High Expression Hampers Horizontal Gene Transfer

Chungoo Park and Jianzhi Zhang*

Department of Ecology and Evolutionary Biology, University of Michigan

*Corresponding author: E-mail: jianzhi@umich.edu.

Abstract

Horizontal gene transfer (HGT) is a common phenomenon in prokaryotic evolution. The rate of HGT is known to vary among genes, currently understand of the cause of this variation summarized by **two rules**:

- Informational genes, which are involved in DNA replication, transcription, and translation, have lower transferabilities than operational genes. (protein function rule)

- Protein interactivity negatively impacts gene transferability. (protein complexity rule)

" hypothesize: high expression hampers HGT

To test the hypothesis, we examined laboratory and natural HGTs to Escherichia coli. We observed lower transferabilities of more highly expressed genes.

Furthermore, expression level predicts gene transferability better than all other factors examined. We confirmed the significant negative impact of gene expression on the rate of HGTs to 127 of 133 genomes of eubacteria and archaebacteria.

Together, these findings establish the gene expression level as a major determinant of horizontal gene transferability.

Introduction

HGT occurs through three cellular mechanisms: transformation, conjugation, and transduction.

Based on five considerations, we propose that hypothesis.
(high expression hampers HGT)

First, expressing an unnecessary gene wastes energy and reduces fitness, and the fitness cost increase with expression level. (energy expenditure in transcription and translation)
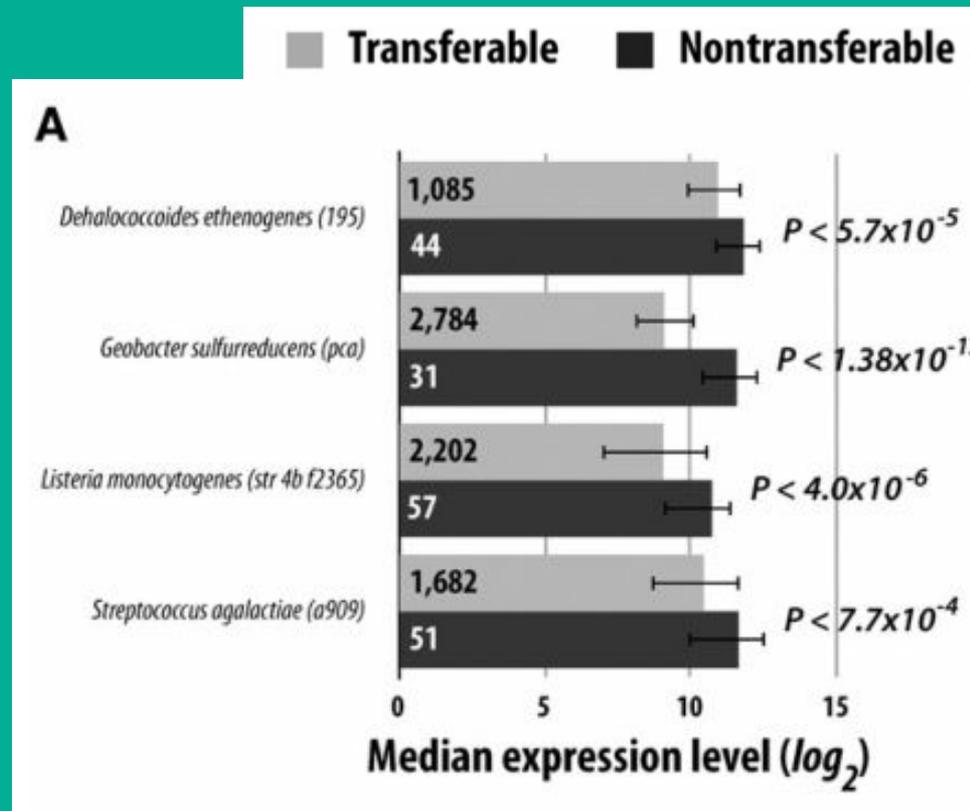
Second, different species prefer different synonymous codons, using unpreferred codons may increase translational errors, which can cause cytotoxic protein misfolding, acquisition of a strongly expressed gene leads to more misfolded protein molecules and do more harm to the recipient.(cytotoxic protein misfolding)
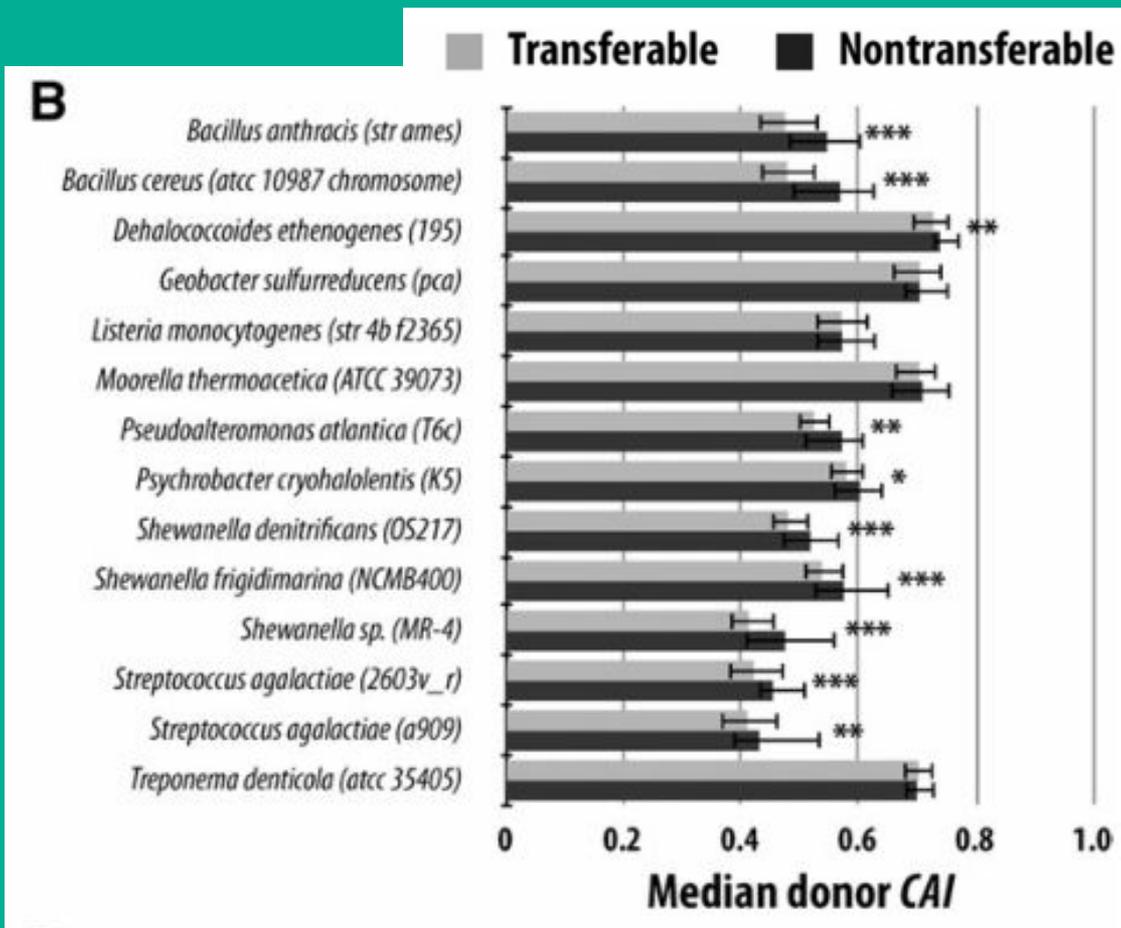
Third, the expression of a foreign gene with a codon usage that is suboptimal in the recipient cell reduces the overall translational efficiency and the fitness of the recipient, and this fitness reduction is amplified when the foreign gene is strongly expressed. (reduction in cellular translational efficiency)

Results

Transferable (light gray) **Nontransferable** (black)

A

Species	
Dehalococcoides ethenogenes (195)	1,085 / 44, $P < 5.7 \times 10^{-5}$
Geobacter sulfurreducens (pca)	2,784 / 31, $P < 1.38 \times 10^{-1}$
Listeria monocytogenes (str 4b f2365)	2,202 / 57, $P < 4.0 \times 10^{-6}$
Streptococcus agalactiae (a909)	1,682 / 51, $P < 7.7 \times 10^{-4}$

Median expression level (log_2)

"We test the impact of gene expression level on the rate of HGT by first using a data set of laboratory HGTs to E. coli. Among the 79 donor genomes analyzed, 14 genomes are amenable to statistical analysis because they each contain at least 30 so-called nontransferable genes. Of these 14 species, four have publicly available microarray-based genome-wide gene expression data. In all four cases, the median expression level of nontransferable genes is 1.6–5.3 times that of transferable genes.

B

Transferable   Nontransferable

Species	
Bacillus anthracis (str ames)	***
Bacillus cereus (atcc 10987 chromosome)	***
Dehalococcoides ethenogenes (195)	**
Geobacter sulfurreducens (pca)	
Listeria monocytogenes (str 4b f2365)	
Moorella thermoacetica (ATCC 39073)	
Pseudoalteromonas atlantica (T6c)	**
Psychrobacter cryohalolentis (K5)	*
Shewanella denitrificans (OS217)	***
Shewanella frigidimarina (NCMB400)	***
Shewanella sp. (MR-4)	***
Streptococcus agalactiae (2603v_r)	***
Streptococcus agalactiae (a909)	**
Treponema denticola (atcc 35405)	

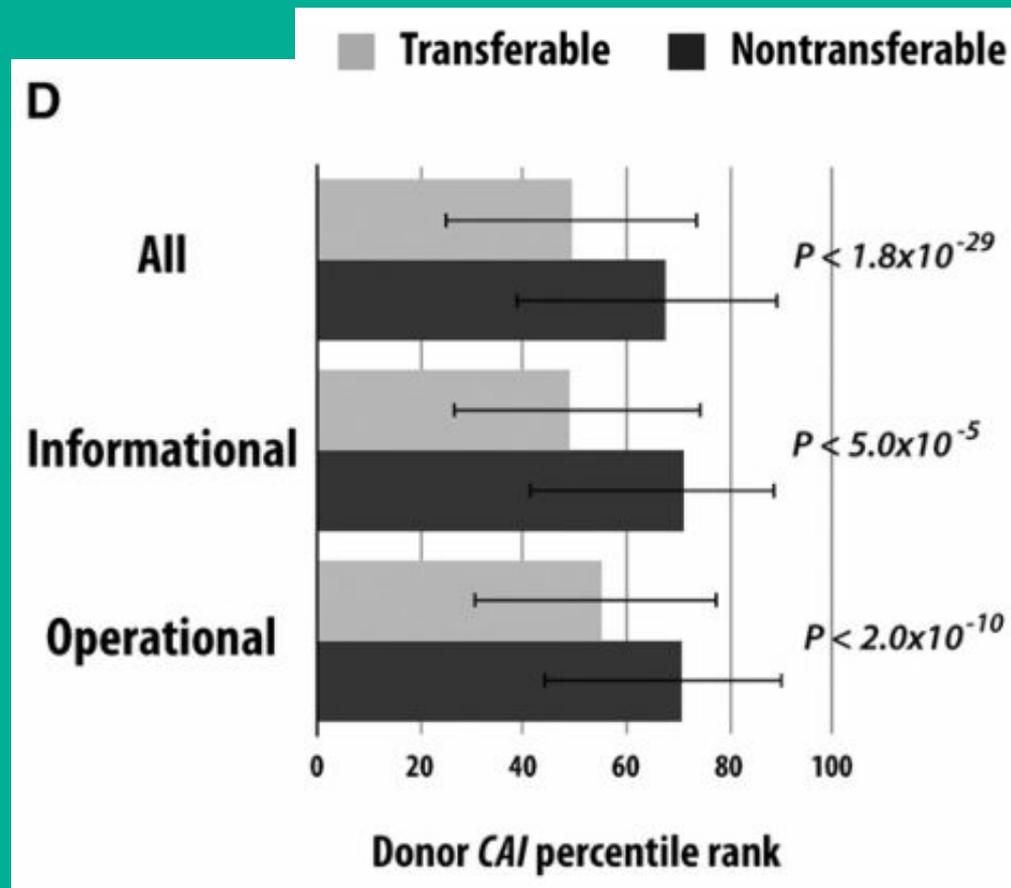Median donor *CAI*

*P < 0.05
**P < 0.01
***P < 0.001

We calculated the CAIs of all genes in each of the 14 donor species. The median CAI is higher for nontransferable genes than transferable genes in 12 of the 14 donors. Ten species show a significant difference in median CAI between transferable and nontransferable genes.

C

Informational   Operational

Transferable   Nontransferable

*P < 0.05
**P < 0.01
***P < 0.001

Median donor *CAI*
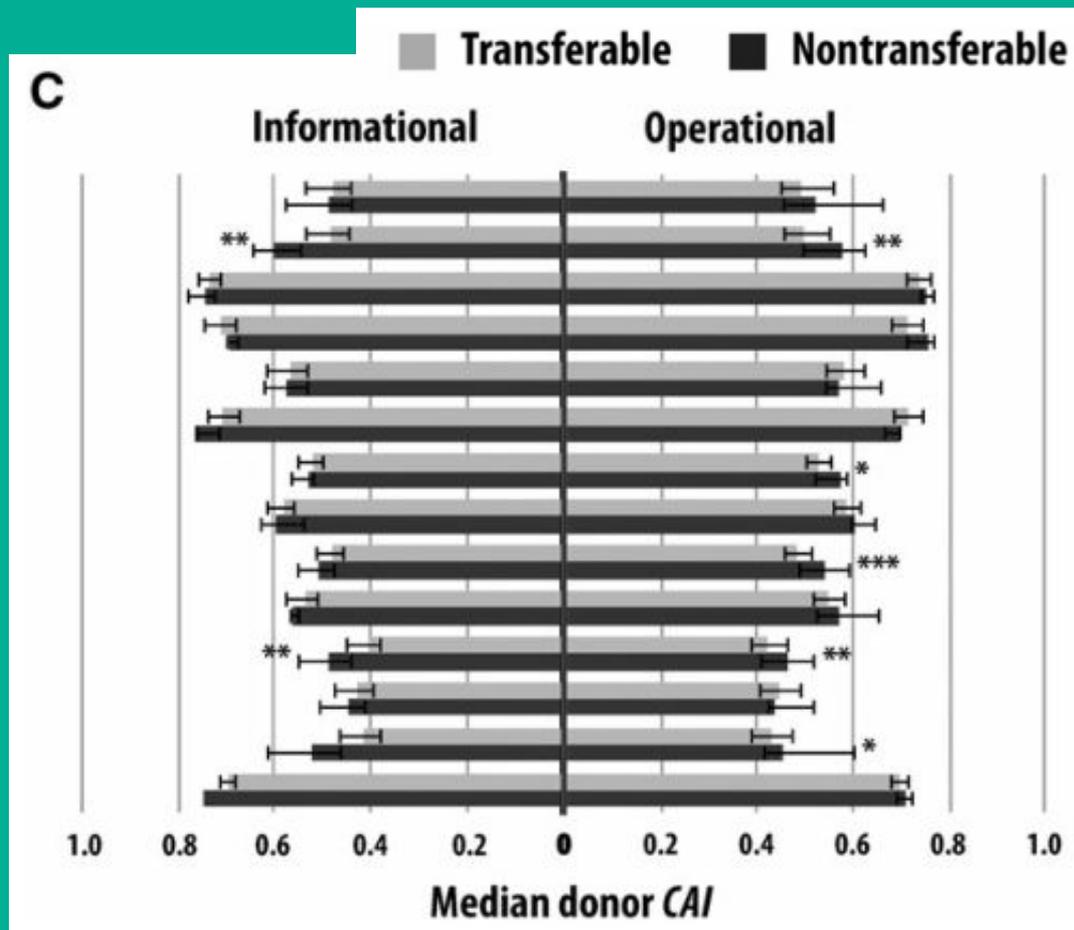
To exclude the possibility that the above observation is a byproduct of the protein function rule, we separately analyzed informational genes and operational genes. Yet, the general pattern of higher expressions or higher CAIs of nontransferable genes than transferable genes remains valid for both informational genes and operational genes.

D Transferable ■ Nontransferable

All — $P < 1.8 \times 10^{-29}$

Informational — $P < 5.0 \times 10^{-5}$

Operational — $P < 2.0 \times 10^{-10}$

Donor *CAI* percentile rank

We conducted a combined analysis of all 14 species. We first converted the CAIs of all genes in a genome to percentile ranks; the highest CAI has a percentile rank of 100 and the lowest has a percentile rank of 0. We then combined all the genes from the 14 species. We observed significantly higher CAI percentile ranks for nontransferable genes than transferable genes, and this pattern is true for both informational genes and operational genes.

We analyzed genes that have been naturally transferred to E. coli K12, because the laboratory HGTs were based on only one mechanism — plasmid mediated transformation, while natural HGTs occur by three mechanisms. In addition, laboratory conditions are different from the nature in many aspects, which may influence HGT rates.

We assume that the recently transferred genes have higher HGT rates than the resident genes of E. coli as well as the rest of the genes in various potential donor species.
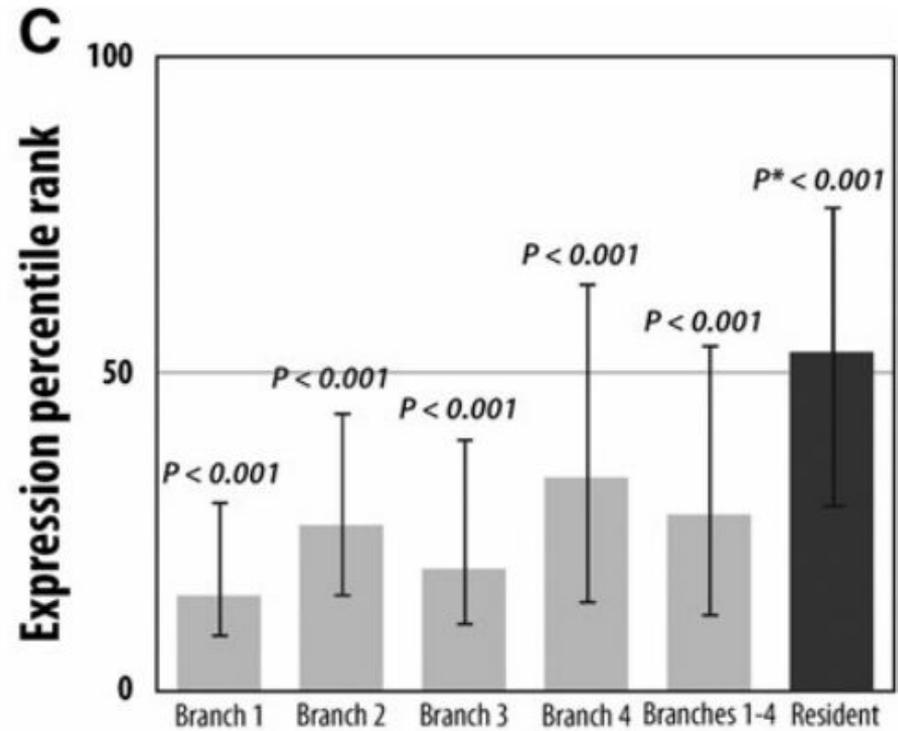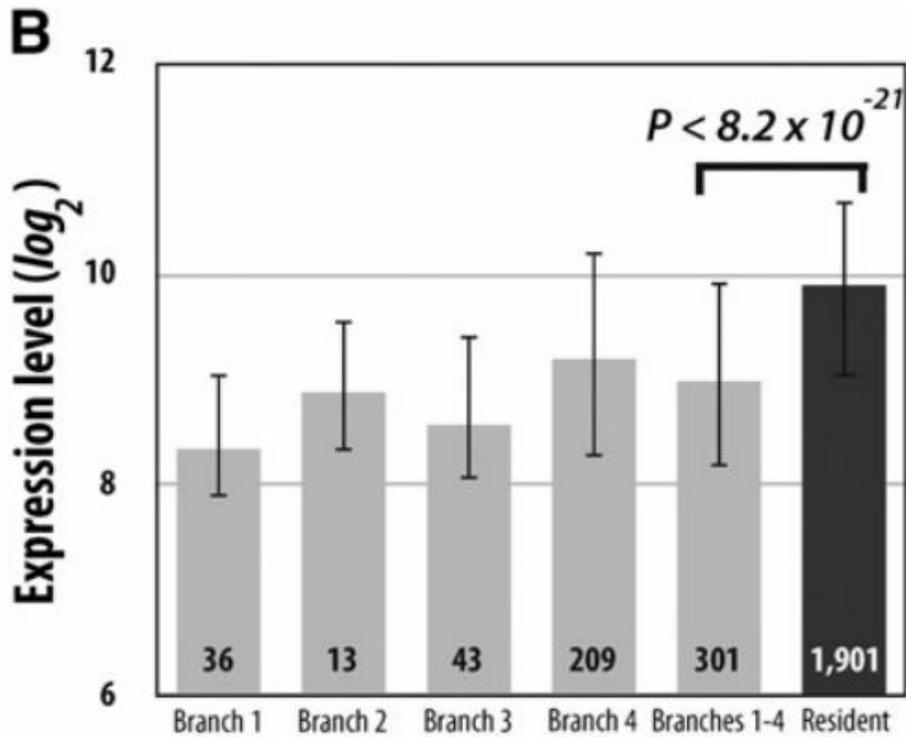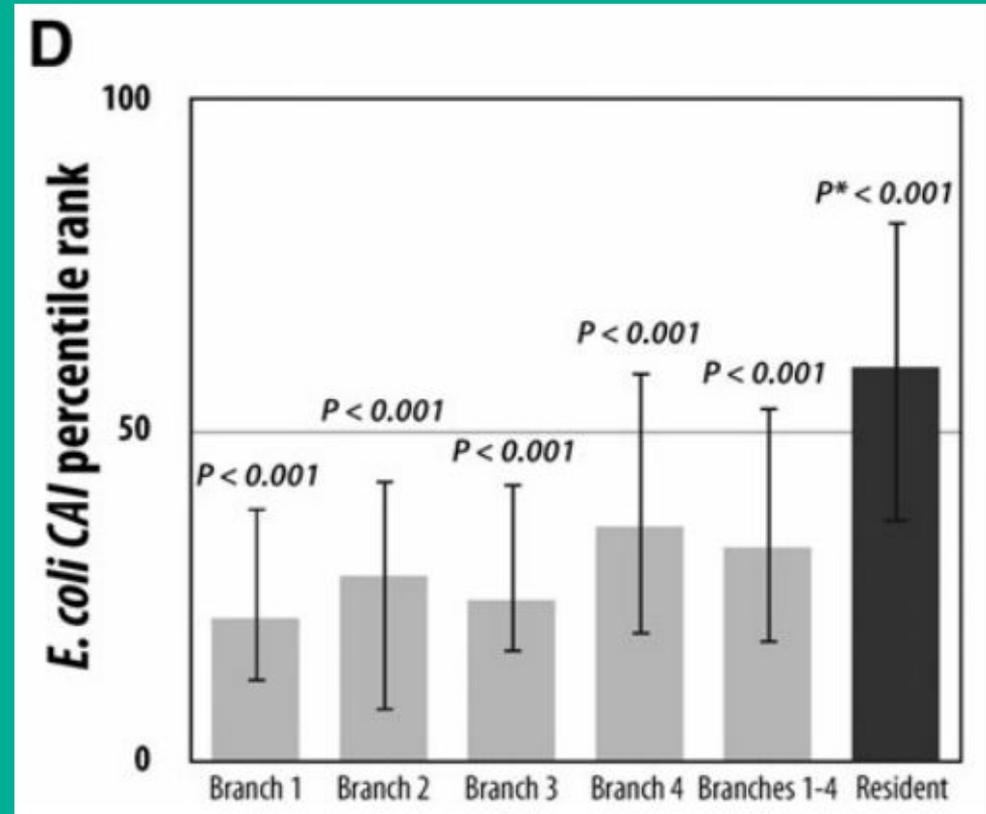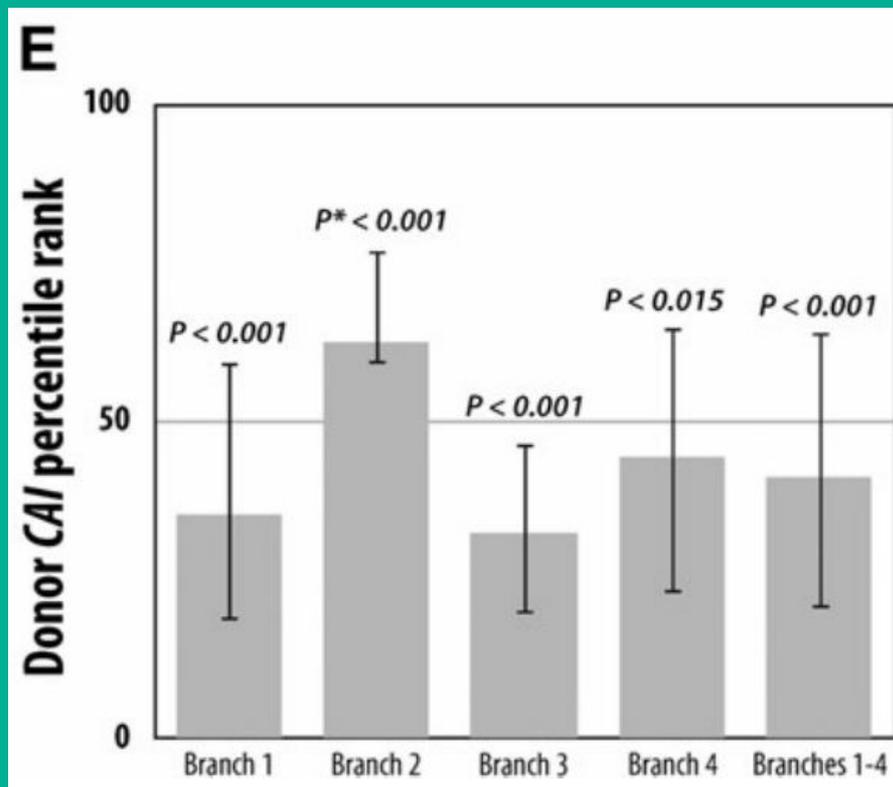
We found the expression levels of the recently acquired genes to be significantly lower than those of resident genes(fig. 2B), The same can be seen in the comparison of expression percentile ranks(fig. 2C).

Analysis of percentile ranks of CAIs calculated based on E. coli codon usage givesa similar result.

The difference in expression level between the recently acquired genes and the resident genes can have two explanations:

- Foreign genes that were recently acquired by E. coli tend to be lowly expressed.

- All genes have reduced expressions when transferred into new hosts.

D

$P^* < 0.001$

$P < 0.001$

$P < 0.001$

$P < 0.001$

$P < 0.001$

$P < 0.001$

E. coli CAI percentile rank

100

50

0

Branch 1   Branch 2   Branch 3   Branch 4   Branches 1-4   Resident

If the difference caused by the second reason, the transferred genes should not be biased toward low expressions in their original hosts. We identified the most likely donor species of each recent HGT to E. coli and then calculated the CAI percentile rank of the transferred gene among all genes in the donor.
Clearly, the horizontally transferred genes have relatively low CAIs among all genes in their donors. Thus, the expression difference must be caused by the first reason that high expression hampers HGT.
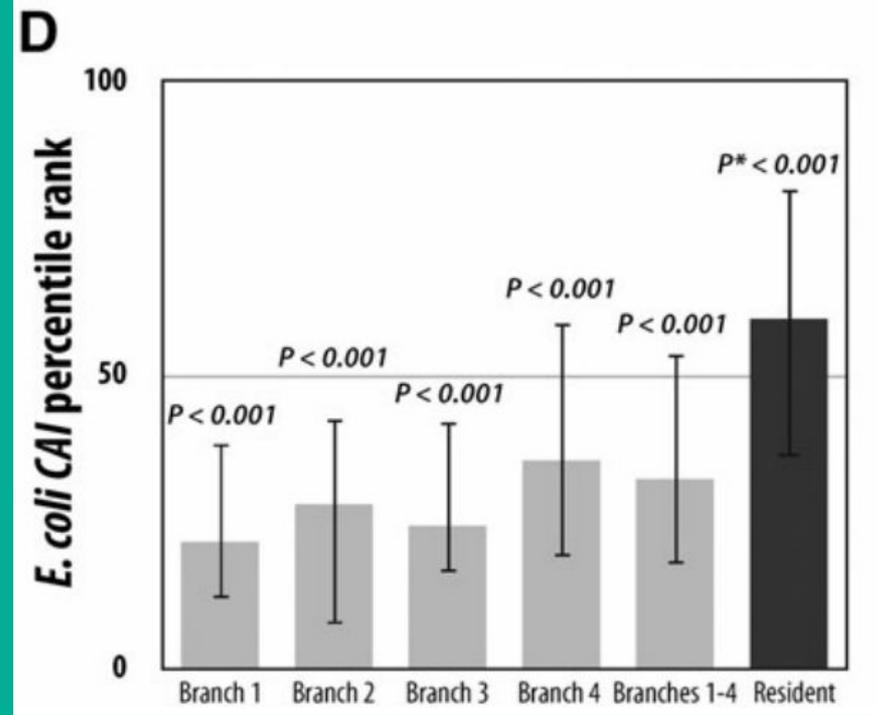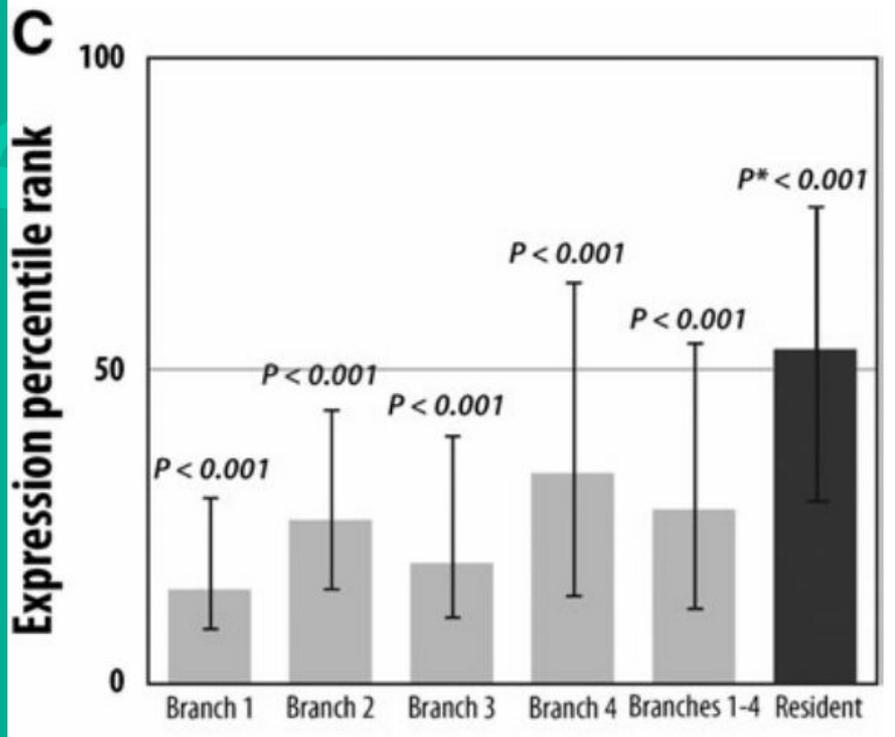
Notable, for the horizontally transferred genes, their expression percentile ranks in the recipient (fig. 2C) appear lower than their CAI percentile ranks in the donor (fig. 2E), suggesting that the second reason is likely at work too.

Table 1

Relative Contributions of Protein Function (Informational vs. Operational), Complexity (Number of Protein Interaction Partners), GC%, and Expression Level on Gene Transferability in Natural HGTs to *Escherichia coli*

Factors Considered	Rank Correlations with Gene Transferability[a]				Multiple Linear Regression[b]	
	Correlation	P Value	Partial Correlation[c]	P Value	RCVE[d]	P Value[e]
Expression level	−0.283 (−0.499)[f]	0.0001 (0.0001)	−0.195	0.0001	0.337	0.0001
Number of protein interactions	−0.250 (−0.432)[f]	0.0001 (0.0001)	−0.148	0.0001	0.191	0.0001
Informational/operational[g]	−0.001 (−0.001)[h]	0.9976 (0.9999)	0.014	0.5280	0.002	0.5281
GC%	−0.147 (−0.259)[f]	0.0001 (0.0001)	−0.055	0.0120	0.026	0.0121
Total[i]					0.105	

Using systematically annotated E. coli protein interactions, we found a significant negative correlation between the transferability of a gene and its number of protein interaction partners.

Table 1

Relative Contributions of Protein Function (Informational vs. Operational), Complexity (Number of Protein Interaction Partners), GC%, and Expression Level on Gene Transferability in Natural HGTs to *Escherichia coli*

Factors Considered	Rank Correlations with Gene Transferability[a]				Multiple Linear Regression[b]	
	Correlation	P Value	Partial Correlation[c]	P Value	RCVE[d]	P Value[e]
Expression level	−0.283 (−0.499)[f]	0.0001 (0.0001)	−0.195	0.0001	0.337	0.0001
Number of protein interactions	−0.250 (−0.432)[f]	0.0001 (0.0001)	−0.148	0.0001	0.191	0.0001
Informational/operational[g]	−0.001 (−0.001)[h]	0.9976 (0.9999)	0.014	0.5280	0.002	0.5281
GC%	−0.147 (−0.259)[f]	0.0001 (0.0001)	−0.055	0.0120	0.026	0.0121
Total[i]					0.105	

Assigning a function score of 1 to informational genes and 0 to operational genes, we found no correlation between function and transferability.

Table 1

Relative Contributions of Protein Function (Informational vs. Operational), Complexity (Number of Protein Interaction Partners), GC%, and Expression Level on Gene Transferability in Natural HGTs to *Escherichia coli*

Factors Considered	Rank Correlations with Gene Transferability[a]				Multiple Linear Regression[b]	
	Correlation	P Value	Partial Correlation[c]	P Value	RCVE[d]	P Value[e]
Expression level	−0.283 (−0.499)[f]	0.0001 (0.0001)	−0.195	0.0001	0.337	0.0001
Number of protein interactions	−0.250 (−0.432)[f]	0.0001 (0.0001)	−0.148	0.0001	0.191	0.0001
Informational/operational[g]	−0.001 (−0.001)[h]	0.9976 (0.9999)	0.014	0.5280	0.002	0.5281
GC%	−0.147 (−0.259)[f]	0.0001 (0.0001)	−0.055	0.0120	0.026	0.0121
Total[i]					0.105	

It has been observed that genes acquired by HGT tend to have low frequencies of guanine (G) and cytosine (C). We thus considered the GC% of a gene as an additional factor potentially impacting HGT. Indeed, we found a significantly negative correlation between gene transferability and GC%.

Note, this correlation may be a byproduct of the correlation between expression level and transferability because highly expressed proteins tend to use metabolically cheap amino acids, which are encoded by GC-rich codons.

Table 1

Relative Contributions of Protein Function (Informational vs. Operational), Complexity (Number of Protein Interaction Partners), GC%, and Expression Level on Gene Transferability in Natural HGTs to *Escherichia coli*

Factors Considered	Rank Correlations with Gene Transferability[a]				Multiple Linear Regression[b]	
	Correlation	P Value	Partial Correlation[c]	P Value	RCVE[d]	P Value[e]
Expression level	−0.283 (−0.499)[f]	0.0001 (0.0001)	−0.195	0.0001	0.337	0.0001
Number of protein interactions	−0.250 (−0.432)[f]	0.0001 (0.0001)	−0.148	0.0001	0.191	0.0001
Informational/operational[g]	−0.001 (−0.001)[h]	0.9976 (0.9999)	0.014	0.5280	0.002	0.5281
GC%	−0.147 (−0.259)[f]	0.0001 (0.0001)	−0.055	0.0120	0.026	0.0121
Total[i]					0.105	

After the controls of protein function, complexity, and GC%, the correlation between gene expression and transferability remains significant, indicating that expression level affects gene transferability independent of the other three factors. Among the four factors examined here, expression level has the strongest correlation with transferability.
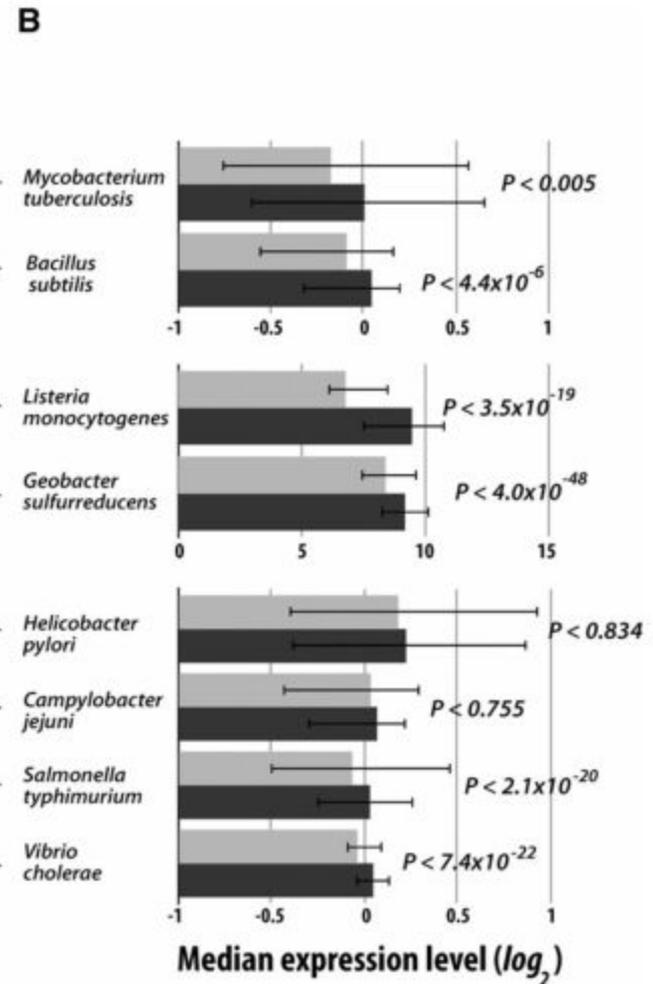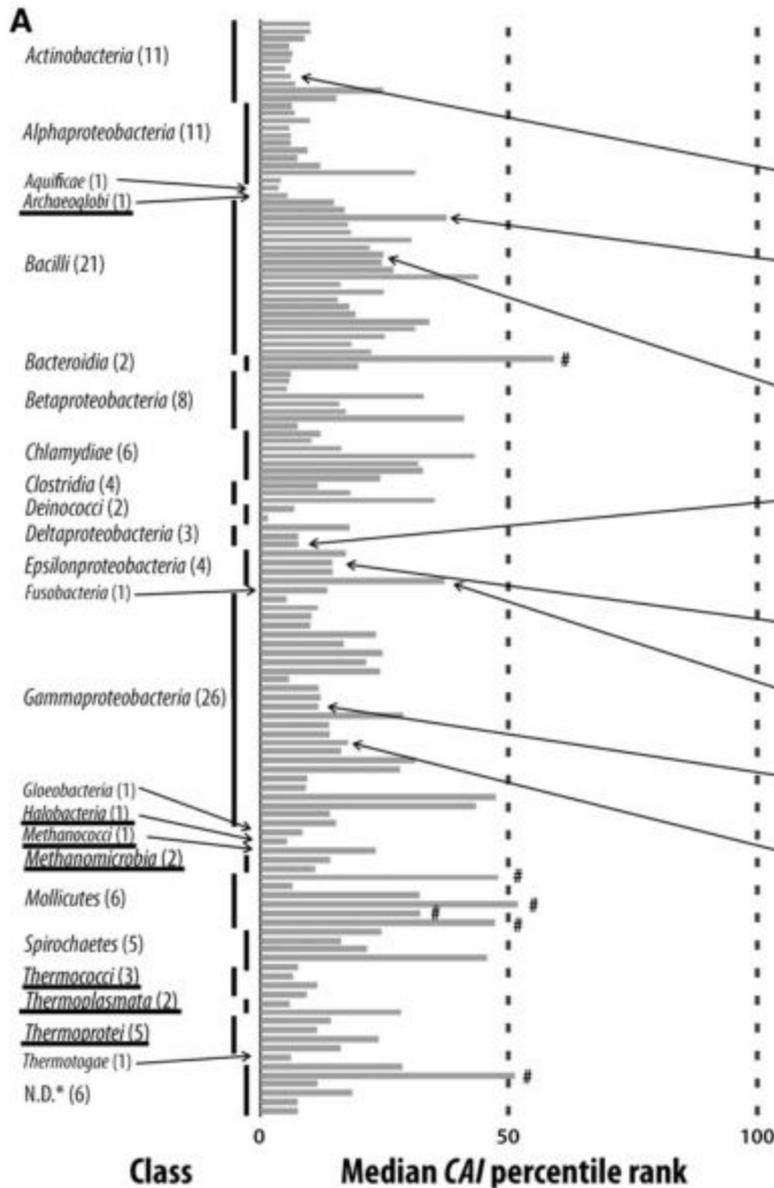
To confirm that the patterns observed in laboratory and natural HGTs to E.coli are not unique to E.coli, we examined HGTs to many other prokaryotes. We analyzed 133 prokaryotic genomes, where recently acquired genes were previously identified.

We found the horizontally acquired genes to have a median CAI percentile rank significantly below 50 in 127 species and insignificantly below 50 in three species (fig. 3A), but none exceeds 50 significantly.

Eight of the 133 species have publicly available microarray expression data. In all eight species, expression levels of horizontally acquired genes are lower than those of resident genes, and the difference is significant in six of the eight species.

Thus, we conclude that the phenomenon of lower HGT rates for more highly expressed genes isg eneral among prokaryotes.

A

Class | Median *CAI* percentile rank

- *Actinobacteria* (11)
- *Alphaproteobacteria* (11)
- *Aquificae* (1)
- *Archaeoglobi* (1)
- *Bacilli* (21)
- *Bacteroidia* (2)
- *Betaproteobacteria* (8)
- *Chlamydiae* (6)
- *Clostridia* (4)
- *Deinococci* (2)
- *Deltaproteobacteria* (3)
- *Epsilonproteobacteria* (4)
- *Fusobacteria* (1)
- *Gammaproteobacteria* (26)
- *Gloeobacteria* (1)
- *Halobacteria* (1)
- *Methanococci* (1)
- *Methanomicrobia* (2)
- *Mollicutes* (6)
- *Spirochaetes* (5)
- *Thermococci* (3)
- *Thermoplasmata* (2)
- *Thermoprotei* (5)
- *Thermotogae* (1)
- N.D.* (6)

0   50   100

B

Median expression level (*log$_2$*)

- *Mycobacterium tuberculosis* — $P < 0.005$
- *Bacillus subtilis* — $P < 4.4 \times 10^{-6}$
- *Listeria monocytogenes* — $P < 3.5 \times 10^{-19}$
- *Geobacter sulfurreducens* — $P < 4.0 \times 10^{-48}$
- *Helicobacter pylori* — $P < 0.834$
- *Campylobacter jejuni* — $P < 0.755$
- *Salmonella typhimurium* — $P < 2.1 \times 10^{-20}$
- *Vibrio cholerae* — $P < 7.4 \times 10^{-22}$

Disscussion

Examining laboratory and natural HGTs to E. coli and natural HGTs to many other prokaryotes, we showed that high expression hinders HGT. Furthermore, we found gene expression level to be a more important determinant of gene transferability than three known factors: protein function, protein complexity, and GC%.

Acquisition of a highly expressed gene is expected to have a greater fitness effect than that of a lowly expressed gene. If most gene acquisitions by HGT are beneficial, strongly expressed genes would confer higher benefits and be more transferable. The observation that it is weakly expressed genes that are more transferable suggests that most HGTs are not beneficial.

The transferred gene must be useful to the recipient for it to be stably retained in the recipient's genome during evolution. When the transferred gene brings in a new function that is useless or even deleterious to the recipient, the new function may become beneficial when the environment or the genetic background is altered. These processe explain how a horizontally acquired gene, even with a nearly neutral origin via HGT, can later become indispensable to the recipient and/or facilitate its adaptation.

A gene can evolve in three aspects: its product function, its expression level and pattern, and its genomic environment. HGT is a common mechanism for gene evolution in the last-named aspect. Compared with lowly expressed genes, highly expressed genes are known to be slower in coding sequence evolution and expression-profile evolution. The present study showed that highly expressed genes are also slower in HGT. Thus, high expression constrains gene evolution in all three aspects.

Thank you !