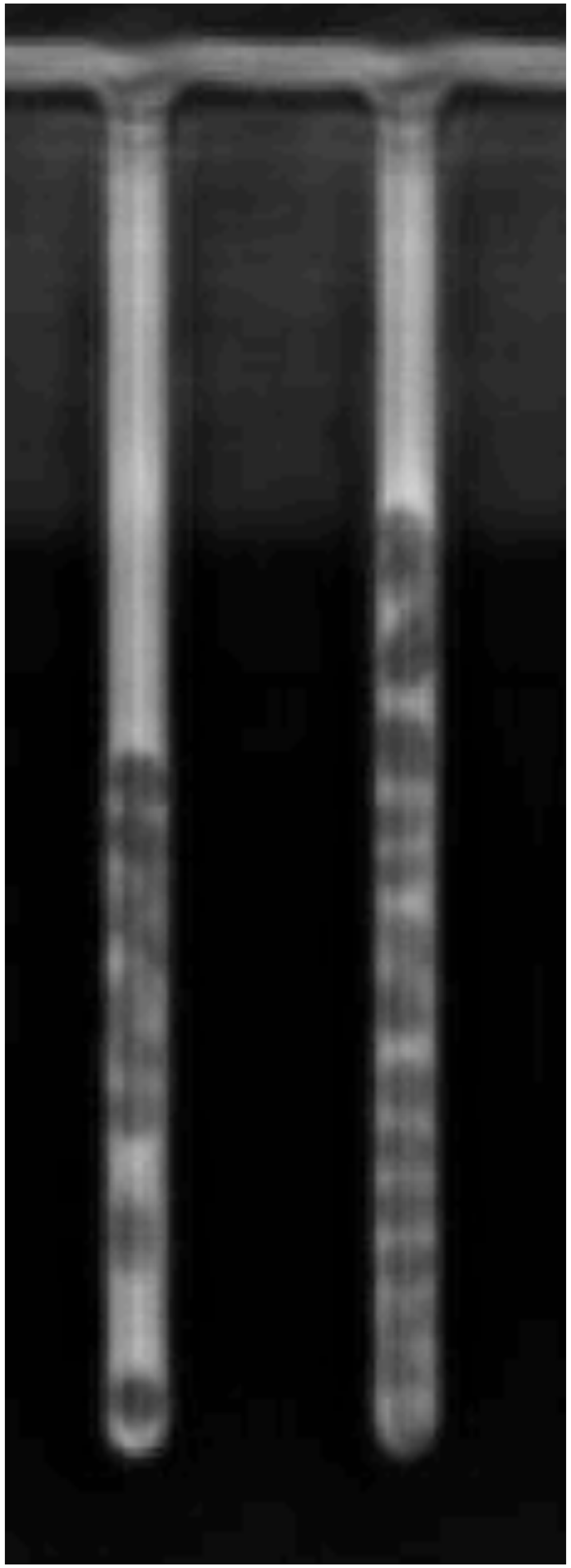
Hypothesis: there is an “underground” metabolism of enzymes that can perform the functions of other enzymes, but much less efficiently

How close are we to the limit of no growth?

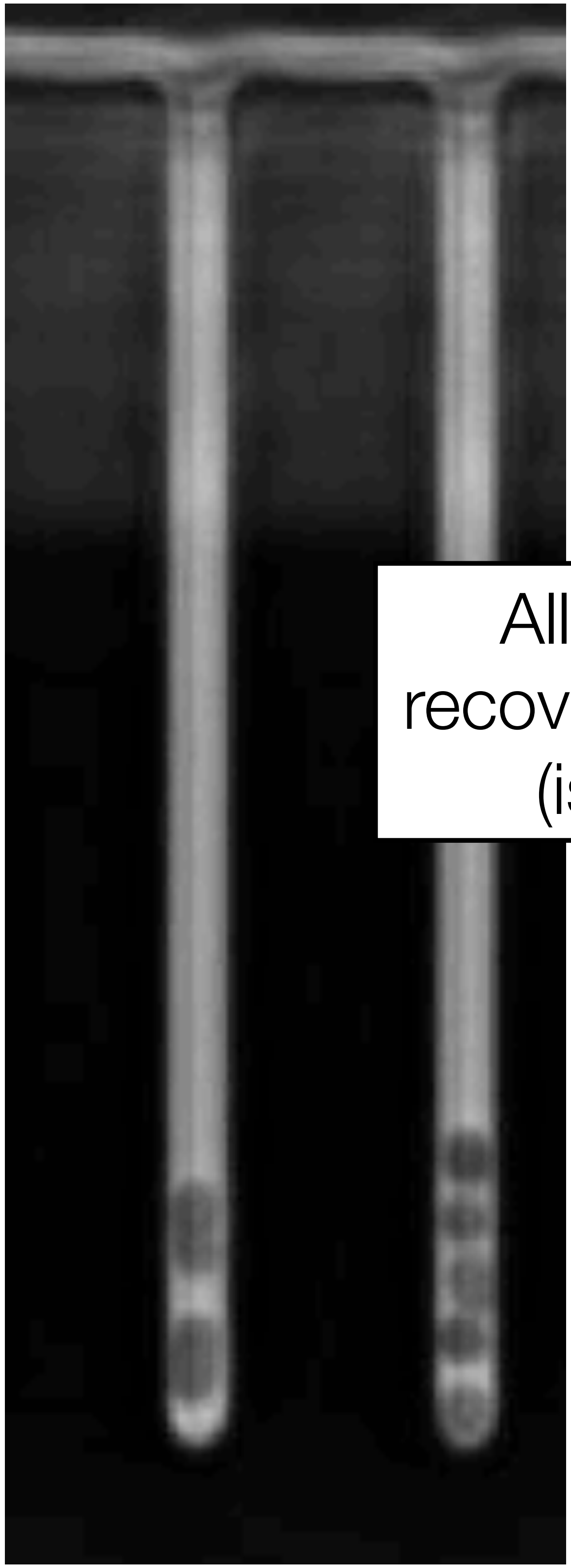


ppc - phosphoenolpyruvate carboxylase

Is a gene conditionally essential in M9 glucose?

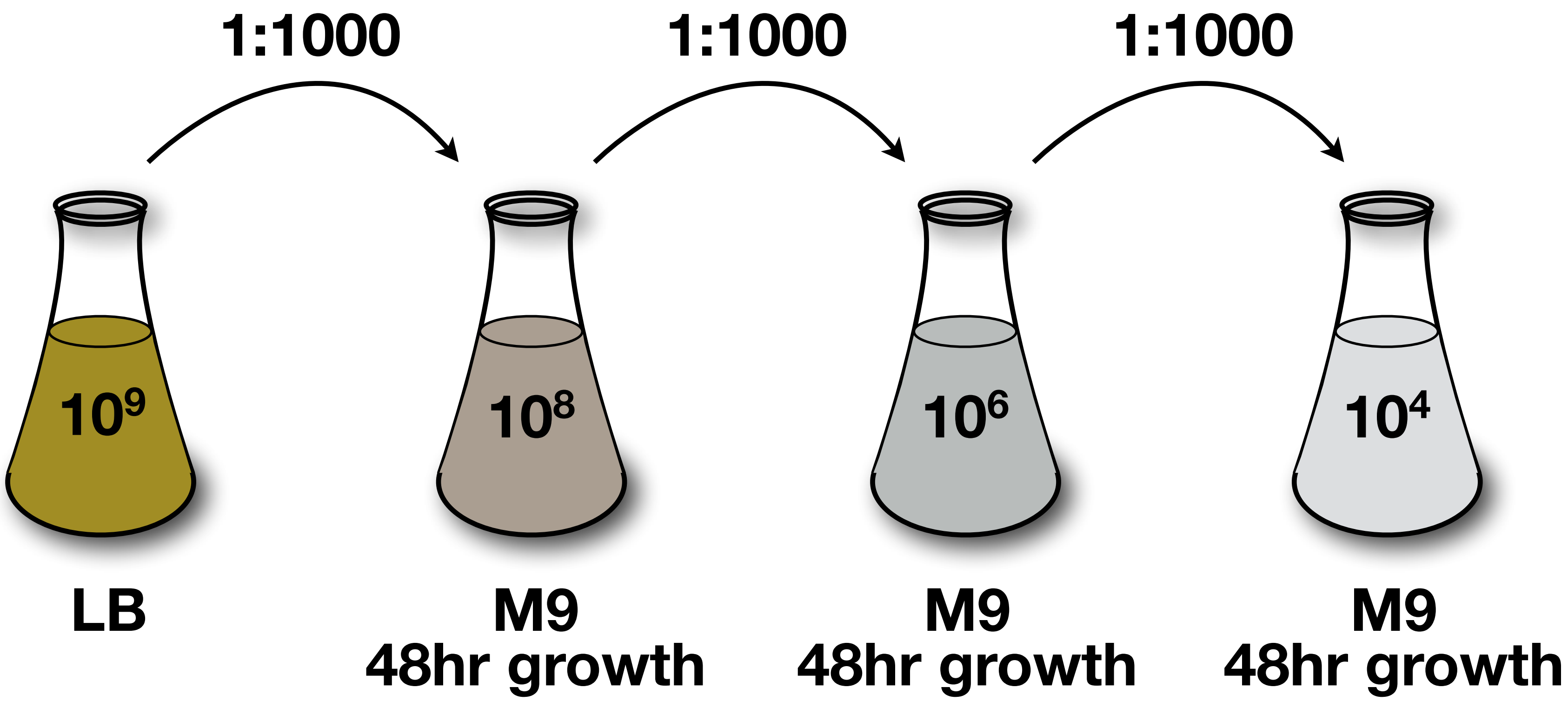


Is a gene conditionally essential in M9 glucose?

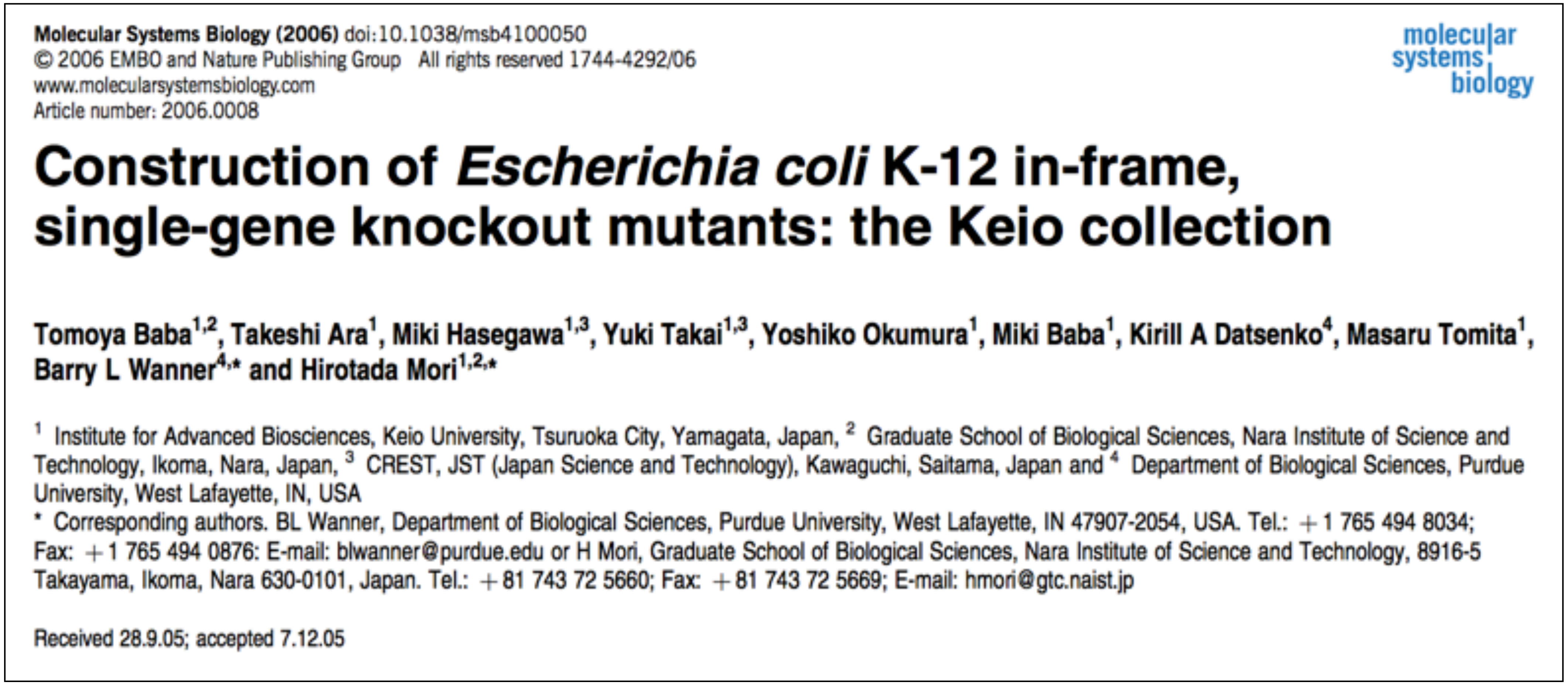


All lineages with this deletion recovered, all had a mutation in *icd* (isocitrate dehydrogenase)

A gene is “essential” if no *sustained* growth is observed in the deletion strain



A gene is “essential” if no *sustained* growth is observed in the deletion strain



Molecular Systems Biology (2006) doi:10.1038/msb4100050
© 2006 EMBO and Nature Publishing Group All rights reserved 1744-4292/06
www.molecularsystemsbiology.com
Article number: 2006.0008

Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection

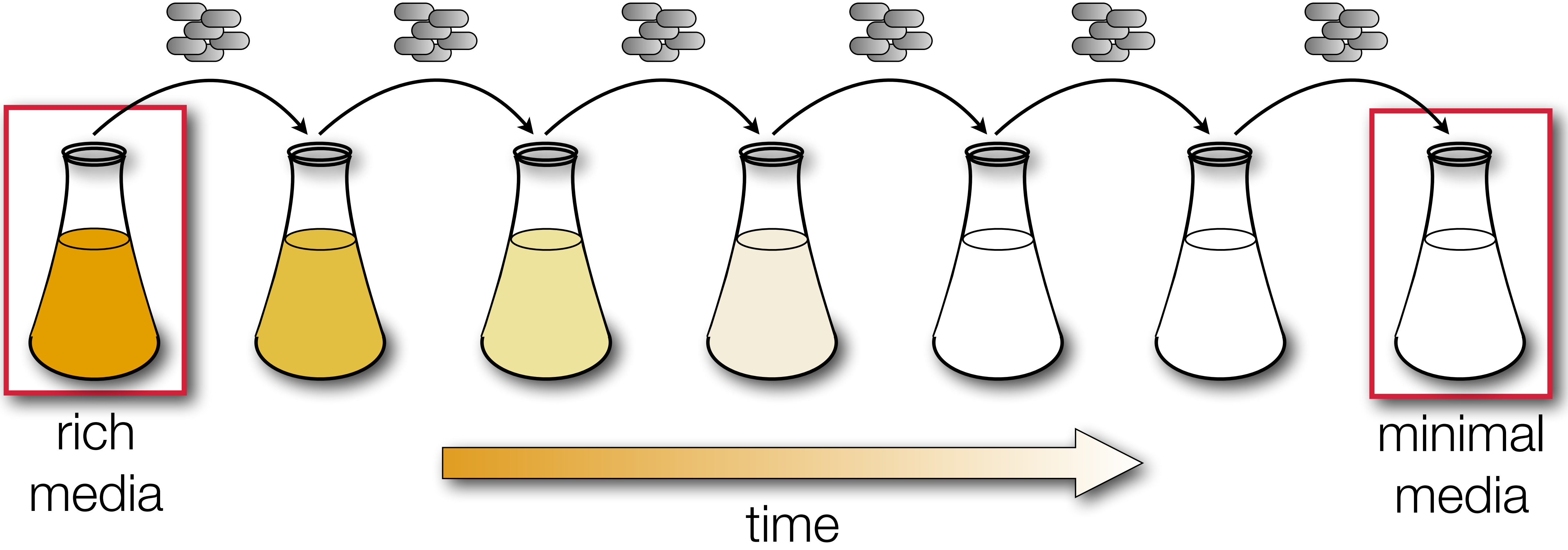
Tomoya Baba[1,2], Takeshi Ara[1], Miki Hasegawa[1,3], Yuki Takai[1,3], Yoshiko Okumura[1], Miki Baba[1], Kirill A Datsenko[4], Masaru Tomita[1], Barry L Wanner[4,*] and Hirotada Mori[1,2,*]

[1] Institute for Advanced Biosciences, Keio University, Tsuruoka City, Yamagata, Japan, [2] Graduate School of Biological Sciences, Nara Institute of Science and Technology, Ikoma, Nara, Japan, [3] CREST, JST (Japan Science and Technology), Kawaguchi, Saitama, Japan and [4] Department of Biological Sciences, Purdue University, West Lafayette, IN, USA
[*] Corresponding authors. BL Wanner, Department of Biological Sciences, Purdue University, West Lafayette, IN 47907-2054, USA. Tel.: +1 765 494 8034; Fax: +1 765 494 0876: E-mail: blwanner@purdue.edu or H Mori, Graduate School of Biological Sciences, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0101, Japan. Tel.: +81 743 72 5660; Fax: +81 743 72 5669; E-mail: hmori@gtc.naist.jp

Received 28.9.05; accepted 7.12.05

We selected **87 different deletion strains**
from the Keio collection

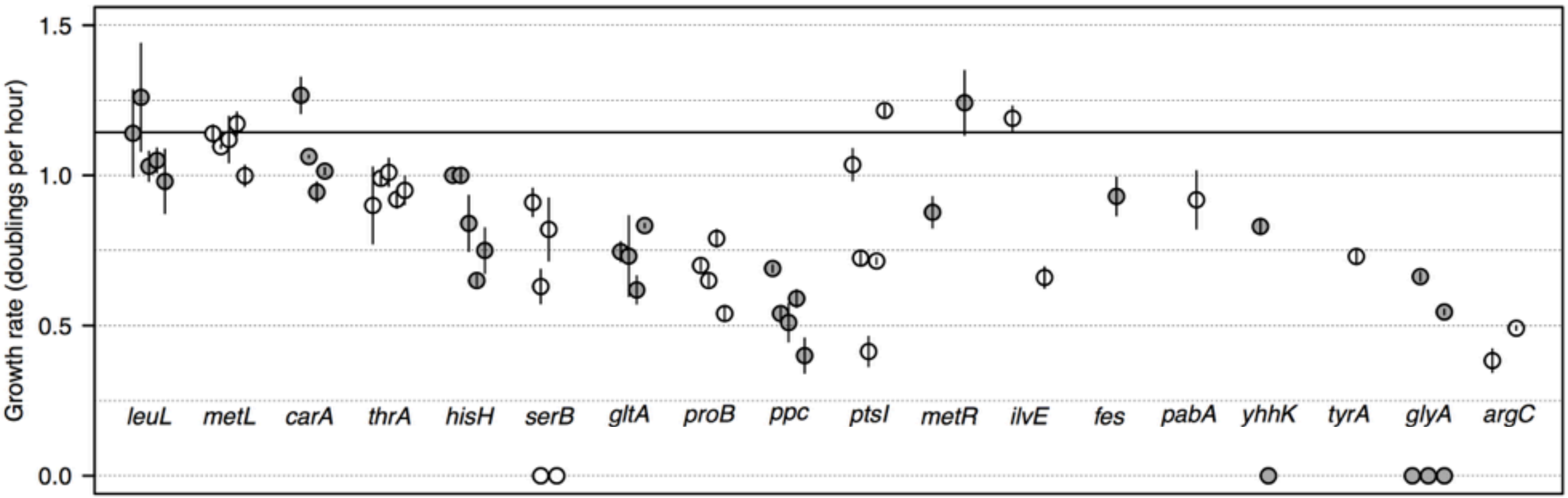
Evolution starts through slow dilution into minimal glucose media



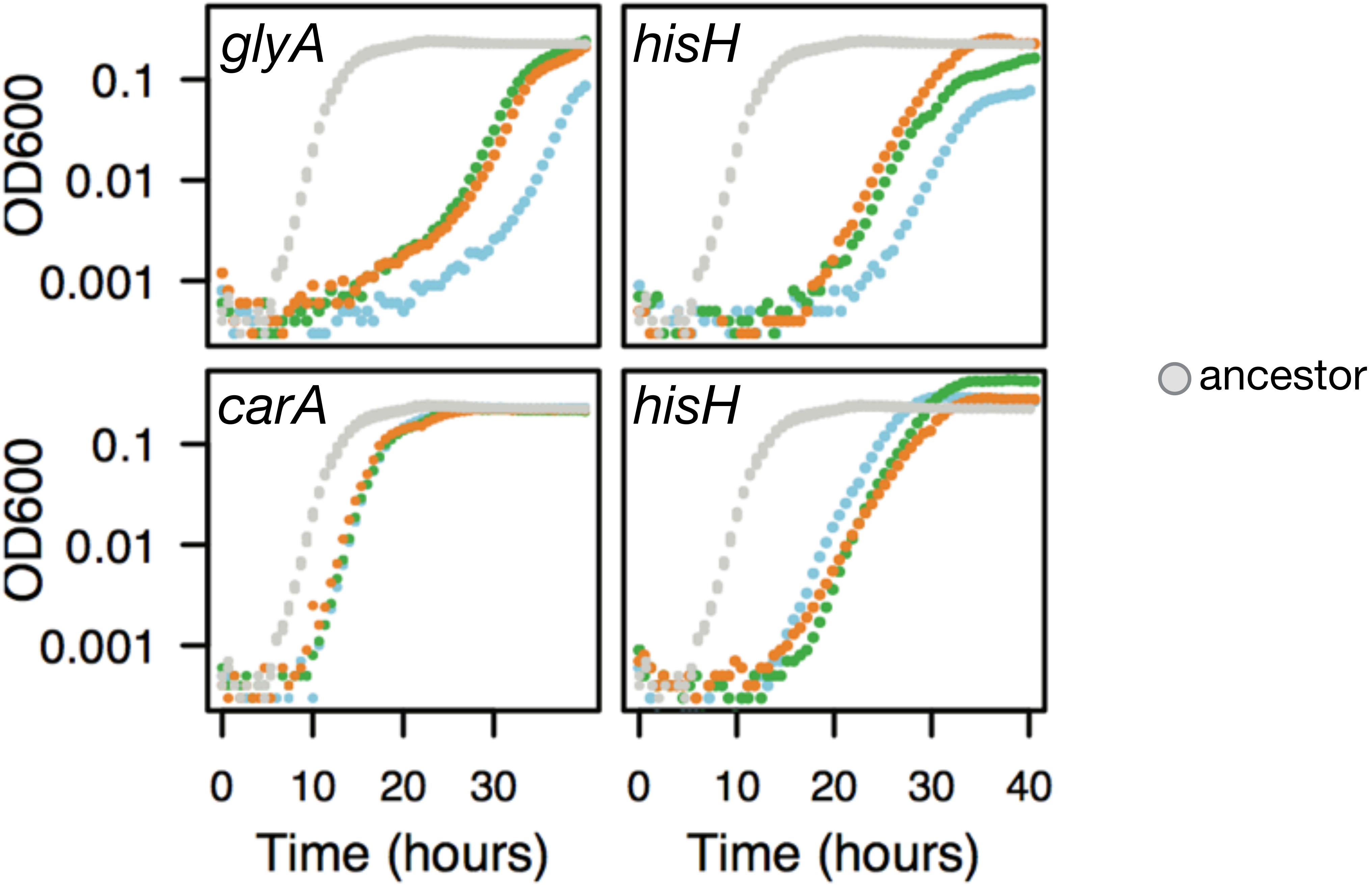
- 5 replicate evolving lineages for each deletion genotype
- 435 lineages in total
- 5-fold dilutions for first ten transfers; 100-fold for next 18 transfers
- ~150 generations of evolution in total

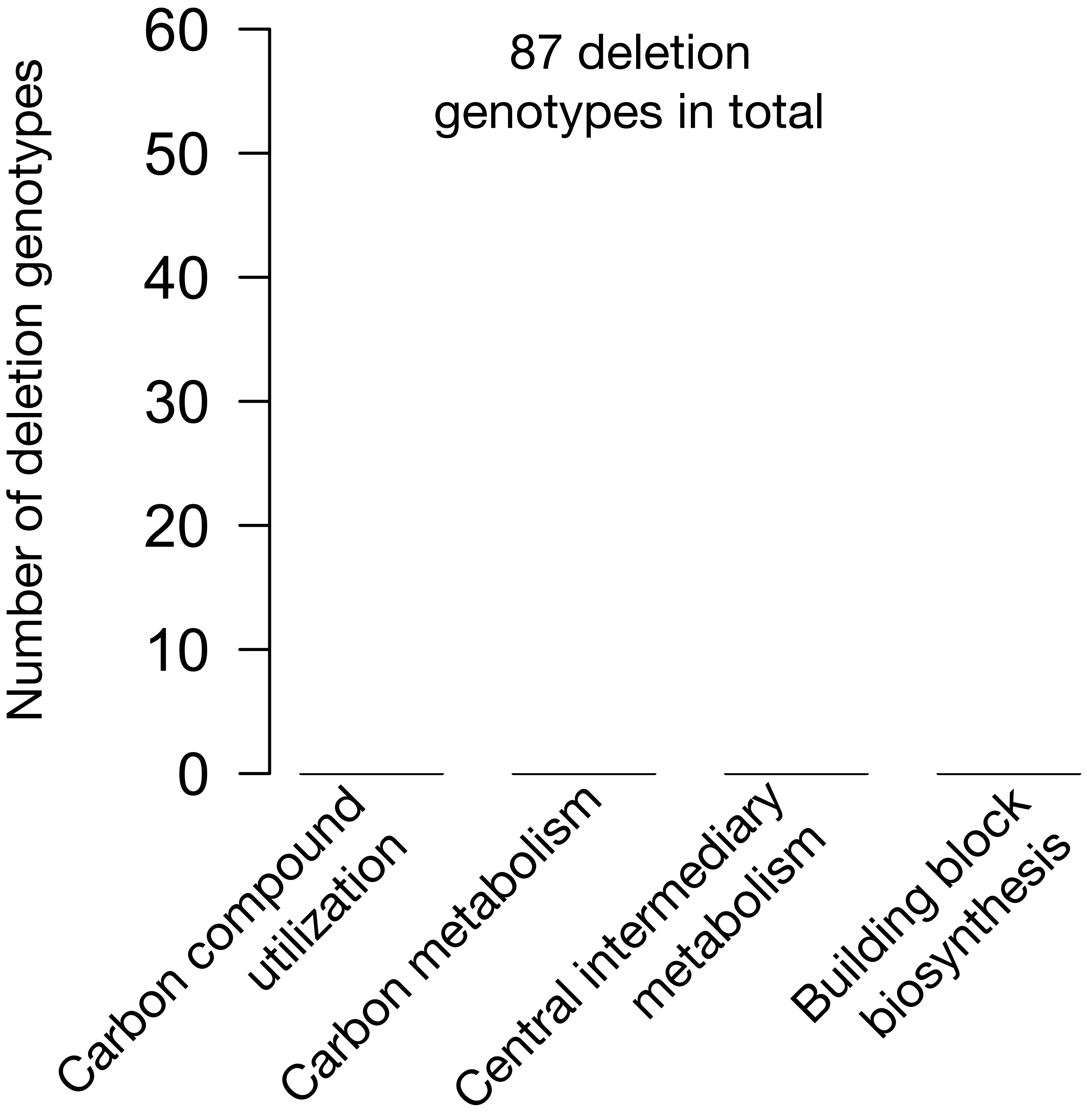
We define “new function” as cases in which deletion strains can now **grow on minimal media**

68 lineages evolved "new functions"



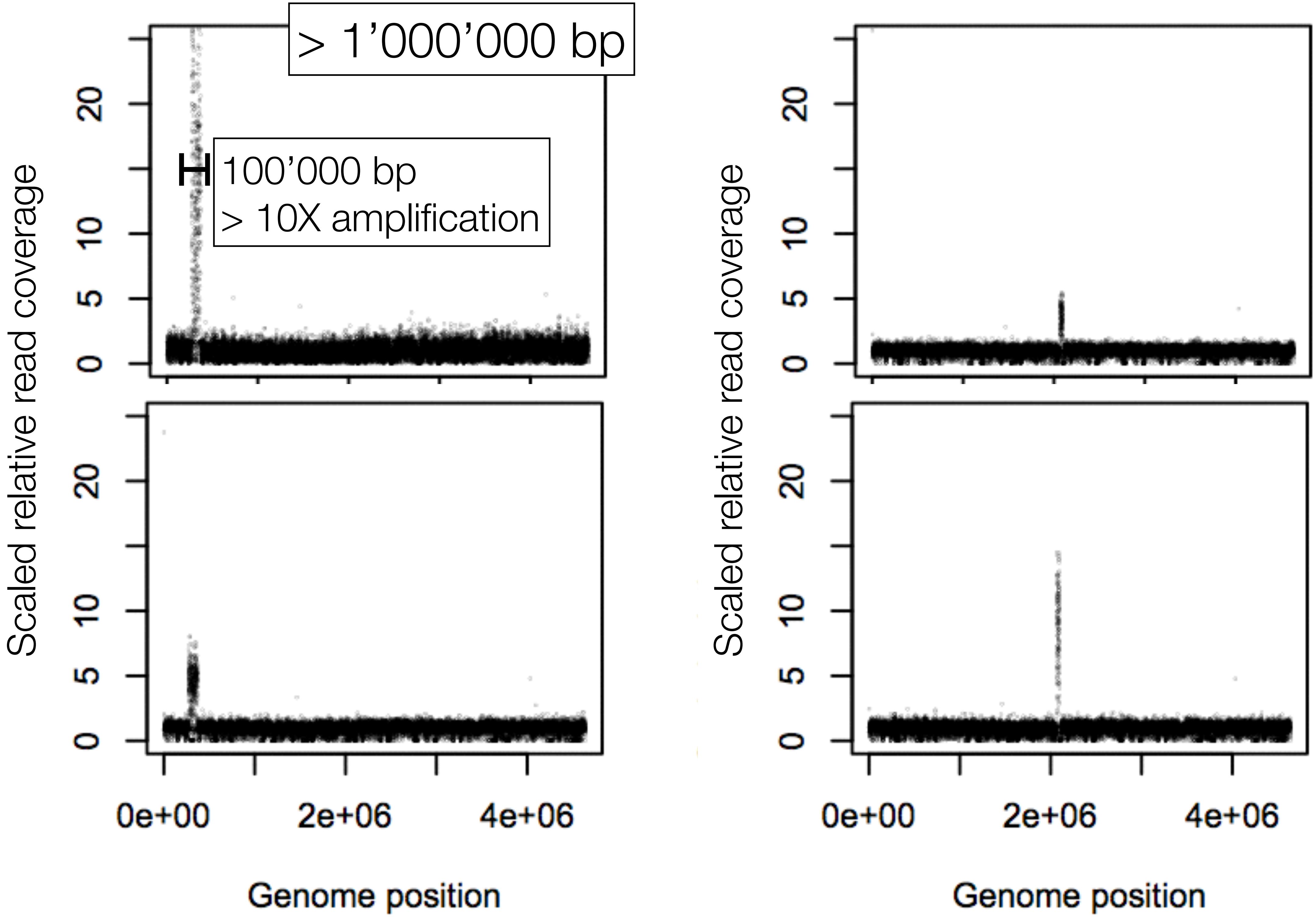
Lag times were consistently longer than the ancestor (no deletion strain)





~10 times as hard to evolve new function for biosynthesis (odds ratio)

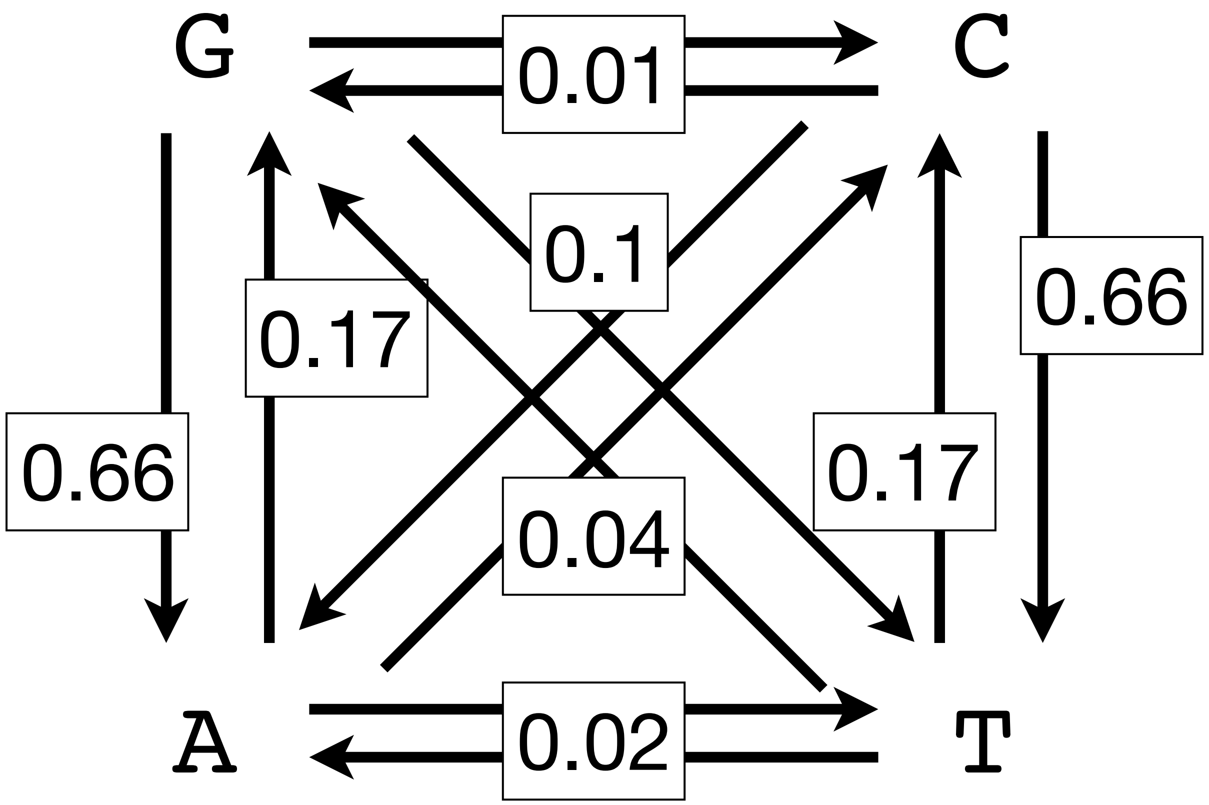
Genomic amplifications are common, often paralleled, and extensive



> 1’000’000 bp

H 100’000 bp
> 10X amplification

How does novel function evolve?



Neutral expectation using empirical matrix
(Petrov and Hershberg 2010):

66% of all mutations are `GC:AT`
83% of all mutations are transitions

How does novel function evolve?

Non-synonymous point mutations highly enriched:

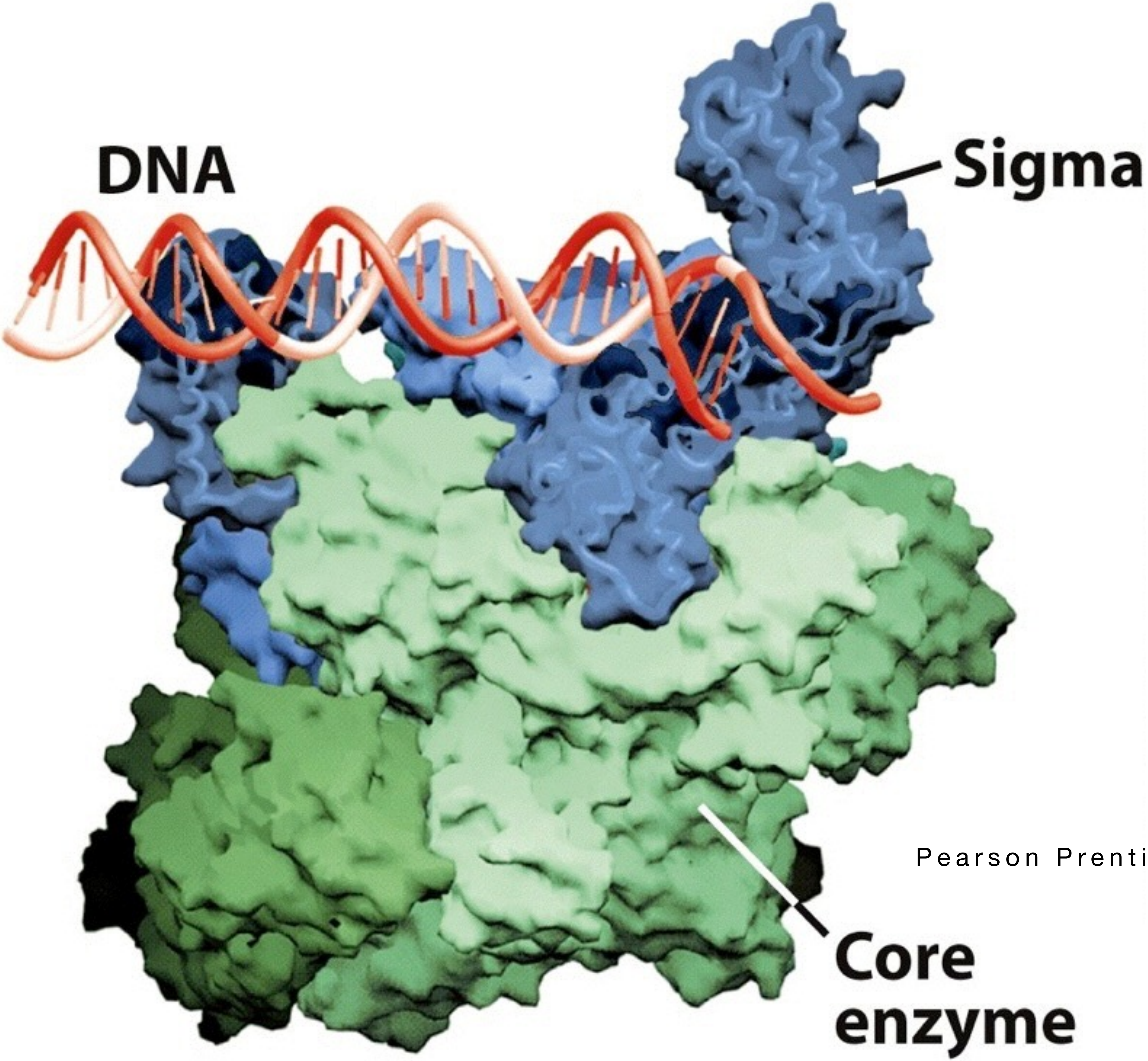
K_a/K_s: 6.0

Intergenic point mutations highly enriched:

K_i/K_s: 11.8

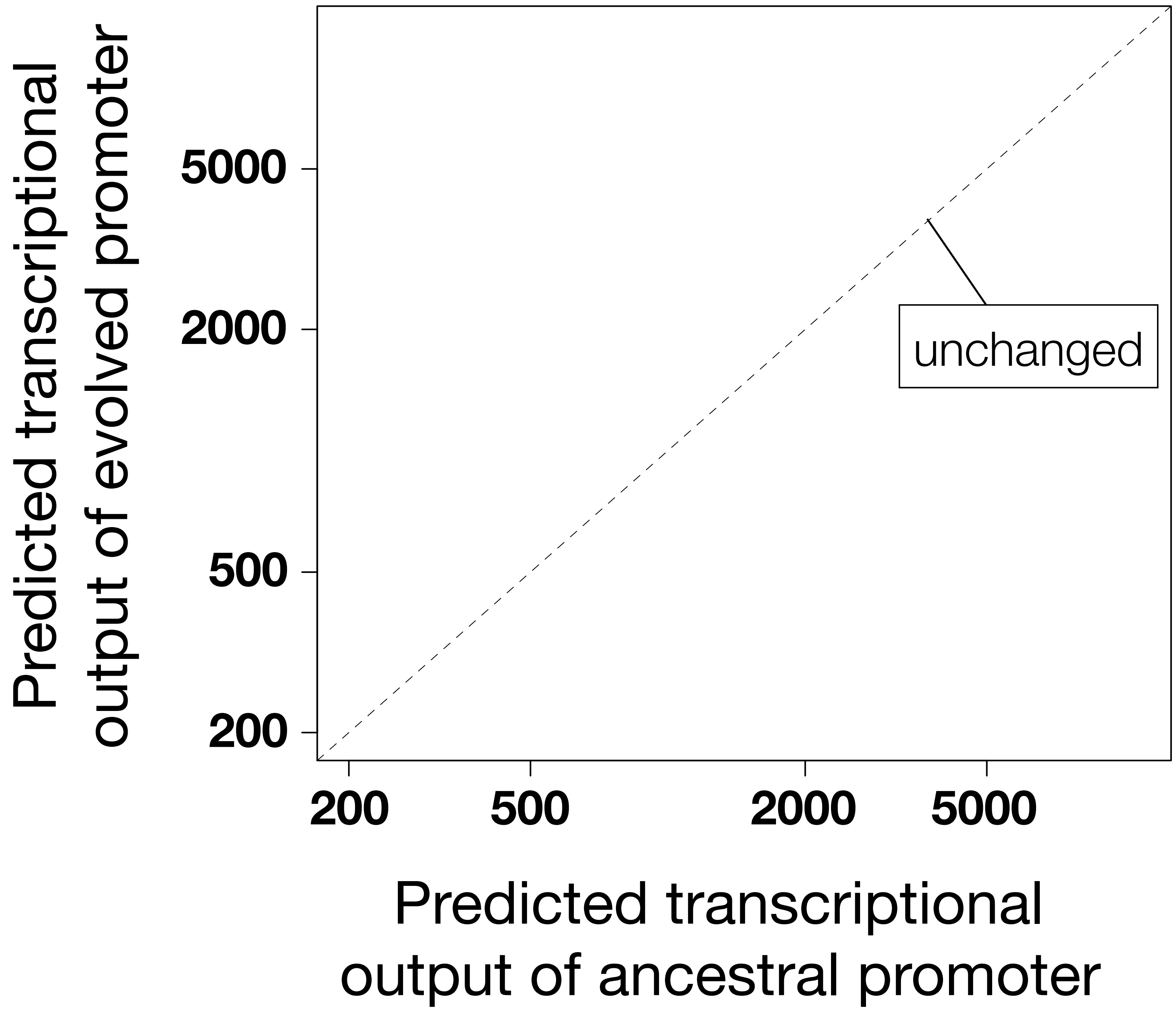
Point mutations at intergenic sites are roughly **twice as likely to be selected for** as point mutations that change an amino acid

Intergenic indel are roughly **3.5 times as likely to be selected for** as coding indels

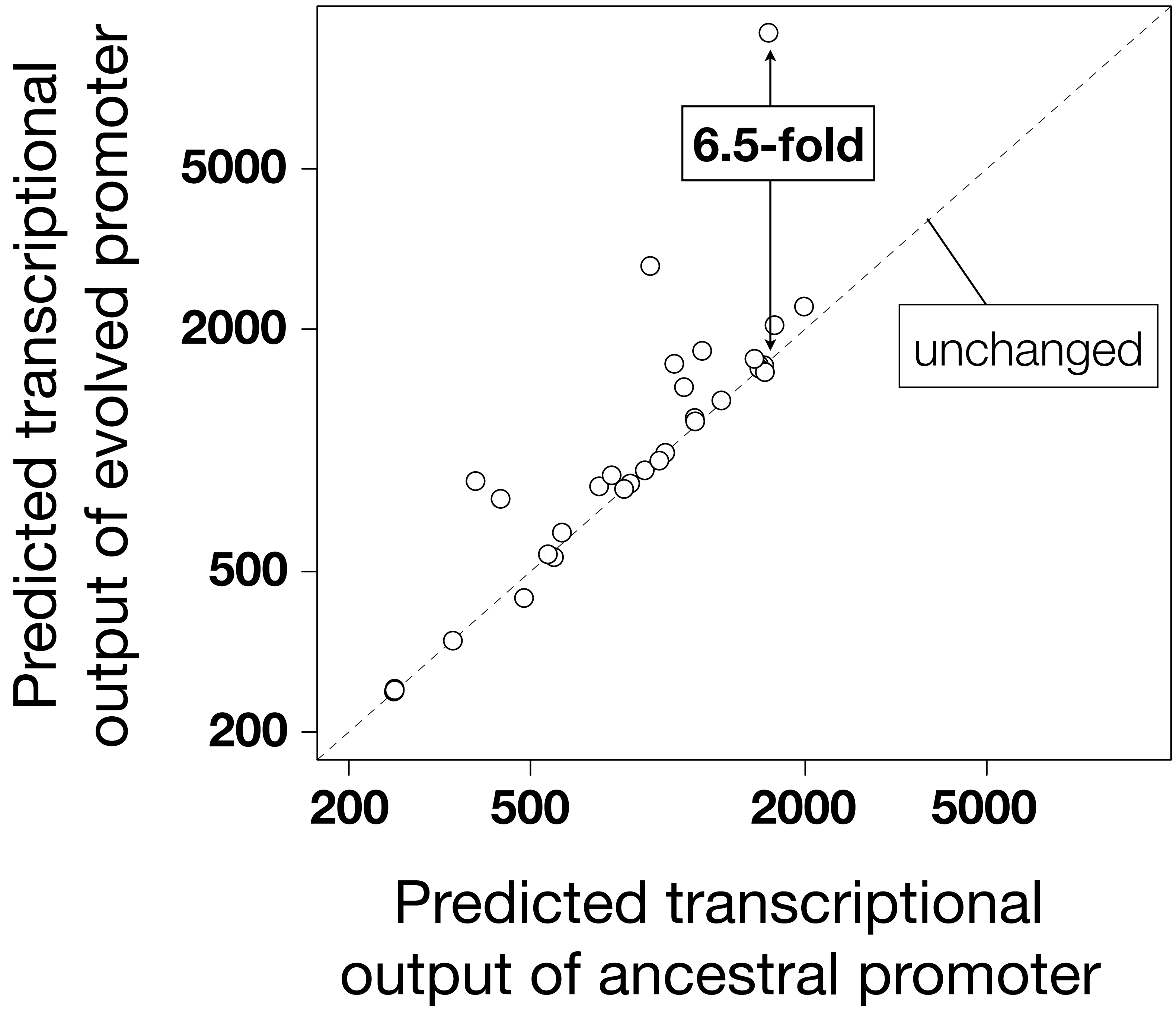


Pearson Prentice Hall

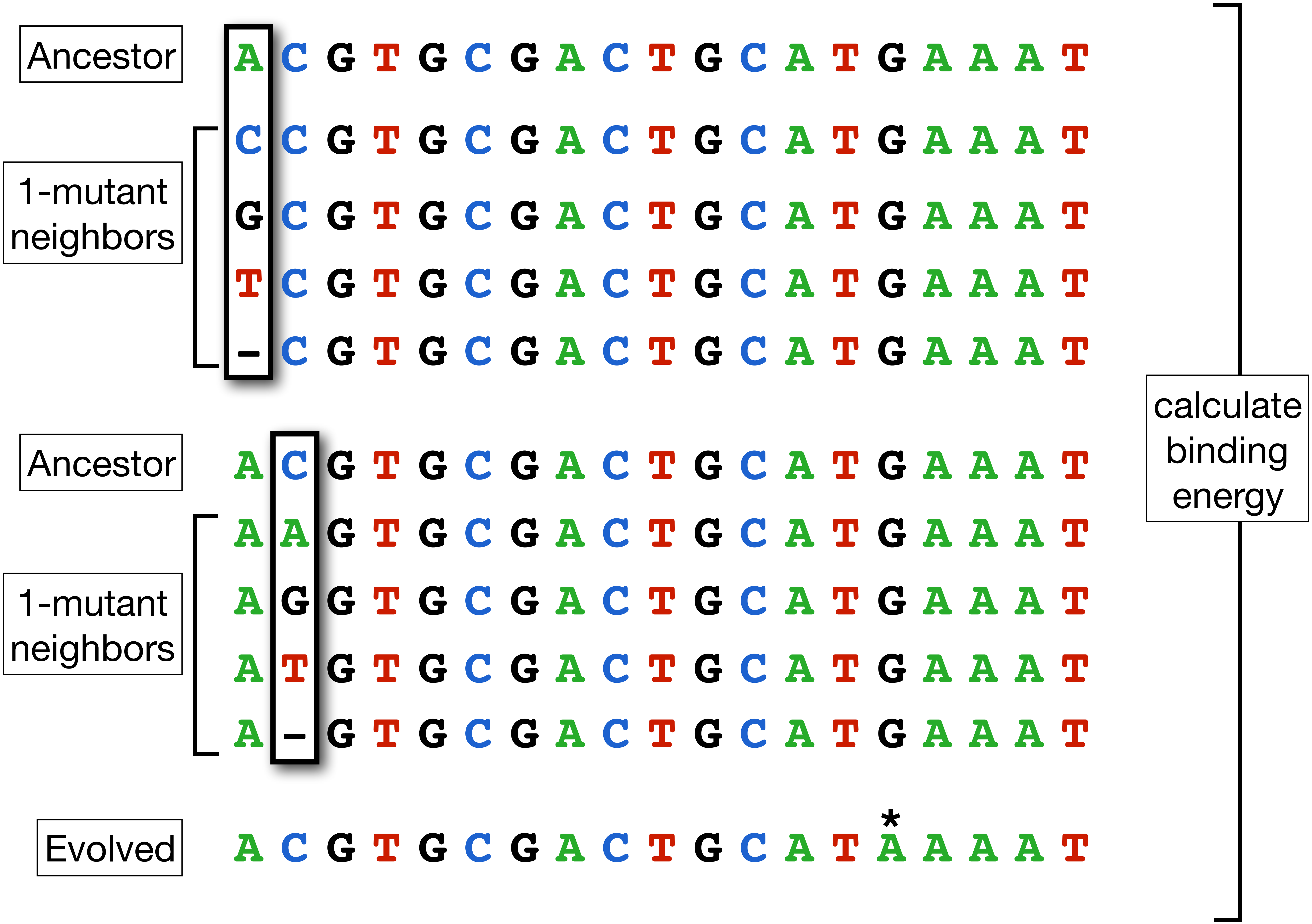
The phenotypic effects of intergenic mutations



The phenotypic effects of intergenic mutations



Is this change meaningful?



Ancestor

A C G T G C G A C T G C A T G A A A T

1-mutant neighbors

C C G T G C G A C T G C A T G A A A T

G C G T G C G A C T G C A T G A A A T

T C G T G C G A C T G C A T G A A A T

– C G T G C G A C T G C A T G A A A T

Ancestor

A C G T G C G A C T G C A T G A A A T

1-mutant neighbors

A A G T G C G A C T G C A T G A A A T

A G G T G C G A C T G C A T G A A A T

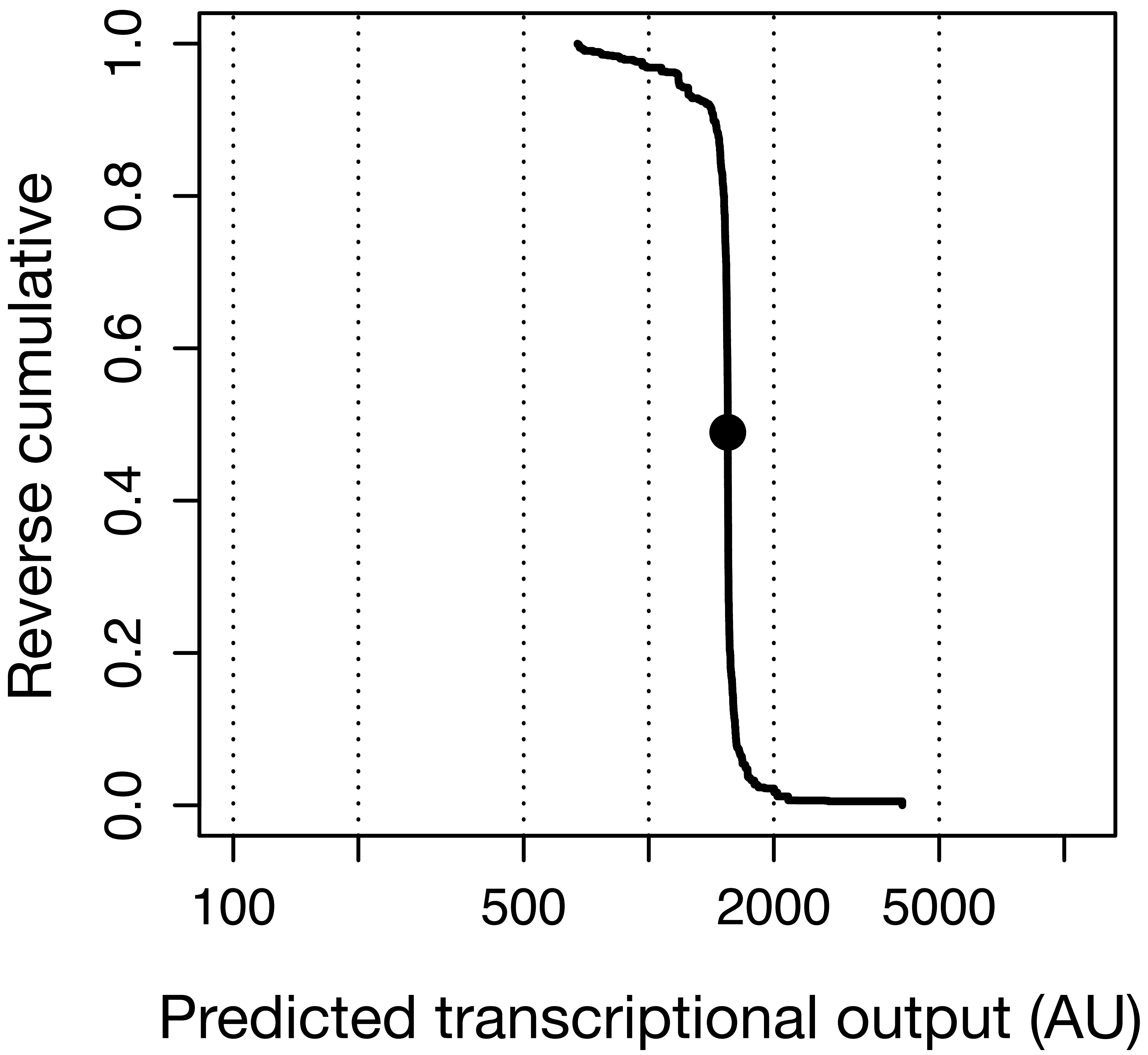
A T G T G C G A C T G C A T G A A A T

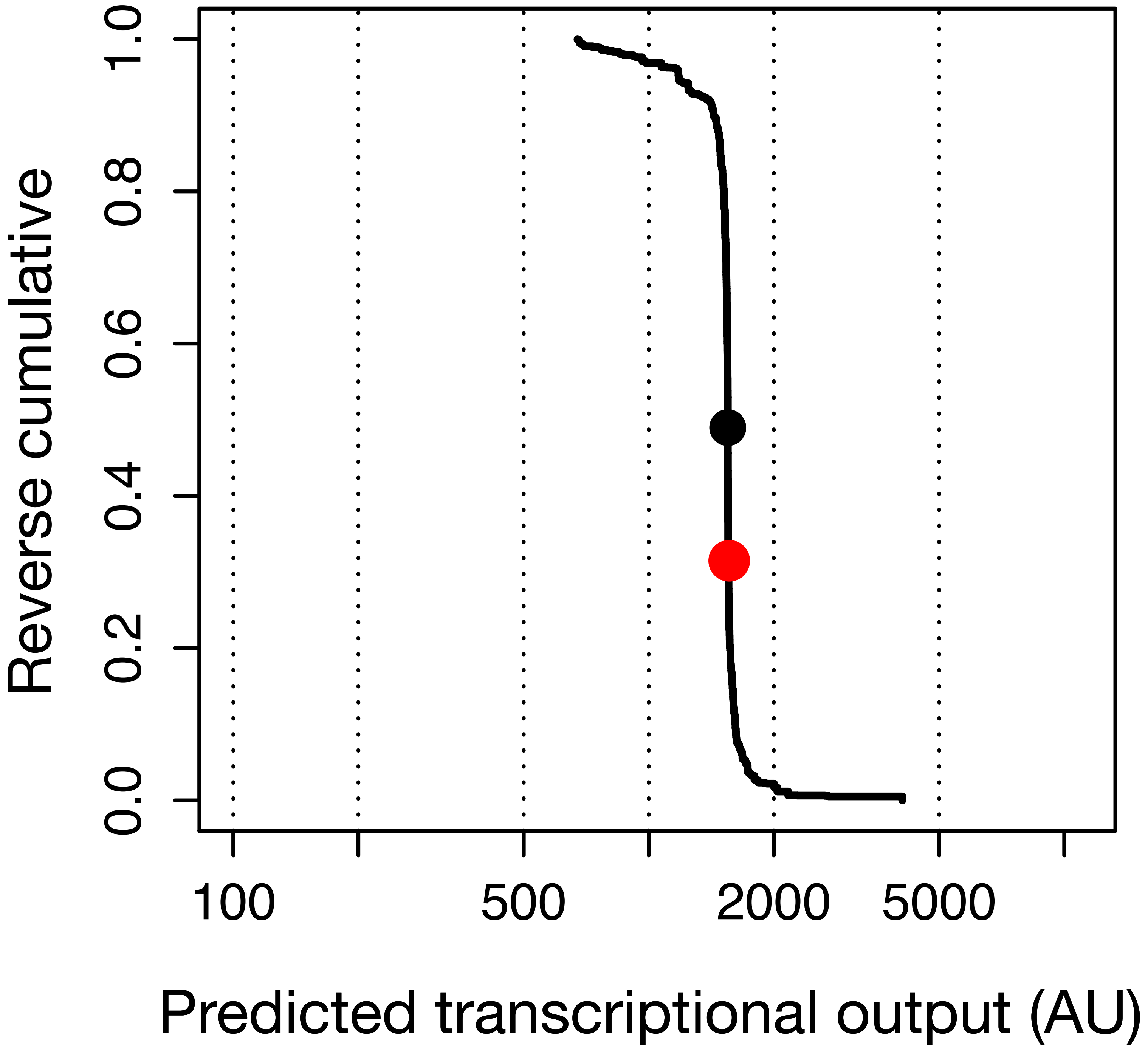
A – G T G C G A C T G C A T G A A A T

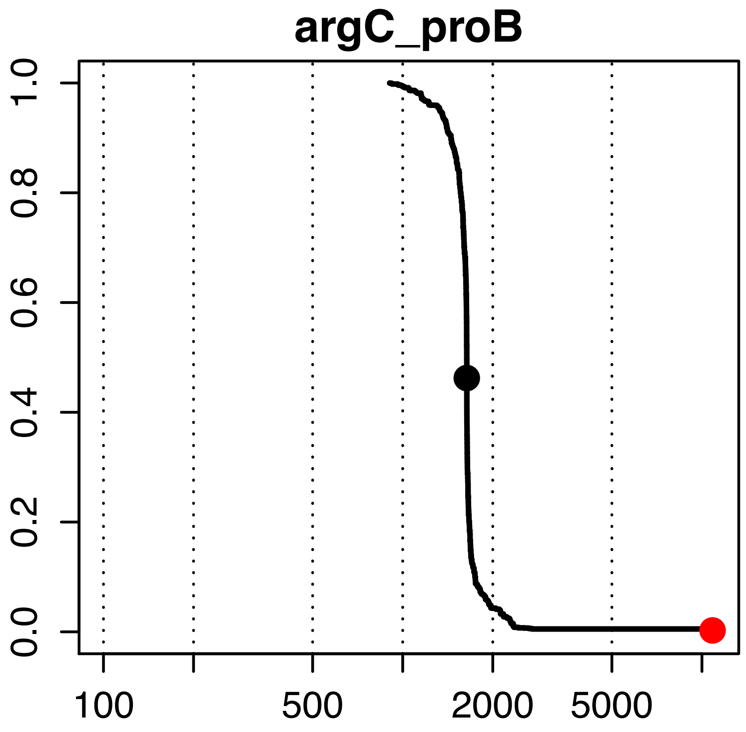
Evolved

A C G T G C G A C T G C A T $\overset{*}{A}$ A A A T

calculate binding energy


Reverse cumulative
0.0
0.2
0.4
0.6
0.8
1.0
100
500
2000
5000
Predicted transcriptional output (AU)


Reverse cumulative
1.0
0.8
0.6
0.4
0.2
0.0
100
500
2000
5000
Predicted transcriptional output (AU)

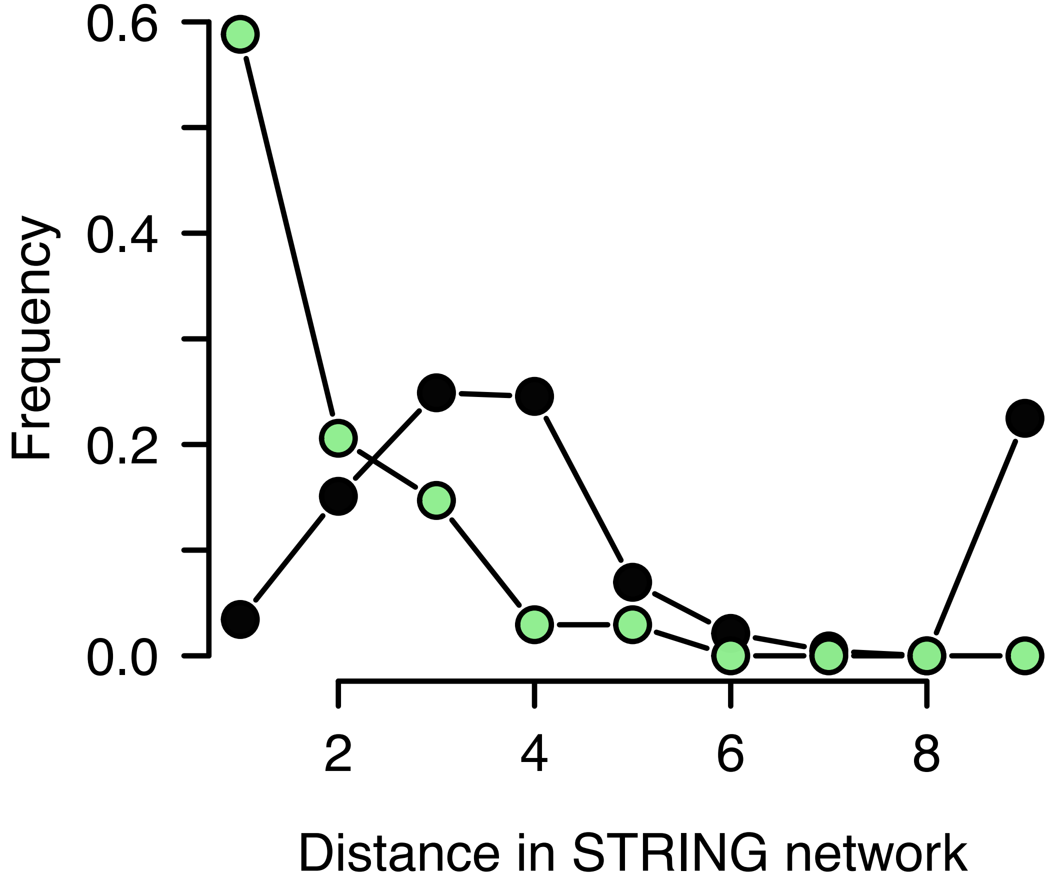

argC_proB
1.0
0.8
0.6
0.4
0.2
0.0
100
500
2000
5000

30% of intergenic mutations
increase s70 binding by >10%

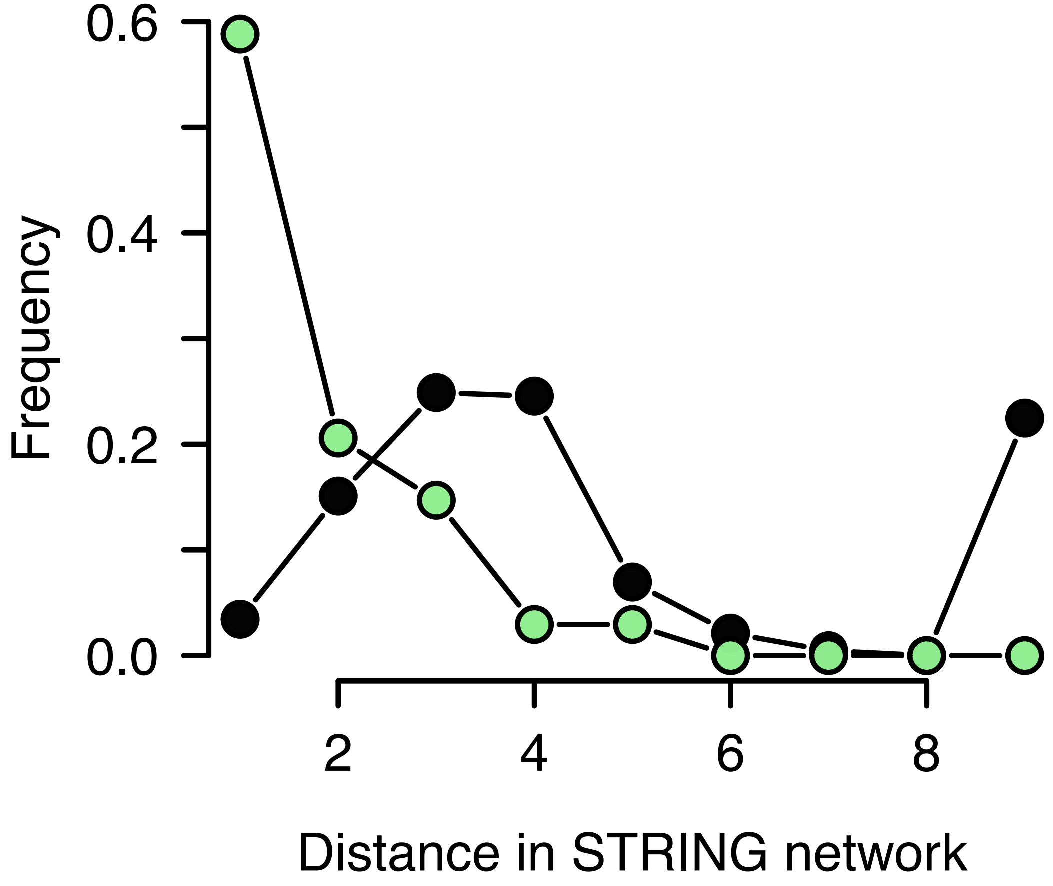
Only 3.5% of random mutations
do the same

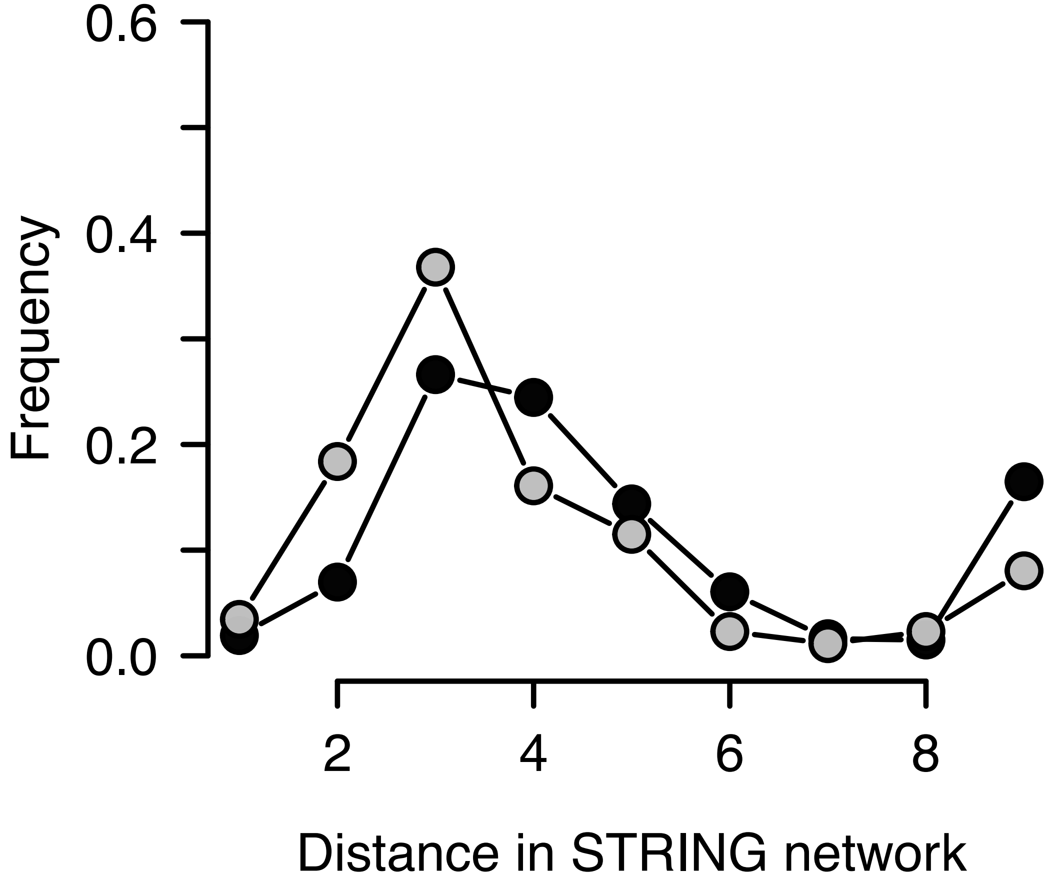
What are the functions of the recruited genes?

What are the functions of the recruited genes?



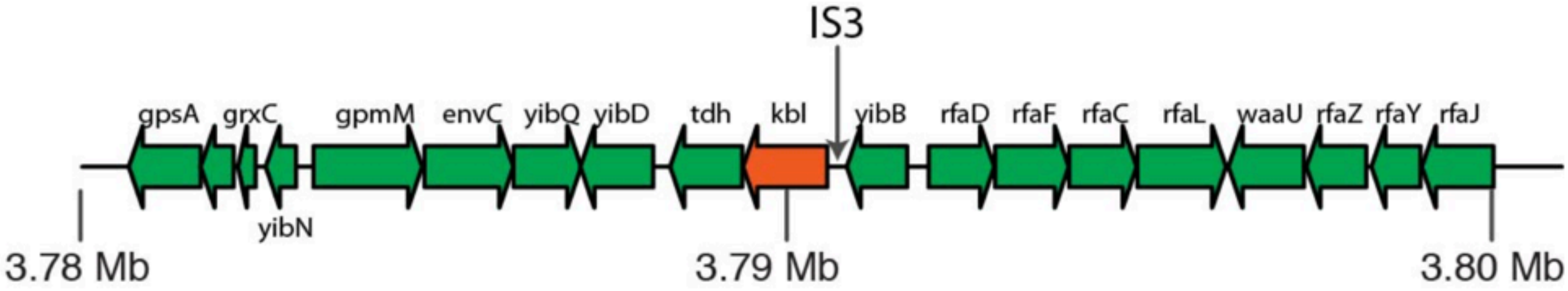
What are the functions of the recruited genes?





Transposon insertions are a frequent mechanism of new function

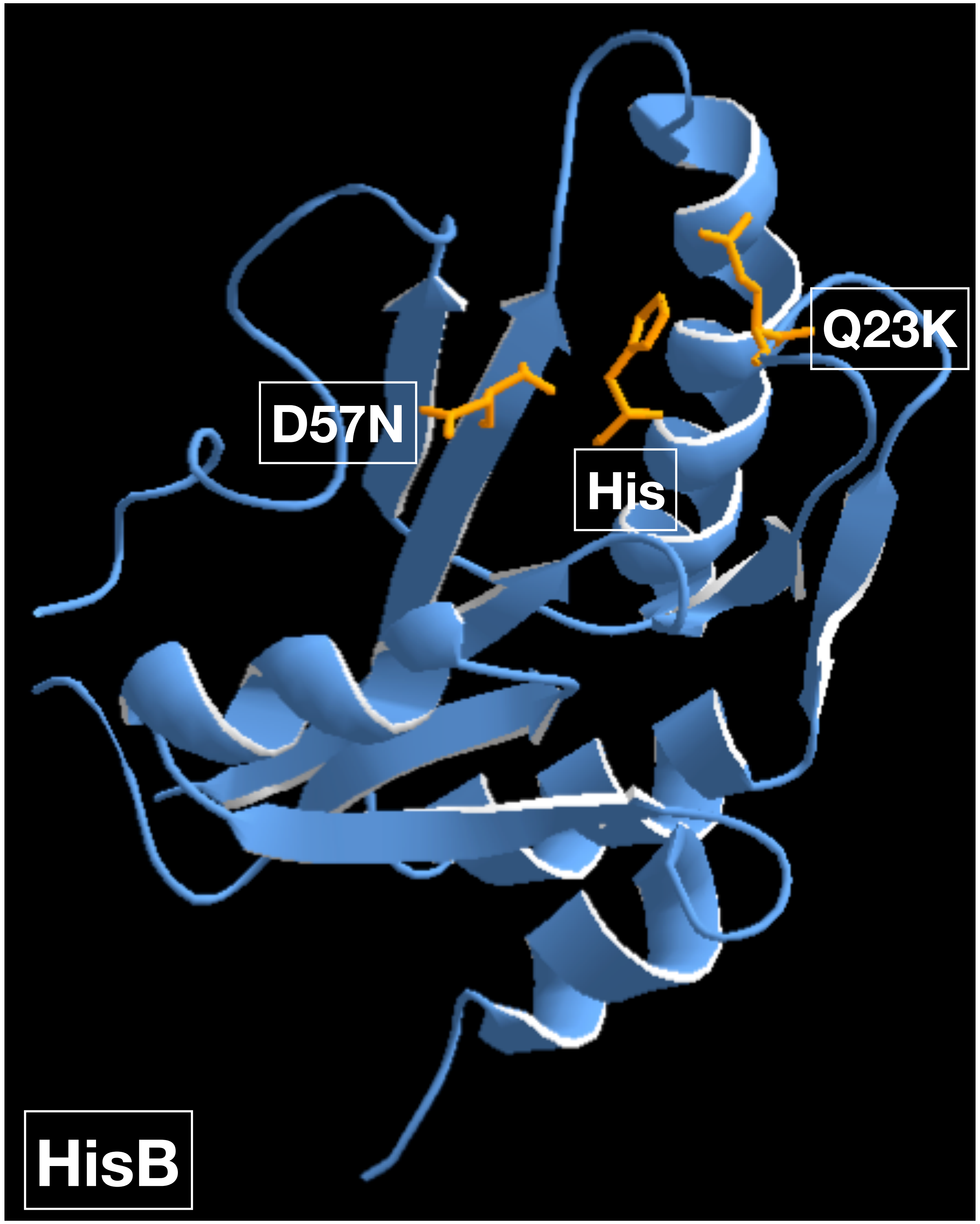
3/5 of *glyA* lines

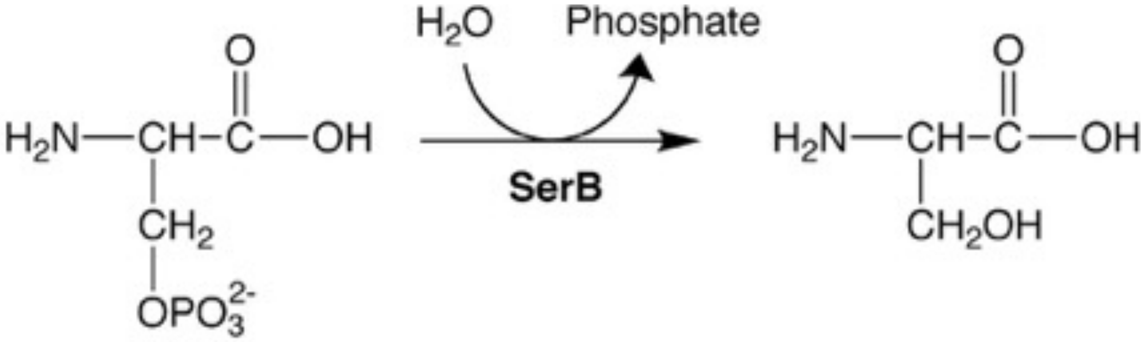


glyA: conversion of serine to glycine

kbl: conversion of threonine to glycine

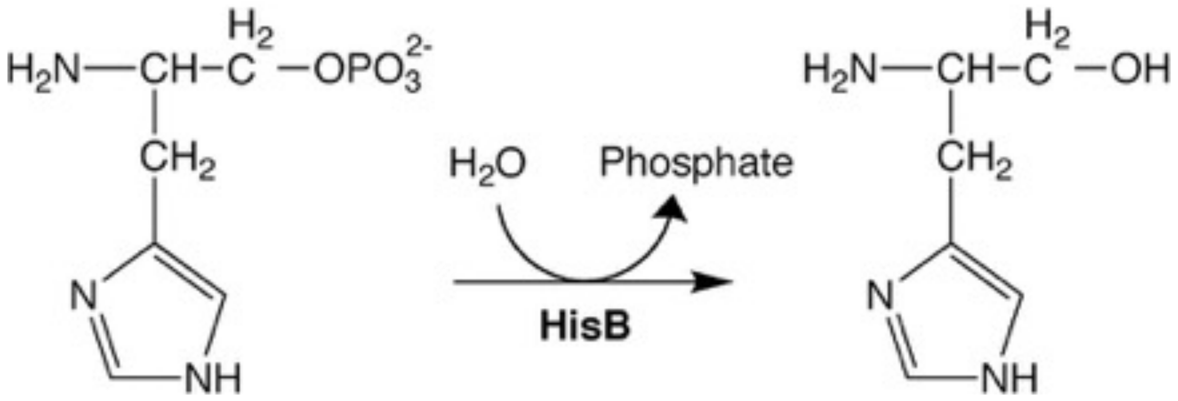
HisB mutations are frequent in *serB* deletion lines





phospho-serine

serine



histidinol-phosphate

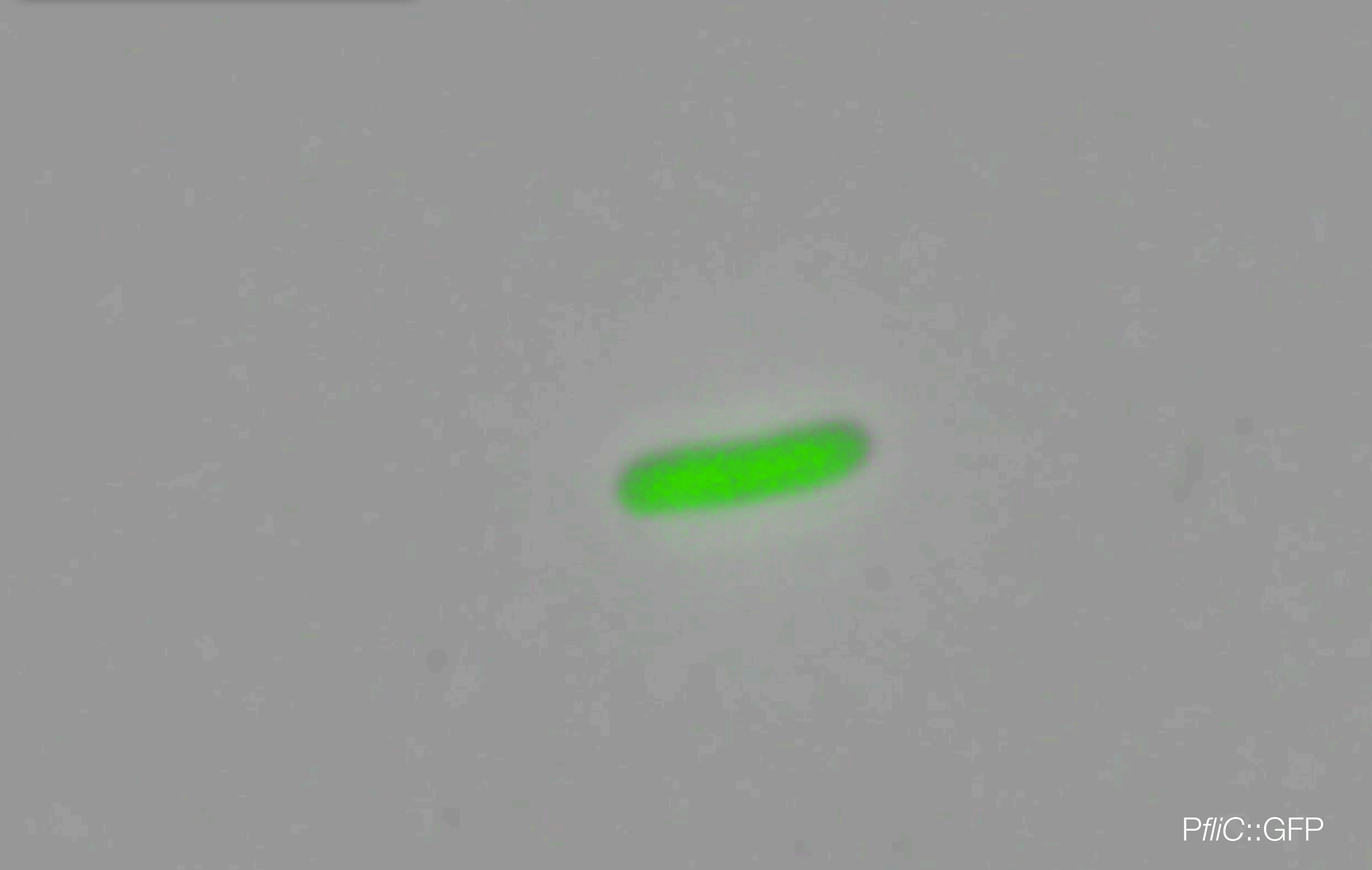
histidinol

increases in mean expression level
are critical for evolving new functions

Single genotype

Homogenous environment

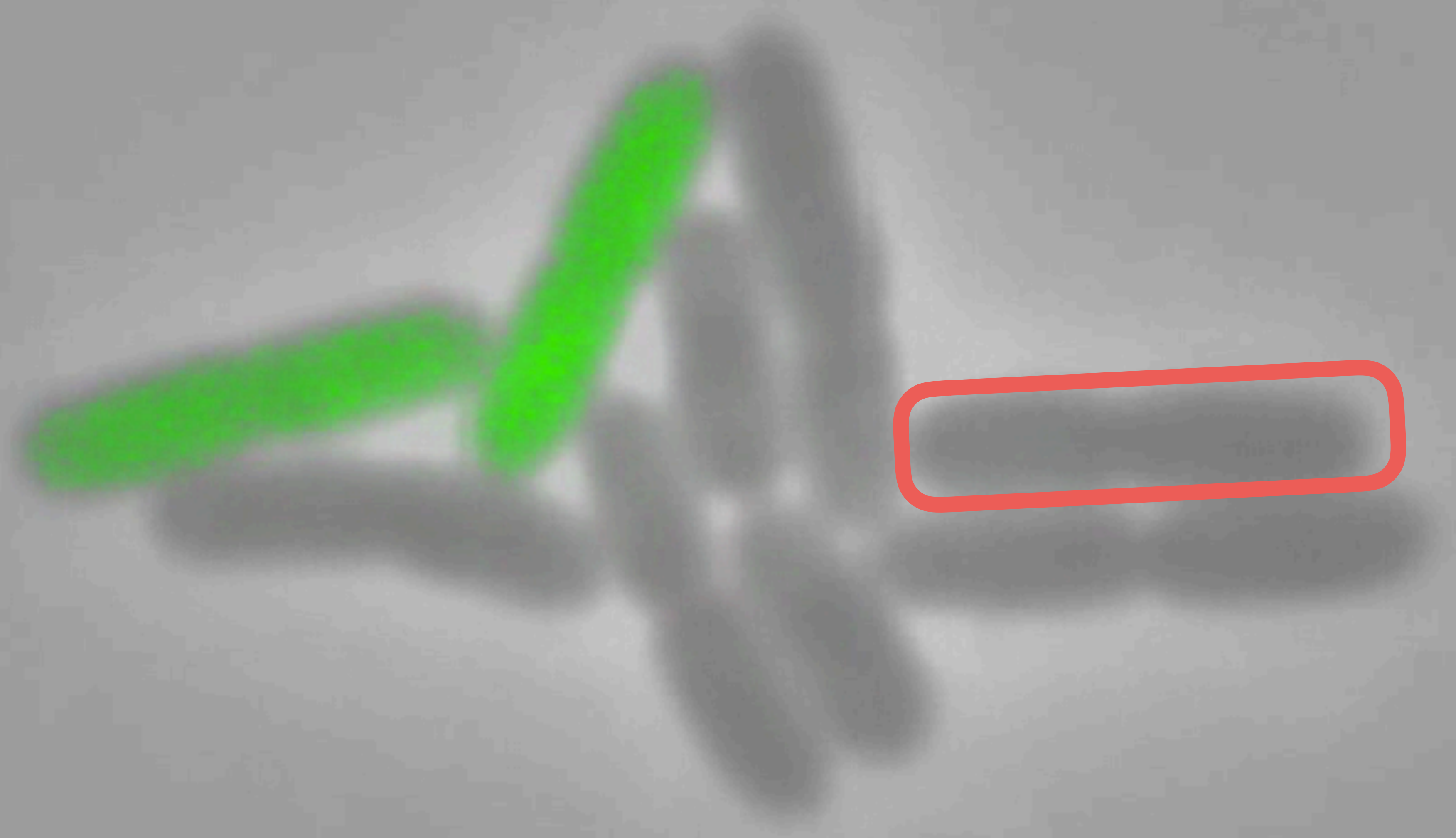
Phenotypic noise



Single genotype

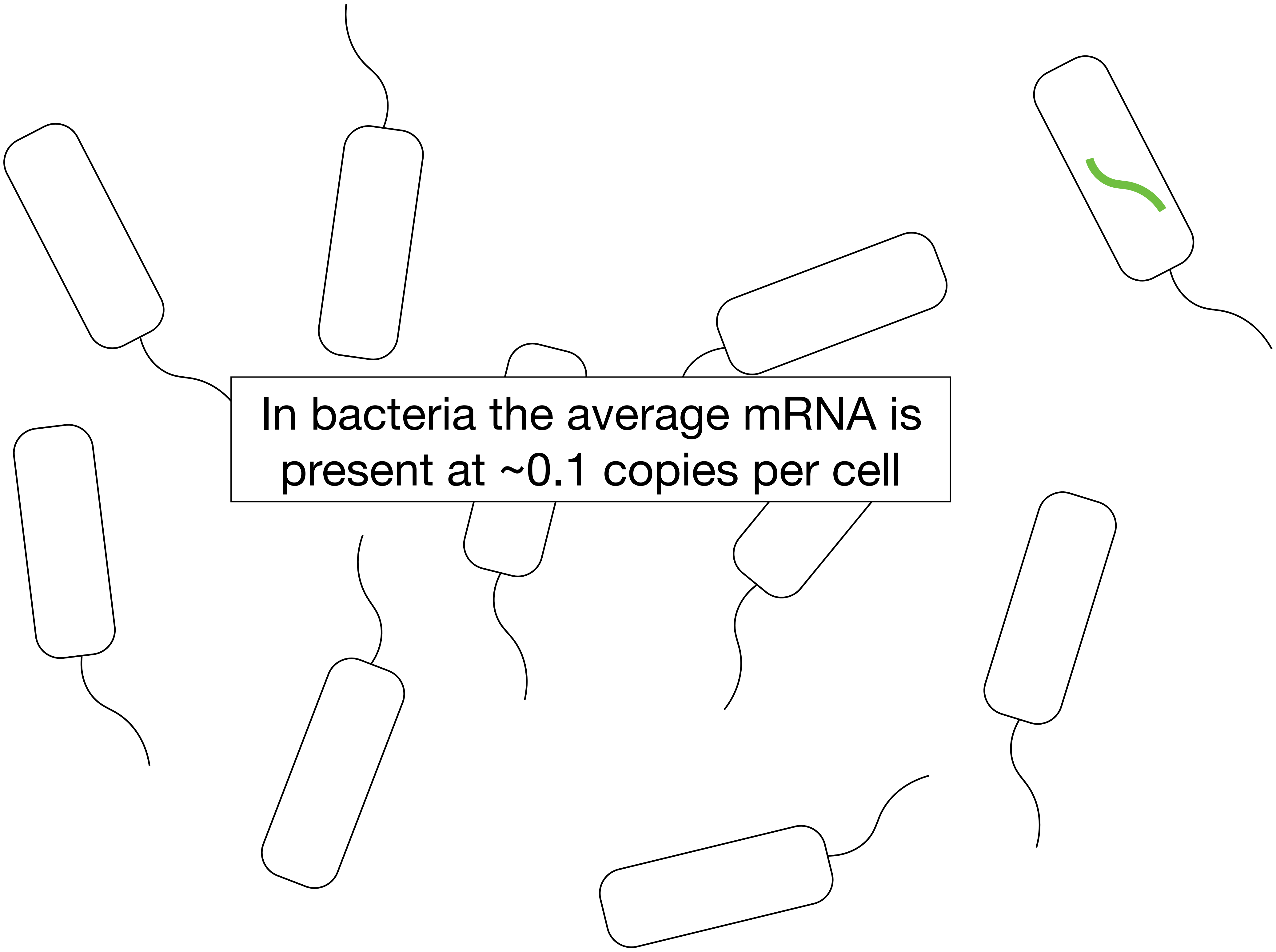
Homogenous environment

Phenotypic noise



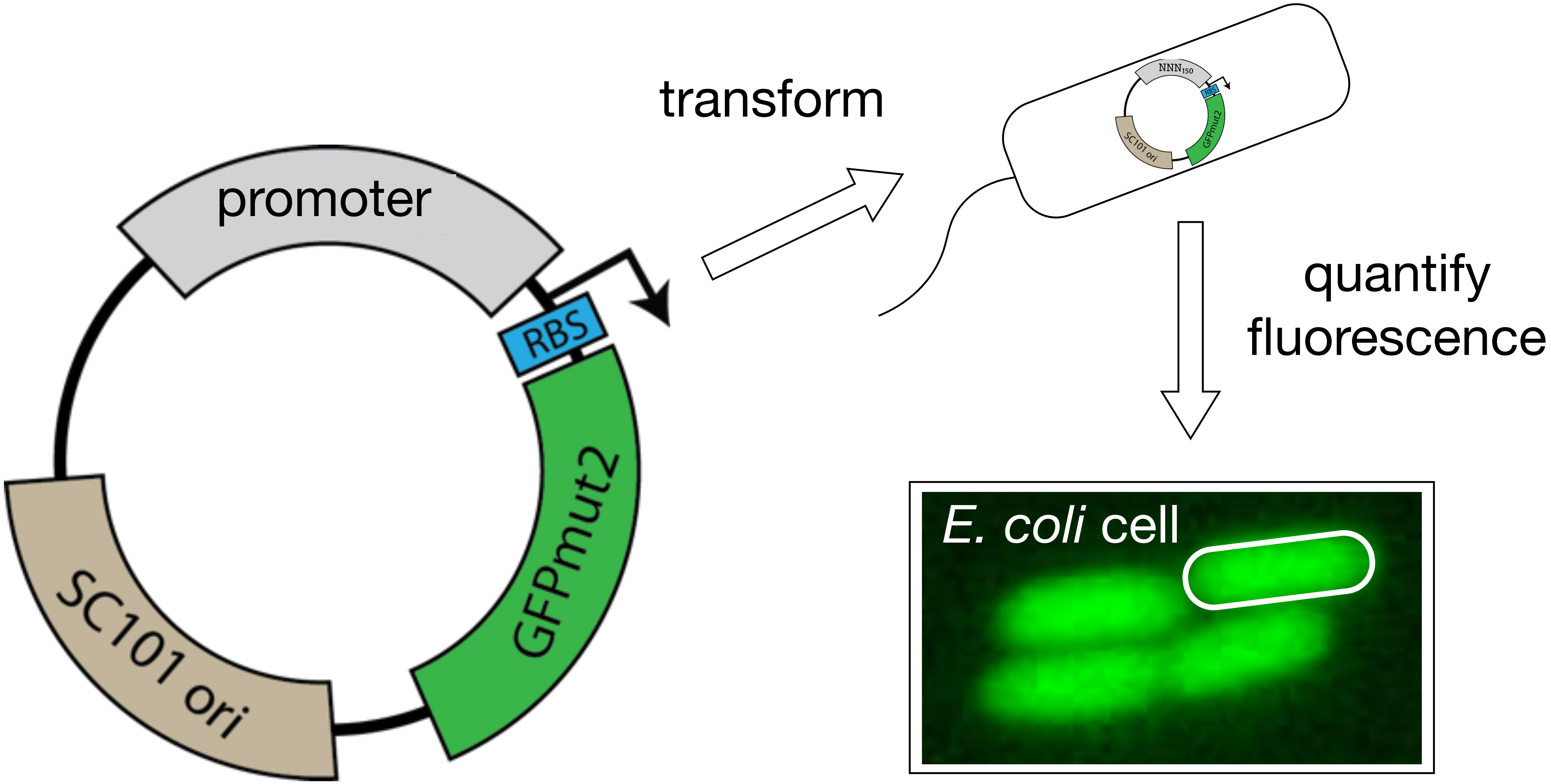
P*fliC*::GFP

Variation is inevitable if processes involve small numbers of molecules



In bacteria the average mRNA is present at ~0.1 copies per cell

What does transcriptional noise
in bacteria (*E. coli*) look like?



transform

quantify fluorescence

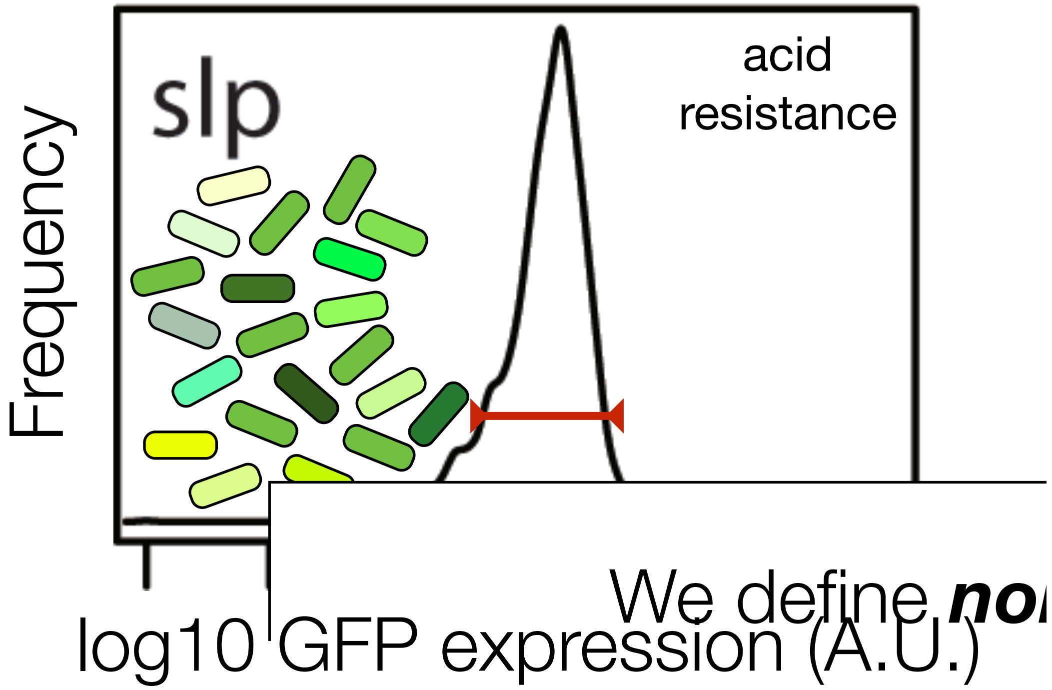
Fluorescence depends on the transcriptional activity of the promoter

What does transcriptional noise in bacteria (*E*. *coli*) look like?

Assay noise in all *E*. *coli* promoters (~1’800) using flow cytometry

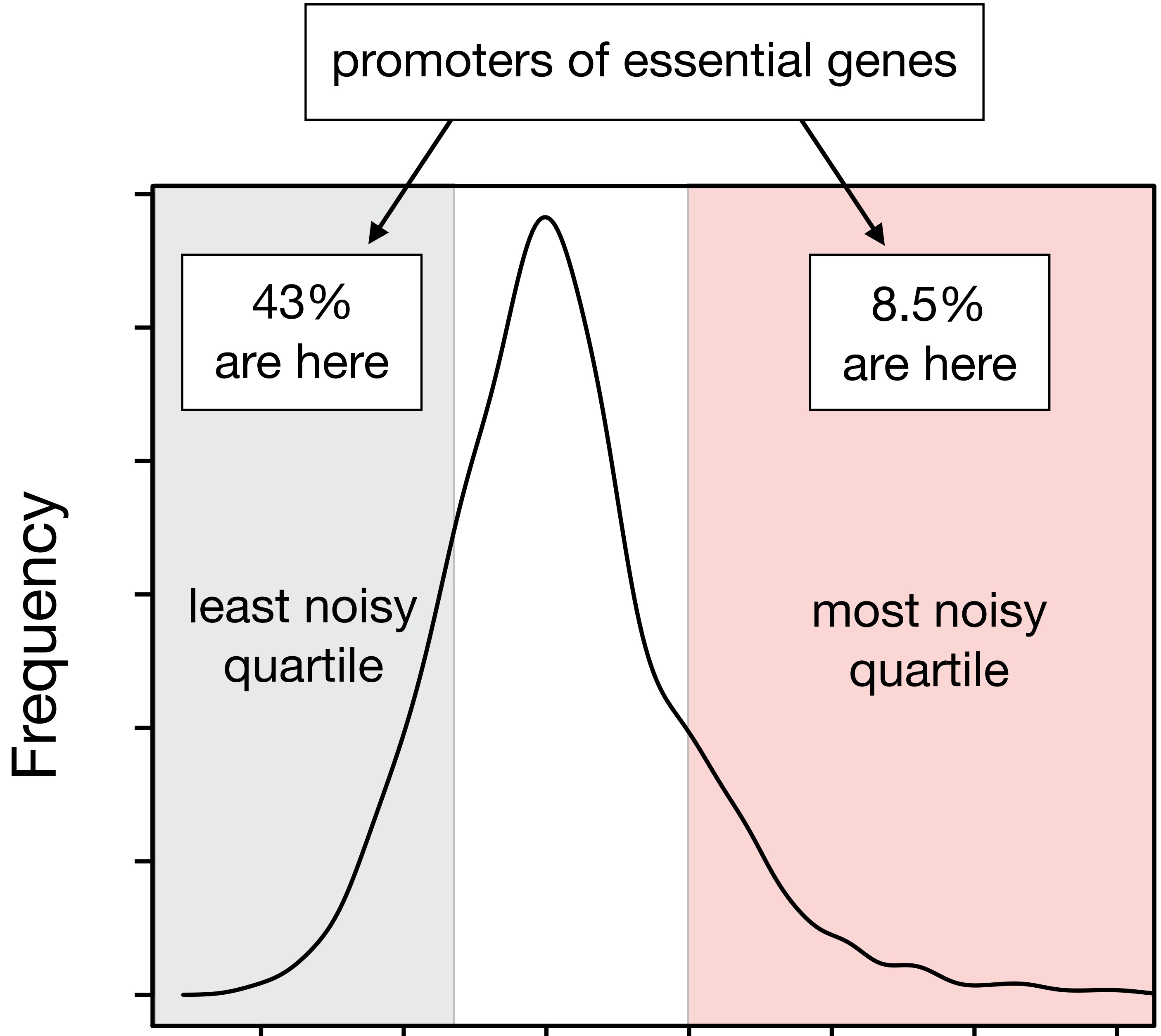
Some promoters in *E. coli* have high noise, some have low noise

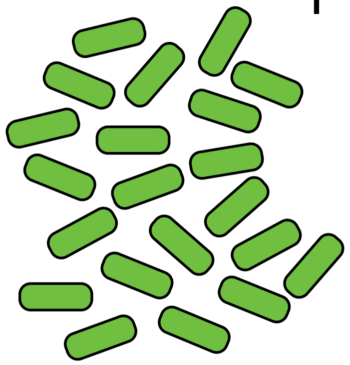
Histograms of expression levels
(100’000 cells measured for each of 1’800 promoters)



We define ***no***

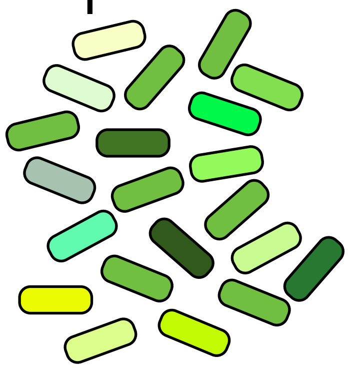
Some promoters in *E. coli* have high noise, some have low noise



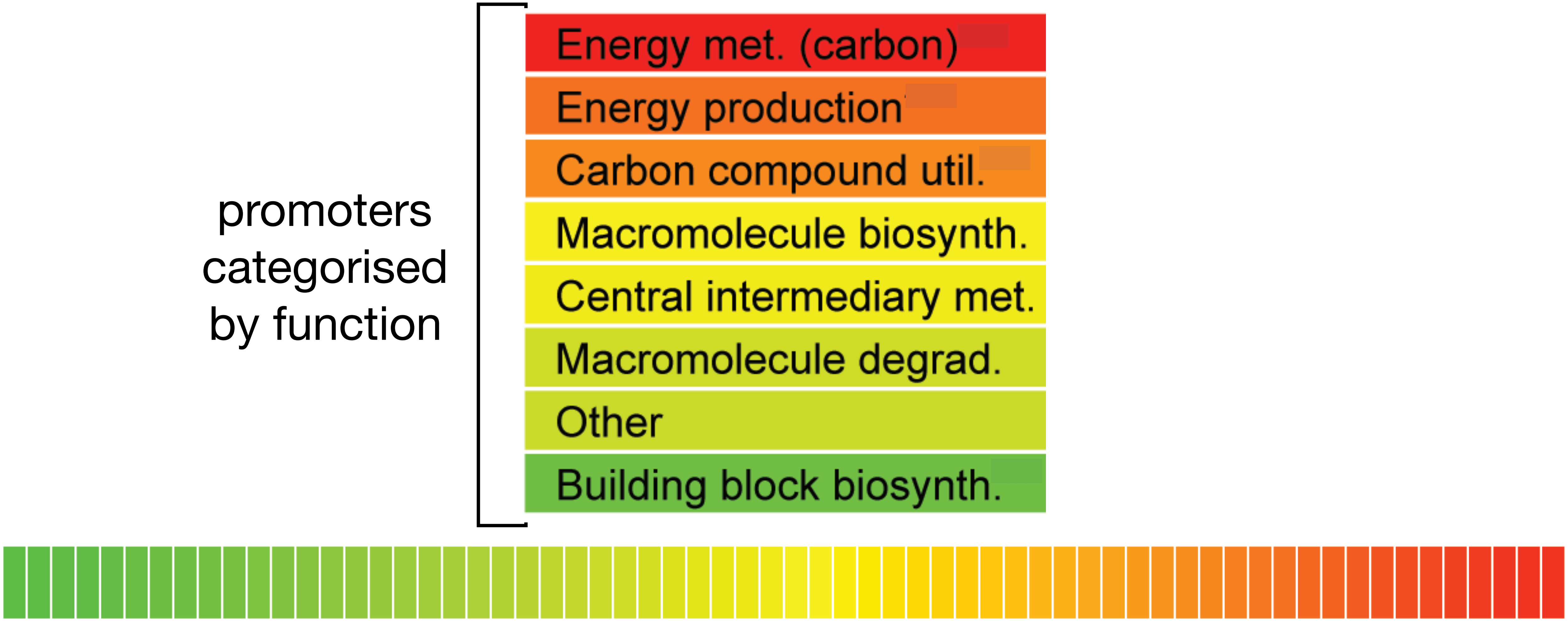


low ⟵ noise ⟶ high

Noise levels for 1,800 promoters in E. coli



Some promoters in *E. coli* have high noise, some have low noise

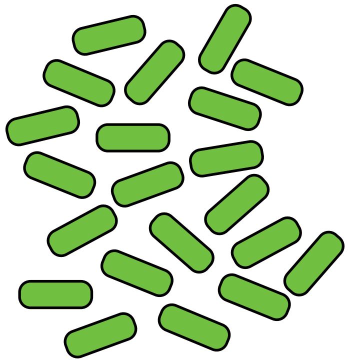


promoters categorised by function

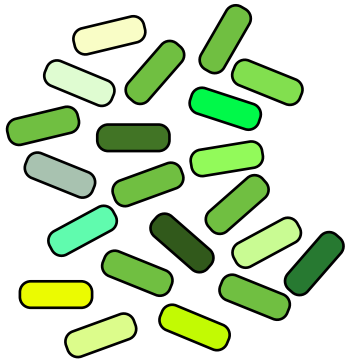
- Energy met. (carbon)
- Energy production
- Carbon compound util.
- Macromolecule biosynth.
- Central intermediary met.
- Macromolecule degrad.
- Other
- Building block biosynth.

low noise

high noise



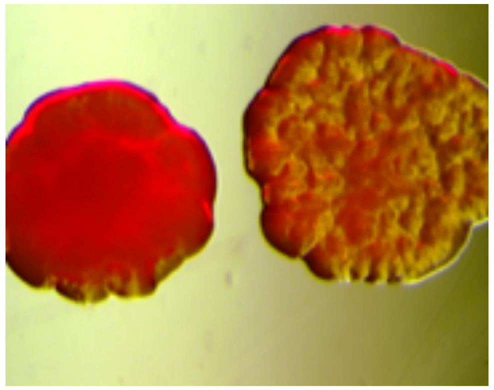
Why?



DRIFT : Some promoters have high noise because **selection doesn't care** about noise

genetic drift

SELECTION : Some promoters have high noise due to **direct selection** for high noise



bet hedging

CONSTRAINT : Some promoters have high noise as a byproduct of **direct selection** on *plasticity*

genetic correlation of plasticity and noise

Luise Wolf

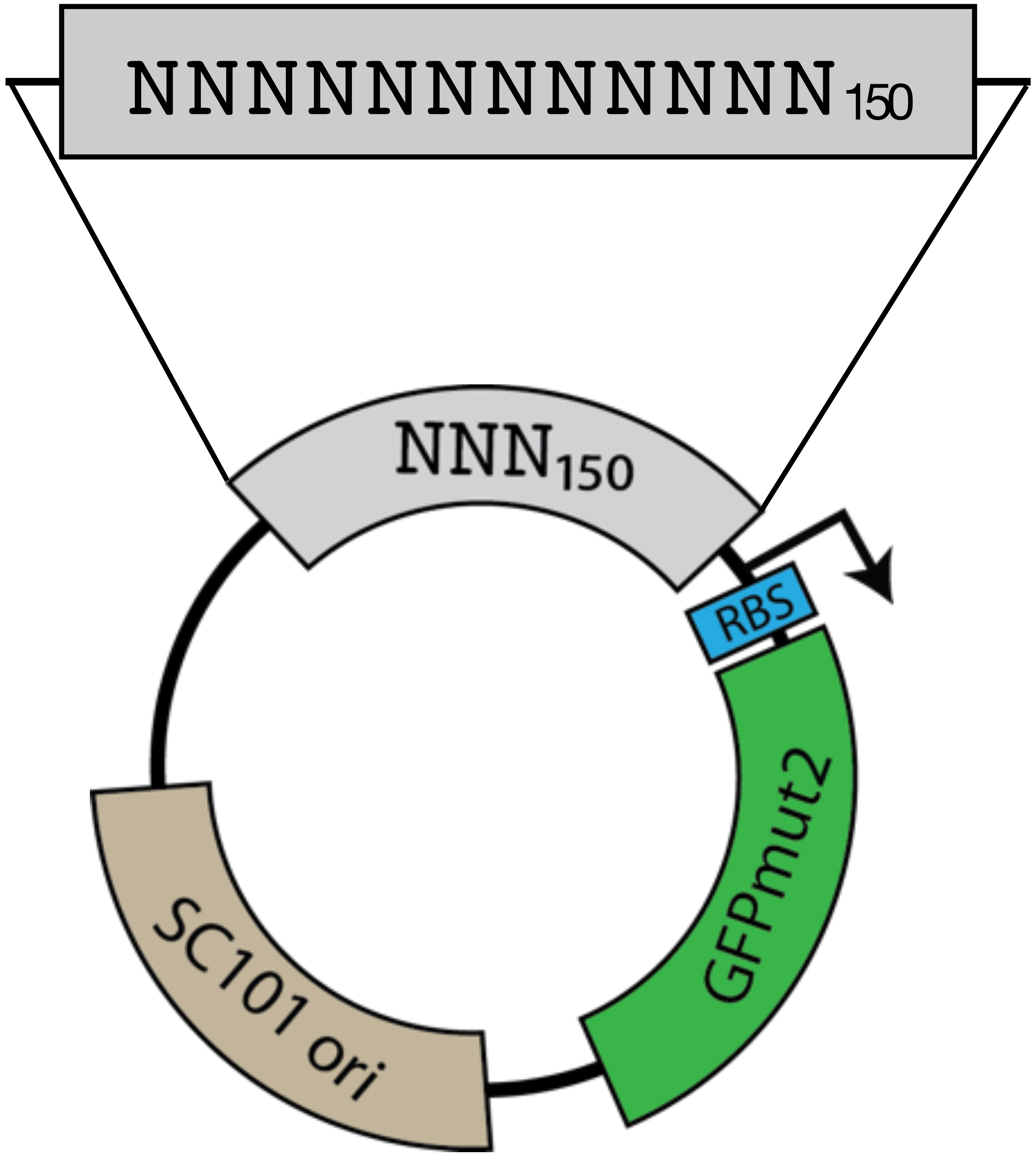
Erik van Nimwegen

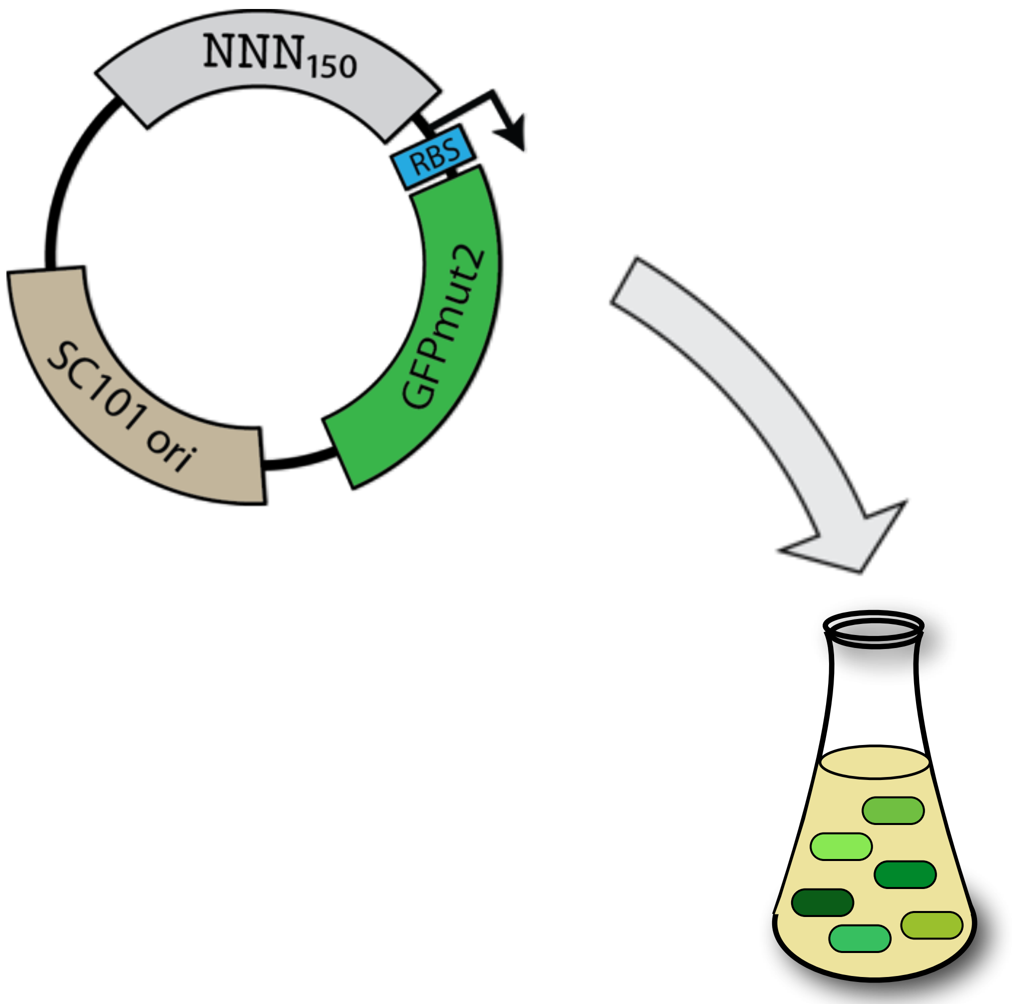
FNSNF
SWISS NATIONAL SCIENCE FOUNDATION

Use controlled selection in the
lab to generate a null hypothesis

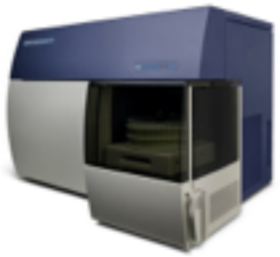
Synthetic 150bp random sequences of DNA

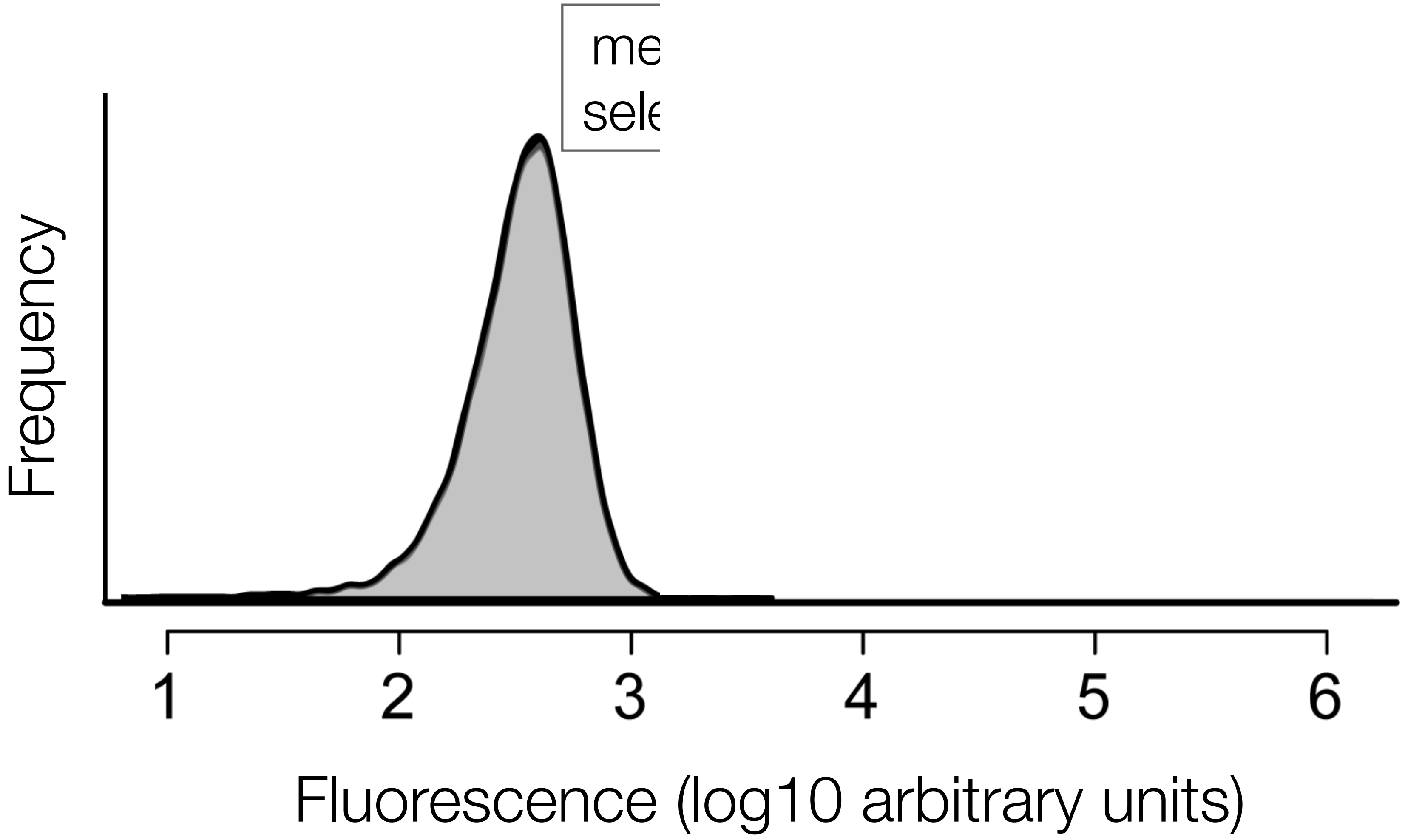
PURIMEX
DNA / RNA Oligonucleotides

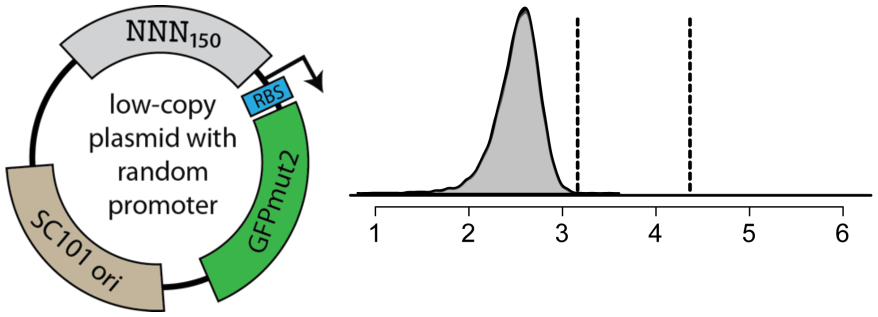




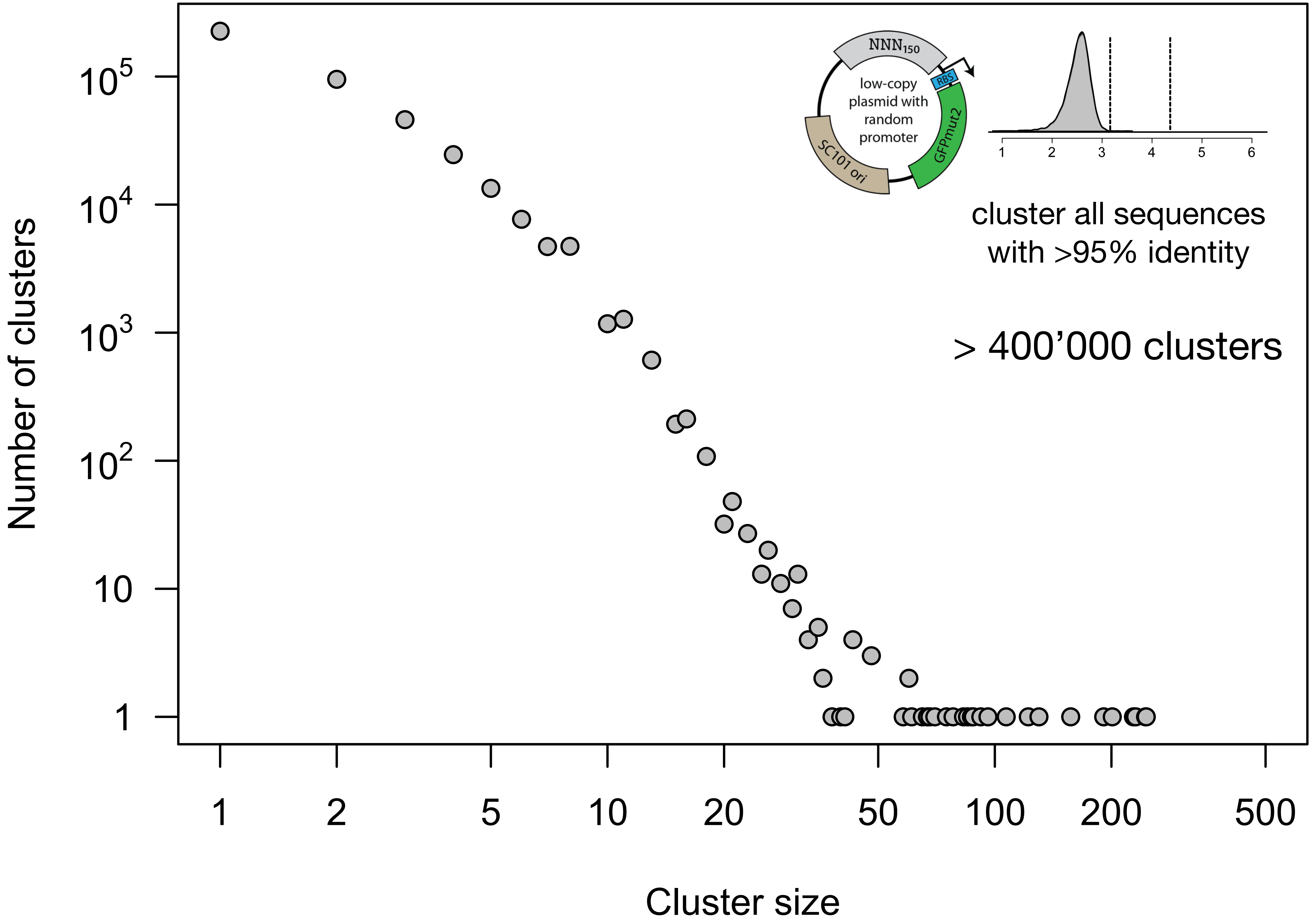
Transform cells




me
sele
Frequency
1
2
3
4
5
6
Fluorescence (log10 arbitrary units)


NNN150
RBS
low-copy plasmid with random promoter
GFPmut2
SC101 ori
1
2
3
4
5
6

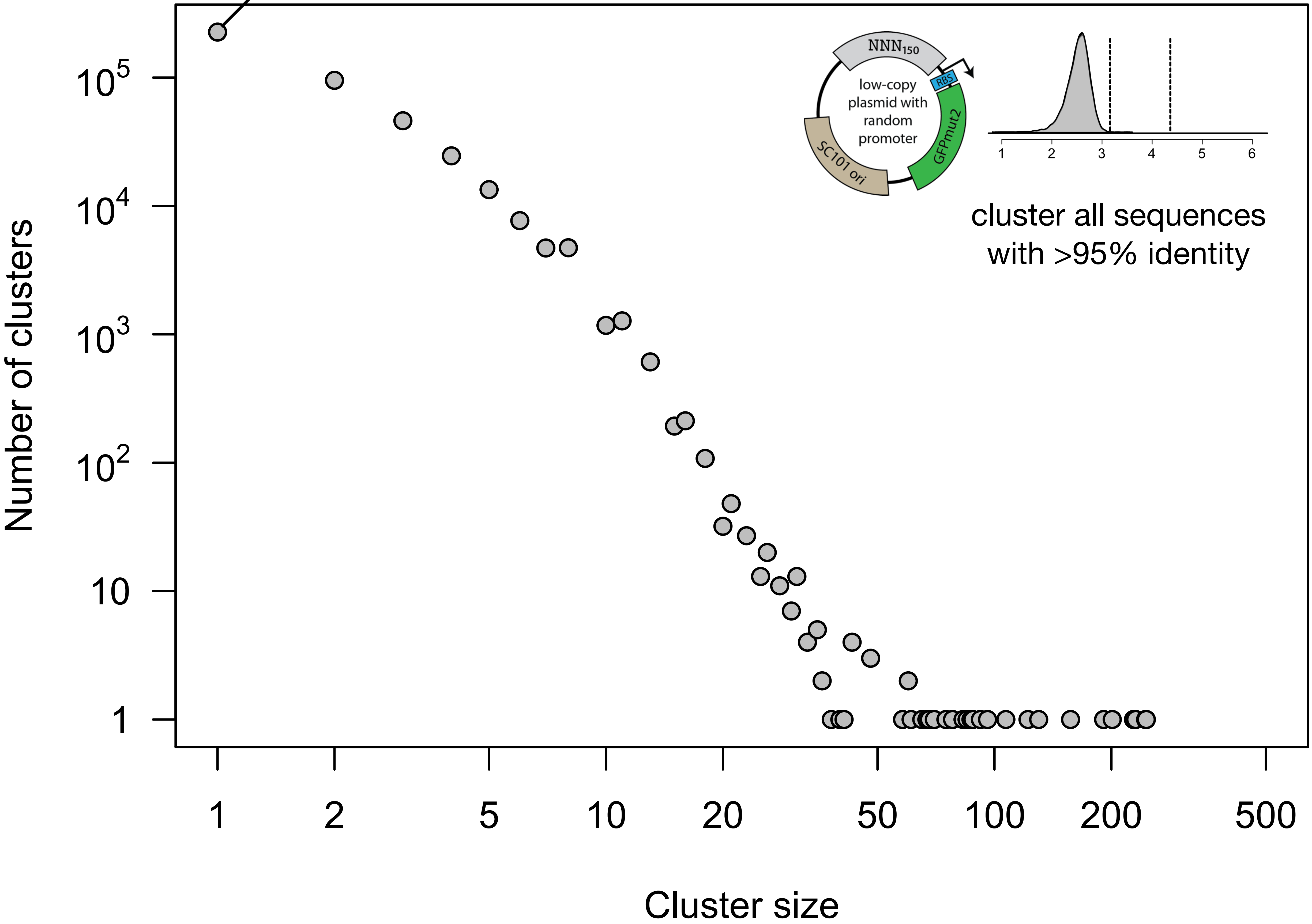
We start with a diverse initial promoter library



cluster all sequences with >95% identity

> 400’000 clusters

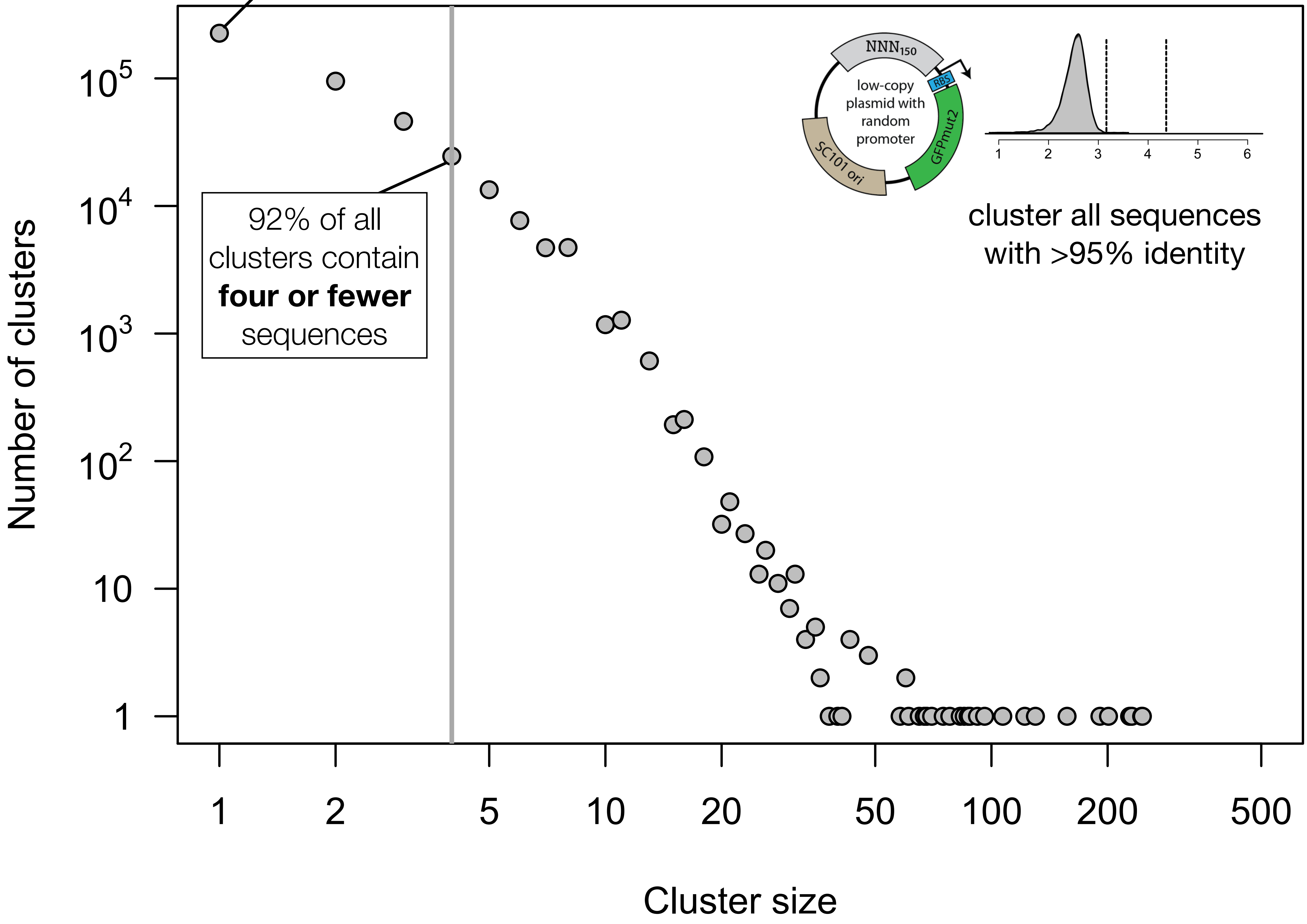
We start with a diverse initial promoter library



226'038 clusters contain **a single** sequence

cluster all sequences with >95% identity

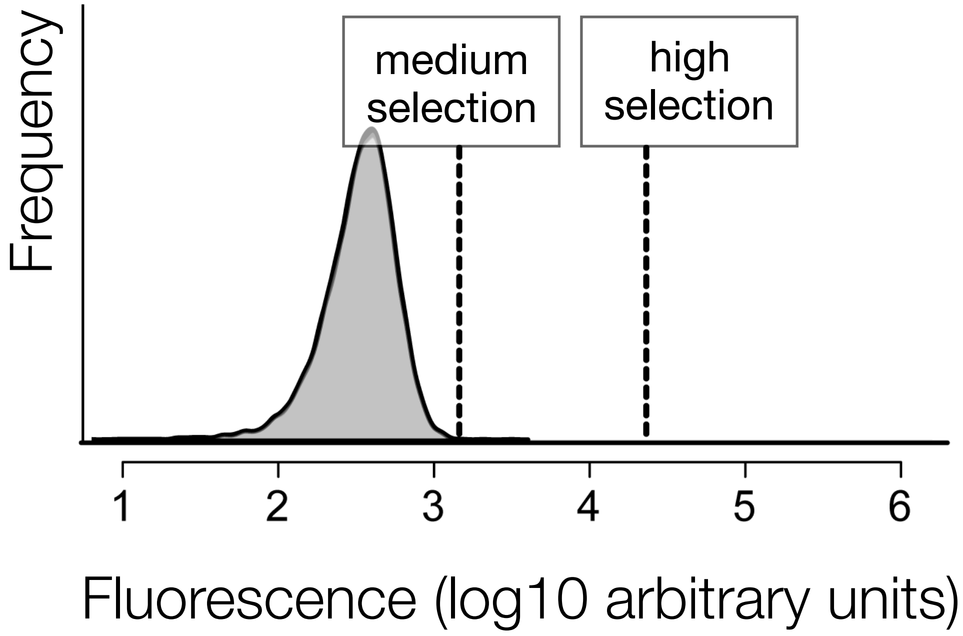
We start with a diverse initial promoter library



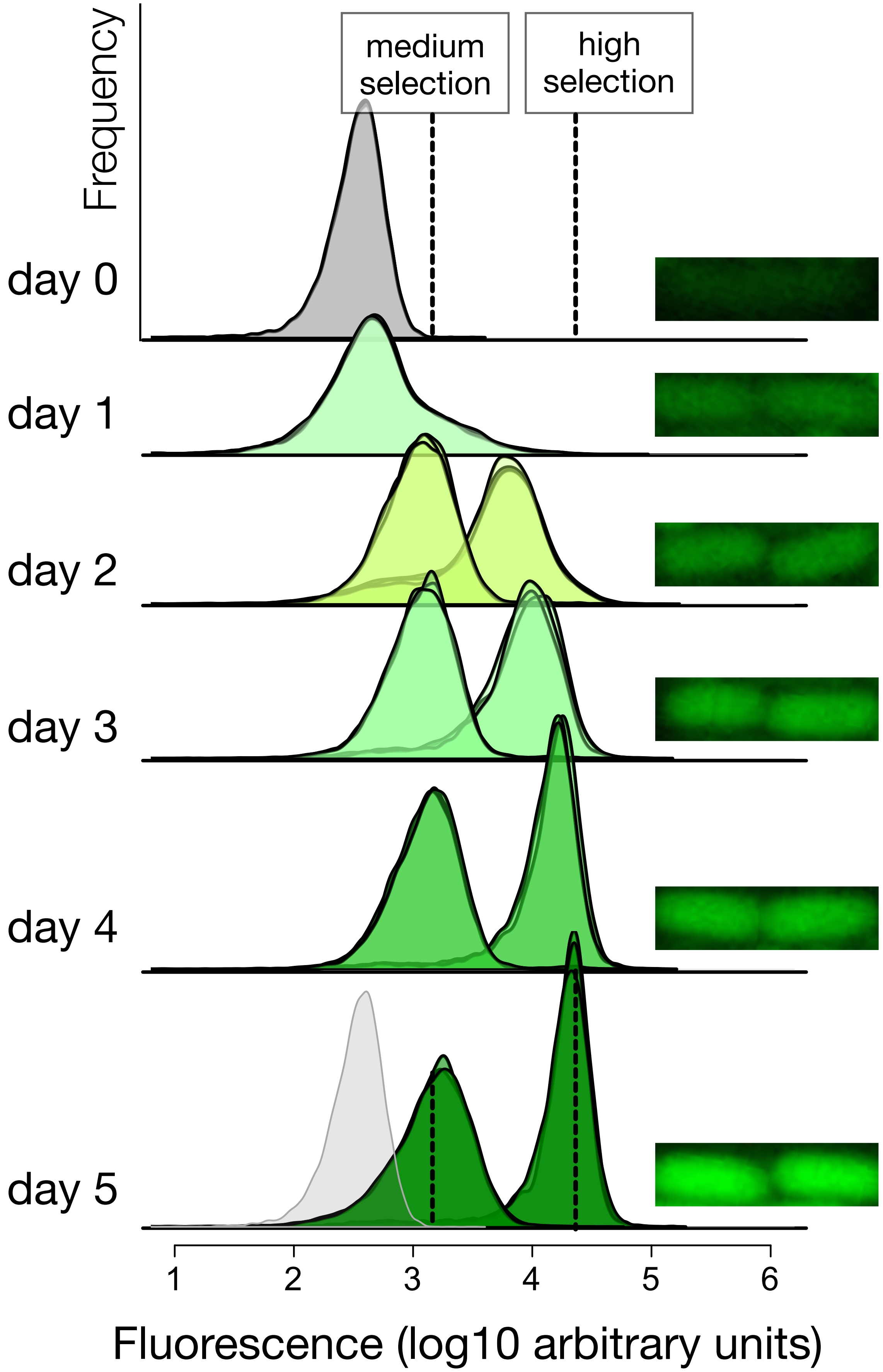
226’038 clusters contain **a single** sequence

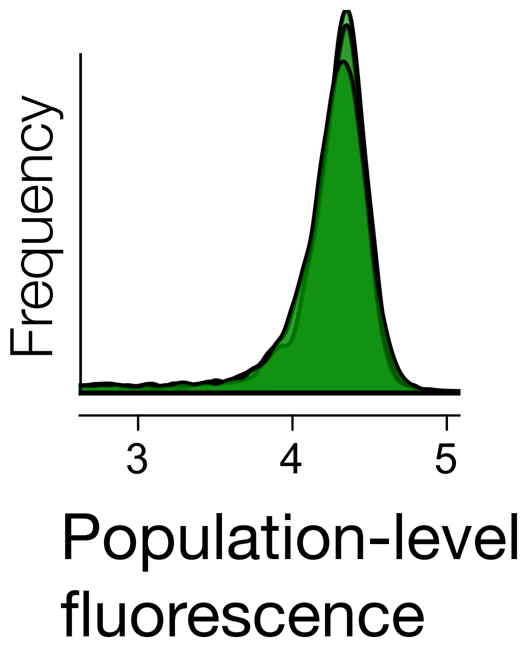
92% of all clusters contain **four or fewer** sequences

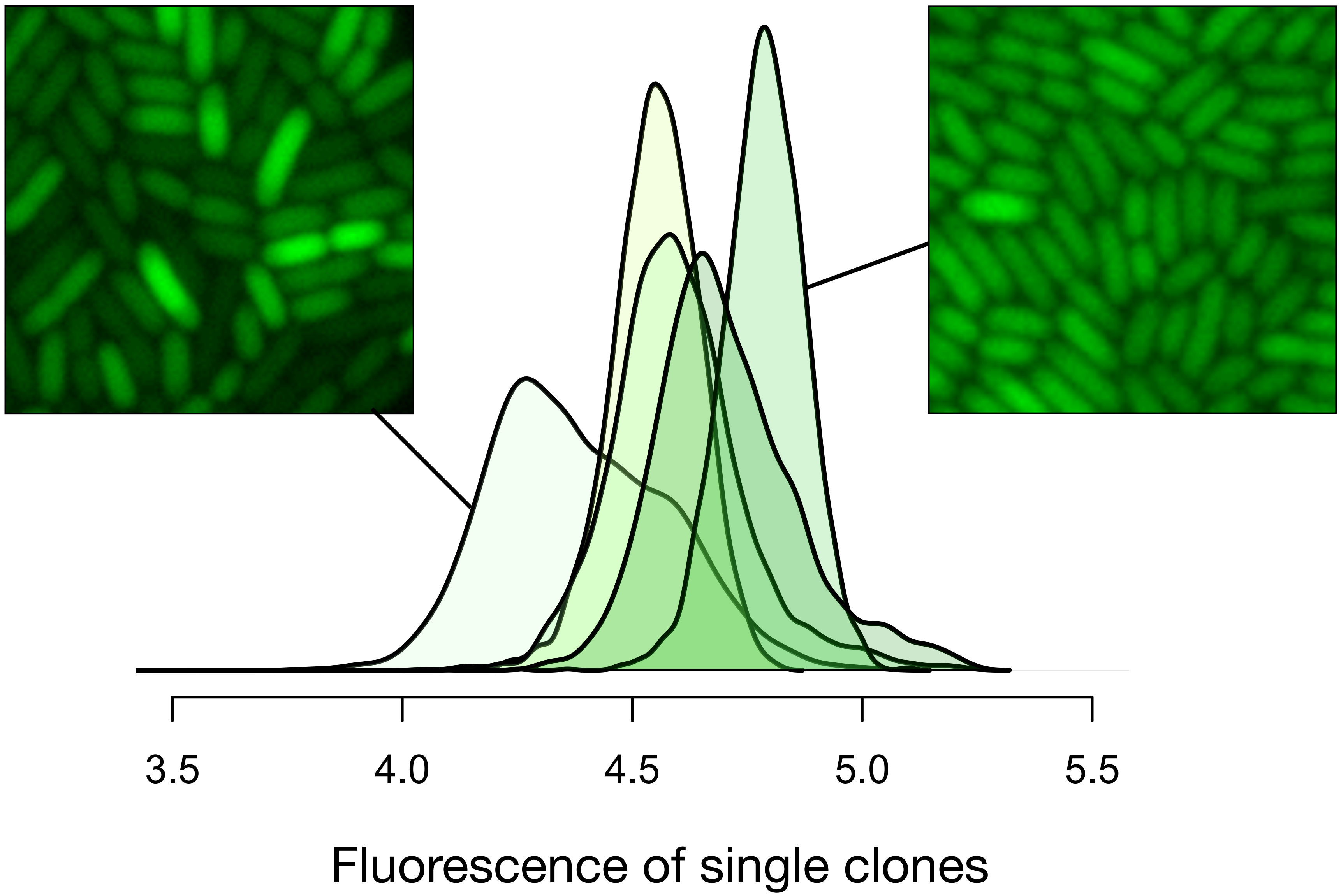
cluster all sequences with >95% identity


Frequency
medium selection
high selection
1
2
3
4
5
6
Fluorescence (log10 arbitrary units)

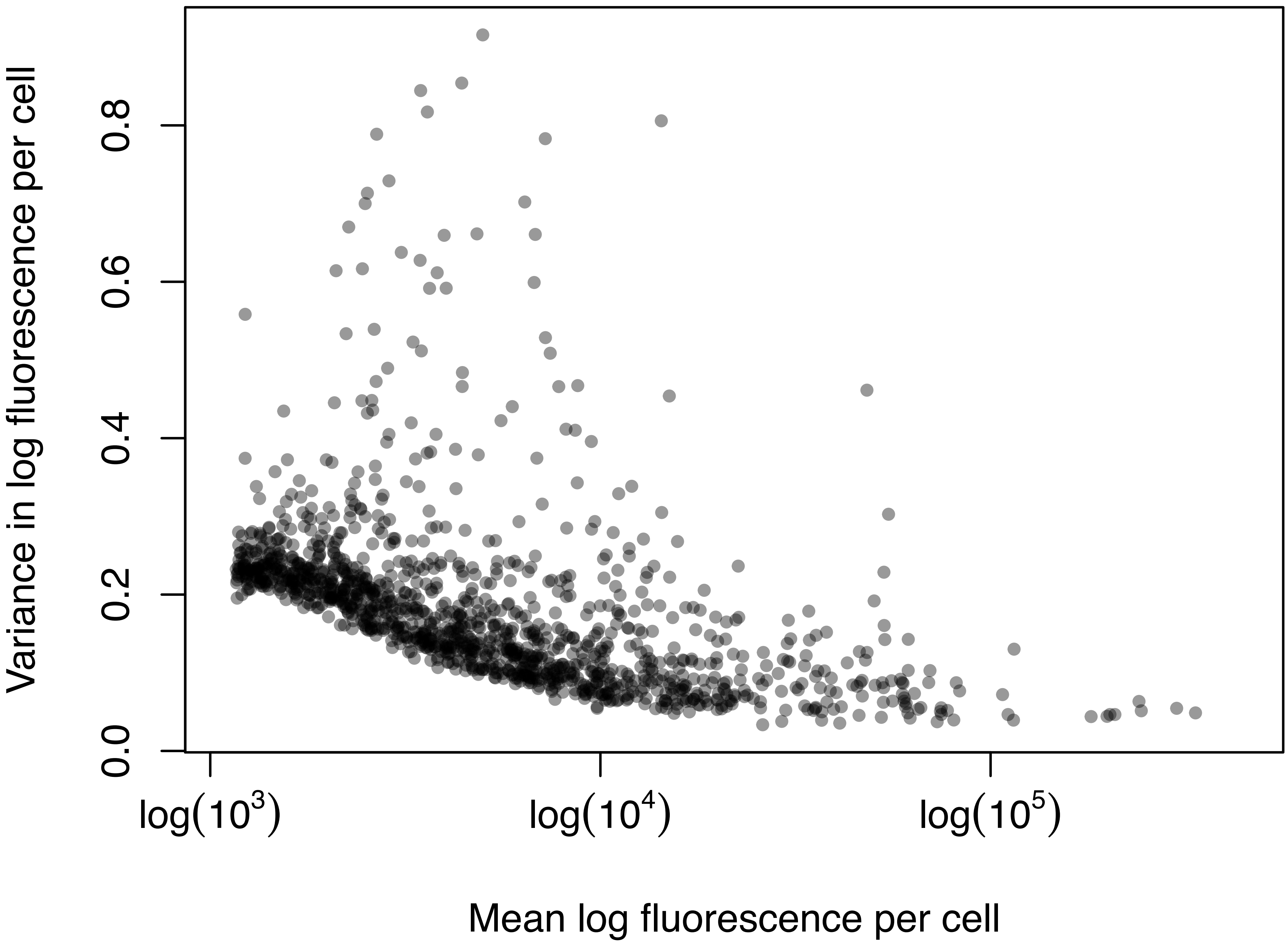
We quickly evolve functional promoters



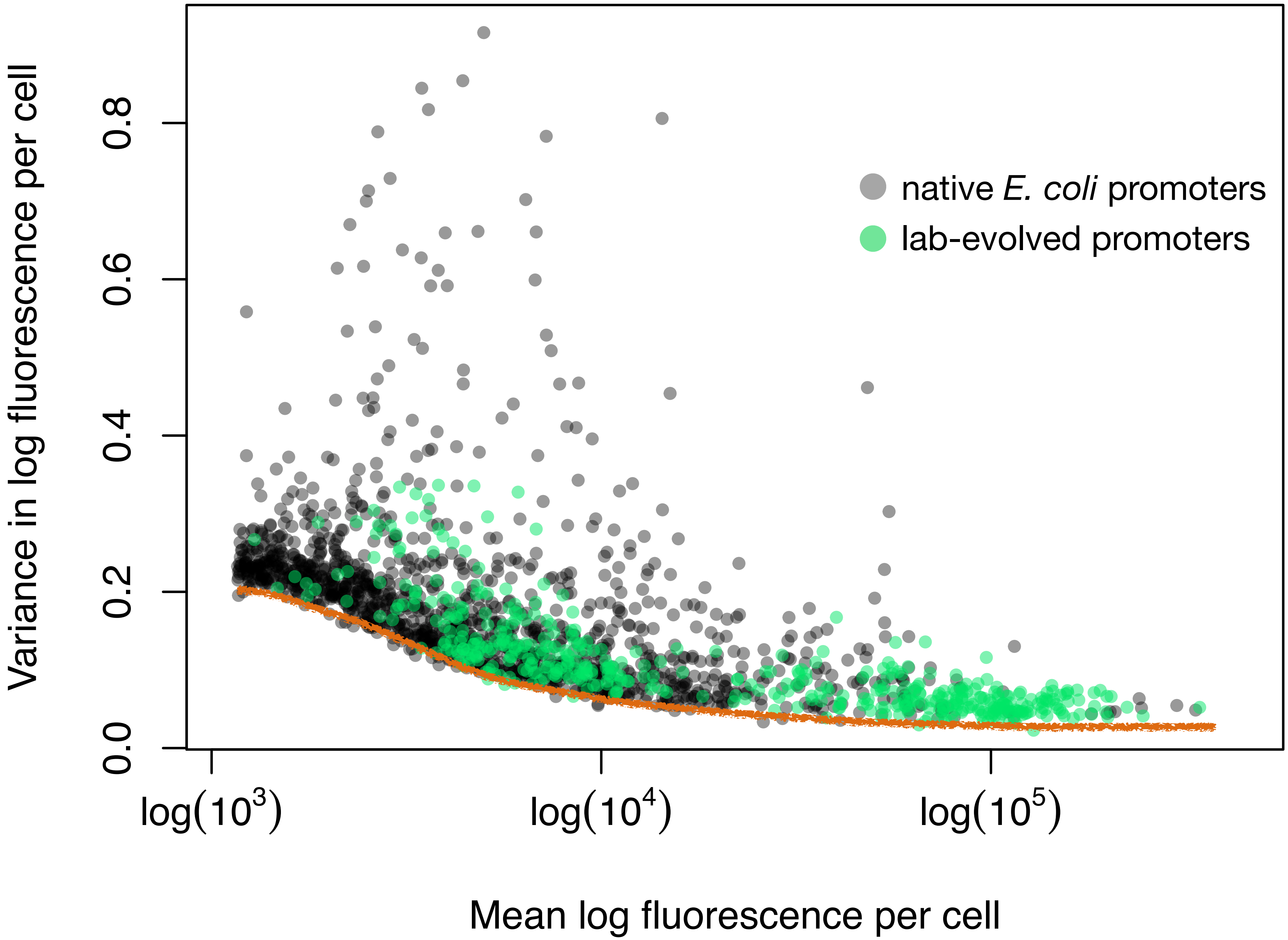

Frequency
3
4
5
Population-level
fluorescence


3.5
4.0
4.5
5.0
5.5
Fluorescence of single clones

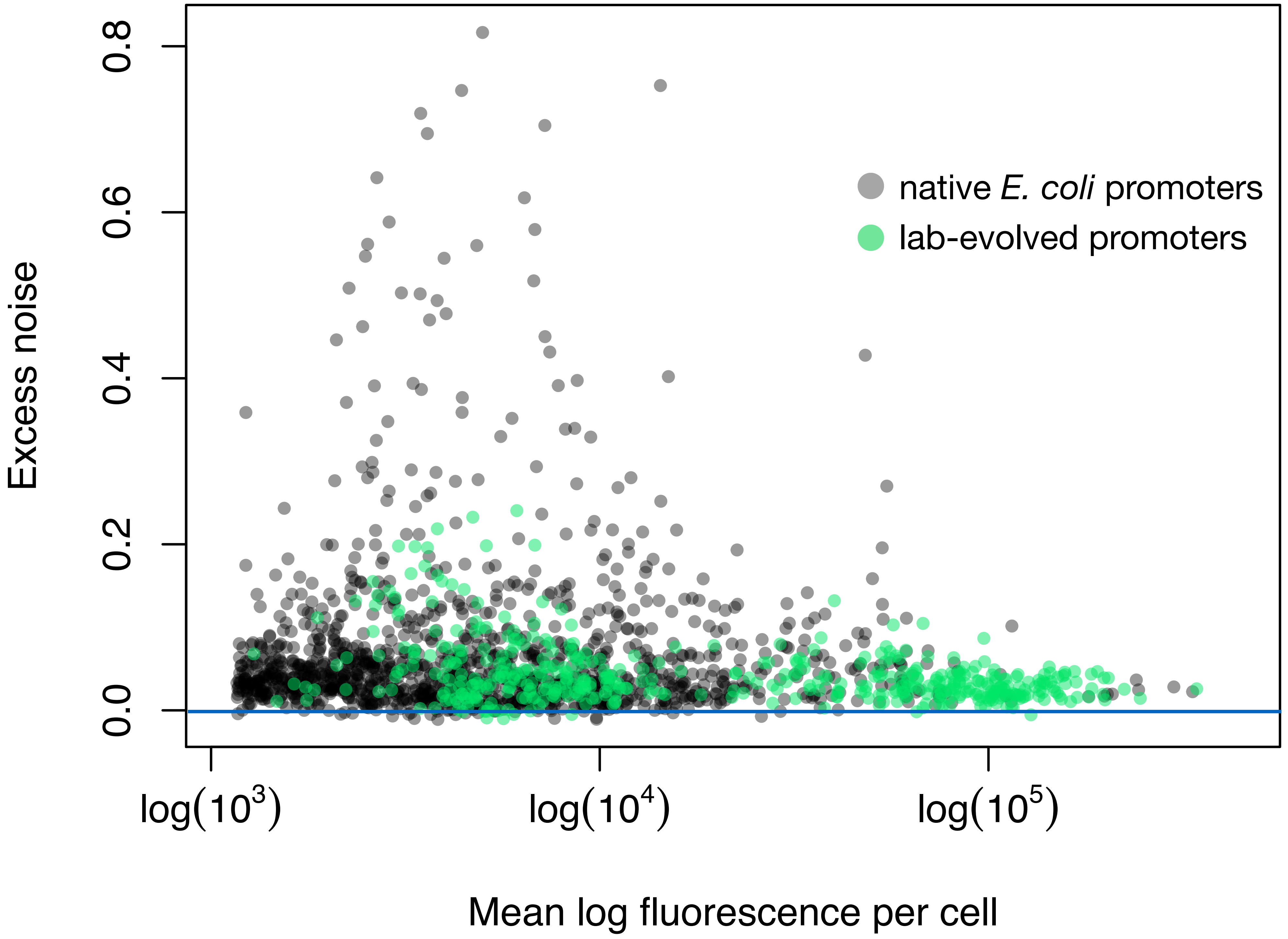
Noise is correlated with mean expression



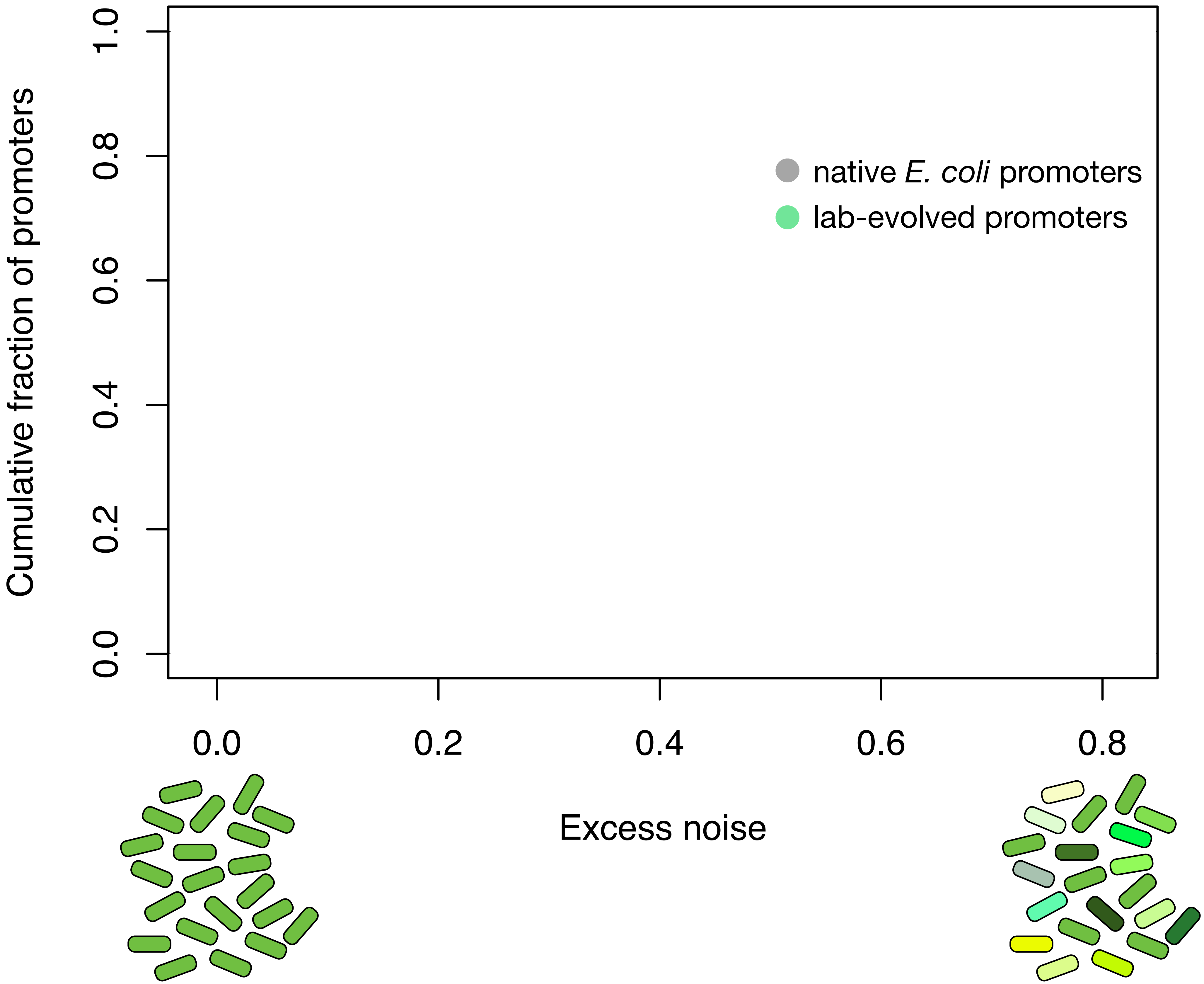
Noise is correlated with mean expression



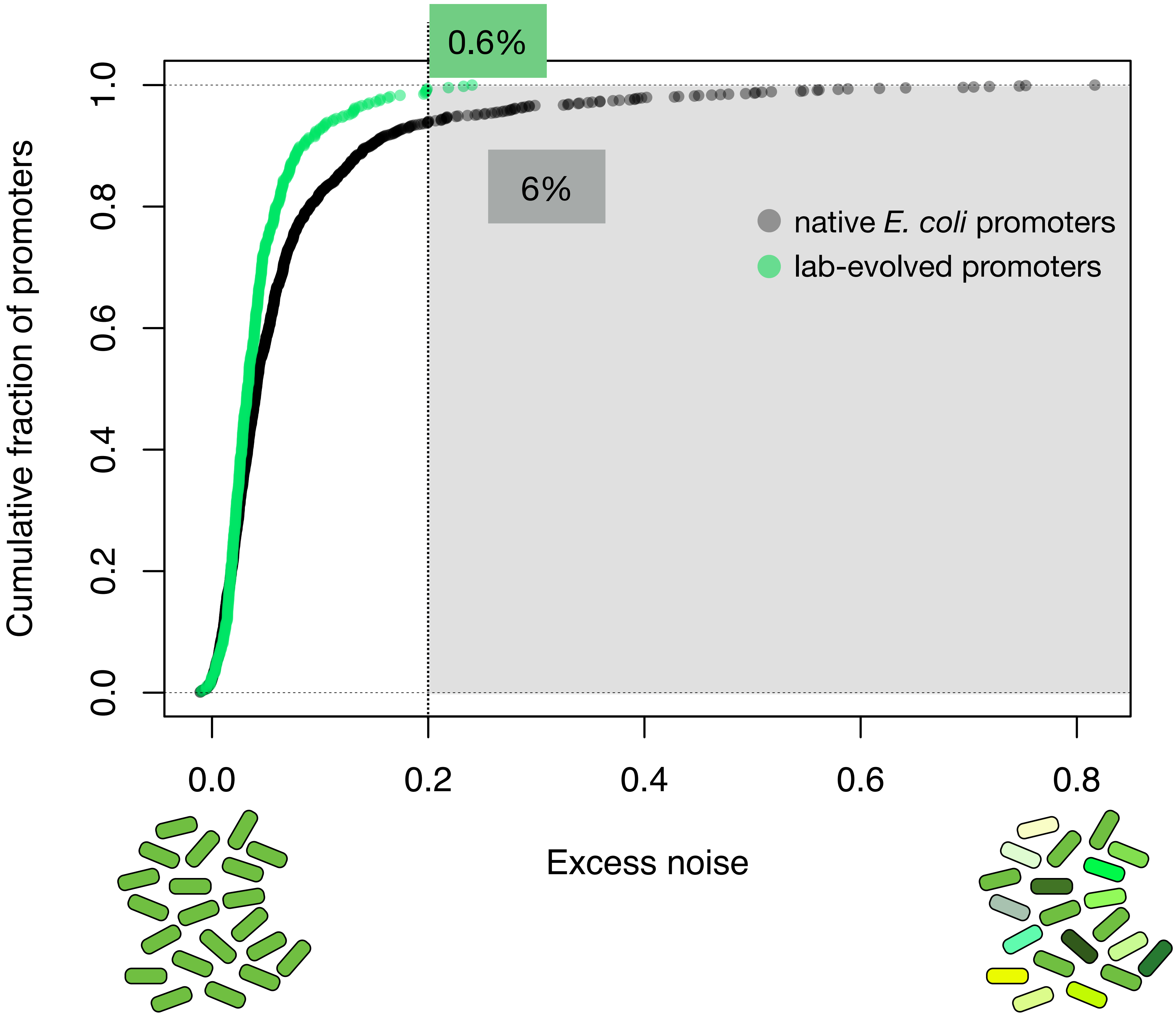
Excess noise is the variance above the lower bound



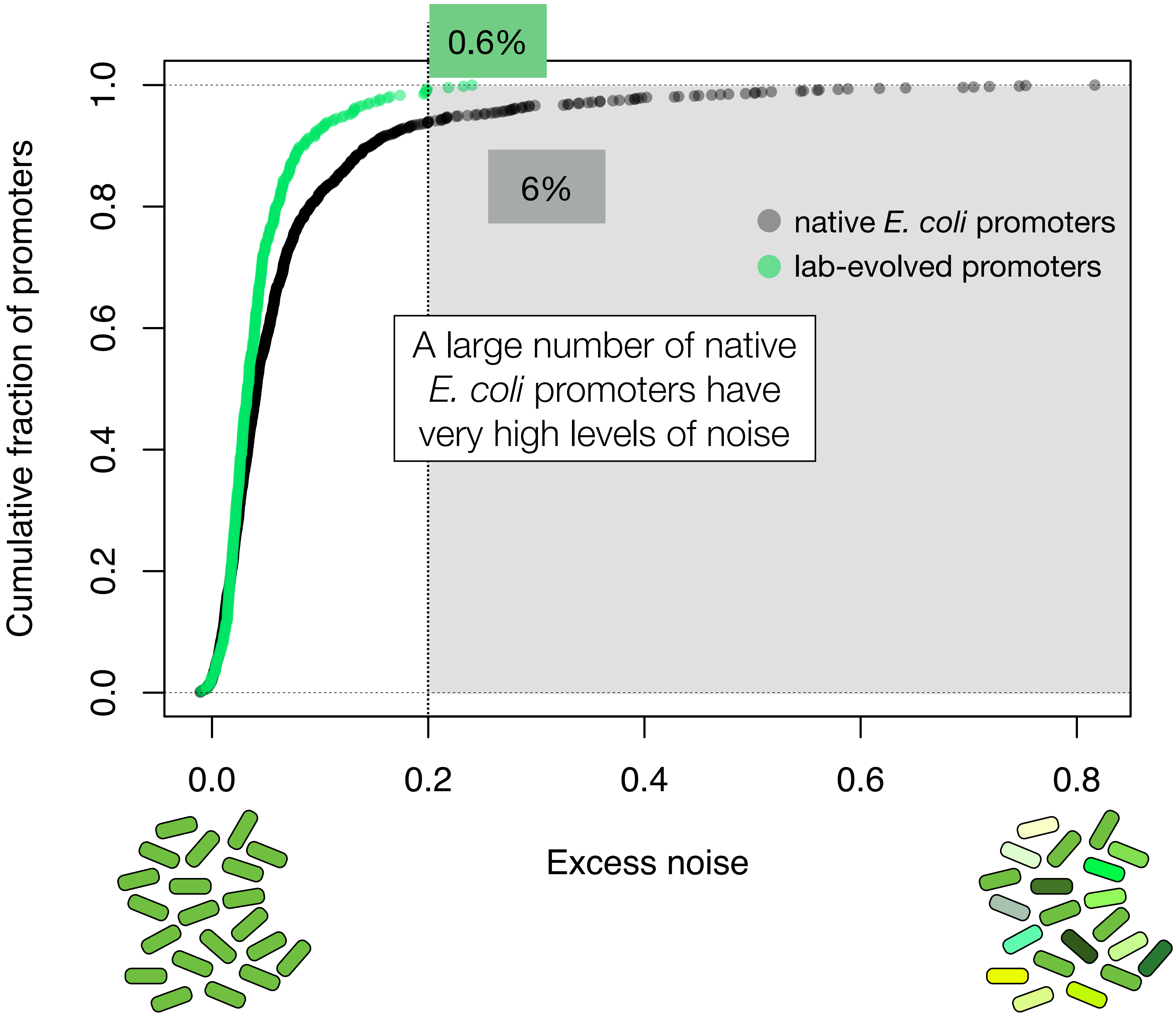
Native *E. coli* promoters exhibit higher noise



Native *E. coli* promoters exhibit higher noise

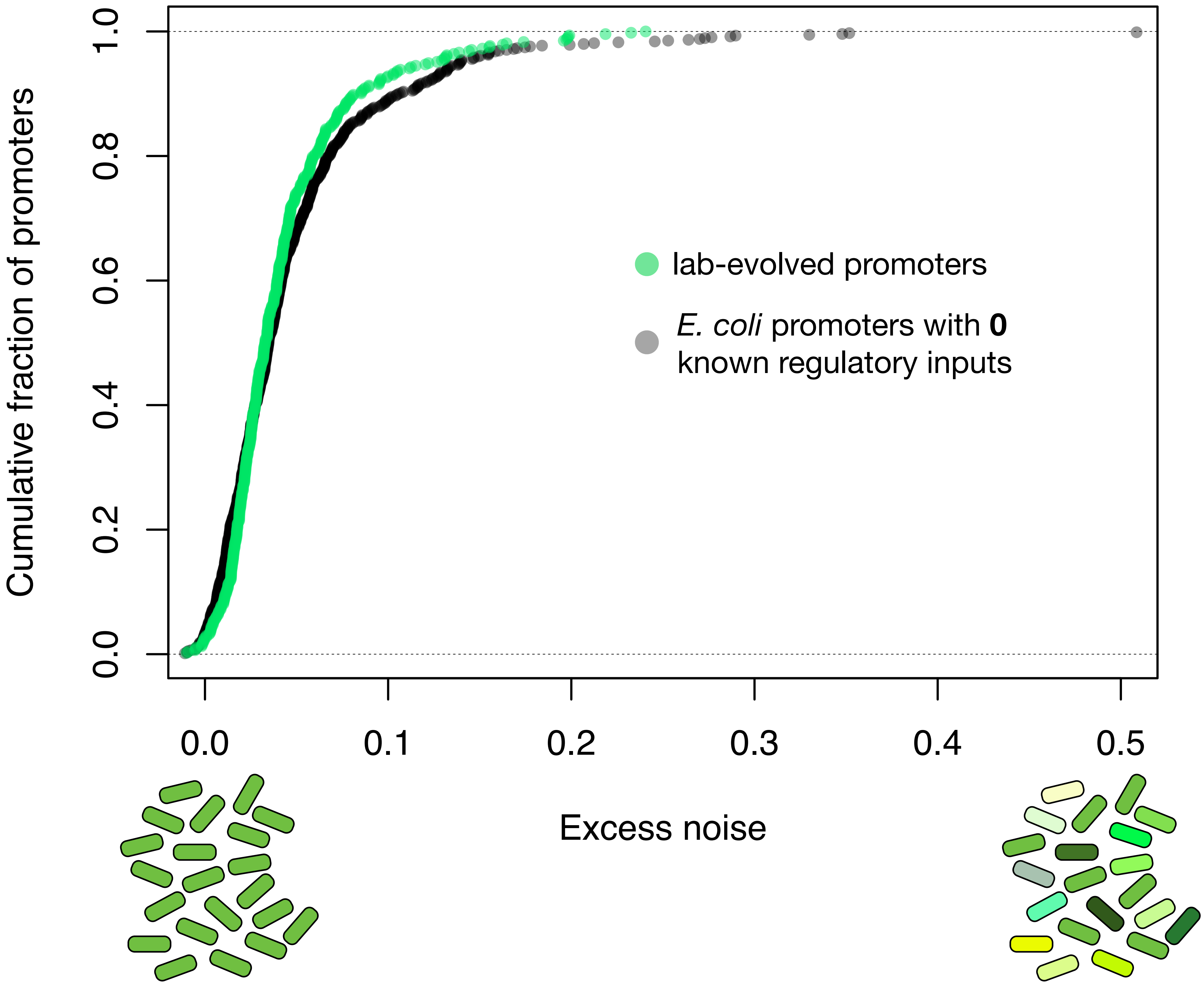


Native *E. coli* promoters exhibit higher noise

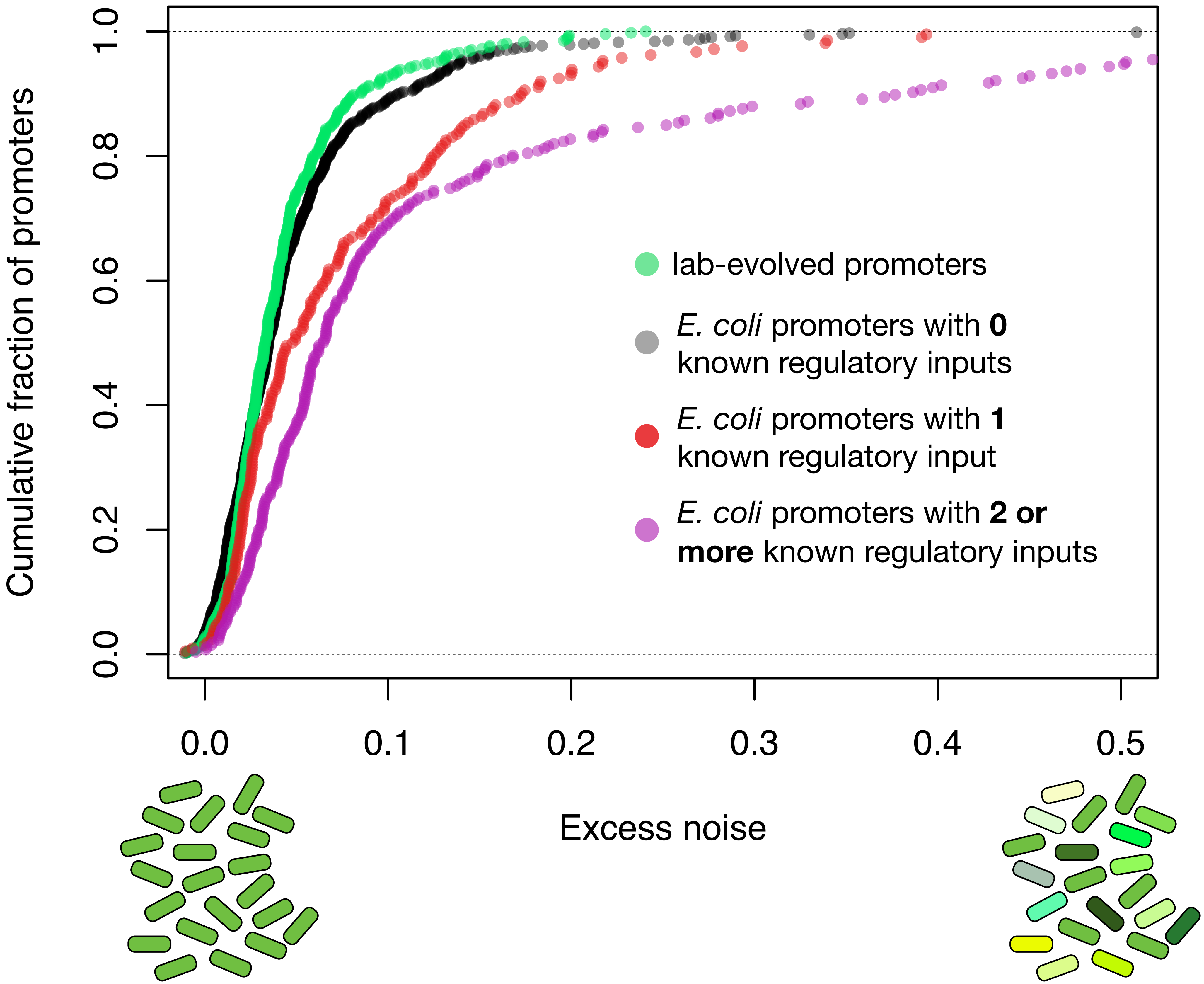


A large number of native *E. coli* promoters have very high levels of noise

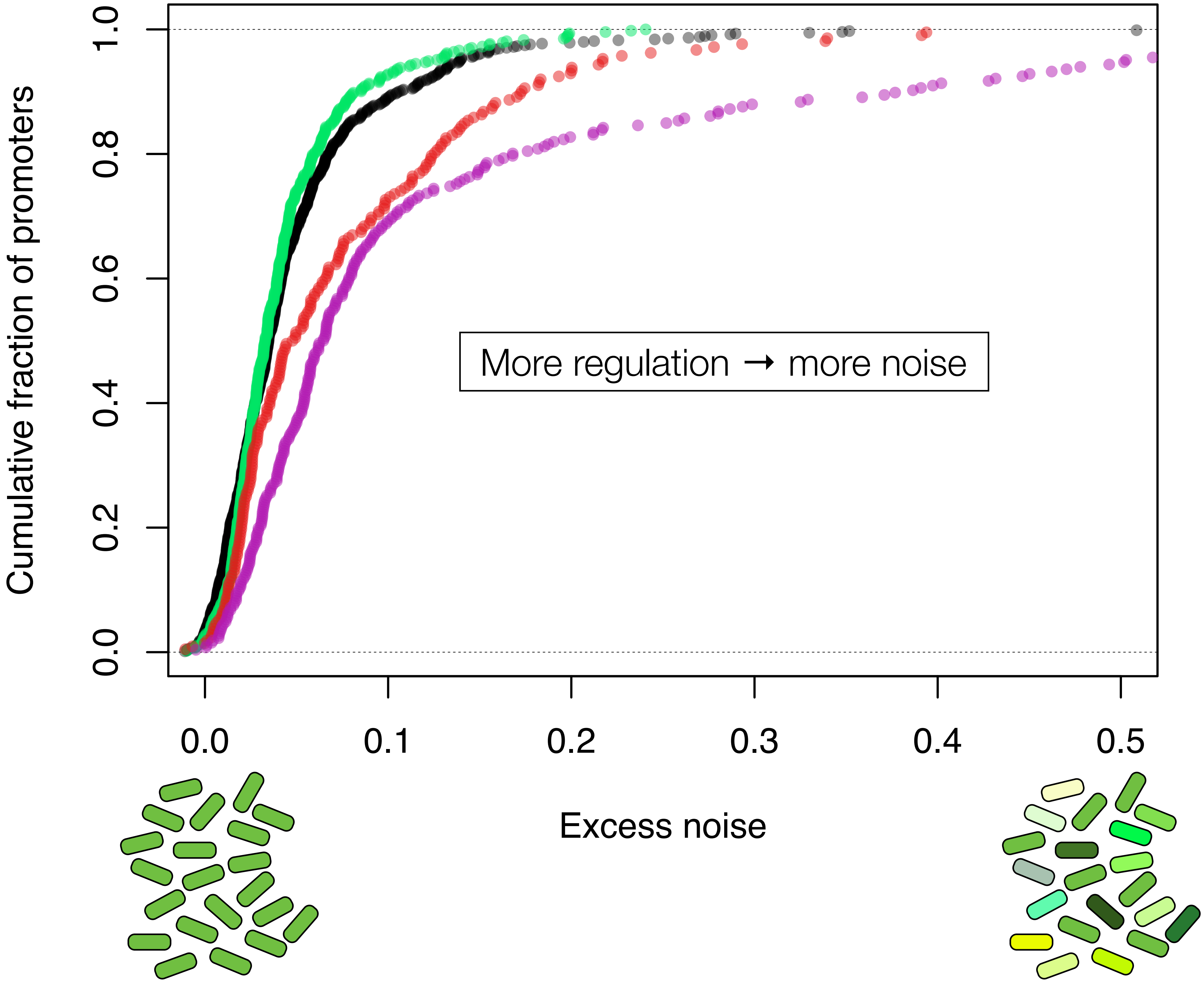
Noise levels correlate with regulatory inputs



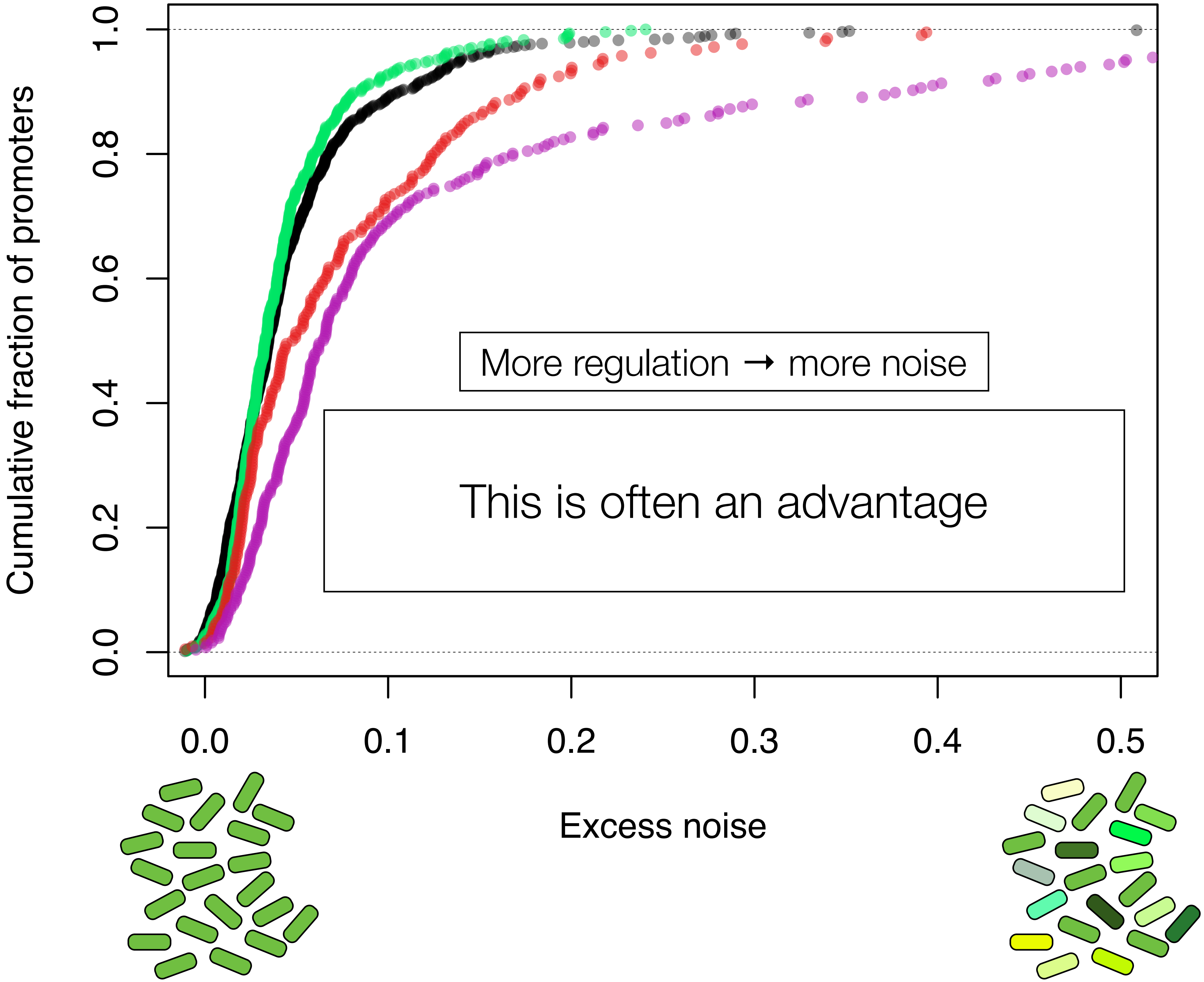
Noise levels correlate with regulatory inputs



Noise levels correlate with regulatory inputs

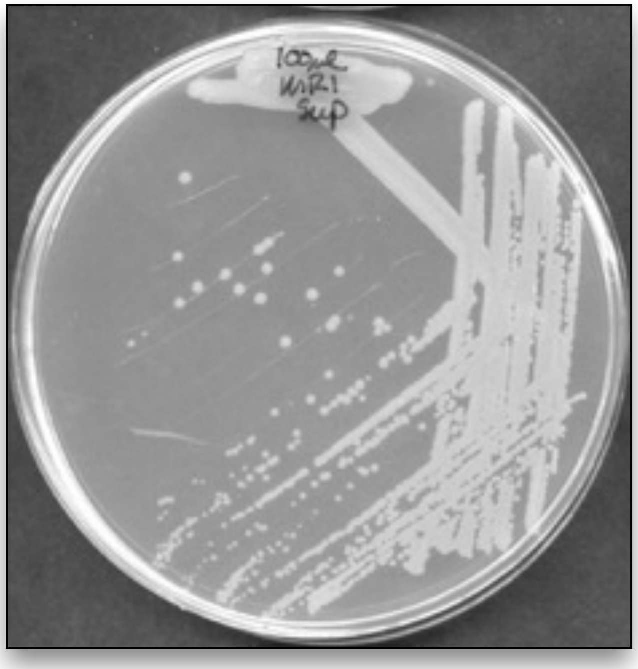


Noise levels correlate with regulatory inputs



More regulation → more noise

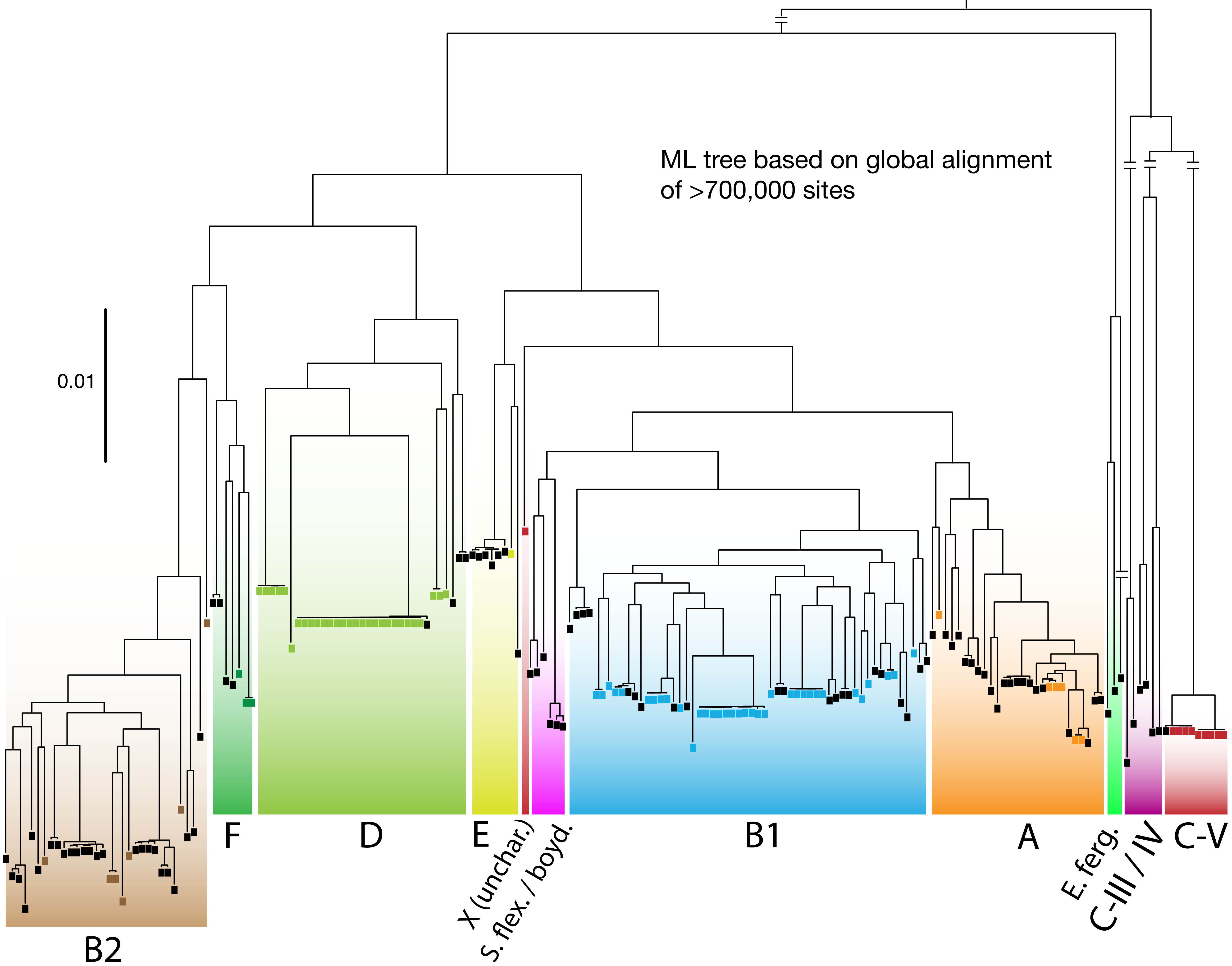
This is often an advantage



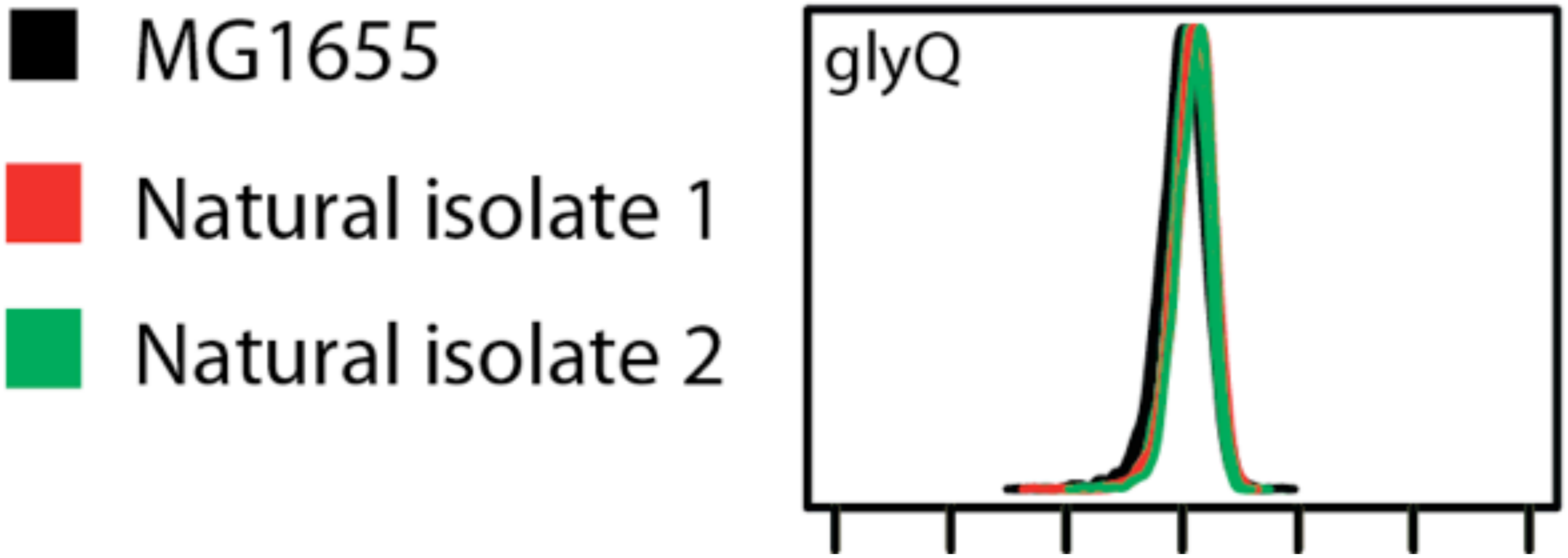
Escherichia coli K12 MG1655
model laboratory strain since 1922

Duluth

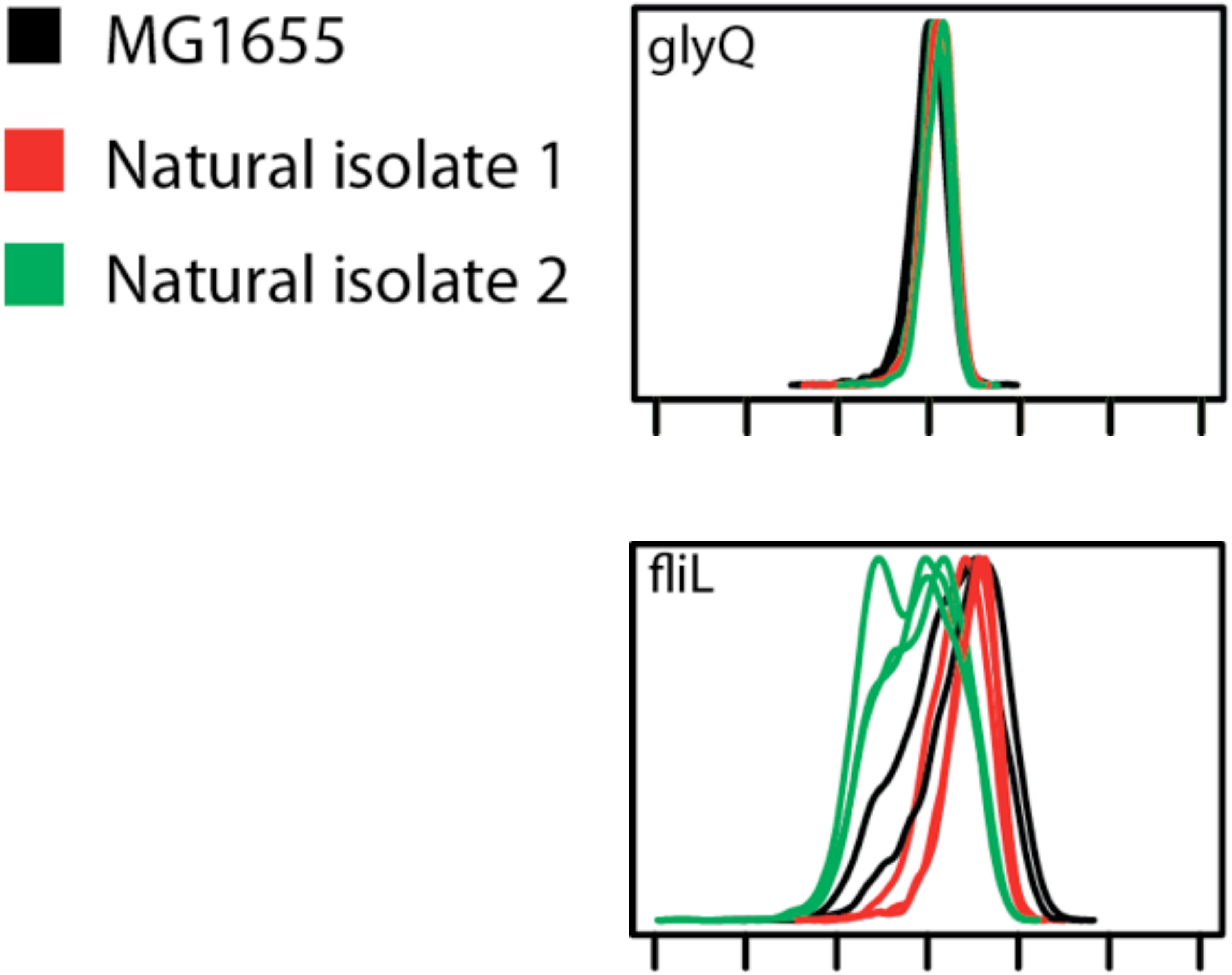
Mike Sadowsky
(Univ. Minnesota)


ML tree based on global alignment of >700,000 sites
0.01
B2
F
D
E
X (unchar.)
S. flex. / boyd.
B1
A
E. ferg.
C-III / IV
C-V

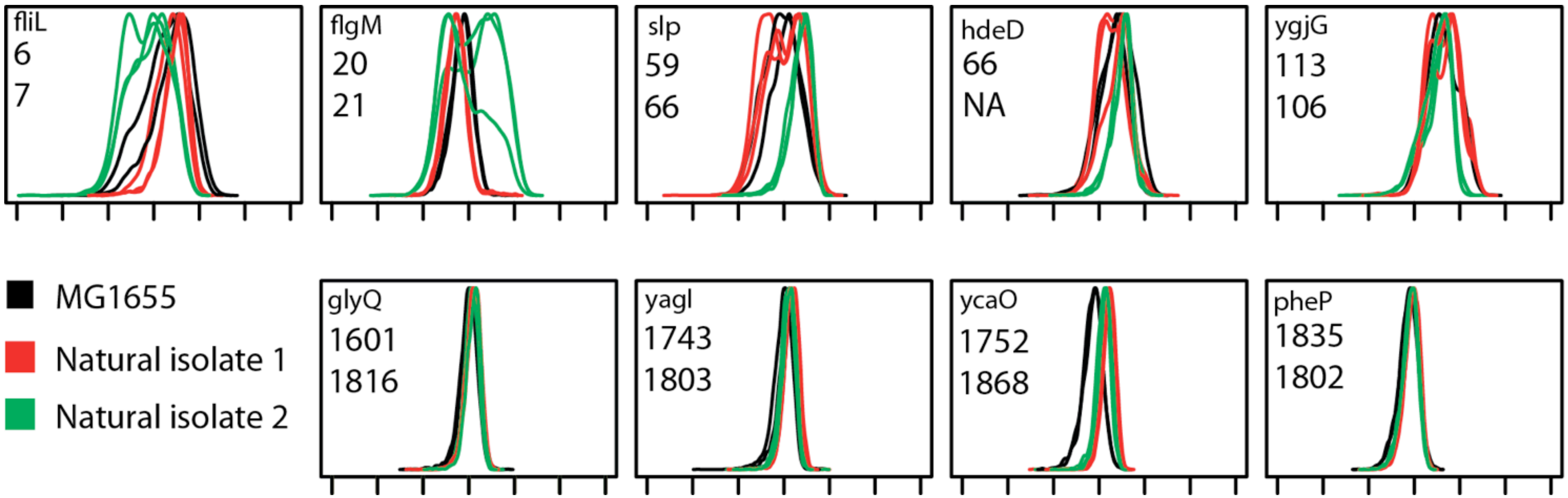
How does transcriptional control of “noisy” genes evolve?

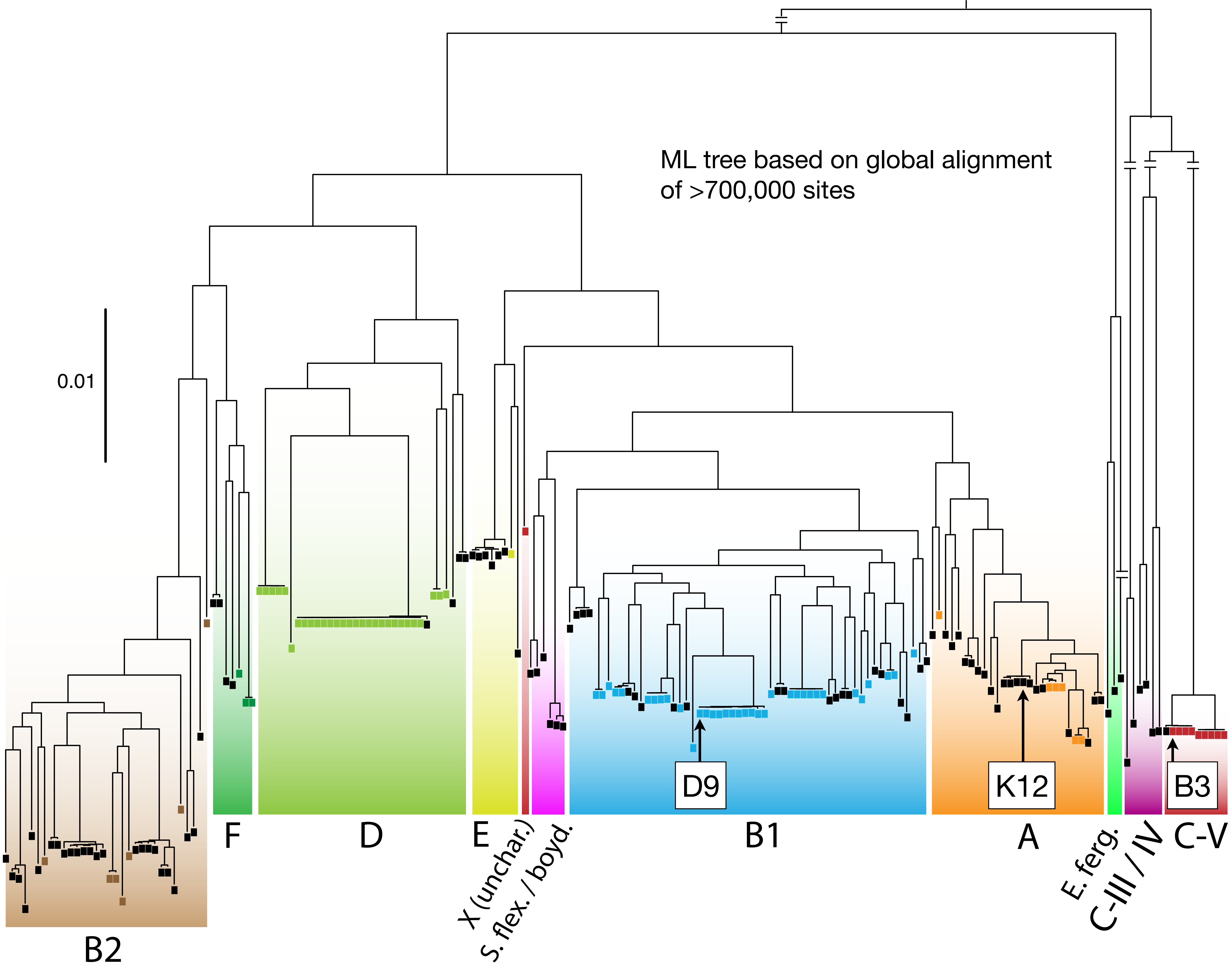


How does transcriptional control of “noisy” genes evolve?



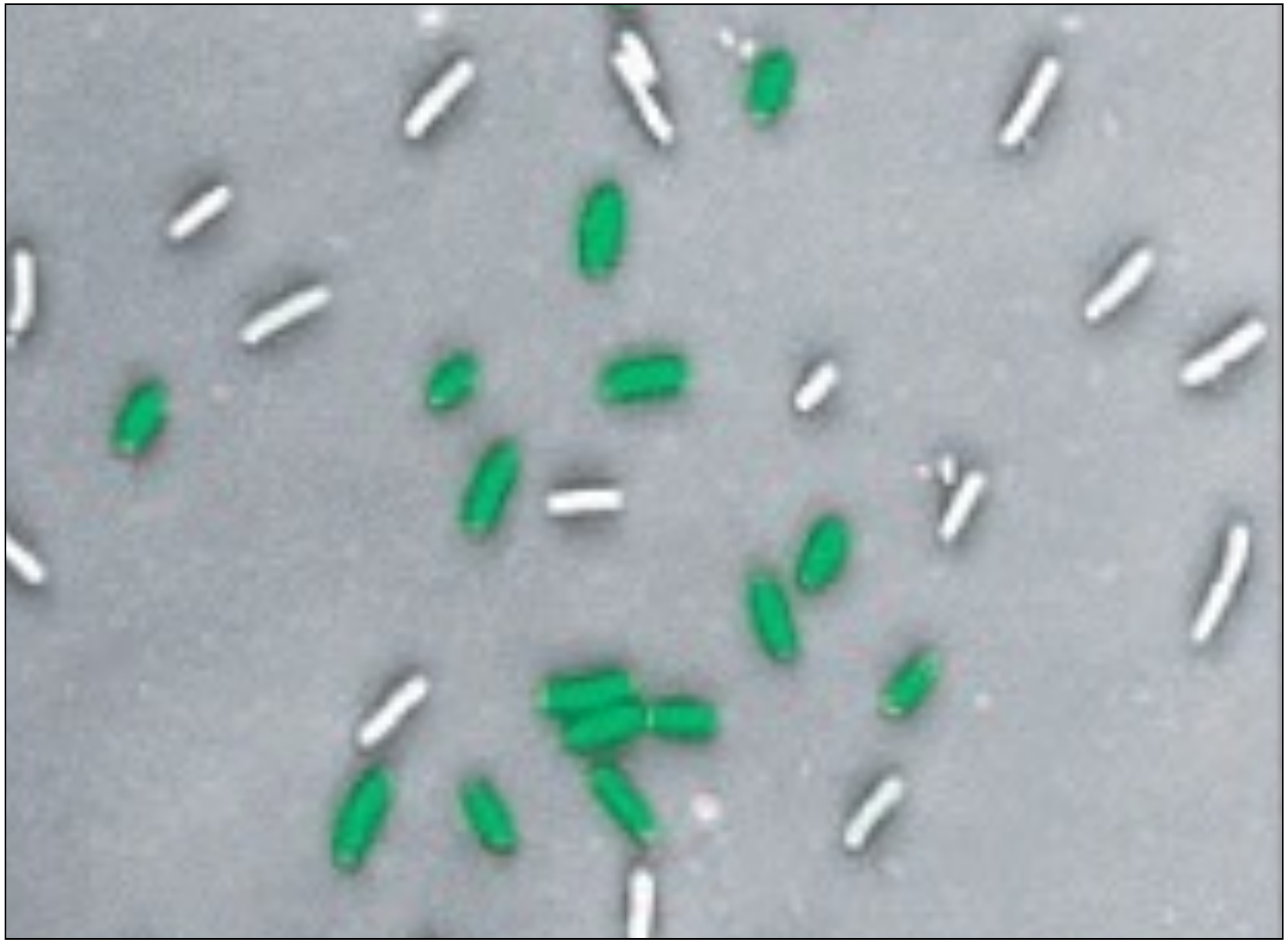
How does transcriptional control of “noisy” genes evolve?




ML tree based on global alignment of >700,000 sites
0.01
D9
K12
B3
B2
F
D
E
X (unchar.)
S. flex. / boyd.
B1
A
E. ferg.
C-III / IV
C-V

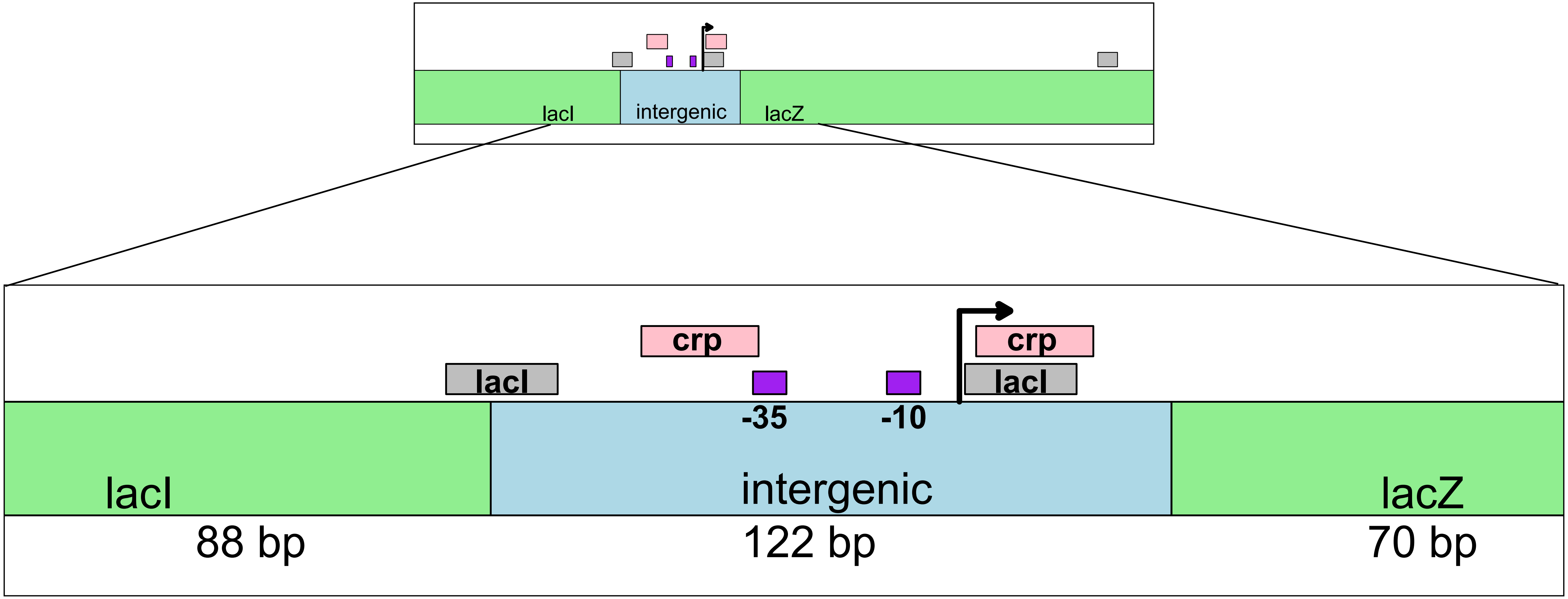
A canonical example of noisy regulation

IPTG induction of the lac operon



Ozbudak et al. (2004) *Nature Genetics*

The lac operon structure



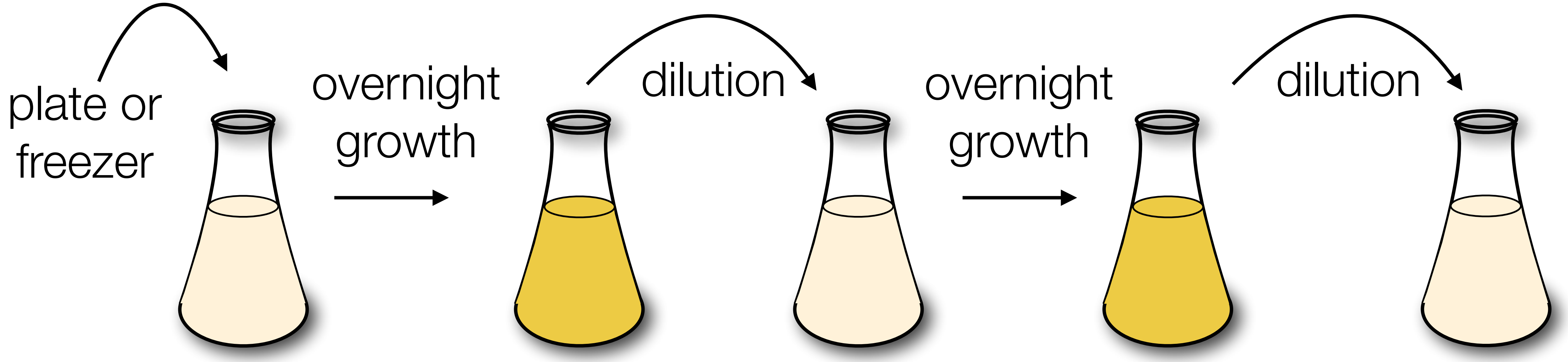


plate or freezer

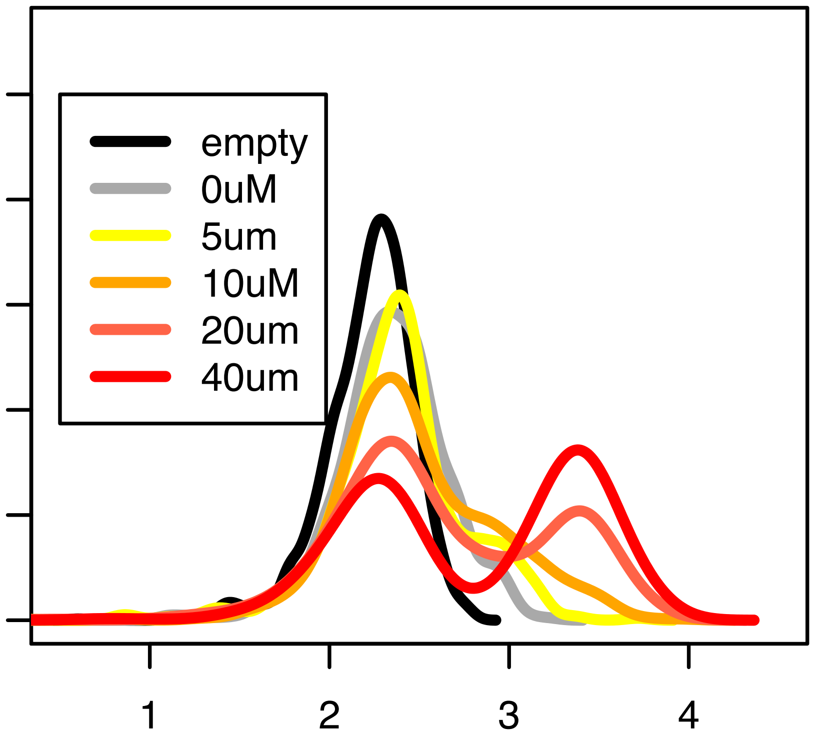
overnight growth

dilution

overnight growth

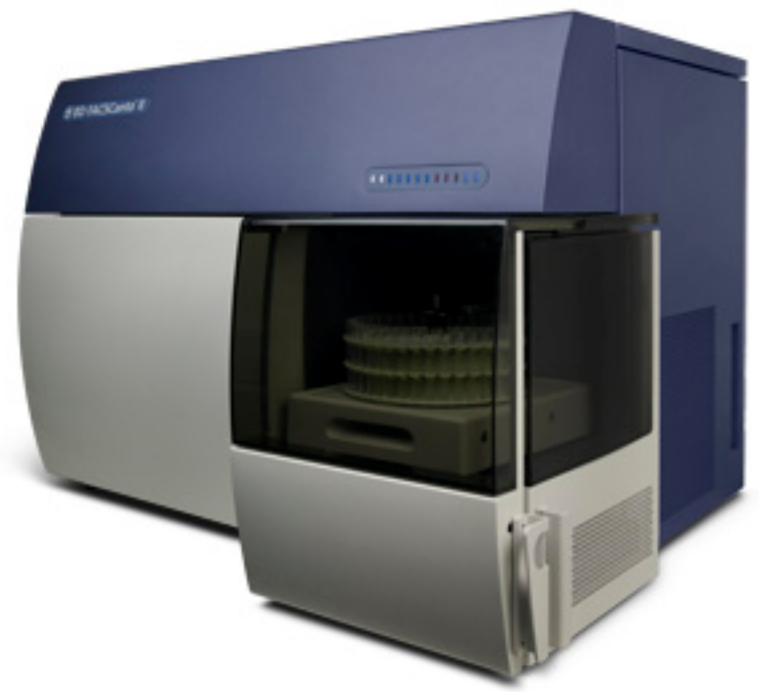
dilution

growth in inducing condition



information for 100,000 cells

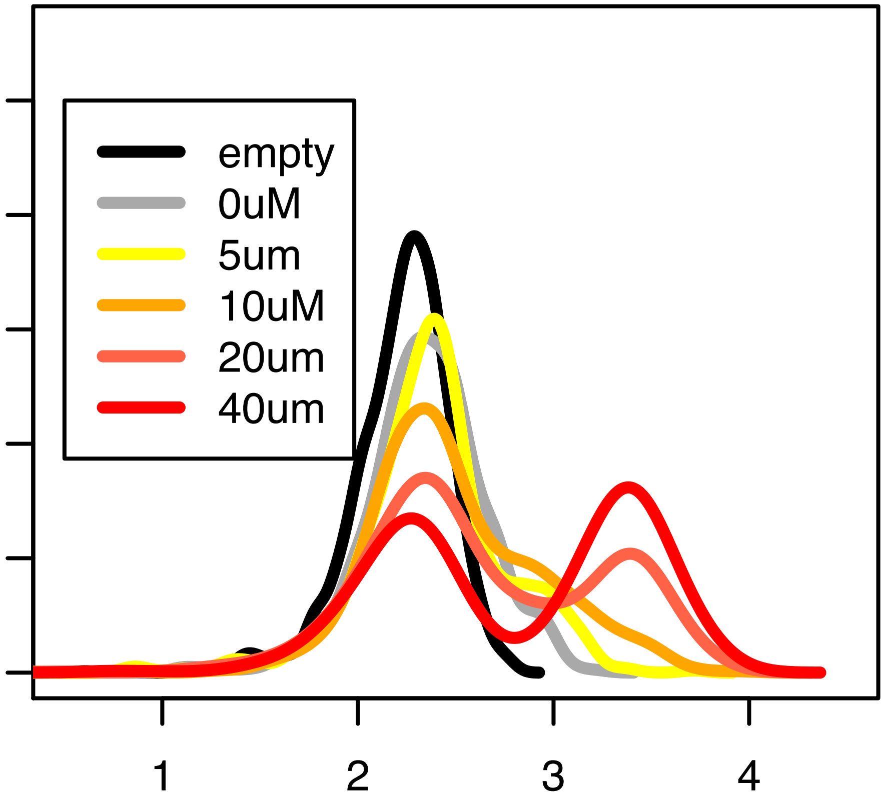
filter based on FSC and SSC



flow cytometry

The K12 lac operon activity exhibits bimodality

E. coli **K12 lac operon** (p*lacZ*::GFP)
Minimal glucose media
6 hr induction with lactose analogue IPTG

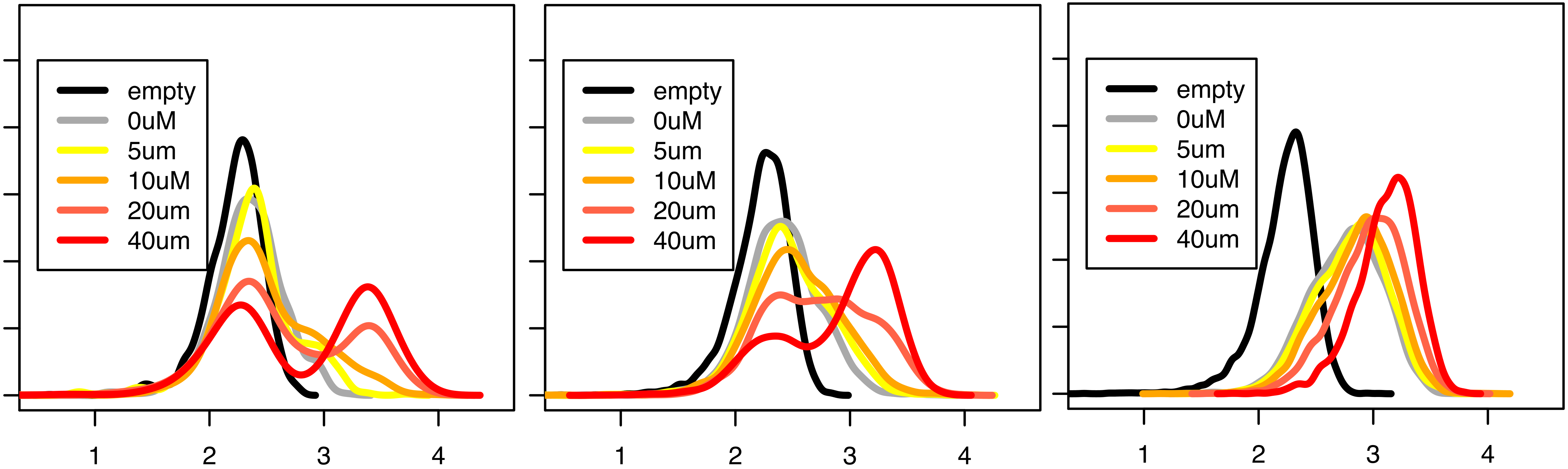


K12

strong bimodal shift

Noise levels differ between natural isolates

E. coli K12 lac operon (p*lacZ*::GFP)
Minimal glucose media
6 hr induction with lactose analogue IPTG

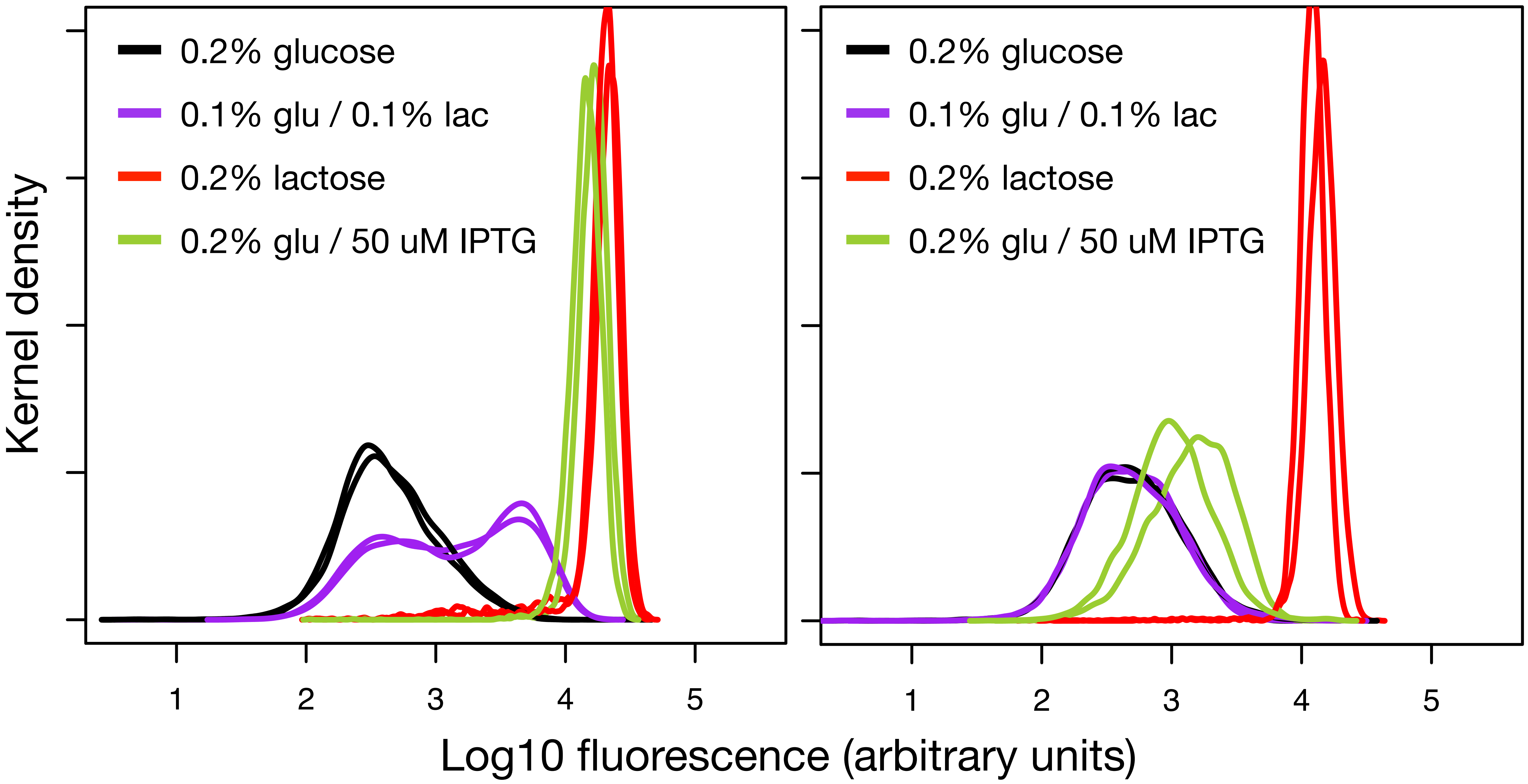


K12
strong bimodal shift

B1
weak bimodal shift

C-V
weak monomodal shift

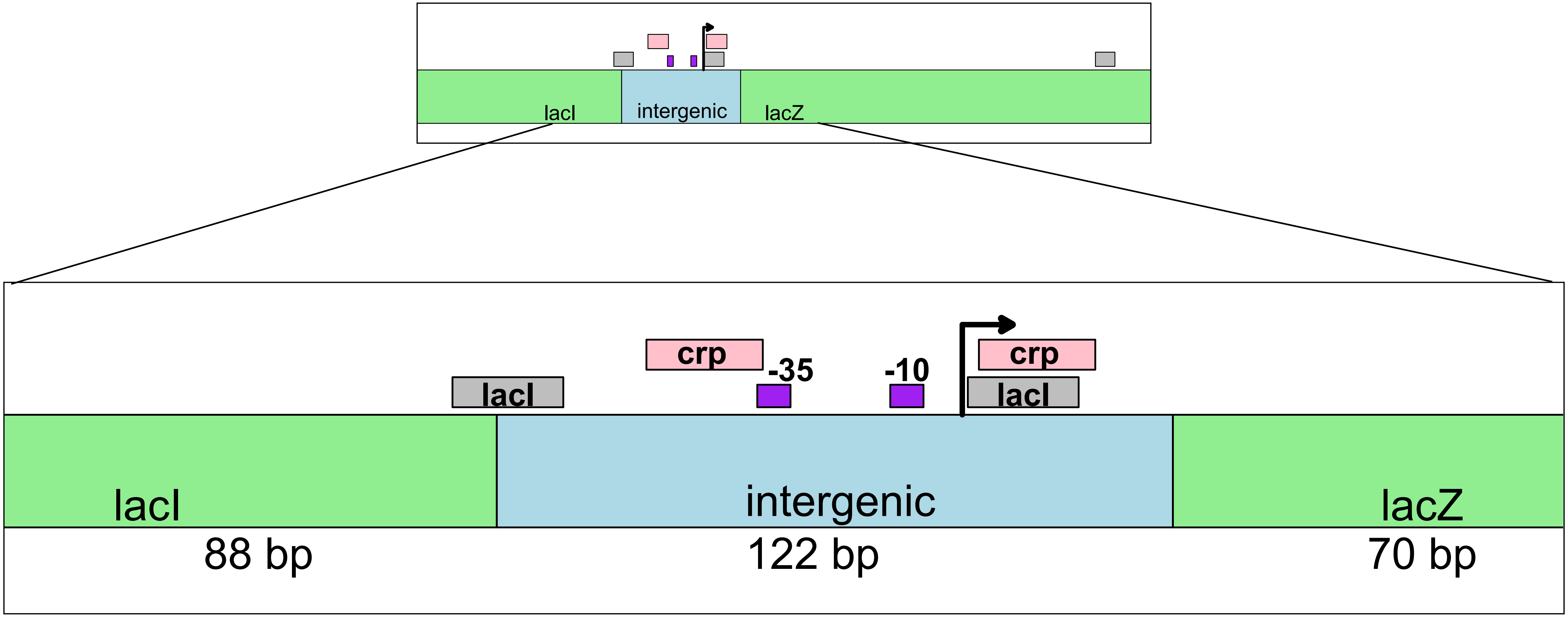
Lactose sensitivity differs between strains



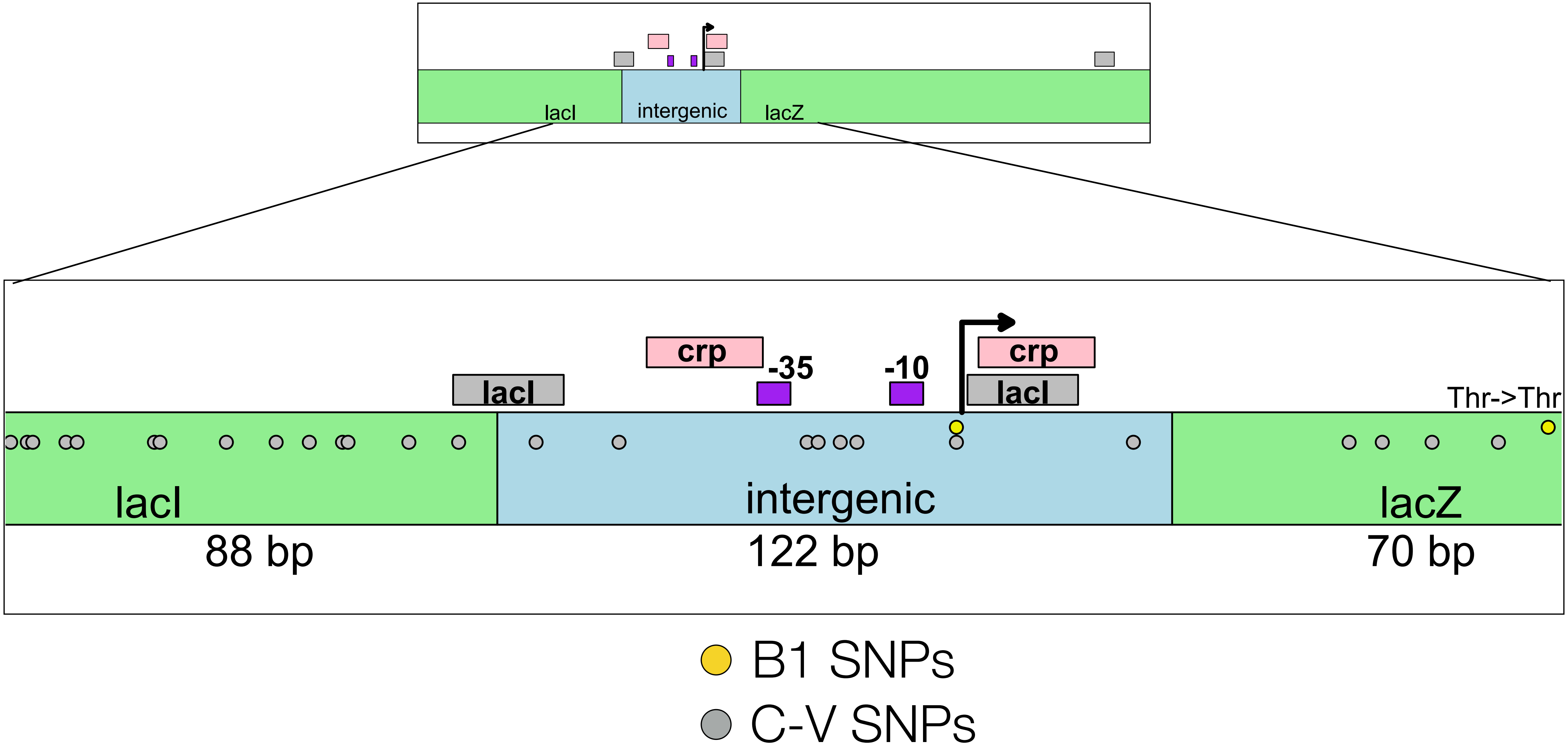
B1 with native p*lacZ*
sensitive

C-V with native p*lacZ*
insensitive

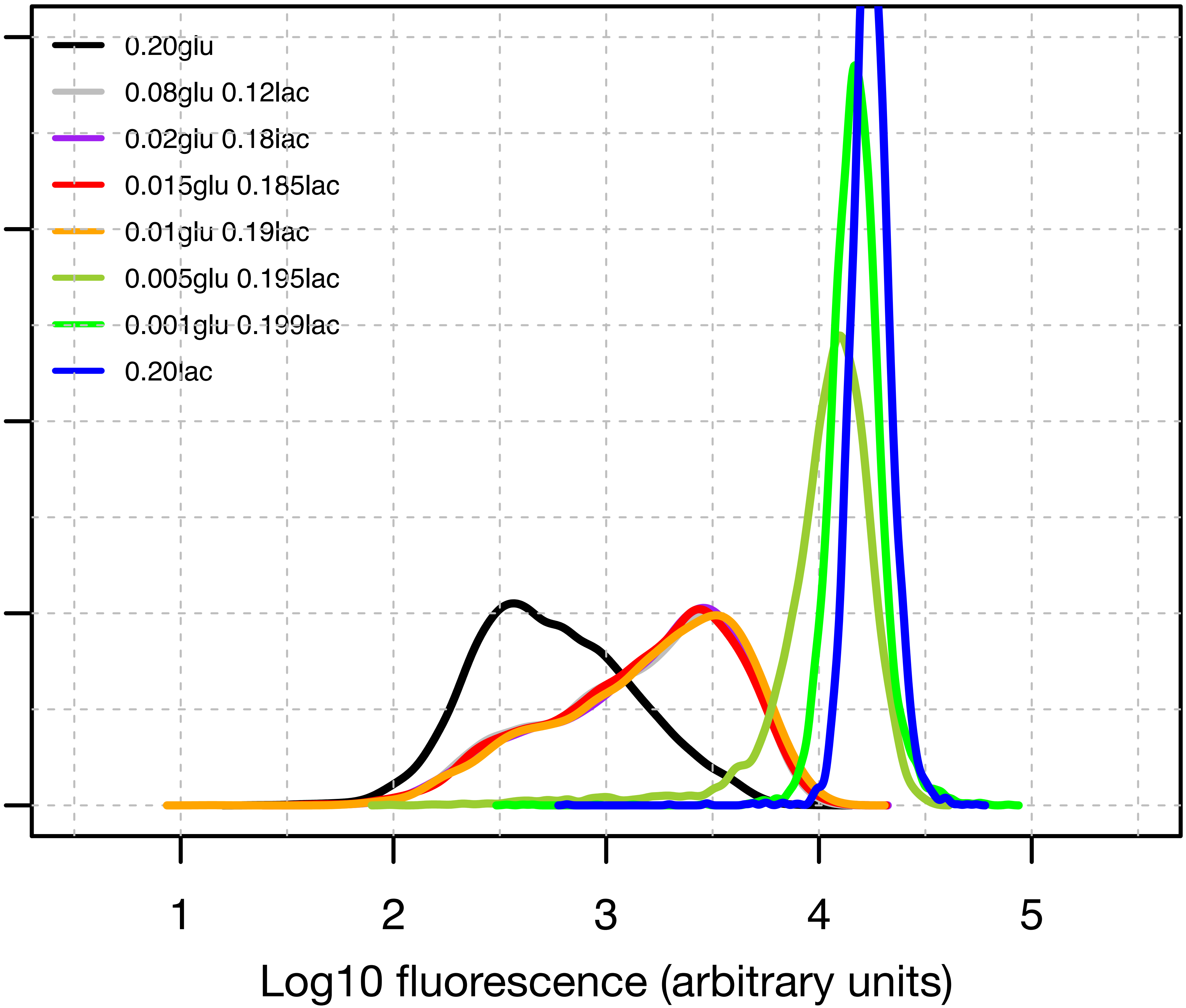
What about cis effects?



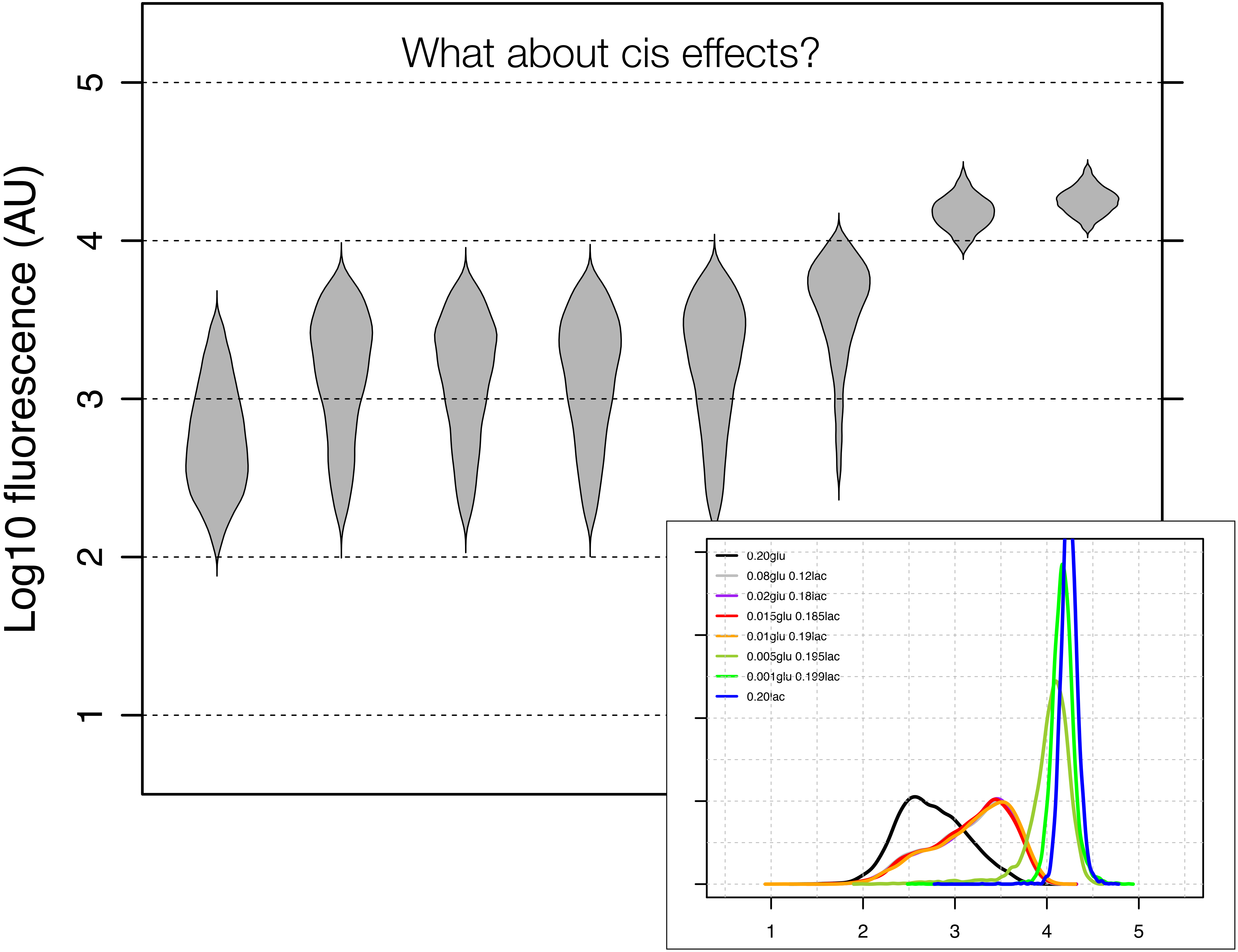
What about cis effects?

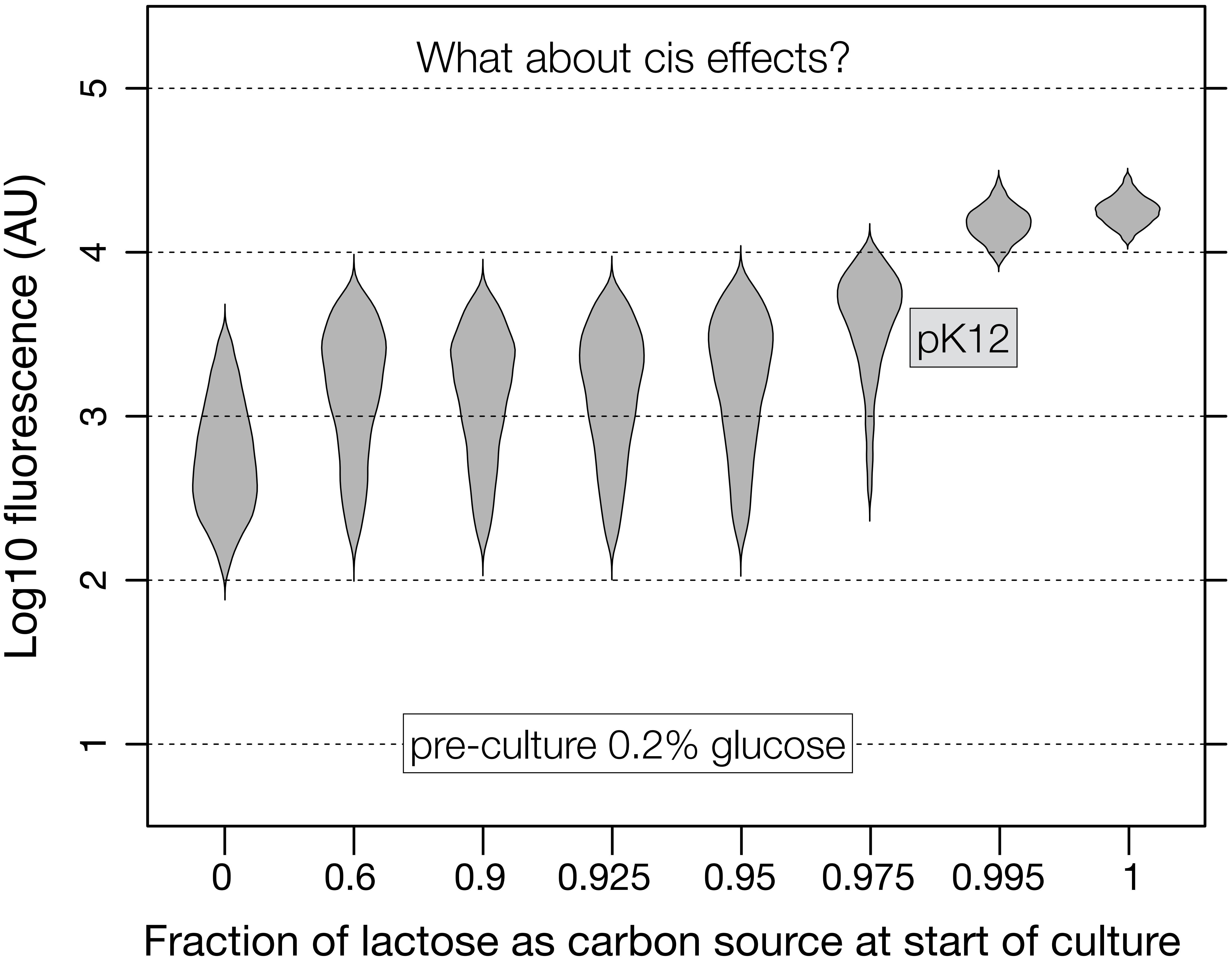


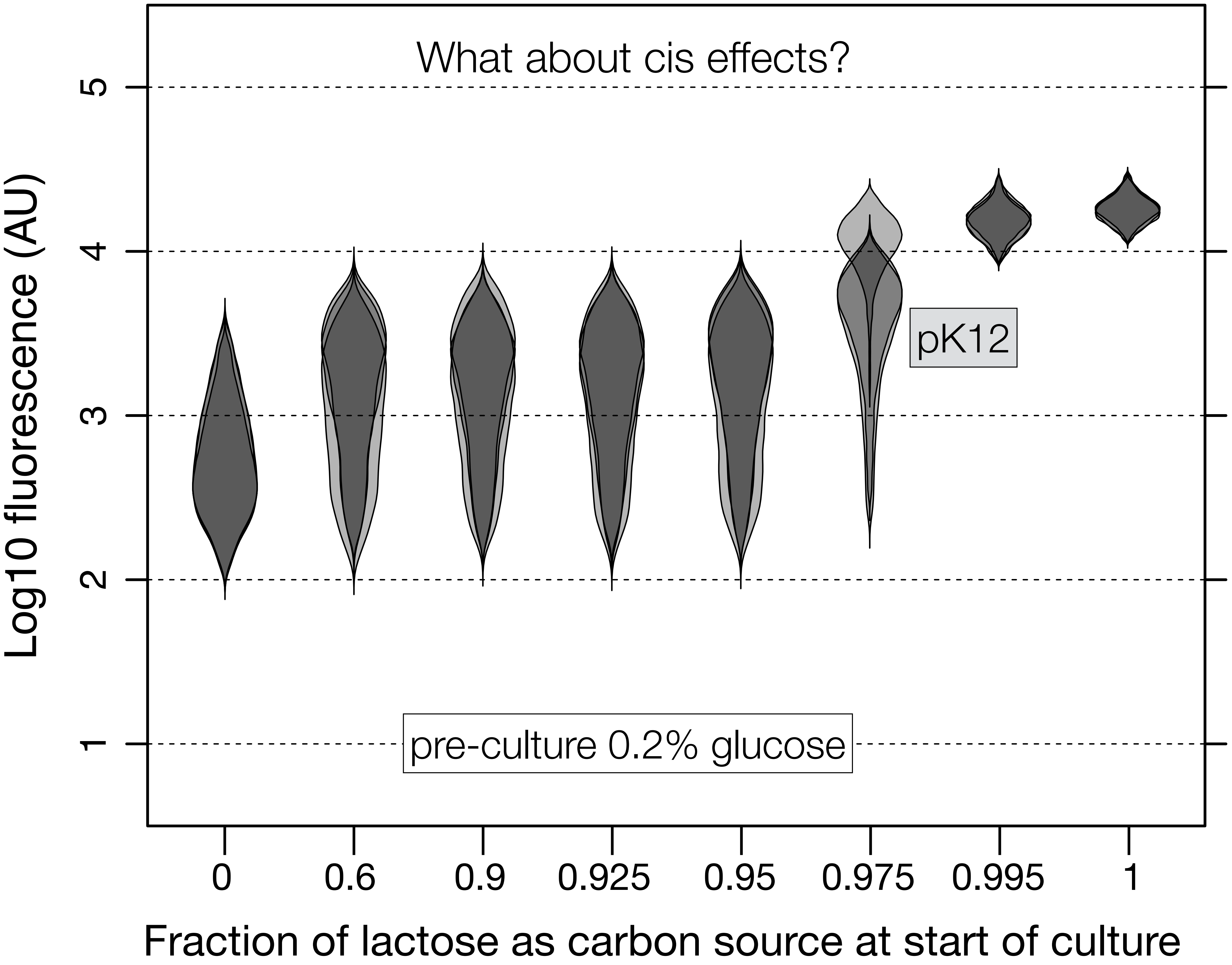
What about cis effects?

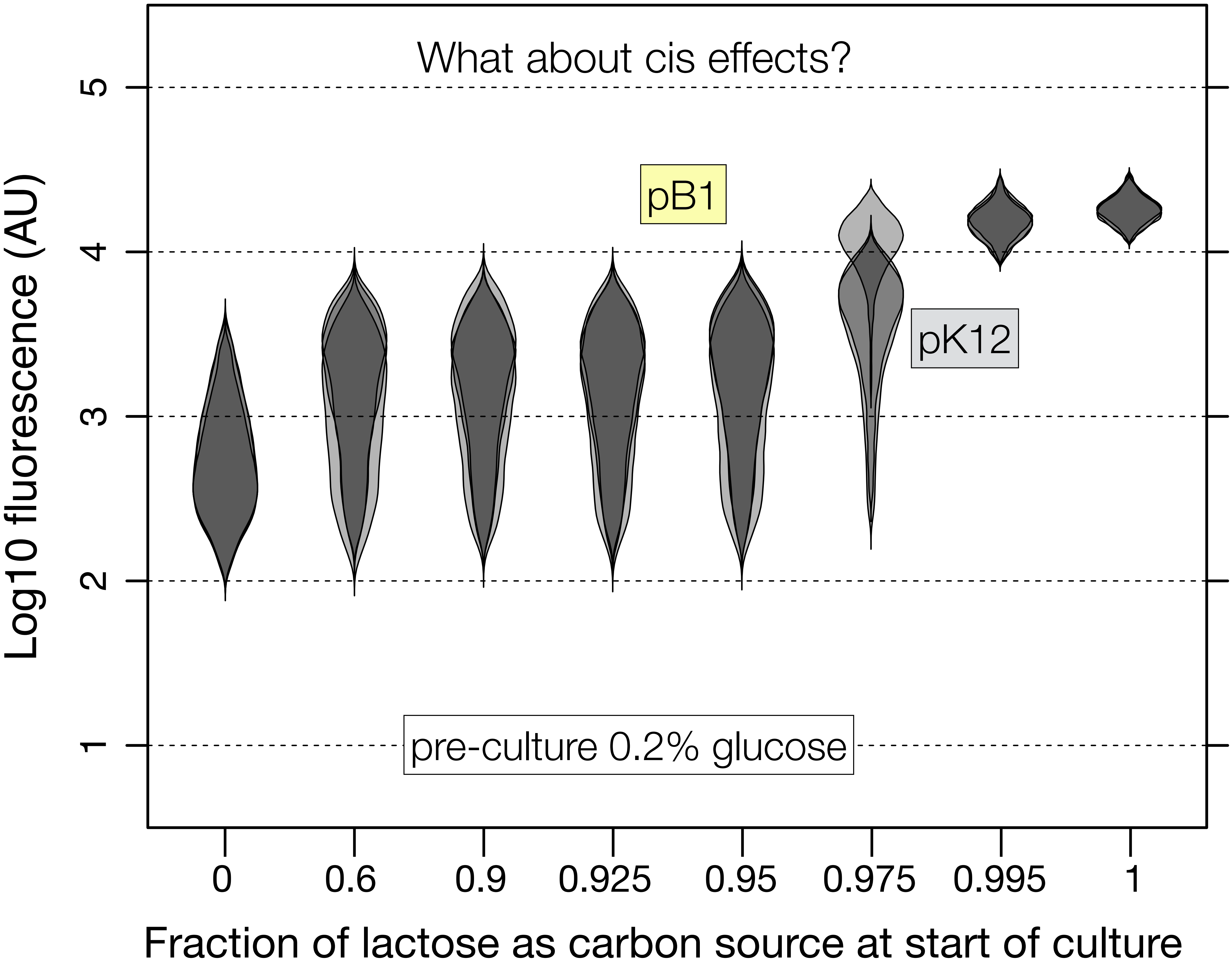


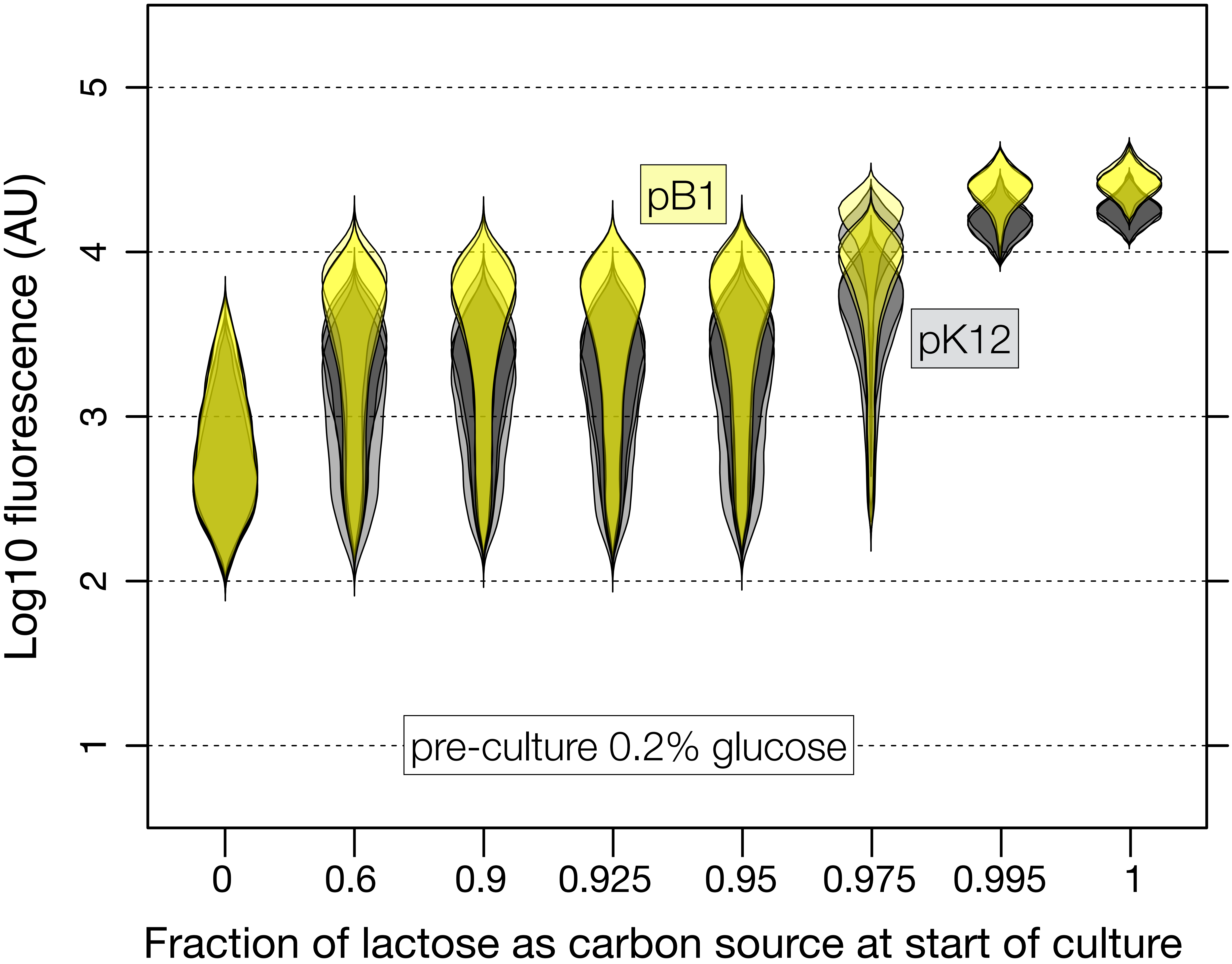
What about cis effects?

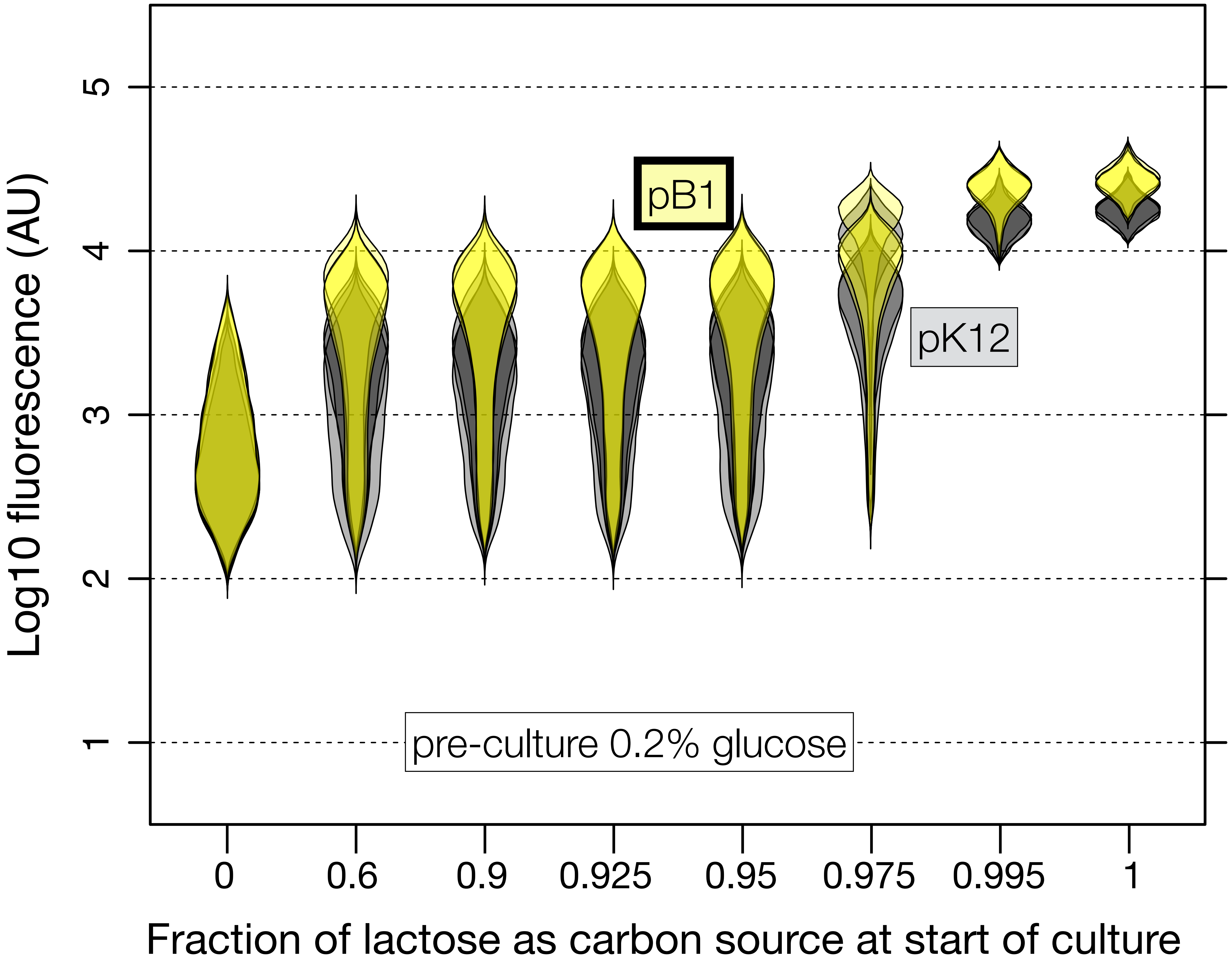


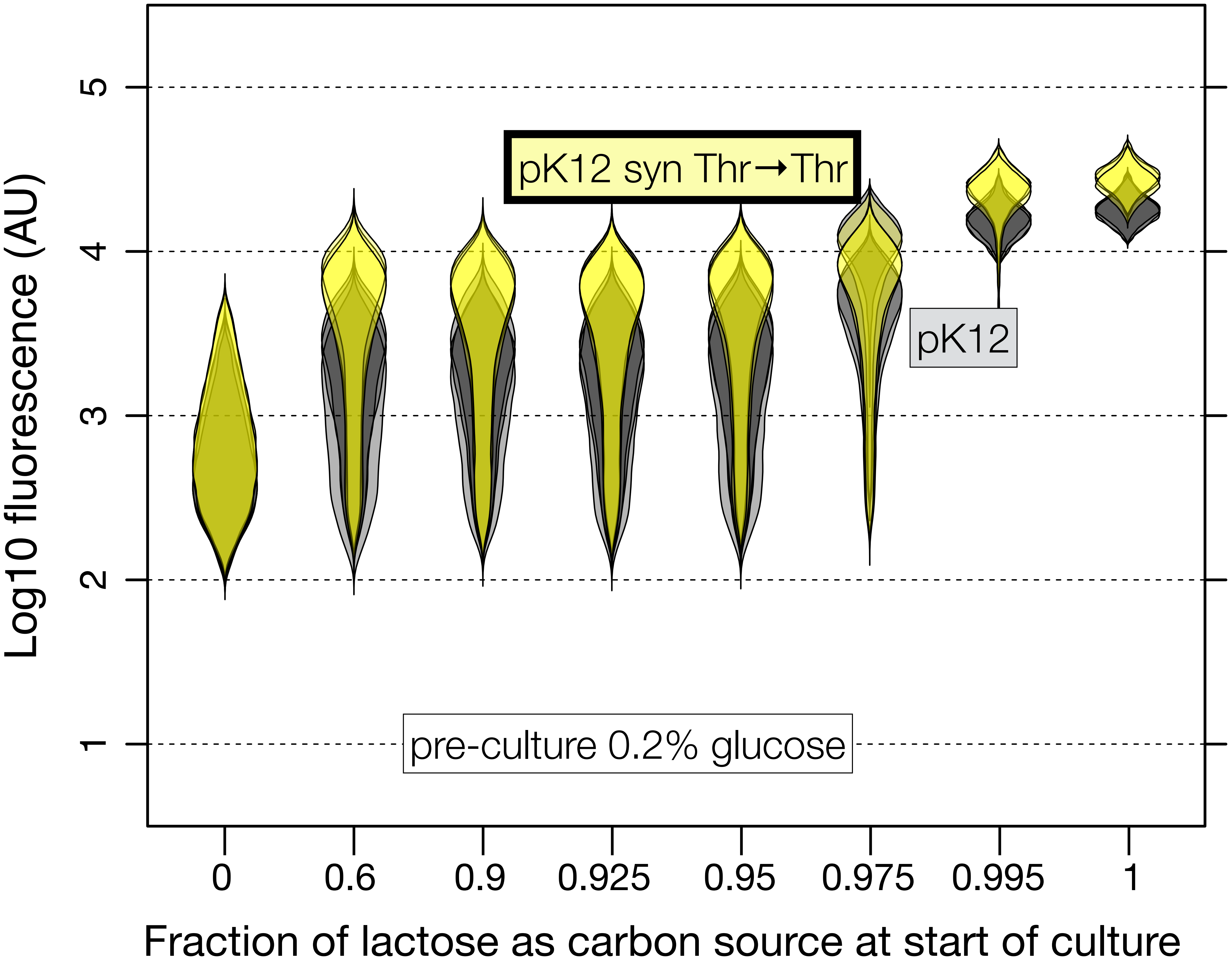

What about cis effects?
pK12
pre-culture 0.2% glucose
Log10 fluorescence (AU)
5
4
3
2
1
0
0.6
0.9
0.925
0.95
0.975
0.995
1
Fraction of lactose as carbon source at start of culture

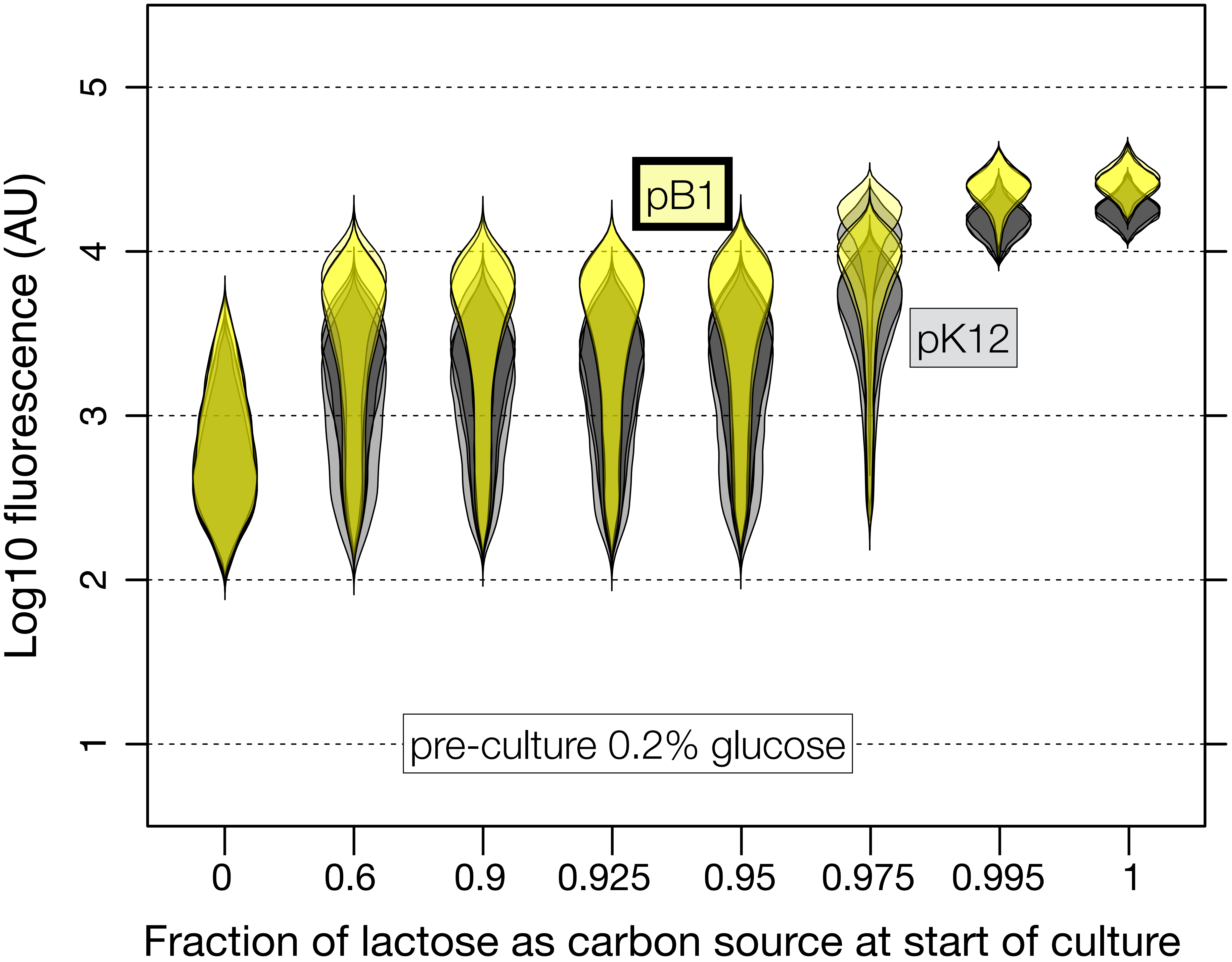

What about cis effects?
pK12
pre-culture 0.2% glucose
Log10 fluorescence (AU)
1
2
3
4
5
0
0.6
0.9
0.925
0.95
0.975
0.995
1
Fraction of lactose as carbon source at start of culture

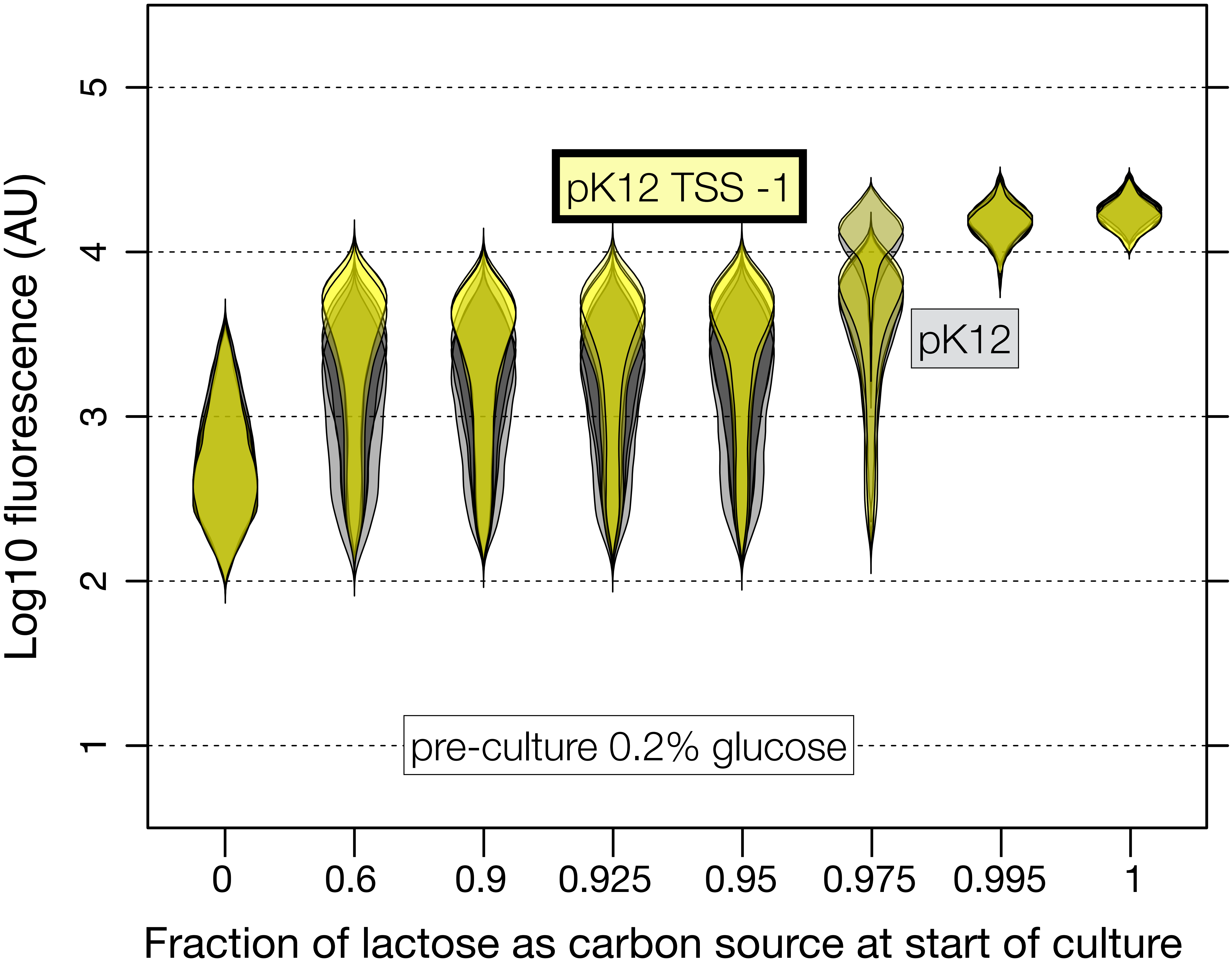

What about cis effects?
pB1
pK12
pre-culture 0.2% glucose
Log10 fluorescence (AU)
1
2
3
4
5
0
0.6
0.9
0.925
0.95
0.975
0.995
1
Fraction of lactose as carbon source at start of culture


Log10 fluorescence (AU)
1
2
3
4
5
pB1
pK12
pre-culture 0.2% glucose
0
0.6
0.9
0.925
0.95
0.975
0.995
1
Fraction of lactose as carbon source at start of culture

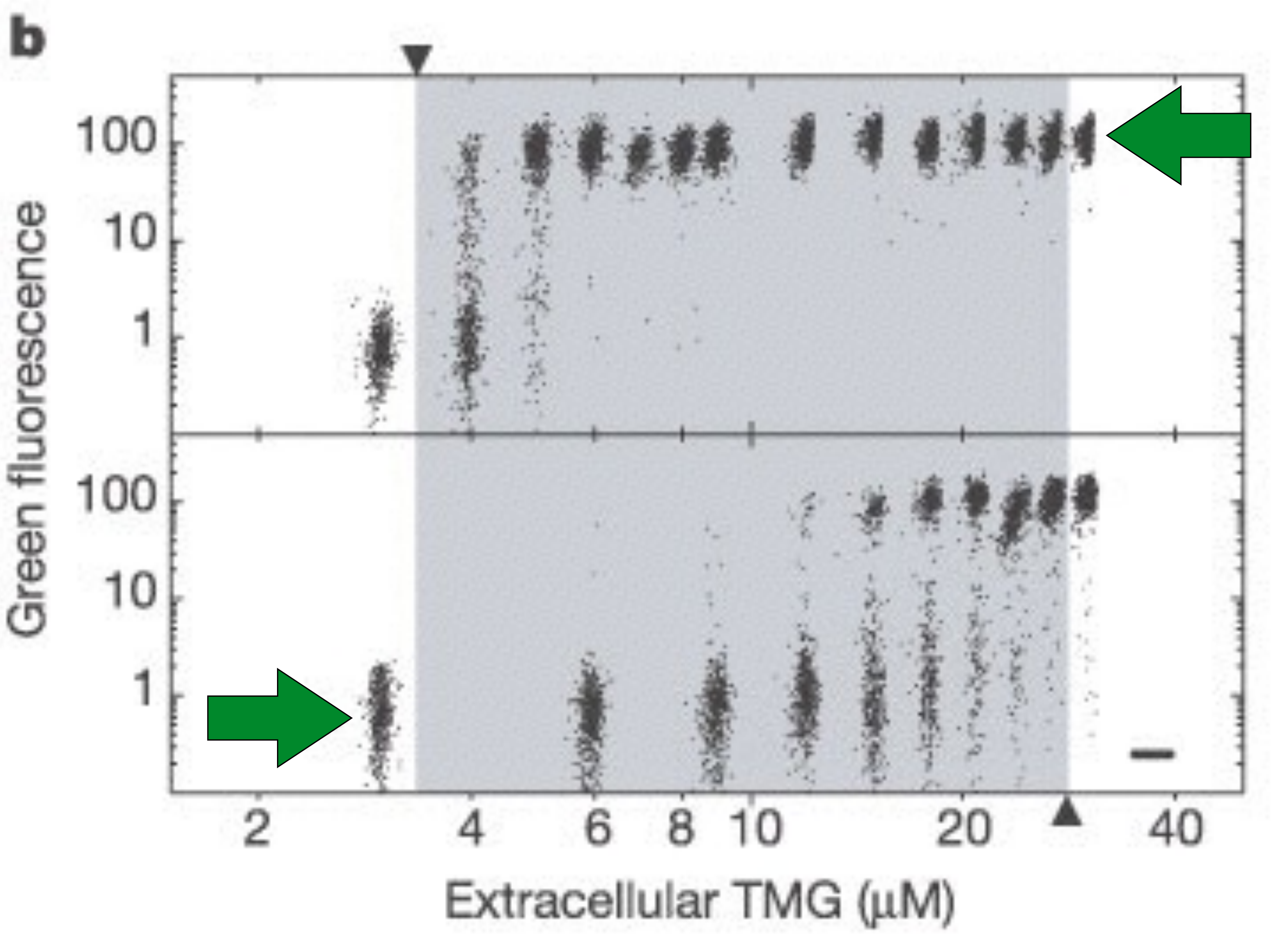

Log10 fluorescence (AU)
1
2
3
4
5
pB1
pK12
pre-culture 0.2% glucose
0
0.6
0.9
0.925
0.95
0.975
0.995
1
Fraction of lactose as carbon source at start of culture


pK12 syn Thr→Thr
pK12
pre-culture 0.2% glucose
Log10 fluorescence (AU)
1
2
3
4
5
0
0.6
0.9
0.925
0.95
0.975
0.995
1
Fraction of lactose as carbon source at start of culture

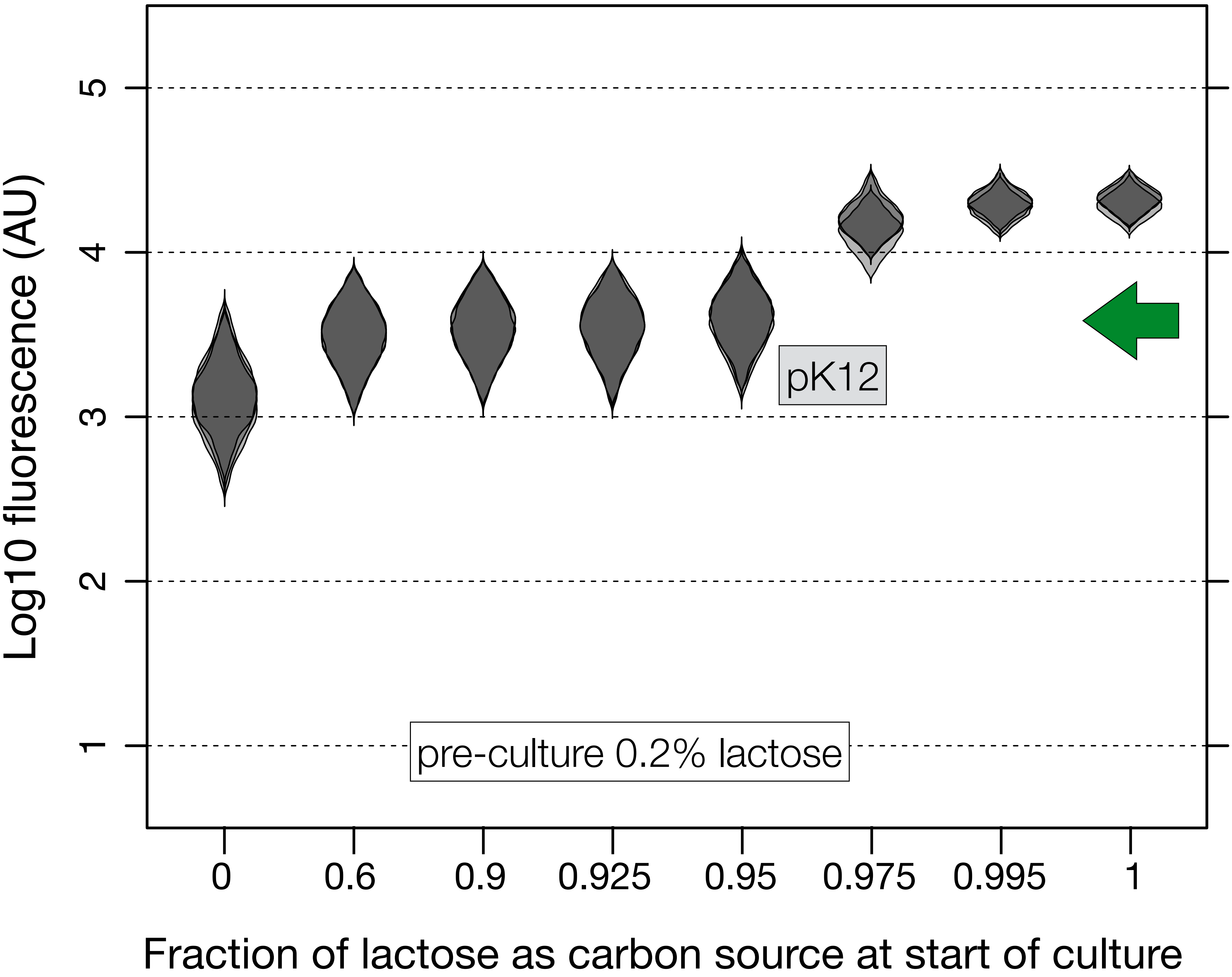

Log10 fluorescence (AU)
5
4
3
2
1
pB1
pK12
pre-culture 0.2% glucose
0
0.6
0.9
0.925
0.95
0.975
0.995
1
Fraction of lactose as carbon source at start of culture

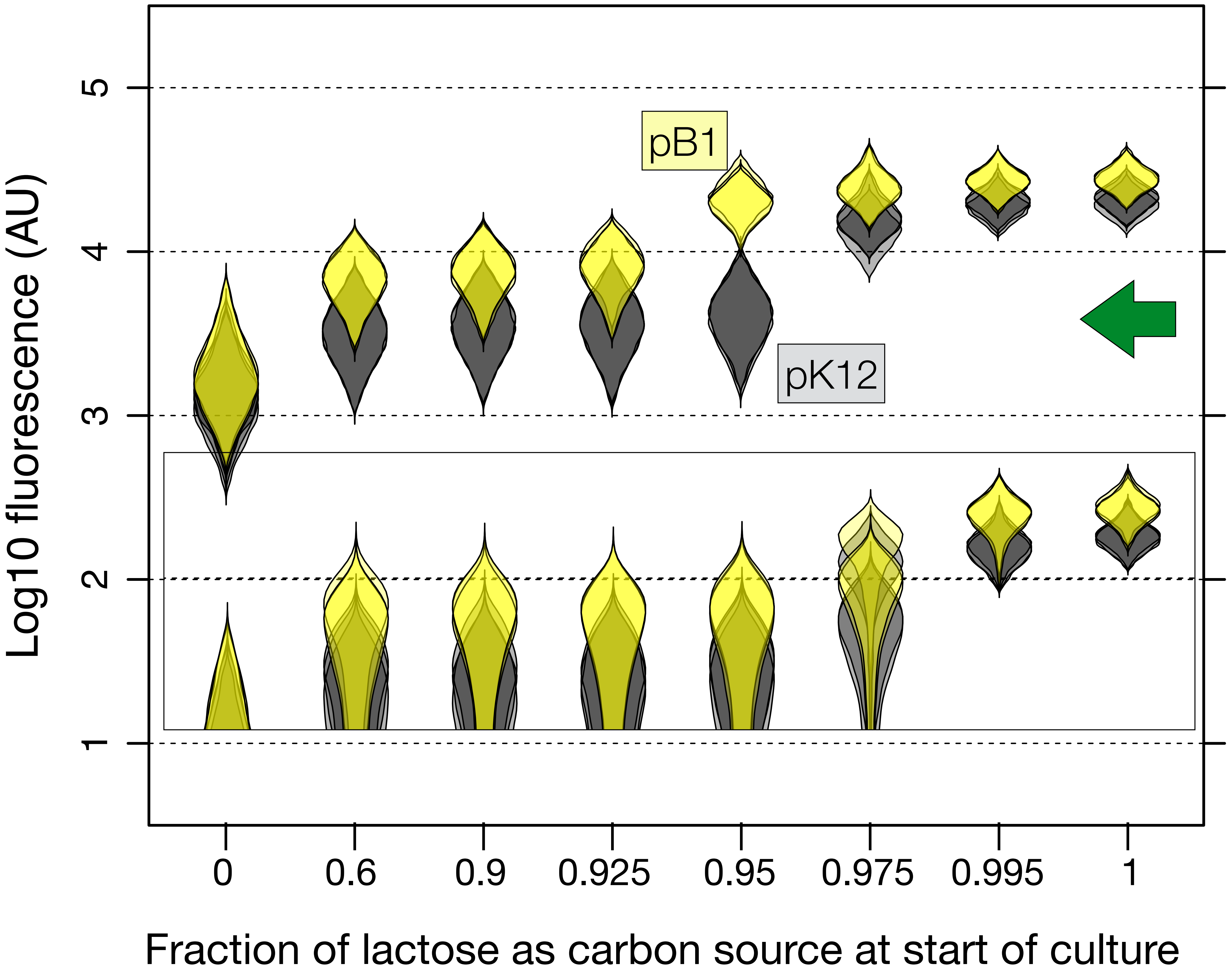

pK12 TSS -1
pK12
pre-culture 0.2% glucose
Log10 fluorescence (AU)
1
2
3
4
5
0
0.6
0.9
0.925
0.95
0.975
0.995
1
Fraction of lactose as carbon source at start of culture

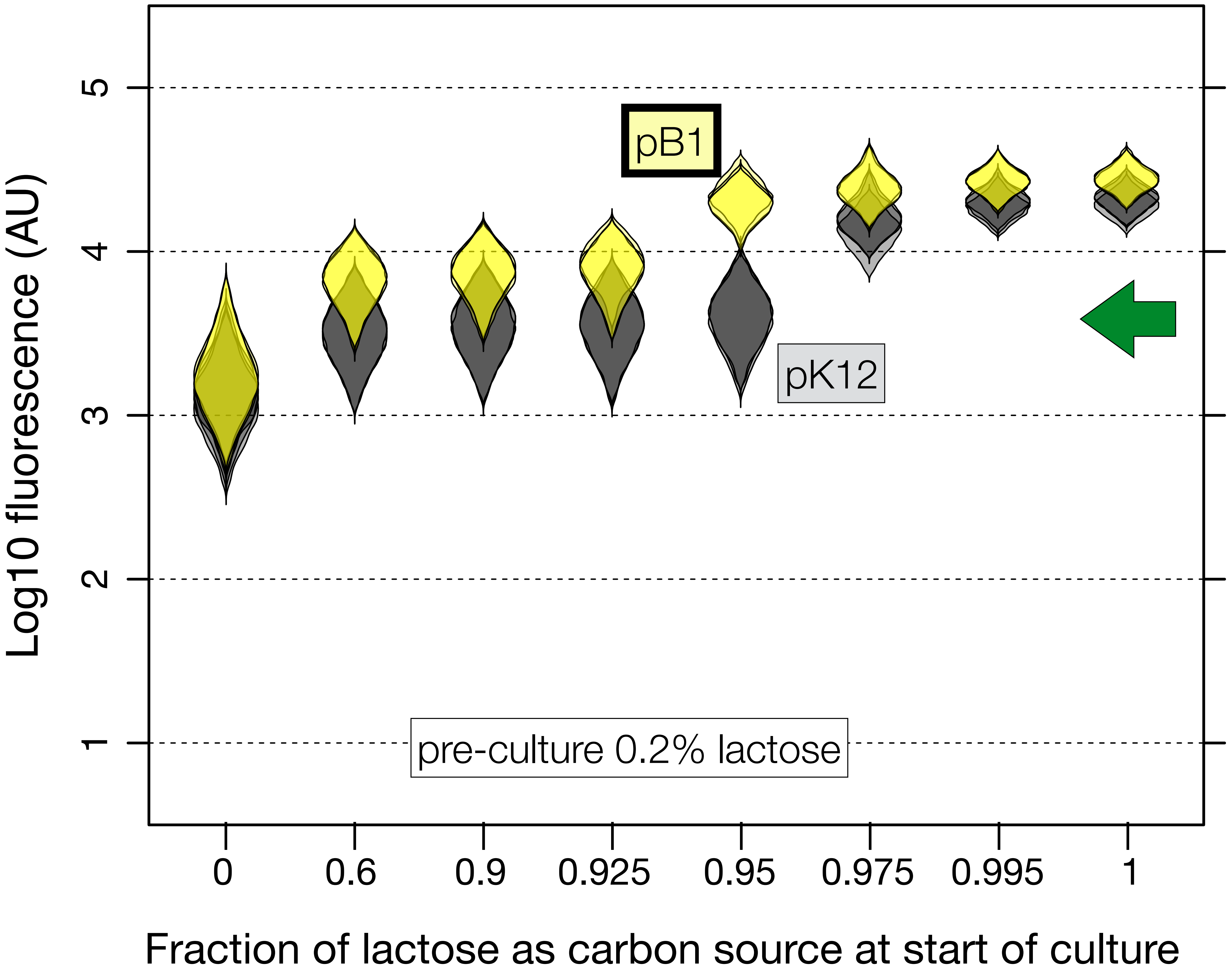
The lac operon exhibits hysteresis

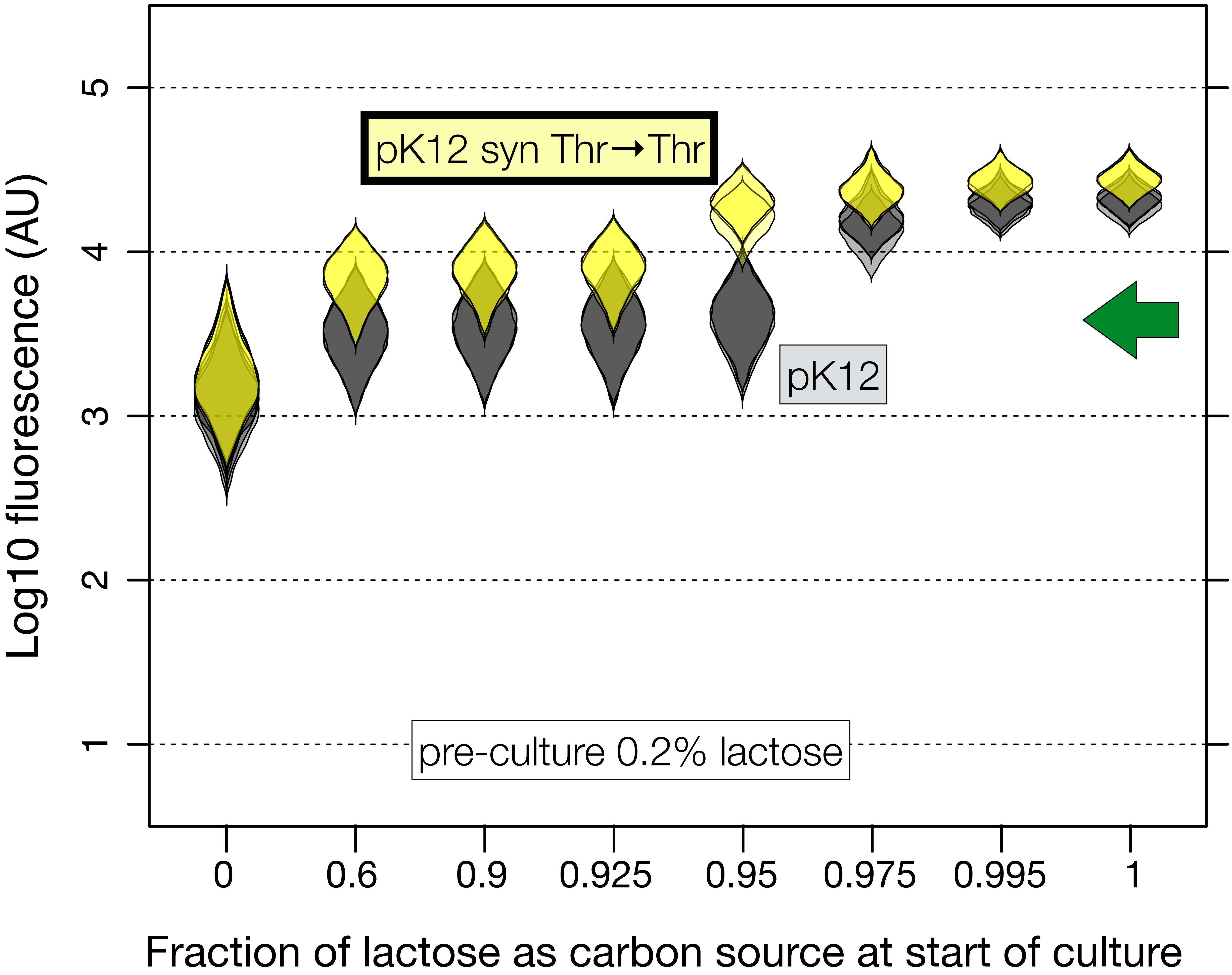


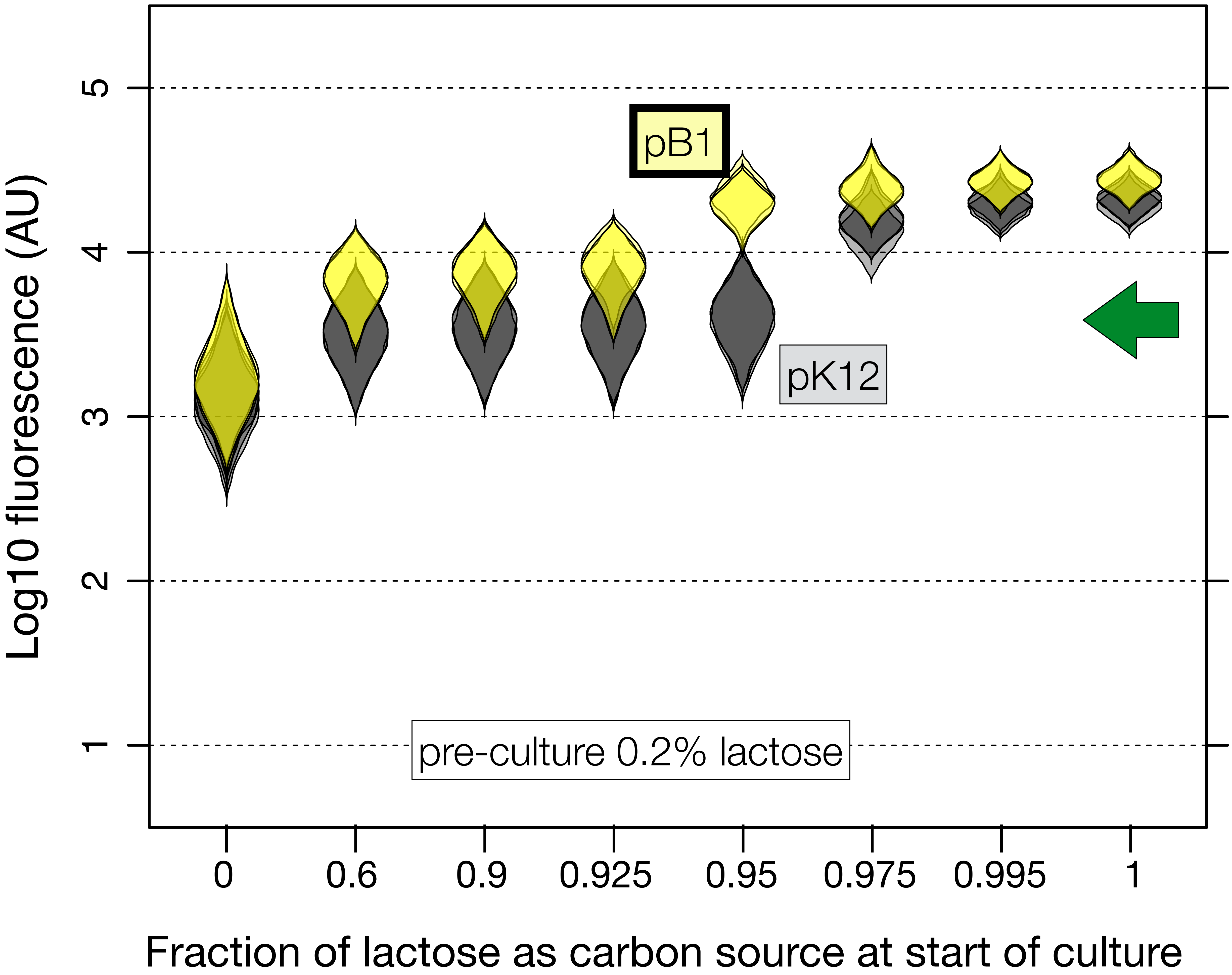
Ozbudak et al. (2004) *Nature Genetics*

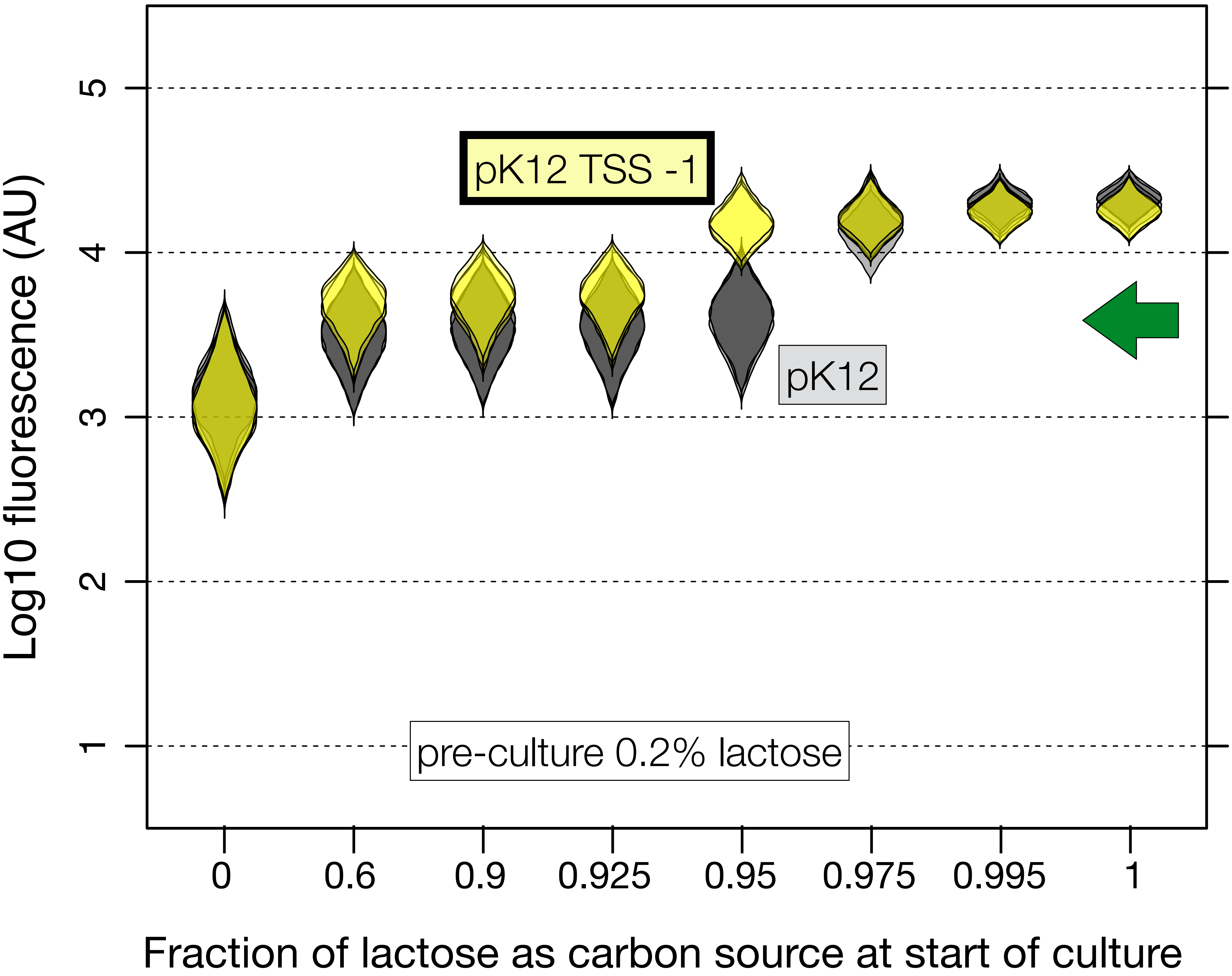

Log10 fluorescence (AU)
5
4
3
2
1
pK12
pre-culture 0.2% lactose
0
0.6
0.9
0.925
0.95
0.975
0.995
1
Fraction of lactose as carbon source at start of culture

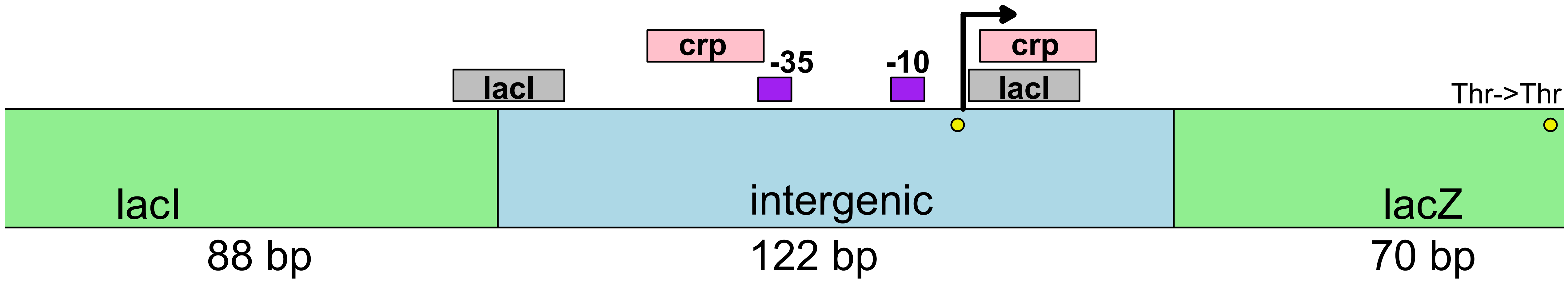

pB1
pK12
Log10 fluorescence (AU)
1
2
3
4
5
0
0.6
0.9
0.925
0.95
0.975
0.995
1
Fraction of lactose as carbon source at start of culture

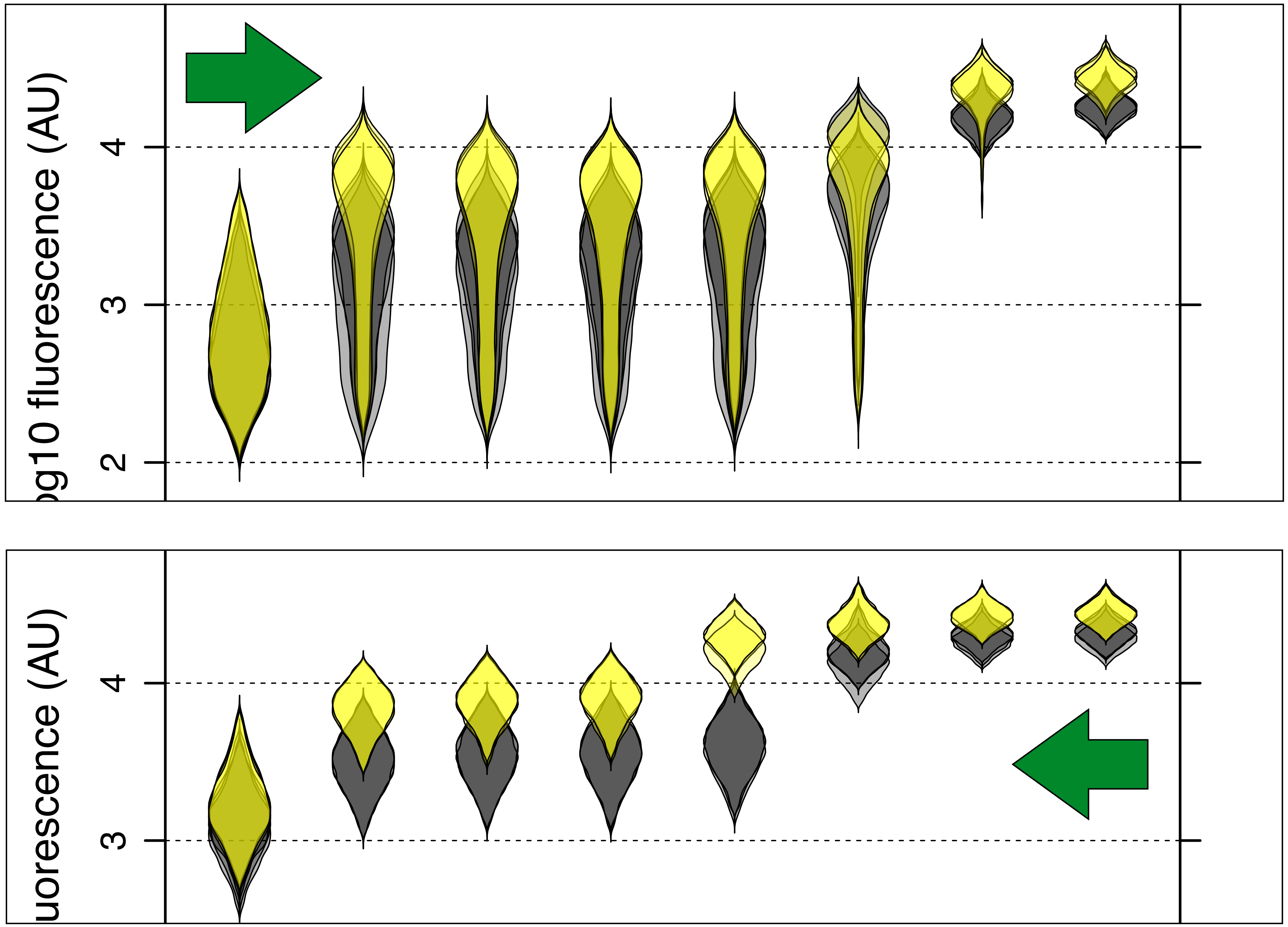

Log10 fluorescence (AU)
5
4
3
2
1
pB1
pK12
pre-culture 0.2% lactose
0
0.6
0.9
0.925
0.95
0.975
0.995
1
Fraction of lactose as carbon source at start of culture

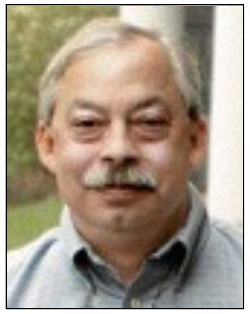

pK12 syn Thr→Thr
pK12
pre-culture 0.2% lactose
Log10 fluorescence (AU)
1
2
3
4
5
0
0.6
0.9
0.925
0.95
0.975
0.995
1
Fraction of lactose as carbon source at start of culture


Log10 fluorescence (AU)
5
4
3
2
1
pB1
pK12
pre-culture 0.2% lactose
0
0.6
0.9
0.925
0.95
0.975
0.995
1
Fraction of lactose as carbon source at start of culture


pK12 TSS -1
pK12
pre-culture 0.2% lactose
Log10 fluorescence (AU)
1
2
3
4
5
0
0.6
0.9
0.925
0.95
0.975
0.995
1
Fraction of lactose as carbon source at start of culture

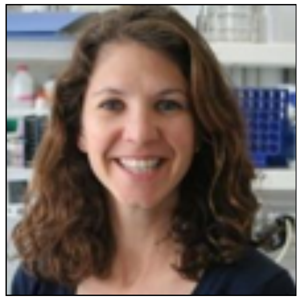

crp
-35
-10
crp
lacI
lacI
Thr->Thr
lacI
intergenic
lacZ
88 bp
122 bp
70 bp


og10 fluorescence (AU)
2
3
4
uorescence (AU)
3
4



Mike Sadowsky
(Univ. Minnesota)

Daying Wen
(Massey Uni)



Nikki Freed
(Massey Uni)

Georgia Breckell
(Massey Uni)

Diana Blank
(Uni Basel)

Erik van Nimwegen
(Uni Basel)

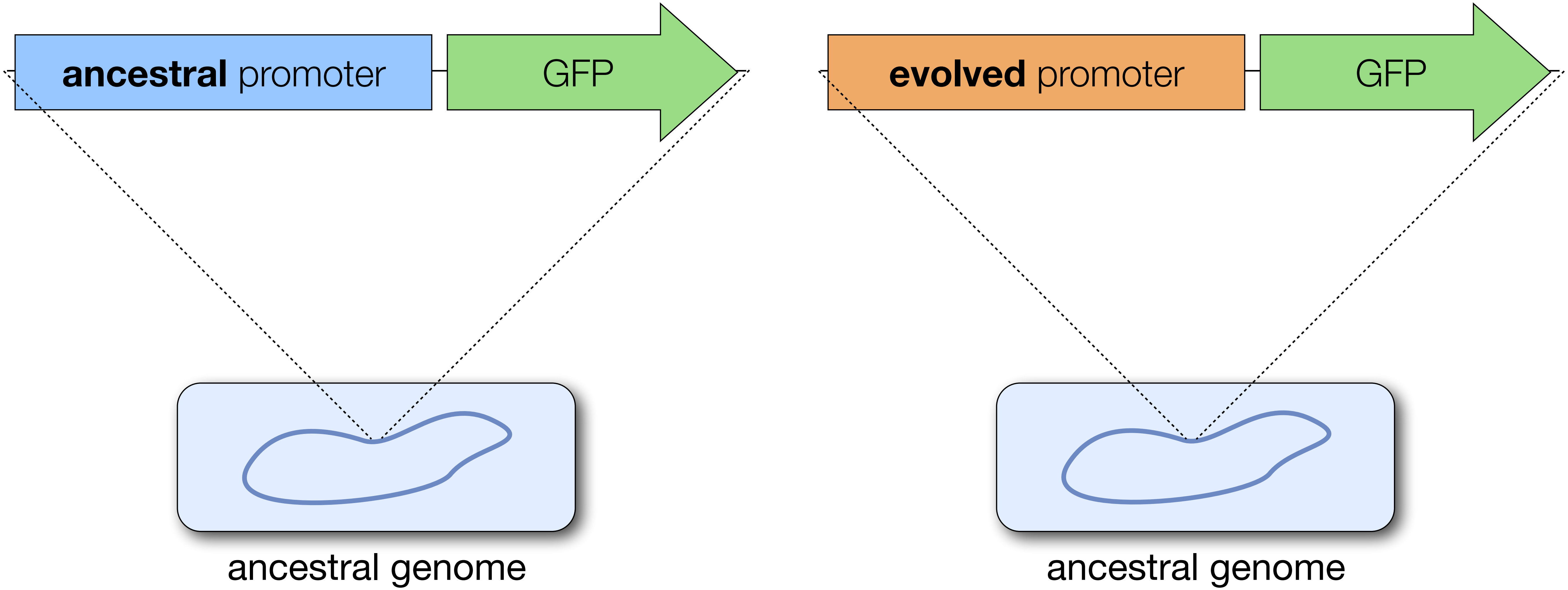
Chris Field
(Uni Basel)

Frederic Bertels
(Uni Basel)

FNSNF
FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION

Massey University

experimental confirmation



ancestral promoter

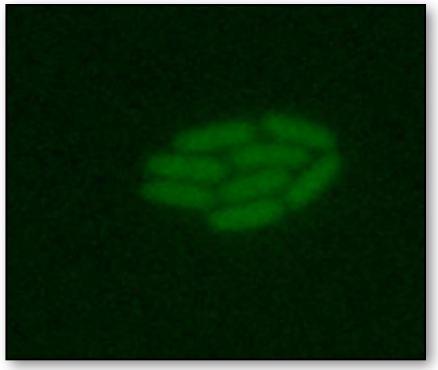
GFP

evolved promoter

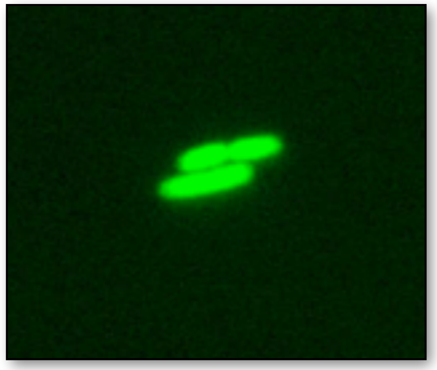
GFP

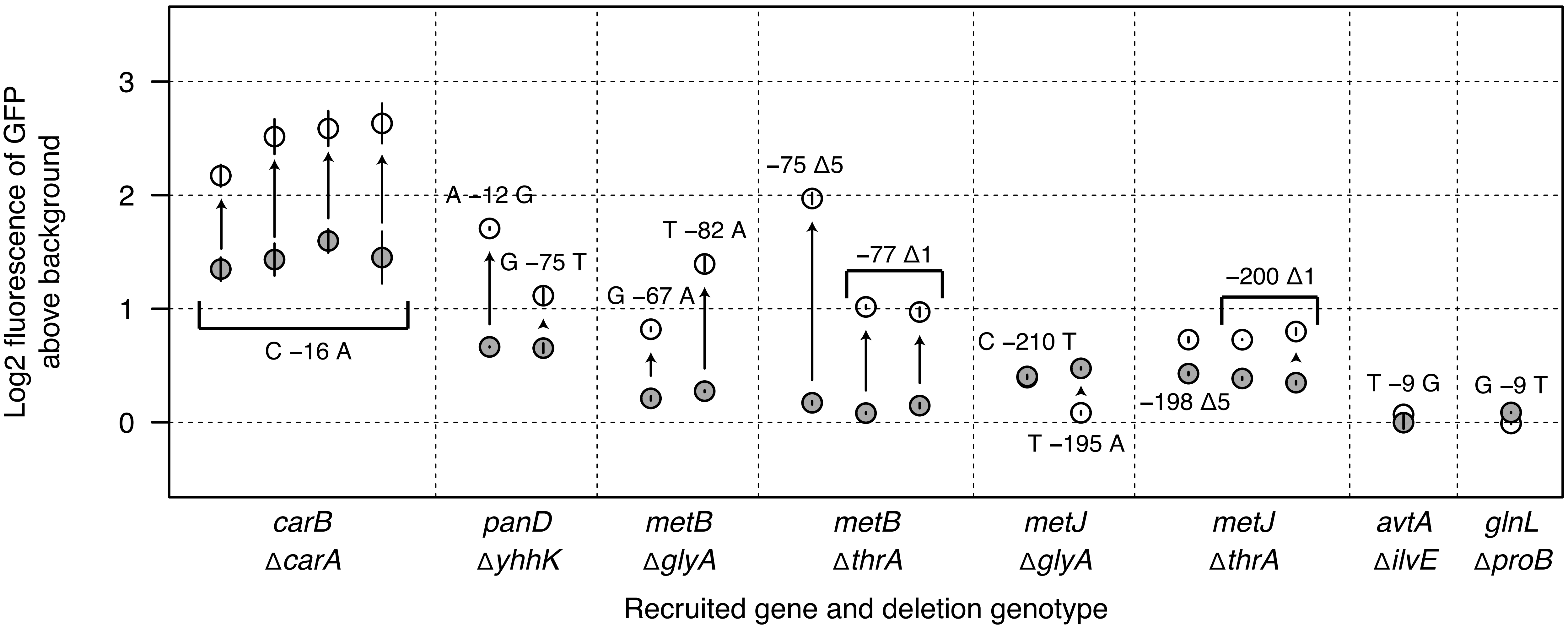
ancestral genome

ancestral genome



compare expression
of ancestral and evolved alleles




Log2 fluorescence of GFP above background
3
2
1
0
C −16 A
A −12 G
G −75 T
G −67 A
T −82 A
−75 Δ5
−77 Δ1
C −210 T
T −195 A
−200 Δ1
−198 Δ5
T −9 G
G −9 T
carB Δ*carA*
panD Δ*yhhK*
metB Δ*glyA*
metB Δ*thrA*
metJ Δ*glyA*
metJ Δ*thrA*
avtA Δ*ilvE*
glnL Δ*proB*
Recruited gene and deletion genotype